IBM SPSS Modeler
도움말 1
(소장용)

원문:
https://www.ibm.com/docs/ko/spss-modeler/SaaS?topic=started-how-use-spss-modeler

2022. 05.

국제개발연구소

<목차>

I. IBM SPSS Modeler 구독 문서 ····· 24
II. 구독 개요 ····· 25
1. IBM SPSS Modeler Subscription ····· 25
2. IBM SPSS Modeler Subscription 의 새로운 기능 ····· 26
3. IBM SPSS Modeler Subscription 구매 ····· 30
4. IBM SPSS Modeler Subscription 관리 ····· 31
5. 다운로드 및 설치 지시사항 ····· 34
6. 로그온 및 업데이트 다운로드 ····· 36
III. IBM SPSS Modeler 도움말 ····· 37
1. 사용자 안내서 ····· 37
1) IBM SPSS Modeler 정보 ····· 37
(1) 문서 ····· 38
① IBM SPSS Modeler Subscription 문서 ····· 38
(2) 애플리케이션 예제 ····· 39
(3) Demos 폴더 ····· 39
2) 제품 개요 ····· 39
(1) 시작하기 ····· 39
① 데이터 마이닝 및 모델링 ····· 40
② 예측 분석 정보 ····· 40
③ 모델링 기술 ····· 41
④ 사용할 기술을 파악하는 방법 ····· 41
⑤ IBM SPSS Modeler 의 효용 ····· 42
⑥ IBM SPSS Modeler 를 사용할 수 있는 경우 ····· 42
⑦ IBM SPSS Modeler 작동 방법 ····· 43
⑧ IBM SPSS Modeler 사용 방법 ····· 44

(2) IBM® SPSS® Modeler 시작 ····· 44
① 명령행에서 시작 ····· 44
(3) IBM SPSS Modeler 인터페이스 살펴보기 ····· 45
① IBM SPSS Modeler 스트림 캔버스 ····· 45
② 노드 팔레트 ····· 45
③ IBM SPSS Modeler 관리자 ····· 47
④ IBM SPSS Modeler 프로젝트 ····· 47
⑤ IBM SPSS Modeler 도구 모음 ····· 47
⑥ 도구 모음 사용자 정의 ····· 48
⑦ IBM SPSS Modeler 창 사용자 정의 ····· 49
⑧ 스트림 아이콘 크기 변경 ····· 50
⑨ IBM SPSS Modeler 에서 마우스 사용 ····· 50
⑩ 단축키 사용 ····· 51
(4) 인쇄 ····· 52
(5) IBM SPSS Modeler 자동화 ····· 53
3) 데이터 마이닝 이해 ····· 53
(1) 모델 유형 ····· 53
(2) 데이터 마이닝 예제 ····· 58
4) 스트림 작성 ····· 59
(1) 스트림 작성 개요 ····· 59
(2) 데이터 스트림 작성 ····· 59
① 노드에 대한 작업 ····· 60
가. 스트림에 노드 추가 ····· 60
나. 스트림의 노드 연결 ····· 61
다. 스트림의 노드 우회 ····· 62
라. 스트림의 노드 사용 안함 ····· 62
마. 기존 연결의 노드 추가 ····· 63

바. 노드 간의 연결 삭제 ····· 63
사. 노드 옵션 설정 ····· 63
아. 노드 캐싱 옵션 ····· 64
자. 노드에서 데이터 미리보기 ····· 66
차. 노드 잠금 ····· 66
② 스트림에 대한 작업 ····· 67
가. 스트림의 옵션 설정 ····· 67
ㄱ. 스트림의 일반 옵션 설정 ····· 68
ㄴ. 스트림의 날짜 및 시간 옵션 설정 ····· 70
ㄷ. 스트림의 숫자 형식 옵션 설정 ····· 71
ㄹ. 스트림의 최적화 옵션 설정 ····· 72
ㅁ. 스트림의 SQL 로깅 및 레코드 상태 옵션 설정 ····· 74
ㅂ. 스트림의 레이아웃 옵션 설정 ····· 74
ㅅ. 스트림의 지리 공간적 옵션 설정 ····· 75
• 지리 공간적 좌표계 선택 ····· 76
나. 스트림 작업 메시지 보기 ····· 76
다. 실행 시간 보기 ····· 77
라. 스트림과 세션 모수 설정 ····· 77
마. 모수 값의 런타임 프롬프트 지정 ····· 79
바. 모수 유형의 값 제한조건 지정 ····· 79
사. 스트림의 실행 반복 ····· 80
아. 스트림의 글로벌 값 보기 ····· 81
자. 스트림에서 노드 검색 ····· 81
차. 스트림 이름 변경 ····· 83
③ 스트림 설명 ····· 83
가. 스트림 설명 미리보기 ····· 84
나. 스트림 설명 내보내기 ····· 85

④ 스트림 실행 ····· 85
⑤ 모델에 대한 작업 ····· 86
⑥ 노드와 스트림에 설명과 주석 추가 ····· 86
가. 설명 ····· 86
ㄱ. 주석 관련 작업 ····· 88
ㄴ. 스트림 주석 나열 ····· 90
ㄷ. 주석(Annotation)을 주석(Comment)으로 변환 ····· 91
나. 주석(Annotation) ····· 92
⑦ 데이터 스트림 저장 ····· 93
가. 상태 저장 ····· 94
나. 노드 저장 ····· 94
다. 여러 스트림 오브젝트 저장 ····· 94
라. 출력 저장 ····· 94
마. 정보 암호화 및 복호화 ····· 95
⑧ 파일 로드 ····· 95
⑨ 데이터 스트림 맵핑 ····· 96
가. 데이터를 템플리트에 맵핑 ····· 97
나. 스트림 사이에 맵핑 ····· 98
다. 본질적 필드 지정 ····· 98
라. 맵핑된 필드 탐색 ····· 99
(3) 팁 및 단축키 ····· 99
5) 데이터에 대한 작업 ····· 100
(1) 차트 작성 ····· 101
① 차트 카탈로그 ····· 101
② 레이아웃 및 용어 ····· 102
③ 차트 유형 갤러리에서 차트 작성 ····· 102
④ 차트 유형 ····· 103

가. 3 차원 차트 ····· 105
나. 막대형 차트 ····· 106
다. 상자도표 ····· 109
라. 거품형 차트 ····· 110
마. 원통형 차트 ····· 111
바. 원 채우기 차트 ····· 112
사. 사용자 정의 차트 ····· 113
아. 이중 Y 축 차트 ····· 114
자. 오차 막대 차트 ····· 115
차. 평가 차트 ····· 117
카. 히트 맵 차트 ····· 120
타. 히스토그램 차트 ····· 121
파. 선형 차트 ····· 123
하. 맵 차트 ····· 124
ㄱ. 맵 서비스 옵션 ····· 126
거. 수학 곡선 차트 ····· 128
너. 다중 차트 ····· 129
더. 다중 계열 차트 ····· 130
러. 평형 차트 ····· 131
머. 원형 차트 ····· 132
버. 인구 피라미드 차트 ····· 134
서. Q-Q 도표 ····· 135
어. 방사형 차트 ····· 137
저. 관계 차트 ····· 138
처. 산점도 및 점도표 ····· 139
커. 산점도 행렬 차트 ····· 141
터. 햇살 차트 ····· 142

퍼. 시간 도표 ····· 144
허. 테마 리버 차트 ····· 145
고. 트리 차트 ····· 146
노. 트리 맵 차트 ····· 148
도. t-SNE 차트 ····· 149
로. 단어 클라우드 차트 ····· 150
(2) 대시보드 ····· 151
(3) 글로벌 시각화 기본 설정 ····· 151
6) 출력 작업 ····· 153
(1) 뷰어 ····· 154
① 결과 표시 및 숨김 ····· 154
가. 테이블 및 차트 숨김 ····· 154
나. 프로시저 결과 숨김 ····· 155
② 출력 이동, 삭제, 복사 ····· 155
가. 뷰어에서 출력 이동 ····· 155
나. 뷰어에서 출력 삭제 ····· 155
③ 초기 맞춤 변경 ····· 155
④ 출력 항목의 맞추기 변경 ····· 155
⑤ 뷰어 개요 ····· 156
가. 개요 보기 접기 및 펼치기 ····· 156
나. 개요 수준 변경 ····· 156
다. 개요 항목의 크기 변경 ····· 157
라. 개요에서 글꼴 변경 ····· 157
⑥ 항목 편집 및 뷰어에 추가 ····· 157
가. 제목 또는 텍스트 추가 ····· 157
나. 텍스트 파일 추가 ····· 158
다. 뷰어에 오브젝트 붙여넣기 ····· 158

⑦ 뷰어에서 찾기 및 바꾸기 ····· 158
(2) 다른 애플리케이션으로 출력 복사 ····· 159
(3) 대화형 출력 ····· 161
(4) 출력 내보내기 ····· 162
① HTML 옵션 ····· 163
② 웹 보고서 옵션 ····· 164
③ Word 옵션 ····· 165
④ Excel 옵션 ····· 166
⑤ PowerPoint 옵션 ····· 167
⑥ PDF 옵션 ····· 168
⑦ 텍스트 옵션 ····· 169
⑧ 이미지만 옵션 ····· 170
⑨ 그래픽 형식 옵션 ····· 171
가. JPEG 차트 내보내기 옵션 ····· 171
나. BMP 차트 내보내기 옵션 ····· 171
다. PNG 차트 내보내기 옵션 ····· 171
라. EMF 및 TIFF 차트 내보내기 옵션 ····· 172
마. EPS 차트 내보내기 옵션 ····· 172
(5) 뷰어 인쇄 ····· 172
① 출력과 차트 인쇄 ····· 173
② 인쇄 미리보기 ····· 173
③ 페이지 속성: 머리말 및 꼬리말 ····· 173
가. 페이지 헤더 및 꼬리말 삽입 ····· 174
④ 페이지 속성: 옵션 ····· 174
가. 차트 인쇄 크기, 페이지 매김, 인쇄된 항목 간 공백 변경 방법 ····· 174
(6) 출력 저장 ····· 175
① 뷰어 문서 저장 ····· 175

(7) 피벗 테이블 ····· 176
① 피벗 테이블 ····· 177
② 피벗 표 조작 ····· 177
가. 피벗 테이블 활성화 ····· 177
나. 테이블 피벗 ····· 177
다. 차원 내에서 요소의 표시 순서 변경 ····· 177
라. 차원 요소 내에서 열 및 행 이동 ····· 178
마. 행과 열 전치 ····· 178
바. 행 또는 열 그룹화 ····· 178
사. 행이나 열 그룹 해제 ····· 178
아. 행이나 열 레이블 회전 ····· 178
자. 행 정렬 ····· 179
차. 행 및 열 삽입 ····· 179
카. 변수 및 변수값 레이블 표시 제어 ····· 180
타. 출력 언어 변경 ····· 180
파. 큰 테이블 탐색 ····· 181
하. 변경 실행 취소 ····· 181
③ 레이어 작업 ····· 181
가. 레이어 작성 및 표시 ····· 181
나. 레이어 범주로 이동 ····· 182
④ 항목 표시 및 숨김 ····· 182
가. 테이블에서 행과 열을 숨김 ····· 182
나. 테이블에서 숨겨진 행과 열을 표시 ····· 182
다. 차원 레이블 숨김 및 표시 ····· 182
라. 테이블 제목 숨김 및 표시 ····· 183
⑤ 테이블모양 ····· 183
가. 테이블모양 적용 ····· 183

나. 테이블모양 편집 또는 작성 ····· 183
⑥ 테이블 특성 ····· 184
가. 피벗 표 특성 변경 ····· 184
나. 테이블 특성: 일반 ····· 184
ㄱ. 표시할 행 수 설정 ····· 185
다. 테이블 특성: 참고 ····· 186
라. 테이블 특성: 셀 형식 ····· 186
마. 테이블 특성: 경계선 ····· 187
바. 테이블 특성: 인쇄 ····· 188
⑦ 셀 특성 ····· 188
가. 셀 특성: 글꼴 및 배경 ····· 188
나. 셀 특성: 형식 값 ····· 189
다. 셀 특성: 맞춤 및 여백 ····· 189
⑧ 꼬리말 및 캡션 ····· 189
가. 꼬리말 및 캡션 추가 ····· 189
나. 캡션 숨김 또는 표시 ····· 189
다. 테이블에서 꼬리말 숨김 또는 표시 ····· 190
라. 꼬리말 표식 ····· 190
마. 꼬리말에 다시 번호 매기기 ····· 191
바. 레거시 테이블의 꼬리말 편집 ····· 191
ㄱ. 꼬리말 글꼴 및 색상 설정 ····· 192
⑨ 데이터 셀 너비 ····· 192
⑩ 열 너비 변경 ····· 192
⑪ 피벗 표에서 숨겨진 경계선 표시 ····· 193
⑫ 피벗 표에서 행, 열과 셀을 선택 ····· 193
⑬ 피벗 테이블 인쇄 ····· 193
가. 넓거나 긴 테이블에 대한 테이블 구분 제어 ····· 193

⑭ 피벗 표에서 차트 작성 ····· 195
⑮ 레거시 테이블 ····· 195
(8) 옵션 ····· 195
① 일반 옵션 ····· 196
가. 일반 옵션 설정 ····· 196
② 뷰어 옵션 ····· 196
가. 뷰어 옵션 설정 ····· 197
나. 새 출력의 초기 표시 상태 제어 ····· 197
다. 새 결과의 초기 맞추기 변경 ····· 198
③ 피벗 테이블 옵션 ····· 198
가. 새 피벗 테이블의 테이블모양 설정 ····· 199
나. 새 피벗 표의 열 너비 제어 ····· 199
다. 사용자 정의 기본 테이블모양 작성 ····· 200
④ 출력 옵션 ····· 200
가. 출력 옵션 설정 ····· 200
7) 결측값 처리 ····· 201
(1) 결측값의 개요 ····· 201
(2) 결측값 처리 ····· 202
① 결측값이 있는 레코드 처리 ····· 202
② 결측값이 있는 필드 처리 ····· 203
③ 시스템 결측값이 있는 레코드 처리 ····· 203
(3) 결측값 대치 또는 채우기 ····· 206
(4) CLEM 함수 결측값 ····· 207
8) CLEM 표현식 작성 ····· 208
(1) CLEM 정보 ····· 208
(2) CLEM 예제 ····· 209
(3) 값 및 데이터 유형 ····· 211

(4) 표현식 및 조건 ····· 212
(5) 스트림, 세션, 수퍼 노드 모수 ····· 213
(6) 문자열에 대한 작업 ····· 214
(7) 공백 및 결측값 처리 ····· 215
(8) 숫자에 대한 작업 ····· 215
(9) 시간 및 날짜에 대한 작업 ····· 216
(10) 다중 필드 요약 ····· 217
(11) 다중반응 데이터에 대한 작업 ····· 218
(12) 표현식 작성기 ····· 219
① 표현식 작성기에 액세스 ····· 220
② 표현식 작성 ····· 220
③ 함수 선택 ····· 220
가. 데이터베이스 함수 ····· 221
④ 필드, 모수 및 글로벌 변수 선택 ····· 224
⑤ 값 보기 또는 선택 ····· 225
⑥ CLEM 표현식 검사 ····· 225
(13) 찾기 및 바꾸기 ····· 226
9) CLEM 언어 참조 ····· 229
(1) CLEM 참조 개요 ····· 229
(2) CLEM 데이터 유형 ····· 229
① 정수 ····· 230
② 실수 ····· 230
③ 문자 ····· 230
④ 문자열 ····· 231
⑤ 목록 ····· 231
⑥ 필드 ····· 231
⑦ 날짜 ····· 232

⑧ 시간 ····· 233
(3) CLEM 연산자 ····· 234
(4) 함수 참조 ····· 236
① 함수 설명 규칙 ····· 237
② 정보 함수 ····· 238
③ 변환 함수 ····· 239
④ 비교 함수 ····· 241
⑤ 논리 함수 ····· 244
⑥ 숫자 함수 ····· 244
⑦ 삼각 함수 ····· 246
⑧ 확률 함수 ····· 247
⑨ 공간 함수 ····· 247
⑩ Bitwise 정수 연산 ····· 249
⑪ 랜덤 함수 ····· 250
⑫ 문자열 함수 ····· 251
⑬ SoundEx 함수 ····· 255
⑭ 날짜 및 시간 함수 ····· 256
가. 날짜 및 시간 값 변환 ····· 261
⑮ 시퀀스 함수 ····· 261
⑯ 글로벌 함수 ····· 267
⑰ 공백 및 널값 처리 함수 ····· 269
⑱ 특수 필드 ····· 270
10) IBM Cloud Pak for Data 에 스트림 저장 ····· 271
11) 외부 애플리케이션으로 내보내기 ····· 273
(1) 외부 애플리케이션으로 내보내기 정보 ····· 273
(2) 스트림을 IBM SPSS Modeler Advantage 에서 열기 ····· 274
12) 프로젝트 및 보고서 ····· 274

(1) 프로젝트에 대한 소개 ····· 274
① CRISP-DM 보기 ····· 275
가. 기본 프로젝트 단계 설정 ····· 275
② 클래스 보기 ····· 276
(2) 프로젝트 작성 ····· 276
① 새 프로젝트 작성 ····· 277
② 프로젝트에 추가 ····· 277
③ 프로젝트를 IBM SPSS Collaboration and Deployment Services Repository 에 전송 ····· 278
④ 프로젝트 특성 설정 ····· 279
⑤ 프로젝트에 주석달기 ····· 279
가. 폴더 특성 및 주석(Annotation) ····· 280
⑥ 오브젝트 특성 ····· 280
⑦ 프로젝트 닫기 ····· 281
(3) 보고서 생성 ····· 281
① 생성된 보고서 저장 및 내보내기 ····· 283
13) IBM® SPSS® Modeler 사용자 정의 ····· 284
(1) IBM SPSS Modeler 옵션 사용자 정의 ····· 284
(2) IBM SPSS Modeler 옵션 설정 ····· 284
① 시스템 옵션 ····· 285
가. 메모리 관리 ····· 286
② 기본 디렉토리 설정 ····· 286
③ 사용자 옵션 설정 ····· 286
가. 알림 옵션 설정 ····· 287
나. 표시 옵션 설정 ····· 289
다. 구문 표시 옵션 설정 ····· 290
라. PMML 내보내기 옵션 설정 ····· 290

마. 사용자 정보 설정 ····· 291
(3) 노드 팔레트 사용자 정의 ····· 291
① 팔레트 관리자 사용자 정의 ····· 291
가. 팔레트 탭 작성 ····· 292
나. 노드 팔레트에 팔레트 탭 표시 ····· 293
다. 팔레트 탭에 하위 팔레트 표시 ····· 293
라. 하위 팔레트 작성 ····· 294
② 팔레트 탭 보기 변경 ····· 295
14) 스트림과 노드의 성능 고려사항 ····· 295
(1) 노드의 순서 ····· 295
(2) 노드 캐시 ····· 296
(3) 성능: 프로세스 노드 ····· 298
(4) 성능: 모델링 노드 ····· 299
(5) 성능: CLEM 표현식 ····· 300
15) IBM® SPSS® Modeler 의 내게 필요한 옵션 ····· 300
(1) IBM SPSS Modeler 의 내게 필요한 옵션 개요 ····· 300
(2) 내게 필요한 옵션 지원 유형 ····· 300
① 시각 장애인을 위한 내게 필요한 옵션 ····· 300
② 맹인 사용자를 위한 내게 필요한 옵션 ····· 302
③ 내게 필요한 옵션의 키보드 기능 ····· 302
가. 기본 창 탐색을 위한 단축키 ····· 303
나. 대화 상자 및 테이블 단축키 ····· 306
다. 주석 단축키 ····· 307
라. 군집 뷰어 및 모델 뷰어 단축키 ····· 308
마. 단축키 예제: 스트림 빌드 ····· 310
바. 단축키 예제: 노드 편집 ····· 310
④ 스크린 리더 사용 ····· 311

가. HTML 출력에 스크린 리더 사용 ····· 312
나. 대화식 트리 창에서 내게 필요한 옵션 ····· 312
(3) 사용 팁 ····· 312
① 다른 소프트웨어와의 충돌 ····· 313
② JAWS 및 Java ····· 314
③ IBM® SPSS® Modeler 에서 그래프 사용 ····· 314
16) 유니코드 지원 ····· 314
(1) IBM SPSS Modeler 에서 유니코드 지원 ····· 314
2. 소스, 프로세스 및 출력 노드 ····· 315
1) 소스 노드 ····· 315
(1) 개요 ····· 315
(2) 필드 저장 공간 및 형식화 설정 ····· 317
① 목록 저장 공간 및 연관된 측정 수준 ····· 320
(3) 지원되지 않는 제어 문자 ····· 321
(4) 데이터베이스 소스 노드 ····· 322
① 데이터베이스 노드 옵션 설정 ····· 323
② 데이터베이스 연결 추가 ····· 324
③ 잠재적 데이터베이스 문제 ····· 327
④ 데이터베이스 연결에 대한 사전 설정된 값 지정 ····· 328
가. SQL Server 에 대한 설정 ····· 328
나. Oracle 에 대한 설정 ····· 328
다. z/OS 용 IBM Db2, IBM Db2 LUW 및 Teradata 에 대한 설정 ····· 329
⑤ 데이터베이스 테이블 선택 ····· 331
⑥ 데이터베이스 조회 ····· 332
가. SQL 쿼리에서 스트림 모수 사용 ····· 333
⑦ 사용자 정의 데이터베이스 구성 파일 사용 ····· 334
(5) 가변파일 노드 ····· 336

① 가변파일 노드의 옵션 설정 ····· 337
② 가변파일 노드에 지리 공간적 데이터 가져오기 ····· 340
(6) 고정 파일 노드 ····· 341
① 고정 파일 노드의 옵션 설정 ····· 341
(7) 통계량 파일 노드 ····· 343
(8) Data Collection 노드 ····· 345
① Data Collection 파일 가져오기 옵션 ····· 345
② Data Collection 가져오기 메타데이터 특성 ····· 348
③ 데이터베이스 연결 문자열 ····· 349
④ 고급 특성 ····· 349
⑤ 다중 응답 세트 가져오기 ····· 349
⑥ Data Collection 열 가져오기 참고 ····· 350
(9) IBM Cognos 소스 노드 ····· 350
① Cognos 오브젝트 아이콘 ····· 351
② Cognos 데이터 가져오기 ····· 352
③ Cognos 보고서 가져오기 ····· 353
④ Cognos 연결 ····· 354
⑤ Cognos 위치 선택 ····· 355
⑥ 데이터 또는 보고서에 대한 매개변수 지정 ····· 355
(10) IBM Cognos TM1 소스 노드 ····· 355
① IBM Cognos TM1 데이터 가져오기 ····· 356
(11) TWC 소스 노드 ····· 357
(12) SAS 소스 노드 ····· 359
① SAS 소스 노드에 대한 옵션 설정 ····· 359
가. 멤버 선택 ····· 360
(13) Excel 소스 노드 ····· 360
(14) XML 소스 노드 ····· 361

① 여러 루트 요소에서 선택 ····· 363
② XML 소스 데이터에서 원하지 않는 공백 제거 ····· 363
(15) 사용자 입력 노드 ····· 364
① 사용자 입력 노드의 옵션 설정 ····· 365
(16) 시뮬레이션 생성 노드 ····· 370
① 시뮬레이션 생성 노드에 대한 옵션 설정 ····· 372
② 복제 필드 ····· 377
③ 세부사항 적합 ····· 378
④ 매개변수 지정 ····· 380
가. 반복 ····· 381
⑤ 분포 ····· 382
(17) 확장 가져오기 노드 ····· 385
① 확장 가져오기 노드 - 명령문 탭 ····· 385
② 확장 가져오기 노드 - 콘솔 출력 탭 ····· 385
③ 필드 필터링 또는 이름 바꾸기 ····· 386
④ 유형에 대한 정보 보기 및 설정 ····· 386
(18) 지리 공간적 소스 노드 ····· 387
① 지리 공간적 소스 노드에 대한 옵션 설정 ····· 387
(19) JSON 소스 노드 ····· 388
(20) 공통 소스 노드 탭 ····· 389
① 소스 노드에서 측정 수준 설정 ····· 389
가. 소스 노드에서 인스턴스화할 시기 ····· 390
② 소스 노드에서 필드 필터링 ····· 390
2) 레코드 작업 노드 ····· 391
(1) 레코드 작업 개요 ····· 391
(2) 선택 노드 ····· 393
(3) 표본 노드 ····· 394

① 표본 노드 옵션 ····· 395
② 군집 및 층화 설정 ····· 398
③ 계층에 대한 표본 크기 ····· 399
(4) 균형 노드 ····· 400
① 균형 노드의 옵션 설정 ····· 400
(5) 통합 노드 ····· 401
① 통합 노드의 옵션 설정 ····· 402
② 최적화 설정 통합 ····· 405
(6) RFM 통합 노드 ····· 405
① RFM 통합 노드에 대한 옵션 설정 ····· 406
(7) 정렬 노드 ····· 407
① 정렬 최적화 설정 ····· 408
(8) 합치기 노드 ····· 408
① 결합의 유형 ····· 409
② 병합 방법 및 키 지정 ····· 411
③ 부분 결합에 대한 데이터 선택 ····· 412
④ 병합을 위한 조건 지정 ····· 413
⑤ 병합을 위한 순위화된 조건 지정 ····· 413
⑥ 병합 노드의 필드 필터링 ····· 415
⑦ 입력 순서 및 태그 지정 설정 ····· 416
⑧ 병합 최적화 설정 ····· 417
(9) 붙여쓰기 노드 ····· 418
① 붙여쓰기 옵션 설정 ····· 419
(10) 고유 노드 ····· 419
① 고유 최적화 설정 ····· 422
② 고유 복합 설정 ····· 422
가. 고유 복합 - 사용자 정의 탭 ····· 423

(11) 스트리밍 시계열 노드 ····· 424
① 스트리밍 시계열 노드 - 필드 옵션 ····· 425
② 스트리밍 시계열 노드 - 데이터 지정 사항 옵션 ····· 426
가. 스트리밍 시계열 노드 - 관측값 ····· 426
나. 스트리밍 시계열 노드 - 분석 시간 구간 ····· 427
다. 스트리밍 시계열 노드 - 통합 및 분포 옵션 ····· 428
라. 스트리밍 시계열 노드 - 결측값 옵션 ····· 429
마. 스트리밍 시계열 노드 - 추정 기간 ····· 430
③ 스트리밍 시계열 노드 - 작성 옵션 ····· 430
가. 스트리밍 시계열 노드 - 일반 작성 옵션 ····· 430
ㄱ. 전이 및 변환 함수 ····· 434
④ 스트리밍 시계열 노드 - 모델 옵션 ····· 435
(12) SMOTE 노드 ····· 437
① SMOTE 노드 설정 ····· 438
(13) 확장 변환 노드 ····· 439
① 확장 변환 노드 - 명령문 탭 ····· 439
② 확장 변환 노드 - 콘솔 출력 탭 ····· 440
(14) Space-Time-Box 노드 ····· 440
① Space-Time-Box 밀도 정의 ····· 443
(15) 스트리밍 TCM 노드 ····· 444
① 스트리밍 TCM 노드 - 시계열 옵션 ····· 444
가. 스트리밍 TCM 노드 - 차원 값 선택 ····· 445
② 스트리밍 TCM 노드 - 관측 옵션 ····· 445
③ 스트리밍 TCM 노드 - 시간 구간 옵션 ····· 447
④ 스트리밍 TCM 노드 - 통합 및 분포 옵션 ····· 447
⑤ 스트리밍 TCM 노드 - 결측값 옵션 ····· 448
⑥ 스트리밍 TCM 노드 - 일반 데이터 옵션 ····· 448

⑦ 스트리밍 TCM 노드 - 일반 작성 옵션 ····· 449
⑧ 스트리밍 TCM 노드 - 추정 기간 옵션 ····· 449
⑨ 스트리밍 TCM 노드 - 모델 옵션 ····· 450
(16) CPLEX 최적화 노드 ····· 450
① CPLEX 최적화 노드에 대한 옵션 설정 ····· 452
3) 필드 조작 노드 ····· 453
(1) 필드 조작 개요 ····· 453
(2) 자동 데이터 준비 ····· 456
① 필드 탭(자동 데이터 준비) ····· 458
② 설정 탭(자동 데이터 준비) ····· 458
가. 필드 설정 ····· 459
나. 날짜 및 시간 준비(자동 데이터 준비) ····· 459
다. 필드 제외(자동 데이터 준비) ····· 460
라. 입력 및 목표 준비 ····· 461
마. 생성 및 필드선택 ····· 462
바. 필드 이름(자동 데이터 준비) ····· 464
③ 분석 탭(자동 데이터 준비) ····· 464
가. 필드 처리 요약(자동 데이터 준비) ····· 465
나. 필드(자동 데이터 준비) ····· 466
다. 동작 요약(자동 데이터 준비) ····· 468
라. 예측력(자동 데이터 준비) ····· 468
마. 필드 테이블(자동 데이터 준비) ····· 468
바. 필드 세부사항(자동 데이터 준비) ····· 469
사. 조치 세부사항(자동 데이터 준비) ····· 471
④ 파생 노드 생성 ····· 473
(3) 유형 노드 ····· 475
① 유형에 대한 정보 보기 및 설정 ····· 476

② 측정 수준 ····· 476
가. 지리 공간적 측정 수준 ····· 479
③ 연속형 데이터 변환 ····· 480
④ 인스턴스화란? ····· 481
⑤ 데이터 값 ····· 482
가. 값 대화 상자 사용 ····· 483
나. 연속형 데이터의 값 및 레이블 지정 ····· 485
ㄱ. 값 및 레이블 하위 대화 상자 ····· 486
다. 명목 및 순서 데이터의 값 및 레이블 지정 ····· 486
라. 플래그의 값 지정 ····· 487
마. 콜렉션 데이터의 값 지정 ····· 487
바. 지리 공간 데이터의 값 지정 ····· 487
⑥ 결측값 정의 ····· 488
⑦ 유형 값 검사 ····· 488
⑧ 필드 역할 설정 ····· 489
⑨ 유형 속성 복사 ····· 490
⑩ 필드 형식 설정 탭 ····· 490
가. 필드 형식 옵션 설정 ····· 491
(4) 필드 필터링 또는 이름 바꾸기 ····· 492
① 필터링 옵션 설정 ····· 493
가. 필드 이름 자르기 ····· 494
나. 필드 이름 익명화 ····· 495
다. 다중 응답 세트 편집 ····· 495
(5) 파생 노드 ····· 496
① 파생 노드에 대한 기본 옵션 설정 ····· 498
② 다중 필드 파생 ····· 498
가. 다중 필드 선택 ····· 499

③ 수식 파생 옵션 설정 ····· 500
가. 파생된 목록 값 설정 ····· 500
나. 목록 또는 지리 공간적 필드 파생 ····· 501
④ 파생 플래그 옵션 설정 ····· 502
⑤ 파생 명목 옵션 설정 ····· 503
⑥ 파생 상태 옵션 설정 ····· 503
⑦ 파생 개수 옵션 설정 ····· 504
⑧ 파생 조건부 옵션 설정 ····· 504
⑨ 파생 노드를 사용하여 값 코딩변경 ····· 505
(6) 채움 노드 ····· 505
① 채움 노드를 사용한 저장 공간 변환 ····· 506
(7) 재분류 노드 ····· 507
① 재분류 노드에 대한 옵션 설정 ····· 508
② 다중 필드 재분류 ····· 509
③ 재분류 필드에 대한 저장 공간 및 측정 수준 ····· 509
(8) 값 익명화 노드 ····· 510
① 익명화 노드의 옵션 설정 ····· 511
가. 필드 값이 익명화되는 방식 지정 ····· 511
② 필드 값 익명화 ····· 512
(9) 구간화 노드 ····· 513
① 구간화 노드의 옵션 설정 ····· 514
② 고정 너비 구간 ····· 514
③ 분위수(동일 개수 또는 합계) ····· 515
④ 케이스 순위 지정 ····· 517
⑤ 평균/표준 편차 ····· 518
⑥ 최적 구간화 ····· 519
가. 절단점 설정 ····· 519

⑦ 생성된 구간 미리보기 ····· 520
(10) RFM 분석 노드 ····· 520
① RFM 분석 노드 설정 ····· 521
② RFM 분석 노드 구간화 ····· 522
(11) 앙상블 노드 ····· 523
① 앙상블 노드 설정 ····· 524
(12) 파티션 노드 ····· 525
① 파티션 노드 옵션 ····· 526
(13) 플래그로 설정 노드 ····· 527
① 플래그로 설정 노드에 대한 옵션 설정 ····· 527
(14) 구조변환 노드 ····· 528
① 구조변환 노드에 대한 옵션 설정 ····· 530
(15) 전치 노드 ····· 530
① 전치 노드의 옵션 설정 ····· 530
(16) 히스토리 노드 ····· 533
① 히스토리 노드에 대한 옵션 설정 ····· 533
(17) 필드 다시 정렬 노드 ····· 534
① 필드 재정렬 설정 옵션 ····· 534
(18) 시간 구간 노드 ····· 536
① 시간 간격 - 필드 옵션 ····· 536
② 시간 구간 - 작성 옵션 ····· 537
(19) 재투영 노드 ····· 538
① 재투영 노드에 대한 설정 옵션 ····· 539

I. IBM SPSS Modeler 구독 문서

Resources

IBM Support Portal Get assistance for the IBM products, services and software you own.	**IBM Fix Central** Provides fixes and updates for your system's software, hardware, and operating system.
IBM Passport Advantage Provides you with secure access to software downloads and media.	**IBM Redbooks** Find best practices for integrating IBM technologies.
IBM Developer Explore technical topics, find trial software and join the community.	**IBM Training** Explore, learn and succeed with IBM training and certification.

II. 구독 개요

1. IBM SPSS Modeler Subscription

IBM® SPSS® Modeler Subscription에서는 일반적인 IBM SPSS Modeler 제품과 동일한 예측 분석 공정능력을 제공합니다. Subscription 버전을 사용하면 제품 업데이트를 정기적으로 다운로드할 수 있습니다.

이 문서는 Subscription 제품 오퍼링 자체와 같이 필요에 따라 연속적으로 개선되고 업데이트됩니다.

제한사항

일부 경우에는 문서에서 제한된 기능이 언급될 수 있습니다. 해당 절은 무시하십시오. 다음 기능은 Subscription 오퍼링에서 지원되지 않습니다. 이 목록에 포함되지 않은 지원되지 않는 기타 기능이 있을 수 있습니다.

- IBM SPSS Modeler Subscription에는 클라이언트 소프트웨어만 포함됩니다. SPSS Modeler Server, SPSS Statistics Server, SPSS Analytic Server 및 SPSS Collaboration and Deployment Services Server와 연관된 모든 기능(서버 연결 설정, Collaboration and Deployment Services 리포지토리에서 저장/검색, 외부 Analytic Server에서 스트림 실행, 스트림 배치 옵션 설정 등)은 사용 가능하지 않습니다.
- 스코어링 어댑터를 사용하여 SQL을 통해 스코어링
- IBM SPSS Statistics 노드
- 데이터 보기 노드
- Analytic Server 소스 및 내보내기 노드

2. IBM SPSS Modeler Subscription의 새로운 기능

2021년 8월

- 모델링의 불편한 점은 시간이 지남에 따라 데이터가 변경되어 모델이 구식이 된다는 것입니다. 이를 일반적으로 *모델 드리프트* 또는 *개념 드리프트*라고 합니다. 모델 드리프트를 효과적으로 극복하기 위해 SPSS Modeler는 *연속 자동 기계 학습*을 제공합니다. IBM의 연구 결과, 생물학의 자연 선택에서 영감을 받은 연속 기계 학습을 자동 분류자 노드와 자동 숫자 노드에 제공합니다. 추가 정보는 연속 기계 학습의 내용을 참조하십시오.
- 이제 클라이언트에서 직접 스트림을 IBM Cloud Pak for Data Server에 업로드할 수 있습니다. 추가 정보는 IBM Cloud Pak for Data에 스트림 저장의 내용을 참조하십시오.
- 클라이언트에 대해 로깅을 사용으로 설정하려면 텍스트 편집기에서 log4j2.xml 파일을 열고 다음 행에서 level="info"를 level="debug"로 변경하십시오.

```
<Logger name="com.spss" additivity="false" level="info">
```

- **음이 아닌 최소 제곱(NNLS) 수행**이라는 새로운 설정이 모수 추정값 아래 작성 옵션 탭의 GLE 노드에 추가되었습니다. 음이 아닌 최소 제곱(NNLS)은 계수가 음수가 될 수 없는 제약된 최소 제곱 문제 유형입니다. 모든 데이터 세트가 NNLS에 적합한 것은 아니므로, 예측변수와 목표값 간에 양의 상관 관계가 필요하거나 아예 필요하지 않습니다.

- **열과 유형에 대해 스캔할 행**이라는 새 설정이 Excel 소스 노드에 추가되었습니다. 이 새 설정의 기본값은 200입니다. 열 유형 및 저장 유형을 판별하기 위해 Excel 데이터의 더 많은 행을 스캔하도록 이 값을 늘려야 하는 경우 성능이 영향을 받을 수 있습니다.
- Planning Analytics on Cloud를 지원하기 위해 새로운 연결 방법(TM1 서버 연결)을 IBM Cognos TM1 소스 노드와 IBM Cognos TM1 내보내기 노드에 사용할 수 있습니다. 관리 서버는 Planning Analytics on Cloud에서 제거되었으므로, TM1 관리 서버를 통해 Planning Analytics on Cloud에 연결되는 TM1 노드가 이전 스트림에 포함되어 있는 경우 해당 스트림이 TM1 서버를 대신 가리키도록 수정할 수 있습니다.
- Db2 11.5가 이제 지원됩니다.
- Db2 Warehouse가 이제 지원됩니다.
- Cloudera Data Platform 7.1.5의 Db2 Big SQL 7.1.0이 지원됩니다.
- Cloudera Data Platform 7.1.5의 Apache Hive 3.1.3 7.1.0이 지원됩니다.
- Cloudera Data Platform 7.1.5의 Cloudera Impala 3.4.0이 지원됩니다.
- Informix 14.10이 이제 지원됩니다.
- 새 버전의 R이 사용됩니다(4.0.4).

2021년 3월

- Oracle Database 19c가 지원됩니다.
- Microsoft SQL Server 2017이 지원됩니다.
- Vertica 9.3가 지원됩니다.
- Snowflake가 지원됩니다.
- Treasure Data가 포함된 Plazma 데이터베이스가 지원되지만 다음과 같은 제한사항이 있습니다.
  - 문자열이 아니라 정수/실수 형식으로 데이터를 내보낼 수 있습니다.
  - Datetime/date/time 형식은 지원되지 않습니다.
- Mac Catalina의 버전 10.15.3 이상이 지원됩니다.
- 이전에는 Text Analytics 노드 변환이 더 이상 사용되지 않았어도 노드를 포함하는 기존 스트림은 여전히 실행할 수 있었습니다. 이제 노드 변환은 더 이상 지원되지 않습니다.
- 새 버전의 SDAP이 사용 가능합니다(8.1).
- 새 버전의 R이 사용됩니다(3.5.1).
- 새 버전의 Python이 사용됩니다(3.7.9).

2020년 3월

새 노드

다음 노드가 추가되었습니다.

- **E-Plot(베타) 노드.** 새 베타 레벨 E-Plot 노드는 노드 팔레트의 그래프 탭에서 사용할 수

있습니다. 이는 직관적이며 현대적이며 높은 수준의 사용자 정의가 가능하며 대화형 데이터 도표가 있는 새 그래프 인터페이스를 사용합니다. 이 새 노드를 사용하여 새 그래프 기능을 활용할 수 있습니다. 세부사항은 E-Plot 그래프 사용을 참조하십시오.

- **가우스 혼합 노드.** 새로운 가우스 혼합 노드를 Python 탭 및 노드 팔레트의 모델링 탭에서 사용할 수 있습니다. 세부사항은 가우스 혼합 노드를 참조하십시오.
- **KDE(Kernel Density Estimation) 노드.** 새 KDE 모델링 노드를 Python 탭과 노드 팔레트의 모델링 탭에서 사용할 수 있습니다. 새 KDE 시뮬레이션 노드는 Python 탭과 출력 탭에서 사용할 수 있습니다. 세부사항은 KDE 노드를 참조하십시오.
- **JSON 노드.** 새 JSON 노드는 데이터를 JSON 형식으로 가져오고 내보내는 데 사용할 수 있습니다. 세부사항은 JSON Source node 및 JSON 내보내기 노드를 참조하십시오.

새 룩앤필

새로운 최신 인터페이스 테마가 **도구** > **사용자 옵션** > **표시**를 통해 사용 가능합니다. 새 테마로 전환하기 위한 지시사항은 표시 옵션 설정을 참조하십시오.

IBM Data Warehouse

이제 IBM Netezza Analytics를 사용하는 데이터베이스 모델링에서 IBM Data Warehouse를 지원합니다. 노드 팔레트의 데이터베이스 모델링 탭에서 노드를 사용하려면 **도구** > **옵션** > **헬퍼 애플리케이션**으로 이동하여 IBM Data Warehouse 탭에서 IBM Data Warehouse 통합을 사용으로 설정하십시오. 사용 가능한 Netezza 노드 중 하나를 실행하는 경우 만들어진 모델이 이제 IBM DB2 Data Warehouse에 작성됩니다. AIX는 지원되지 않습니다.

IBM SPSS Modeler Text Analytics 개선사항

다음 기능이 향상되었습니다. 이 개선사항의 대부분은 IBM® SPSS® Text Analytics for Surveys에서 확인되는 기능과 유사합니다.

- 이제 텍스트 분석 패키지(.tap)에서 자원을 가져오는 방법과 동일한 방식으로 SPSS Text Analytics for Surveys 프로젝트(.tas)를 가져올 수 있습니다. 텍스트 마이닝 모델링 노드를 구성할 때 추출 중에 사용될 자원을 지정해야 합니다. 자원 템플리트를 선택하는 대신 .tap 또는 .tas(신규)를 선택하여 해당 자원뿐만 아니라 범주 세트를 노드에 복사할 수 있습니다.
- 이제 데이터 분할창에서 플래그를 사용할 수 있습니다. "완료" 플래그 또는 "중요" 플래그를 사용하여 문서에 플래그를 지정할 수 있습니다. 새로운 열에는 사용 가능한 플래그가 표시되며 열 내부를 클릭하여 플래그 유형을 변경할 수 있습니다. 이 기능은 범주 모델의 완성도를 검토할 때 유용합니다. 반응 플래그 지정을 참조하십시오.
- 추출된 개념 결과가 개선되었습니다(이제 SPSS Text Analytics for Surveys 의 추출된 개념 결과와 유사함).
- 빈 레코드는 이제 SPSS Text Analytics for Surveys 에 있을 때와 동일하게 처리됩니다. 예를 들어, Excel 소스 파일을 사용하는 경우 빈 레코드는 이제 텍스트의 일부로 유지됩니다.
- 레코드를 범주에 또는 범주로부터 적용하기 위해 새 자동 설정 및 자동 설정 종료 옵션을 데이터 분할창에서 사용할 수 있습니다. 이 기능은 빈 레코드에 추출된 개념이 없는 경우

그리고 개념 또는 TLA 출력을 통해 적절한 범주를 찾을 수 없는 경우에도 유용합니다. 범주에 반응 강제 적용을 참조하십시오.

- 이제 TRR(Type Reassignment Rule)을 사용할 수 있습니다. TRR은 유형, 매크로 및/또는 토큰으로 구성된 시퀀스를 특정 유형의 새 개념으로 변환합니다. 극성에서 변경된 의견을 예외 처리(catch)하기 위해 의견 템플리트에서 사용할 수 있습니다. 세부사항은 TRR(Type Reassignment Rule)을 참조하십시오.

2018년 2월

IBM SPSS Modeler Text Analytics가 이제 등록 오퍼링에 포함됩니다. IBM SPSS Modeler Text Analytics는 고급 언어 기술 및 NLP(Natural Language Processing)를 사용하여 여러 비정형 텍스트 데이터를 신속히 처리하고 키 개념을 추출하여 조직화하고 이러한 개념을 카탈로그로 그룹화합니다. 추출된 개념 및 범주는 인구 통계와 같은 기존의 구조화된 데이터와 결합되고 전체 IBM SPSS Modeler 데이터 마이닝 도구를 통한 모델링에 적용되어 더 좋고 더 집중된 의사결정을 산출할 수 있습니다. Text Analytics에 대한 자세한 정보는 IBM SPSS Modeler Text Analytics 도움말을 참조하십시오.

2017년 10월(초기 IBM SPSS Modeler Subscription 오퍼링)

- **라이센스 부여.** IBM SPSS Modeler Subscription 라이센스 부여 프로세스가 IBMid라고도 하는 IBM 계정에 의해 대체되었습니다. IBMid는 사용자가 사용권을 가진 모든 IBM 애플리케이션, 커뮤니티 및 지원 채널에 대한 액세스를 제공합니다. 추가 정보는 로그온 및 업데이트 다운로드의 내용을 참조하십시오.먼저 IBM SPSS Modeler Subscription을 열 때, IBMid로 로그온하도록 프롬프트됩니다. 아직 IBMid가 없는 경우, 화면의 지시사항을 따르십시오.
  라이센스 부여가 IBM SPSS Modeler Subscription에서 단순화되었습니다. 다음 라이센스 부여 옵션을 사용할 수 있습니다.

표 1. IBM SPSS Modeler Subscription 라이센스 부여 옵션

IBM SPSS Modeler Subscription, 기본 개정판
IBM SPSS Modeler Subscription, 추가 기능 - SQL 최적화

- **업데이트 확인.** 기본적으로 해당 제품은 새 제품 업데이트가 사용 가능할 때마다 자동으로 이를 사용자에게 알립니다. 즉시 업데이트를 설치하거나 업데이트를 설치하지 않도록 선택할 수 있습니다. 또한 **도움말** 메뉴에서 **업데이트 확인** 옵션을 선택할 수 있습니다. 제품에서 업데이트를 알리는 것을 원하지 않는 경우, **옵션** > **사용자 옵션** > **알림** 아래에서 **선택적 업데이트가 사용 가능한 경우 시작할 때 알림**을 선택 취소하십시오. 추가 정보는 로그온 및 업데이트 다운로드의 내용을 참조하십시오.

- **Spark 노드.** 노드 팔레트의 새 Spark 탭은 Python 알고리즘 사용에 필요한 노드를 제공합니다. 이러한 새 노드는 Windows 64 및 Mac에서 지원됩니다.
  - **Isotonic-AS 노드.** 새 Isotonic-AS 노드는 새 Spark 탭에서 사용할 수 있습니다. 세부사항은 Isotonic-AS 노드를 참조하십시오.
  - **XGBoost-AS 노드.** 새 XGBoost-AS 노드는 새 Spark 탭에서 사용할 수 있습니다. 세부사항은 XGBoost-AS 노드를 참조하십시오.
  - **K-Means-AS 노드.** 새 K-Means-AS 노드는 새 Spark 탭에서 사용할 수 있습니다. 세부사항은 K-Means-AS 노드를 참조하십시오.

- **하이퍼-모수 최적화(Rbfopt 기준).** 단일 클래스 SVM 노드(전문가 탭), XGBoost 선형 노드(작성 옵션 탭) 및 XGBoost 트리 노드(작성 옵션 탭)에 새 옵션이 추가되었습니다. 새 하이퍼-모수 최적화 옵션은 모델이 표본에 대해 기대빈도 또는 하한 오차율을 달성할 수 있도록 모수의 최적 조합을 자동으로 검색합니다.
- **랜덤 포리스트 노드.** 새 랜덤 포리스트 노드는 Python 탭에서 사용할 수 있습니다. 세부사항은 랜덤 포리스트 노드를 참조하십시오.
- **t-SNE 노드.** 새 t-SNE(t-Distributed Stochastic Neighbor Embedding) 노드는 Python 탭 및 그래프 탭에서 사용할 수 있습니다. 세부사항은 t-SNE 노드를 참조하십시오.
- **CPLEX 최적화 노드에 대한 다중 데이터 소스.** 이제 최적화 전문가가 다중 데이터 소스에서 CPLEX 최적화 노드로 데이터 소스를 가져와서 각 데이터 소스를 튜플에 할당할 수 있습니다. 세부사항은 CPLEX 최적화 노드 및 CPLEX 최적화 노드에 대한 옵션 설정을 참조하십시오.

3. IBM SPSS Modeler Subscription 구매

IBM® SPSS® Modeler는 오랜 기간 동안 데이터 패턴을 발견하고 예측 정확도를 얻고 의사결정을 개선하는 데 도움이 되는 예측 가능한 분석을 제공하는 분야에서 리더 역할을 수행해 왔습니다.

이제 더 이상 소프트웨어 사용권이나 버전 업데이트에 대해 걱정할 필요 없이 이전보다 더 빠른 속도와 사용 편의성을 제공하는 구독을 통해 SPSS Modeler를 사용할 수 있습니다.

IBM SPSS Modeler Subscription은 일반적인 SPSS Modeler 라이센스 부여 모델에 의존하지 않으므로 더욱 융통성 있는 구매 옵션을 제공합니다. 기간 약정 없이 매월 구독으로 소프트웨어를 구매할 수 있습니다. 단순화된 패키징으로 인해 필요에 따라 적합한 옵션을 구매할 수 있습니다. IBM Marketplace에서 직접 소프트웨어를 구매할 수 있으며 몇 분 내에 시작하고 실행할 수 있습니다.

다음은 IBM SPSS Modeler Subscription 구매 프로세스에 대한 안내를 제공하는 단계입니다.
1. IBM SPSS Modeler 포털로 이동하십시오. 평가판 버전을 다운로드하거나 IBM SPSS

Modeler Subscription 기본 버전 및 사용 가능한 추가 기능을 구매할 수 있는 옵션이 제시됩니다.

2. **구매**를 클릭하여 구독 옵션을 검토한 다음 원하는 구독 옵션에 대해 **지금 구매**를 클릭하십시오.

> ⓘ **참고:** 모든 추가 기능에 대해 기본 옵션은 필수입니다. 추가 기능은 기본 버전과 함께 사용되는 경우에만 작동합니다.

3. 임의의 구독 오퍼링을 구매하는 경우, 다중 구독 사용권을 구매할 수 있는 옵션이 제시됩니다. 원하는 수의 구독 사용권을 선택한 다음 필요에 따라 사용 가능한 추가 기능에 대한 구독 사용권 수를 선택하십시오.
4. 매월 총 금액을 검토한 다음 계속 **체크아웃**을 클릭하십시오.
5. 이 시점에서 IBMid를 사용하여 IBM 계정에 로그인해야 합니다. IBMid 신임 정보를 입력하거나 **IBMid 작성**을 클릭하여 새 IBM 계정을 작성하십시오. 로그인한 다음 연락처 및 청구 정보를 입력하도록 요청하는 체크아웃 단계가 표시됩니다.
6. 연락처 정보를 검토하고 주문 세부사항을 검토한 다음 **주문 제출**을 클릭하십시오. 주문 번호가 제공되고 주문 확인 이메일이 발송됩니다.
7. 또한 이메일 주소를 검증하도록 요청하는 두 번째 이메일이 발송됩니다. 이메일에서 **이메일 주소 검증**을 클릭하면 사용자가 IBM SPSS Modeler Subscription 관리에서 설명한 IBM 제품 및 서비스 포털로 이동합니다.

4. IBM SPSS Modeler Subscription 관리

IBM 제품 및 서비스 포털은 IBM Marketplace의 확장이며 사용자에게 IBM 제품 구매, 구독, 사용자 및 서비스를 관리하는 데 필요한 중앙 허브를 제공합니다. 제품 및 서비스 포털에서 다음을 수행할 수 있습니다.

- IBM 제품 다운로드 관리
- IBM 제품이 설치되는 디바이스 관리
- IBM 제품 사용자 관리(사용자 역할, 추가 기능 구성요소 및 상태 포함)
- 포럼, 지원 티켓, 전화 또는 이메일을 통해 IBM 지원 센터에 문의

다음 절에서는 IBM 제품 및 서비스 포털 기능을 탐색하고 관리하는 데 필요한 정보를 제공합니다. 해당 정보에서는 사용자가 이미 IBM 제품을 구매하거나 등록했다고 가정합니다. 자세한 정보는 IBM SPSS Modeler Subscription 구매의 내용을 참조하십시오.

> ⓘ **참고:** IBM® SPSS® Modeler Subscription 제품은 제공되는 모든 예에 사용됩니다.

IBM 제품 및 서비스 포털에 로그인

1. IBM Marketplace에서 IBM 계정(IBMid라고도 함)으로 로그인하십시오. 아직 활성 IBM 계정이 없는 경우에는 IBMid에 등록해야 합니다.
   로그인한 후에는 계정 프로파일이 사용자가 자격을 가진 모든 IBM 제품 및 서비스를 표시하는 **제품 및 서비스** 섹션을 제공합니다.

2. **제품 및 서비스** 섹션에서 **다운로드**를 클릭하여 제품을 다운로드하거나 **관리**를 클릭하여 사용 가능한 제품 관리 옵션을 보십시오.

   > **참고:** 제품을 다운로드하여 설치한 후에 구독을 활성화하기 위해 IBMid 신임 정보를 사용하여 로그인해야 합니다.

IBM 제품 및 서비스 기능 관리

IBM 제품 및 서비스 포털은 사용자가 특정 구독 정보 및 사용자 속성을 관리할 수 있는 다음 섹션을 제공합니다.

개요
개요 섹션은 기본 구독 정보를 제공합니다.

소유자
현재 제품 구독 소유자 이름 및 연관된 이메일 주소를 표시합니다.

계획
현재 제품 구독 계획을 식별합니다.

권한 부여된 사용자의 수
지정된 사용자 및 사용 가능한 구독의 현재 수를 포함하여 권한 부여된 사용자의 총 수를 나열합니다.

추가 기능
구독된 IBM SPSS Modeler Subscription 추가 기능의 총 수를 나열합니다. 추가 기능의 총 수에는 지정된 사용자 및 사용 가능한 추가 기능 구독의 현재 수가 포함됩니다.

청구 빈도
제품 구독이 청구되는 방법(매월, 매년)을 식별합니다.

구독 시작 날짜
구독 시작 날짜를 식별합니다.

구독 ID
고유한 구독 ID 번호를 제공합니다.

개요 섹션에서는 현재 청구 주기의 끝에 구독을 효과적으로 취소하는 취소 계획 링크도 제공됩니다.

다운로드

다운로드 섹션은 사용자가 자격을 가진 모든 IBM 제품 및 서비스를 표시합니다. 사용 가능한 다운로드는 각각의 운영 체제를 기준으로 하여 정렬됩니다. 예를 들어, IBM SPSS Modeler Subscription은 Microsoft Windows 64비트 및 Mac OS에 대해 사용 가능합니다.

관련 다운로드 섹션은 기본 제품을 지원하는 다운로드 링크를 제공합니다. IBM SPSS Modeler Subscription의 경우 관련 다운로드에는 Essentials for R 및 데이터 액세스 팩이 포함됩니다.

디바이스 관리

디바이스 관리 섹션은 구독된 제품이 현재 설치되어 활성 상태인 디바이스를 나열합니다. 제공되는 정보에는 컴퓨터 이름, 운영 체제, 구독 활성화 날짜 및 조치가 포함됩니다.

조치 열에는 연관된 컴퓨터 이름에 대한 제품 구독을 비활성화하는 **비활성화** 링크가 포함됩니다. 해당 기능을 사용하면 다른 컴퓨터/랩탑에 구독된 제품을 설치하고 활성화할 수 있습니다. 예를 들어, 작업 영역 컴퓨터에서 구독을 비활성화한 다음 랩탑 컴퓨터에서 구독을 활성화할 수 있습니다.

일반적으로 관리되는 디바이스에서 IBM SPSS Modeler Subscription을 종료하고 로그오프하여 관리되는 디바이스를 비활성화합니다. **비활성화** 링크를 사용하면 원격으로 디바이스를 비활성화할 수 있습니다. 예를 들어, 관리되는 디바이스에서 IBM SPSS Modeler Subscription을 로그오프하는 것을 잊은 경우 등입니다.

사용자 관리

사용자 관리 섹션은 관리자가 IBM SPSS Modeler Subscription 사용자를 관리할 수 있는 옵션을 제공합니다. 관리자는 새 사용자를 추가하고 기존 사용자를 편집할 수 있습니다.

새 사용자 추가

1. 사용자 추가 옆에 있는 드롭 다운 화살표를 클릭하십시오. 단일 사용자 추가 대화 상자를 사용하면 관리자가 새 사용자 이름, 이메일 주소/IBMid 및 역할(관리자, 라이센스 사용자, 또는 둘 다)을 정의할 수 있습니다. 또한 해당 대화 상자를 사용하여 관리자가 새 사용자에게 자격이 부여되는 IBM SPSS Modeler Subscription을 지정할 수 있습니다.

   ⓘ **참고:** 대화 상자는 사용 가능한 구독의 나머지 수 옆에 지정된 구독의 현재 수를 표시합니다.

2. 필수적인 새 사용자 정보를 입력한 뒤에 **사용자 추가**를 클릭하십시오.

관리자가 사용자 이메일 주소를 입력할 때 새 사용자가 IBMid가 없는 경우, 새 사용자에게 IBMid 작성에 대한 지시사항을 제공하는 이메일 메시지가 전송됩니다. 새 사용자가 IBMid를 작성한 후에 IBM SPSS Modeler Subscription에 로그인할 때 IBMid 신임 정보가 사용됩니다.

ⓘ **참고:** 새로 추가된 사용자가 **사용자 관리** 섹션에 표시되는 데 시간이 걸릴 수 있습니다.

기존 사용자 검색 및 필터링

사용자 관리 섹션은 관리자가 기존 사용자를 검색하고 사용자 역할, 프로비저닝된 추가 기능 및 상태를 기준으로 하여 결과를 필터링할 수 있는 검색 및 필터 옵션을 제공합니다.

검색 및 **필터** 필드를 사용하여 검색 기준을 정의하십시오.

기존 사용자 관리

1. 기존 사용자 레코드 옆의 아래로 화살표를 클릭하십시오. 레코드가 확장되어 다음 옵션을 제공합니다.
   - **이름:** 수정할 수 없는 이 필드는 사용자 이름 및 이메일 주소를 표시합니다.
   - **역할:** 사용자 역할 정의에 필요한 옵션을 제공합니다.
   - **상태:** 사용자 상태 옵션을 표시합니다. 관리자가 사용자를 활성화 또는 제거할 수 있습니다.
   - **추가 기능:** 사용 가능한 IBM SPSS Modeler Subscription 추가 기능을 표시하며 관리자가 사용자에게 추가 기능을 프로비저닝할 수 있습니다.

   > **참고:** 사용자 레코드는 사용 가능한 구독의 나머지 수 옆에 지정된 구독의 현재 수를 표시합니다.

2. 원하는 대로 사용자 레코드를 업데이트한 다음 **저장**을 클릭하여 변경사항을 적용하십시오.

별명 지정

별명 지정 섹션에서는 제품 및 서비스 포털에 표시되는 공식 제품 이름에 별명을 지정할 수 있습니다. 별명은 다른 제품 버전과 차별화하는 데 도움이 됩니다. 별명은 공식 제품 이름 바로 밑에 표시됩니다.

제품 지원

제품 지원 섹션은 포럼, 지원 티켓, 전화 또는 이메일을 통해 IBM 지원 센터에 문의하는 데 필요한 옵션을 제공합니다.

5. 다운로드 및 설치 지시사항

> **참고:** 다음 단계에서는 IBM Marketplace에서 이미 IBM® SPSS® Modeler Subscription을 구매했다고 가정합니다.
> IBM SPSS Modeler Subscription은 빈 폴더에 설치해야 합니다.

IBM SPSS Modeler Subscription을 다운로드하여 설치하려면 IBM Marketplace로 이동하여 다음을 수행하십시오.

1. IBM 계정(IBMid라고도 함)으로 로그인하십시오. 활성 IBM 계정이 아직 없는 경우 IBMid에 등록해야 합니다.
2. 로그인한 후, 계정 프로파일은 사용자가 자격을 가진 모든 IBM 제품 및 서비스를 표시하는 **제품 및 서비스** 섹션을 제공합니다.
3. **IBM SPSS Modeler Subscription** 옆의 **다운로드**를 클릭하십시오.
4. 제품 및 서비스 페이지에서 **IBM SPSS Modeler Subscription** 아래의 **다운로드** 링크를 클릭하십시오.
5. 적절한 다운로드 옵션 (예를 들어, Microsoft Windows 64비트 또는 Mac OS)을 선택한 다음 **다운로드**를 클릭하십시오.
6. 프롬프트되면 **파일 저장**을 클릭하십시오.
7. 워크스테이션의 파일 저장 위치로 이동하여 설치 프로그램 파일을 두 번 클릭하십시오. Mac OS의 경우, 디스크 이미지를 마운트한 후에 설치 프로그램 파일을 두 번 클릭해야 합니다.
8. 제품 설치가 완료될 때까지 사용 동의서에 동의하는 것을 포함하여 설치 단계에 따르십시오.IBM SPSS Modeler Subscription은 빈 폴더에 설치해야 합니다.
9. 이제 IBM SPSS Modeler Subscription을 사용할 준비가 되었습니다.

Windows 자동 설치 지시사항

1. 다운로드한 이미지를 추출하십시오.

```
ModelerSub_Setup.exe /s /x /b"{user_extract_path}" /v"/qn"
```

2. 다음 명령을 사용하여 설치 프로그램 폴더로 이동하십시오.

```
cd {user_extract_path}
```

3. 자동 설치에 필요한 다음 명령을 실행하십시오.

```
msiexec.exe /i "IBM SPSS Modeler Subscription.msi" /qn /L*v logfile.txt INSTALLDIR="{InstallPath}"
```

Windows 설치 제거 지시사항

1. 자동 설치 제거에 필요한 다음 명령을 실행하십시오.

```
msiexec /x {FEFFBC36-D5F1-4124-BBEB-CAB3B1823C2C} /q
```

2. ${ALLUSERSPROFILE}₩IBM₩SPSS₩TextAnalyticsSubscription₩18.3을 찾아 이동하십시오.

```
cd ${ALLUSERSPROFILE}₩IBM₩SPSS₩TextAnalyticsSubscription₩18.3
```

3. Text Analytics 데이터베이스가 더 이상 필요하지 않음을 확인한 후에만 tmwb_18.db 관련 파일을 제거하십시오.

Mac OS 자동 설치 지시사항

1. 터미널 세션을 열고 디스크 이미지를 마운트하십시오. 예를 들어, 다음과 같습니다.

```
hdiutil attach {user_download_path}/SPSSModelerSub_Installer.dmg
```

2. 다음 명령을 사용하여 설치 프로그램 폴더로 이동하십시오.

```
cd /Volumes/SPSSModelerSub_Installer
```

3. 자동 설치에 필요한 다음 명령을 실행하십시오. 루트로 실행하거나 sudo 명령으로 실행해야 합니다.

```
sudo installer -pkg IBM₩ SPSS₩ Modeler₩ Subscription.pkg -target /
umount -f /Volumes/SPSSModelerSub_Installer*
```

> **참고:** 설치 프로그램은 기본 설치 경로("/Applications/IBM SPSS Modeler Subscription/1.1")를 사용합니다. 설치 경로를 수정할 수 있는 방법은 없습니다.

Mac OS 설치 제거 지시사항

1. 설치 폴더(기본적으로 "/Applications/IBM SPSS Modeler Subscription/1.1")를 휴지통으로 끄십시오.
2. 홈 폴더에서 Library/Preferences를 찾으십시오.
3. com.ibm.spss.plist를 휴지통으로 끄십시오. 해당 파일은 IBM SPSS Modeler의 일반적인 비구독 버전에서도 사용됩니다. 이 애플리케이션이 여전히 설치되어 있는 경우, 파일을 제거하지 마십시오.
4. 휴지통을 비우십시오.

6. 로그온 및 업데이트 다운로드

처음 제품을 열거나 제품의 온라인 시작 페이지에 액세스하는 경우, IBMid를 사용하여 로그온하십시오. IBMid를 사용하여 IBM 애플리케이션, 서비스, 커뮤니티, 지원, 온라인 구매 페이지 등에 액세스할 수 있습니다. 아직 IBMid가 없는 경우, 화면의 지시사항을 따르십시오. 로그온한 후에 평가판을 시작하거나 제품의 전체 버전을 구매할 수 있습니다.

또한 사용 가능한 경우, 최신 제품 업데이트 및 수정사항을 다운로드할 수 있습니다.

제품 업데이트 다운로드 방법

해당 제품은 새 제품 업데이트가 사용 가능할 때마다 자동으로 이를 사용자에게 알립니다. 즉시 업데이트를 설치하거나 업데이트를 설치하지 않도록 선택할 수 있습니다. 또한 **도움말** 메뉴에서 **업데이트 확인** 옵션을 선택할 수도 있습니다. 업데이트가 사용 가능할 때마다 프롬프트되지 않게 하려면 **옵션** > **사용자 옵션** > **알림**에서 **선택적 업데이트가 사용 가능할 때 시작 시 알림**을 선택 취소하십시오.

III. IBM SPSS Modeler 도움말

1. 사용자 안내서

1) IBM SPSS Modeler 정보

IBM® SPSS® Modeler는 비즈니스 전문 지식을 사용하여 예측 모형을 신속하게 개발하고 이를 비즈니스 운영에 배포하여 의사결정의 정확성을 향상시켜주는 데이터 마이닝 도구 세트입니다. 산업 표준 CRISP-DM 모델을 중심으로 디자인된 IBM SPSS Modeler는 데이터에서 보다 나아진 비즈니스 결과에 이르는 전체 데이터 마이닝 프로세스를 지원합니다.

IBM SPSS Modeler는 기계 학습, 인공지능 및 통계로부터 취한 다양한 모델링 방법을 제공합니다. 모델링 팔레트에서 사용할 수 있는 이러한 방법을 통해 데이터로부터 새로운 정보를 얻어서 예측 모형을 개발할 수 있습니다. 각각의 방법은 그것만의 장점이 있으며 특정한 문제점 유형에 가장 적합합니다.

SPSS Modeler를 사용하여 프로그래밍하지 않고 신속하게 직관적으로 정확한 예측 모델을 작성할 수 있습니다. 고유한 시각적 인터페이스를 사용하면 데이터 마이닝 프로세스를 쉽게 시각화할 수 있습니다. 제품에 포함된 고급 분석 지원을 통해 데이터에서 이전에 숨겨진 패턴과 추세를 발견할 수 있습니다. 결과를 모델링하고 결과에 영향을 주는 요인을 이해하여 비즈니스 기회를 활용하고 위험을 줄일 수 있습니다.

이 IBM Documentation은 제품의 Subscription 버전용입니다. IBM SPSS Modeler Subscription에서는 일반적인 IBM SPSS Modeler 제품과 동일한 예측 분석 공정능력을 제공합니다. Subscription 버전을 사용하면 제품 업데이트를 정기적으로 다운로드할 수 있습니다.

(1) 문서

문서는 SPSS® Modeler의 도움말 메뉴에서 사용할 수 있습니다. 제품 외부에서 항상 사용 가능한 온라인 IBM Documentation가 열립니다.

① IBM SPSS Modeler Subscription 문서

IBM® SPSS® Modeler Subscription 문서 스위트(설치 지시사항은 제외)는 다음과 같습니다.

- **IBM SPSS Modeler Subscription 개요.** 제한 및 설치 정보와 같이 구독 오퍼링에 대한 일반 정보입니다.
- **IBM SPSS Modeler 사용자 안내서.** SPSS Modeler 사용에 관한 일반적인 소개이며, 데이터 스트림 작성, 결측값 처리, CLEM 표현식 작성 및 프로젝트 및 보고서에 대한 작업에 배포하기 위한 스트림 패키지 방법이 포함됩니다.
- **IBM SPSS Modeler 소스, 프로세스 및 출력 노드.** 여러 형식의 데이터를 읽고 처리하며, 출력하는 데 사용하는 모든 노드에 대한 설명입니다. 실질적으로 이는 모델링 노드 이외의 모든 노드를 의미합니다.
- **IBM SPSS Modeler 모델링 노드.** 데이터 마이닝 모델을 작성하는 데 사용하는 모든 노드에 대한 설명입니다. IBM SPSS Modeler는 기계 학습, 인공지능 및 통계로부터 취한 다양한 모델링 방법을 제공합니다.
- **IBM SPSS Modeler 애플리케이션 안내서.** 이 안내서의 예제는 특정 모델링 방법과 기법을 중점적으로 간략히 소개합니다. 자세한 정보는 애플리케이션 예제의 내용을 참조하십시오.
- **IBM SPSS Modeler Python 스크립팅 및 자동화.** 노드와 스트림을 조작하는 데 사용할 수 있는 특성을 포함하여 Python 스크립팅을 통한 시스템 자동화에 대한 정보입니다.
- **IBM SPSS Modeler CLEF 개발자 안내서.** CLEF는 데이터 처리 루틴 또는 모델링 알고리즘과 같은 써드파티 프로그램을 IBM SPSS Modeler의 노드로 통합하는 기능을 제공합니다.
- **IBM SPSS Modeler In-Database 마이닝 안내서.** 데이터베이스의 능력을 사용하여 성능을 향상시키고 써드파티 알고리즘을 통해 분석 기능 범위를 확장하는 방법에 대한 정보입니다.

- **IBM SPSS Modeler CRISP-DM 안내서.** SPSS Modeler에서 데이터 마이닝에 CRISP-DM 방법론을 사용하기 위한 단계별 안내서입니다.
- **SPSS Modeler Text Analytics 사용자 안내서.** SPSS Modeler에서 텍스트 분석 추가 기능 사용에 대한 정보로, 텍스트 마이닝 노드, 대화식 워크벤치, 템플리트 및 기타 자원에 대해 설명합니다.

(2) 애플리케이션 예제

SPSS® Modeler의 데이터 마이닝 도구가 광범위한 비즈니스 및 조직의 문제점을 해결하는데 도움을 주는 가운데, 애플리케이션 예제는 특정 모델링 방법 및 기술에 대해 대상화된 간략한 소개를 제공합니다. 여기서 사용된 데이터 세트는 일부 데이터 마이너에 의해 관리되는 거대한 데이터 스토어보다 훨씬 작지만, 관련된 개념과 방법은 실제 애플리케이션에 대해 확장 가능합니다.

예제에 액세스하려면 SPSS Modeler의 도움말 메뉴에서 **애플리케이션 예제**를 클릭하십시오.

데이터 파일 및 샘플 스트림은 제품 설치 디렉토리 아래에 있는 Demos 폴더에 설치됩니다. 추가 정보는 Demos 폴더의 내용을 참조하십시오.

(3) Demos 폴더

애플리케이션 예제에 사용되는 데이터 파일 및 샘플 스트림은 제품 설치 디렉토리 아래의 Demos 폴더에 설치됩니다. Windows 시작 메뉴의 IBM SPSS® Modeler 프로그램 그룹에서, 또는 파일 > 스트림 열기 대화 상자의 최근 디렉토리 목록에서 Demos를 클릭해서도 이 폴더에 액세스할 수 있습니다.

2) 제품 개요

(1) 시작하기

IBM® Corp.의 모델링 제품군의 멤버인 IBM SPSS® Modeler를 시작합니다. 시작 위치가 궁금한 경우 다음과 같은 몇 가지 제안이 있습니다.

다음을 수행할 수 있습니다.

- 스트림 작성을 시도하십시오.
- 이 릴리스의 새 기능을 참조하십시오.
- IBM Analytics 웹 사이트(https://www.ibm.com/analytics/us/en/technology/spss/)의 IBM SPSS Modeler 섹션을 참조하십시오.
- 통합 지원 사이트(http://www.ibm.com/support)를 방문하고 정보 관리를 지원 유형으로 선택하십시오.

IBM SPSS Modeler에 대한 대화식 자습서도 있습니다. 기본 메뉴에서 다음을 클릭하십시오.

도움말 > 애플리케이션 예

다음에 대해 학습할 수도 있습니다.

① 데이터 마이닝 및 모델링

데이터 마이닝은 비즈니스 데이터를 조사하여 숨겨진 패턴 및 관계를 발견하는 프로세스입니다. 데이터 마이닝은 일반적인 문제점을 해결합니다. 데이터가 많을수록 데이터로부터 의미를 효과적으로 분석하고 도출하는 것이 더 어렵고 시간도 많이 걸립니다. 인원, 시간 또는 전문 지식의 부족으로 인해 금광이 개척되지 않은 경우가 많습니다.

데이터 마이닝은 명확한 비즈니스 프로시저와 강력한 분석 기술을 사용하여 데이터라는 산을 신속하고 철저하게 탐색하여 사용자에게 필요한 가치 있는 유용한 정보(비즈니스 인텔리전스)를 끄집어 내어 사용자에게 제공합니다.

테이터에서 이전에 알려져 있지 않은 이러한 패턴 및 관계는 그 자체로도 흥미롭지만 이것이 이야기의 끝이 아닙니다. 이러한 과거 작동의 패턴을 사용하여 미래에 일어날 일을 예측할 수 있다면 어떨까요? 여기서 모델링이 관여합니다. **모델**은 소스 데이터에서 추출된 규칙, 수식 또는 방정식 세트이며 이에 대한 예측을 생성할 수 있게 합니다. 이것이 예측 분석의 핵심입니다.

② 예측 분석 정보

예측 분석은 현재 상태와 미래 이벤트에 대해 신뢰할 수 있는 결론을 도출하여 데이터를 효과적인 조치와 연결하는 데 도움을 주는 비즈니스 프로세스 및 관련 기술 세트입니다. 예측 분석은 다음의 항목으로 구성됩니다.

- 고급 분석
- 의사결정 최적화

고급 분석에서는 다수의 도구 및 기술을 사용하여 과거 및 현재 이벤트를 분석하고 미래의 결과를 예측합니다. **의사결정 최적화**에서는 가능한 최상의 결과를 생성하는 사용자 파트의 조치를 판별하여 해당 권장 조치가 사용자의 비즈니스 프로세스에 가장 효과적으로 통합될 수 있는 위치에 전달되도록 합니다.

③ 모델링 기술

모델링 기술은 특정 문제점 해결을 위한 일련의 지시사항인 알고리즘의 사용을 기반으로 합니다. 특정 알고리즘을 사용하여 해당 유형의 모델을 작성합니다. 모델링 기술의 세 가지 기본 클래스가 있으며 IBM® SPSS® Modeler는 각각에 대한 여러 예제를 제공합니다.

- 감독
- 연관
- 세분화("군집화"라고도 함)

*감독 모델*은 하나 이상의 출력 또는 **목표** 필드를 예측하기 위해 하나 이상의 **입력** 필드 값을 사용합니다. 이러한 기술의 일부 예는 다음과 같습니다. 의사결정 트리(C&R 트리, QUEST, CHAID 및 C5.0 알고리즘), 회귀분석(1차, 로지스틱, 일반화 선형 및 Cox 회귀 알고리즘), 신경망, 지원 벡터 머신 및 베이지안 네트워크입니다.

*연관 모델*은 하나 이상의 엔티티(예: 이벤트, 구매 또는 속성)가 하나 이상의 다른 엔티티와 연관되어 있는 데이터에서 패턴을 발견합니다. 모델은 이러한 관계를 정의하는 규칙 세트를 구성합니다. 여기에서 데이터 내의 필드는 입력과 목표 둘 모두의 역할을 할 수 있습니다. 이러한 연관을 수동으로 찾을 수 있지만 연관 규칙 알고리즘은 이를 보다 신속하게 수행하므로 더 복잡한 패턴을 탐색할 수 있습니다. Apriori 및 Carma 모델은 이러한 알고리즘 사용의 예입니다. 연관 모델의 또 다른 유형은 순차 발견 모델이며 이는 시간 구조 데이터에서 순차 패턴을 발견합니다.

*세분화 모델*은 데이터를 입력 필드의 패턴이 유사한 레코드의 세그먼트 또는 군집으로 나눕니다. 이들은 입력 필드에만 관심이 있으므로 세분화 모델에는 출력이나 목표 필드와 같은 개념이 없습니다. 세분화 모델의 예는 코호넨 네트워크, K-평균 군집, 이단계 군집 및 이상 항목 발견입니다.

④ 사용할 기술을 파악하는 방법

사용자가 원하는 대로 사용할 수 있는 기술이 매우 광범위하기 때문에 특정 문제를 해결하기 위해 어디서 시작해야 하는지 파악하는 것이 어려울 수 있습니다. 다행히 IBM® SPSS® Modeler는 **자동화된 모델링** 양식으로 이러한 의사결정 중 일부를 수행할 수 있습니다. 이는 다수의 서로 다른 모델링 방법을 평가하고 비교하여 효과순으로 순위화하는 매우 강력한 기술입니다. 이

방식을 사용하면 한 번의 모델링 실행으로 다양한 접근 방식을 시도할 수 있습니다.
다양한 기술에 익숙해지면 특정 문제점에 가장 적합한 알고리즘 및 해당 알고리즘 내 옵션에 대해 알게 됩니다.

⑤ IBM SPSS Modeler의 효용

IBM® SPSS® Modeler는 데이터 마이닝, 모델링 및 보고 도구입니다. 이는 신속하고 편리하게 데이터에서 패턴 및 추세를 파악할 수 있게 하는 고유 그래픽 인터페이스를 제공합니다. 제품의 고급 분석으로 작성된 통계 알고리즘을 사용하여 현재 및 히스토리 데이터로부터 미래의 결과를 예측하는 데이터 모델을 작성할 수 있습니다.

⑥ IBM SPSS Modeler를 사용할 수 있는 경우

IBM® SPSS® Modeler를 사용하여 광범위한 비즈니스 문제점을 해결할 수 있습니다. 데이터와 관련된 문제가 있는 경우 IBM SPSS Modeler가 해법을 제공할 수 있습니다. 여기에서는 IBM SPSS Modeler가 영향을 미칠 수 있는 업계 및 애플리케이션 중 몇 가지 예만 제공합니다.

교육. 커뮤니티 대학에서는 학생 유지를 개선하기 위해 학생 등록 패턴 및 경향에 대한 깊은 이해를 얻길 원합니다. 학교로 돌아올 가능성이 낮은 학생을 예측하여 교수진 및 관리 부서에서는 학업 상담, 학자금 지원 패키지 및 학과 과정에 대해 직간접적으로 개입할 수 있습니다.

금융. 고객 정보를 분석하여 저당 대출자는 할인 또는 고정 금리 기간 종료 시 다른 업체로 이동할 수 있는 고객, 새로운 금리로 전활할 고객 및 표준 금리로 복귀할 고객을 추정합니다.

범죄 발견. 사법당국은 범죄 보고서 파일을 분석하여 데이터에서 패턴을 찾습니다. 사법당국은 데이터를 세그먼트화하여 시간, 위치 및 방법이 동일하여 동일한 범죄자가 저질렀을 수 있는 범죄의 군집을 찾습니다.

비영리 조직. 과거 기부자 및 해당 기부 금액의 목록을 입력으로 사용하여 자선 단체에서는 광고 우편물 캠페인에 대한 긍정적인 응답의 확률을 계산합니다. 가능성 있는 긍정적인 응답자의 목록을 사용하여 자선 단체에서는 기부 금액을 예측하고 캠페인에 대한 수익 확률을 계산하고 잠재적 기부자의 메일링 목록을 구성합니다.

제조. 산업 제조 업체에서는 가능한 고장이 발생하기 전에 이러한 고장을 예측하기 위해 다양한 기계 부품의 작동 상태를 모니터합니다.

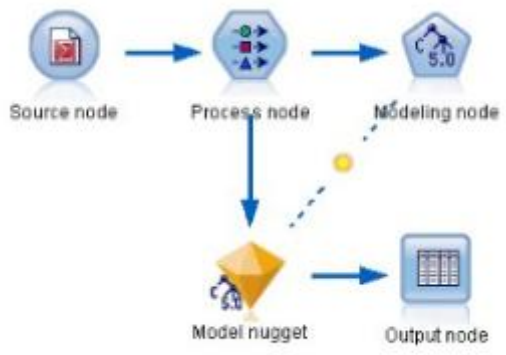
의료. 제약 회사는 여러 처치 중 동일한 질병으로 고통받는 환자들을 위해 가장 효과적일 수 있는 처치를 알고 싶어 합니다. 특정 증상에 대해 더 나은 반응률을 보여주는 약을 분석하기 위해 제약 회사에서는 특정 유형의 환자에게 가장 적합한 약을 예측하는 모델을 작성합니다.

소매. 계산대에서 전자적 POS(Point of Sale) 데이터를 분석하여 수퍼마켓에서는 고객 카드를 가진 고객이 일반적으로 구입하는 제품 조합을 발견할 수 있습니다. 이 장바구니 분석의 결과를 사용하여 소매업체에서는 프로모션 대상 고객을 더 정확하게 지정하거나 우편물의 응답률을 향상시키거나 다른 분점에서 특정 유형의 인구 통계 모집단에 대한 재고 범위를 조정할 수 있습니다.
통신. 고객 이탈(비즈니스 손실) 감소 노력의 일환으로 통신 회사는 다른 서비스로 빠르게 전환하는 고객과 연관된 요인을 판별하기 위해 "이탈 시간"을 모델링하는 데 관심을 가지고 있습니다. 이를 위해 고객의 무작위 표본이 선택되고 이 고객에 대한 다양한 데이터를 데이터베이스에서 가져와서 분석합니다. 회사에서는 이탈 시간을 모델링하여 향후 2년 동안 유지되는 예상 고객 수를 도표화하고 다음 연도에 이탈할 가능성이 가장 높은 개별 고객을 식별합니다.

이 사례 중 일부는 다수의 기타 사례와 함께 *IBM SPSS Modeler 애플리케이션 안내서*에서 완전하게 작동하는 예제로 사용할 수 있습니다.

⑦ IBM SPSS Modeler 작동 방법

그림 1. 스트림의 노드



IBM® SPSS® Modeler의 고유 그래픽 인터페이스는 **노드** 및 스트림을 기반으로 합니다. **노드**는 데이터에 대한 개별 조작을 나타내는 아이콘 또는 모양입니다. 노드는 **스트림**에서 함께 링크되어 각 조작의 데이터 플로우를 나타냅니다.

알고리즘은 **모델링 노드**로 알려져 있는 특별한 유형의 노드에 의해 표시됩니다. IBM SPSS Modeler가 제공하는 각 알고리즘에 대해 다른 모델링 노드가 있습니다. 모델링 노드는 5면체 모양으로 표시됩니다.

기타 유형의 노드는 **소스 노드**, 프로세스 노드 및 출력 노드입니다. 소스 노드는 데이터를 스트림으로 가져오는 노드이며 항상 스트림의 시작 부분에 표시됩니다. **프로세스 노드**는 개별 데이터 레코드 및 필드에 대해 조작을 수행하며 일반적으로 스트림의 중간에서 발견됩니다. **출력 노드**는 데이터, 차트 및 모델 결과에 대해 다양한 출력을 생성하거나 결과를 다른 애플리케이션(데이터베이스 또는 스프레드시트)에 내보낼 수 있게 합니다. 출력 노드는 일반적으로 스트림 또는 스트

림 분기의 마지막 노드로 표시됩니다.
모델링 노드가 포함된 스트림을 실행하면 결과 모델이 스트림에 추가되고 **모델 너깃**으로 알려져 있는 특별한 유형의 노드에 의해 표시됩니다(금색 너깃 같은 모양을 가짐).

⑧ IBM SPSS Modeler 사용 방법

데이터에 대한 임시 분석을 위해 IBM® SPSS® Modeler를 사용하여 이전의 숨겨진 패턴 또는 추세를 발견할 수 있습니다. 하지만 이 강력한 소프트웨어 부분에서 실제 값을 얻으려면 더 주의 깊은 체계적인 접근 방식이 필요합니다.

단순한 구조화된 접근 방식은 다음과 비슷합니다.

- 사용자의 비즈니스 및 데이터 마이닝을 통해 달성하려는 목표에 대한 이해
- 데이터 조사 및 이해
- 분석 및 모델링을 위해 데이터 준비
- 예측 모형 작성
- 모델을 평가하여 올바른지 확인
- 조직에 모델 배포

이 시퀀스는 실제로 **CRISP-DM**(Cross-Industry Standard Process for Data Mining)으로 알려져 있는 데이터 마이닝에 대한 더 상세하고 구조화된 접근 방식의 기초입니다. IBM SPSS Modeler는 특히 이 접근 방식과 함께 작동하도록 설계되어 있습니다.

자세한 정보를 보려면 기본 메뉴에서 다음을 클릭하십시오.

도움말 > CRISP-DM 도움말

(2) IBM® SPSS® Modeler 시작

애플리케이션을 시작하려면 다음을 클릭하십시오.

시작 > [모든] 프로그램 > IBM SPSS Modeler <버전> > IBM SPSS Modeler <버전>

몇 초 후에 기본 창이 표시됩니다.

① 명령행에서 시작

운영 체제의 명령행을 사용하여 다음과 같이 IBM® SPSS® Modeler를 시작할 수 있습니다.

1. IBM SPSS Modeler가 설치된 컴퓨터에서 DOS 또는 명령 프롬프트 창을 여십시오.
2. 대화식 모드로 IBM SPSS Modeler 인터페이스를 시작하려면 modelerclient 명령 뒤에 필수 인수를 입력하십시오. 예를 들어, 다음과 같습니다.

```
modelerclient -stream report.str -execute
```

사용 가능한 인수(플래그)를 사용하면 필요에 따라 서버에 연결하고 스트림을 로드하며, 스크립트를 실행하거나 다른 모수를 지정할 수 있습니다.

(3) IBM SPSS Modeler 인터페이스 살펴보기

데이터 마이닝 프로세스의 각 지점에서 사용하기 용이한 IBM® SPSS® Modeler 인터페이스는 사용자의 특정 비즈니스 전문 지식을 활용합니다. 예측, 분류, 세분화, 연관 발견과 같은 모델링 알고리즘은 강력하고 정확한 모델을 보장합니다. 모델 결과를 쉽게 배포하고 데이터베이스, IBM SPSS Statistics, 다양한 기타 애플리케이션으로 읽어들일 수 있습니다.

IBM SPSS Modeler에 대한 작업은 데이터에 대해 작업하는 3단계 프로세스입니다.

- 먼저, IBM SPSS Modeler로 데이터를 읽어들입니다.
- 그 다음에는 일련의 조작을 통해 데이터를 실행합니다.
- 마지막으로, 데이터를 대상으로 보냅니다.

각 조작을 통해 소스에서 마침내 대상(모델 또는 데이터 출력 유형)으로 레코드별로 데이터가 플로우되기 때문에 이러한 작업 시퀀스를 **데이터 스트림**이라고 합니다.

① IBM SPSS Modeler 스트림 캔버스

스트림 캔버스는 IBM® SPSS® Modeler 창의 가장 큰 영역이며 데이터 스트림을 작성하고 조작하는 위치입니다.

스트림은 인터페이스의 기본 캔버스에서 비즈니스 관련 데이터 연산 다이어그램을 그려서 작성됩니다. 각 연산은 아이콘 또는 **노드**로 표시되며, 노드는 각 연산을 통한 데이터 플로우를 표시하는 **스트림**에서 함께 연결됩니다.

동일한 스트림 캔버스에서 또는 새 스트림 캔버스를 열어 IBM SPSS Modeler에서 동시에 다중 스트림에 대한 작업을 할 수 있습니다. 세션 동안 스트림은 IBM SPSS Modeler 창의 오른쪽 상단에 있는 스트림 관리자에 저장됩니다.

> ⓘ **참고:** 기본 제공되는 트랙 패드의 **포스클릭 및 햅틱 피드백** 설정이 사용 가능한 상태에서 MacBook을 사용하는 경우, 노드 팔레트에서 스트림 캔버스로 노드를 끌어다 놓으면 중복 노드가 캔버스에 추가됩니다. 이 문제를 방지하려면 **포스클릭 및 햅틱 피드백** 트랙 패드 시스템 환경 설정을 사용할 수 없음으로 지정하도록 권장합니다.

② 노드 팔레트

SPSS® Modeler에서 대다수의 데이터와 모델링 도구는 스트림 캔버스 아래의 창 아래에 걸쳐서 노드 팔레트에서 사용할 수 있습니다.

예를 들어, **레코드 Ops** 팔레트 탭에는 선택, 병합 및 붙여쓰기 등과 같이 데이터 레코드에서 작업을 수행하는 데 사용할 수 있는 노드가 포함됩니다.

노드를 캔버스에 추가하려면 노드 팔레트에서 아이콘을 두 번 클릭하거나 이를 캔버스로 끄십시오. 그런 다음 이들을 연결하여 데이터 플로우를 나타내는 스트림을 작성할 수 있습니다.

그림 1. 노드 팔레트에서 레코드 Ops 탭

각 팔레트 탭에는 다음과 같이 스트림 작업의 서로 다른 단계에 사용하는 관련 노드의 콜렉션이 포함됩니다.

- **소스** 노드는 데이터를 SPSS Modeler로 가져옵니다.
- **레코드 작업** 노드는 선택, 병합 및 붙여쓰기 등의 데이터 레코드 작업을 수행합니다.
- **필드 작업** 노드는 필터링, 새 필드 파생 및 주어진 필드의 측정 수준 판별 등의 데이터 필드 작업을 수행합니다.
- **그래프** 노드는 모델링 전후 데이터를 그래프로 표시합니다. 그래프에는 도표, 히스토그램, 웹 노드 및 평가 차트를 포함합니다.
- **모델링** 노드는 신경망, 의사결정 트리, 군집 알고리즘 및 데이터 순차규칙 등과 같이 SPSS Modeler에서 사용 가능한 모델링 알고리즘을 사용합니다.
- **데이터베이스 모델링** 노드는 Microsoft SQL Server, IBM DB2 및 Oracle과 Netezza 데이터베이스에서 사용 가능한 모델링 알고리즘을 사용합니다.
- **출력** 노드는 SPSS Modeler에서 볼 수 있는 데이터, 차트 및 모델 결과의 다양한 출력을 생성합니다.
- **내보내기** 노드는 IBM® SPSS Data Collection 또는 Excel과 같은 외부 애플리케이션에서 볼 수 있는 다양한 출력을 생성합니다.
- **Python** 노드는 Python 알고리즘을 실행하는 데 사용할 수 있습니다.
- **Spark** 노드는 Spark 알고리즘을 실행하는 데 사용할 수 있습니다.

SPSS Modeler와 친숙해짐에 따라 원하는 대로 팔레트 내용을 사용자 정의할 수 있습니다. 자세한 정보는 노드 팔레트 사용자 정의의 내용을 참조하십시오.

노드 팔레트의 왼쪽에서는 감독, 연관 또는 세분화를 선택하여 표시되는 노드를 필터링할 수 있습니다. 자세한 정보는 모델링 노드의 개요의 내용을 참조하십시오.

노드 팔레트 아래에 있는 보고서 분할창은 데이터를 데이터 스트림으로 읽어 올 때 등과 같이 다양한 작업의 진행률에 대한 피드백을 제공합니다. 마찬가지로 노드 팔레트 아래에 있는 상태 분할창은 애플리케이션이 현재 하는 작업뿐만 아니라 사용자 피드백이 필요한 시기 표시에 대한 정보를 제공합니다.

> ⓘ **참고:** 기본 제공되는 트랙 패드의 **포스클릭 및 햅틱 피드백** 설정이 사용 가능한 상태에서 MacBook을 사용하는 경우, 노드 팔레트에서 스트림 캔버스로 노드를 끌어다 놓으면 중복 노드가 캔버스에 추가됩니다. 이 문제를 방지하려면 **포스클릭 및 햅틱 피드백** 트랙 패드 시스템 환경 설정을 사용할 수 없음으로 지정하도록 권장합니다.

③ IBM SPSS Modeler 관리자

창의 오른쪽 상단은 관리자 분할창입니다. 여기에는 스트림, 출력, 모델을 관리하는 데 사용하는 세 개의 탭이 있습니다.

스트림 탭을 사용하여 세션에 작성된 스트림을 열고 이름을 변경하며, 저장 및 삭제할 수 있습니다.

출력 탭은 IBM® SPSS® Modeler에서 스트림 작업을 통해 생성된 그래프, 테이블과 같은 다양한 파일을 포함합니다. 이 탭에 나열된 테이블, 그래프, 보고서를 표시하고 저장하며, 이름을 변경하고 닫을 수 있습니다.

모델 탭은 관리자 탭 중 가장 강력합니다. 이 탭은 현재 세션에 대해 IBM SPSS Modeler에 생성된 모델을 포함하는 모든 모델 **너깃**을 포함합니다. 이러한 모델을 모델 탭에서 직접 찾아보거나 캔버스의 스트림에 추가할 수 있습니다.

④ IBM SPSS Modeler 프로젝트

창의 오른쪽 하단은 데이터 마이닝 **프로젝트**(데이터 마이닝 작업과 관련된 파일 그룹)를 작성하고 관리하는 데 사용하는 프로젝트 분할창입니다. IBM® SPSS® Modeler에서 사용자가 작성하는 프로젝트를 볼 수 있는 두 가지 방법이 있는데, 클래스 보기와 CRISP-DM 보기입니다.

CRISP-DM 탭은 업계에서 입증된 비독점 방법론인 CRISP-DM(Cross-Industry Standard Process for Data Mining)에 따라 프로젝트를 구성하는 방법을 제공합니다. 숙련된 데이터 마이너와 초보 데이터 마이너 모두를 위해 CRISP-DM 도구를 사용하면 데이터 마이닝을 위한 노력을 쉽게 체계화하고 전달할 수 있습니다.

클래스 탭은 IBM SPSS Modeler에서 작업을 범주적으로(작성하는 오브젝트 유형별로) 구성하는 방법을 제공합니다. 이 보기는 데이터, 스트림, 모델 인벤토리를 만들 때 유용합니다.

⑤ IBM SPSS Modeler 도구 모음

IBM® SPSS® Modeler 창의 맨 위에서 다양한 유용한 기능을 제공하는 아이콘 도구 모음을 볼 수 있습니다. 다음은 도구 모음 단추 및 해당 함수입니다.

	새 스트림 작성		스트림 열기
	스트림 저장		현재 스트림 인쇄
	잘라내기 & 클립보드로 이동		클립보드에 복사
	선택 붙여넣기		마지막 작업 실행 취소
	다시 실행		노드 검색
	스트림 특성 편집		SQL 생성 미리보기
	현재 스트림 실행		스트림 선택 실행
	스트림 중지(스트림이 실행 중인 동안에만 활성)		수퍼노드 추가
	확대(수퍼노드 전용)		축소(수퍼노드 전용)
	스트림에 마크업 없음		주석 삽입
	스트림 마크업 숨김(있는 경우)		숨겨진 스트림 마크업 표시
	IBM SPSS Modeler Advantage에서 스트림 열기		

스트림 마크업은 스트림 주석, 모델 링크 및 스코어링 분기 표시로 구성됩니다.

⑥ 도구 모음 사용자 정의

다음과 같이 도구 모음의 다양한 측면을 변경할 수 있습니다.

- 표시되는지 여부
- 아이콘에 사용 가능한 도구 팁이 있는지 여부
- 큰 아이콘 또는 작은 아이콘을 사용하는지 여부

도구 모음 표시를 켜고 끄기

1. 기본 메뉴에서 다음을 클릭하십시오.
   보기 > 도구 모음 > 표시

도구 팁이나 아이콘 크기 설정 변경

1. 기본 메뉴에서 다음을 클릭하십시오.
   보기 > 도구 모음 > 사용자 정의

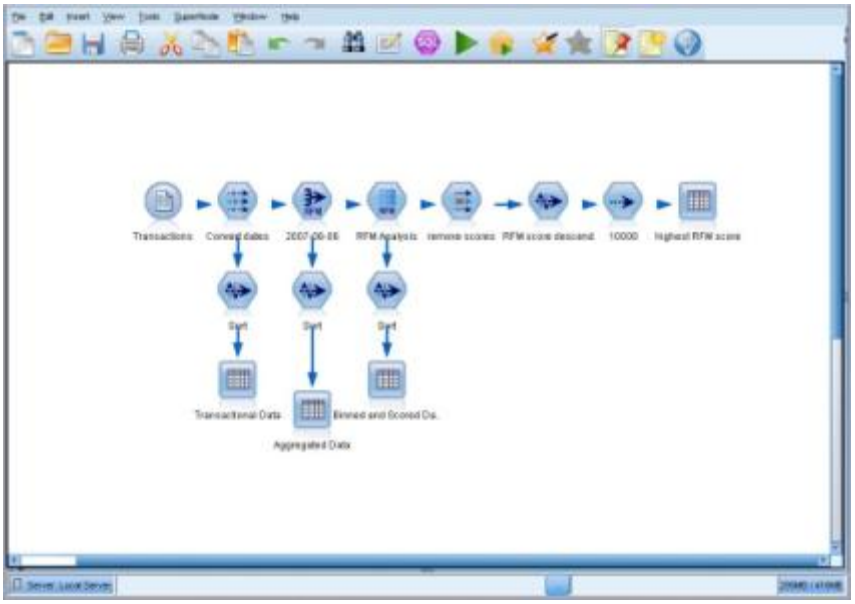
필요에 따라 **도구 모음 팁 표시** 또는 **큰 단추**를 클릭하십시오.

⑦ IBM SPSS Modeler 창 사용자 정의

SPSS® Modeler 인터페이스의 다양한 부분 사이에 디바이더를 사용할 때 기본 설정을 충족하도록 도구의 크기를 조정하거나 닫을 수 있습니다. 예를 들어, 대용량 스트림에 대해 작업 중인 경우 각 디바이더에 있는 작은 화살표를 사용하여 노드 팔레트, 관리자 분할창 및 프로젝트 분할창을 닫을 수 있습니다. 이는 스트림 캔버스를 최소화하여 대용량 또는 다량의 스트림을 위한 충분한 작업 공간을 제공합니다.

또는 보기 메뉴에서 **노드 팔레트**, **관리자** 또는 **프로젝트**를 클릭하여 이러한 항목을 켜거나 끌 수도 있습니다.

그림 1. 최대 스트림 캔버스



노드 팔레트, 관리자 및 프로젝트 분할창을 닫는 대신에 SPSS Modeler 창의 측면과 아래쪽에 있는 스크롤바를 가로 세로로 이동하여 스트림 캔버스를 스크롤 가능한 페이지로 사용할 수 있습니다.

또한 스트림 설명, 모델 링크 및 스코어링 분기 표시로 구성된 화면 마크업 표시를 제어할 수도 있습니다. 이 표시를 켜거나 끄려면 다음을 클릭하십시오.

보기 > 스트림 마크업

⑧ 스트림 아이콘 크기 변경

다음 방법으로 스트림 아이콘 크기를 변경할 수 있습니다.

- 스트림 특성 설정을 통해
- 스트림의 팝업 메뉴를 통해
- 키보드 사용

전체 스트림 보기를 표준 아이콘 크기의 8%-200% 사이의 많은 크기 중 하나로 스케일링할 수 있습니다.

전체 스트림 스케일링 방법(스트림 특성 방법)

1. 메뉴에서 다음을 선택하십시오.
   도구 > 스트림 특성 > 옵션 > 레이아웃
2. 아이콘 크기 메뉴에서 원하는 크기를 선택하십시오.
3. **적용**을 클릭하여 결과를 확인하십시오.
4. **확인**을 클릭하여 변경사항을 저장하십시오.

전체 스트림 스케일링 방법(메뉴 방법)

1. 캔버스에서 스트림 배경을 마우스 오른쪽 단추로 클릭하십시오.
2. 아이콘 크기를 선택하고 원하는 크기를 선택하십시오.

전체 스트림 스케일링 방법(키보드 방법)

1. 기본 키보드에서 Ctrl + [-]를 눌러 크기를 축소하십시오.
2. 기본 키보드에서 Ctrl + Shift + [-]를 눌러 크기를 확대하십시오.

이 방식의 확대는 사용하는 운영 체제 및 키보드에 따라 작동하지 않을 수 있습니다.

이 기능은 복합 스트림에 대한 전체 보기를 확보할 경우 특히 유용합니다. 또한 이를 사용하여 스트림을 인쇄하는 데 필요한 페이지 수를 최소화할 수 있습니다.

⑨ IBM SPSS Modeler에서 마우스 사용

IBM® SPSS® Modeler에서 마우스의 가장 흔한 용도는 다음과 같습니다.

- **단일 클릭.** 마우스 오른쪽이나 왼쪽 단추를 사용하여 메뉴에서 옵션을 선택하고, 팝업 메뉴를 열고, 다양한 기타 표준 제어와 옵션에 액세스할 수 있습니다. 노드를 이동하거나 끌려면 단추를 클릭한 후 계속 누르십시오.
- **두 번 클릭.** 마우스 왼쪽 단추를 두 번 클릭하여 노드를 스트림 캔버스에 놓고 기존 노드를 편집하십시오.
- **가운데 클릭.** 마우스 가운데 단추를 클릭하여 커서를 끌어서 스트림 캔버스에 있는 노드를 연결하십시오. 노드를 연결 해제하려면 마우스 가운데 단추를 두 번 클릭하십시오. 버튼이 3개인 마우스가 아닌 경우에는 마우스를 클릭하고 끄는 중에 Alt 키를 눌러서 이 기능을 시뮬레이션할 수 있습니다.

⑩ 단축키 사용

IBM® SPSS® Modeler의 많은 비주얼 프로그래밍 작업에는 연관된 단축키가 있습니다. 예를 들어, 노드를 클릭하고 키보드에서 삭제 키를 눌러 노드를 삭제할 수 있습니다. 마찬가지로, Ctrl 키를 누른 상태에서 S 키를 눌러서 스트림을 빠르게 저장할 수 있습니다. 이와 같은 제어 명령은 Ctrl 및 다른 키 조합(예: Ctrl+S)으로 표시됩니다.

표준 Windows 작업에 사용하는 많은 단축키가 있습니다(예: 잘라내기는 Ctrl+X). 이러한 단축키는 다음 애플리케이션 특정 단축키와 함께 IBM SPSS Modeler에서 지원됩니다.

> ⓘ **참고:** 어떤 경우에는 IBM SPSS Modeler에서 사용한 이전 단축키가 표준 Windows 단축키와 충돌할 수 있습니다. 이러한 이전 단축키는 Alt 키의 추가와 함께 지원됩니다. 예를 들어, Ctrl+Alt+C는 캐시를 켜기/끄기를 토글하는 데 사용할 수 있습니다.

표 1. 지원되는 단축키

단축키	기능
Ctrl+A	모두 선택
Ctrl+X	잘라내기
Ctrl+N	새 스트림
Ctrl+O	스트림 열기
Ctrl+P	인쇄
Ctrl+C	복사
Ctrl+V	붙여넣기
Ctrl+Z	실행 취소

단축키	기능
Ctrl+Q	선택한 노드의 모든 노드 다운스트림 선택
Ctrl+W	모든 다운스트림 노드 선택 취소(Ctrl+Q로 토글)
Ctrl+E	선택한 노드에서 실행
Ctrl+S	현재 스트림 저장
Alt+화살표 키	선택한 노드를 캔버스 스트림에서 사용된 화살표의 방향으로 이동
Shift+F10	선택한 노드의 팝업 메뉴 열기

표 2. 이전 단축키의 지원되는 단축키

단축키	기능
Ctrl+Alt+D	노드 중복
Ctrl+Alt+L	노드 로드
Ctrl+Alt+R	노드 이름 변경
Ctrl+Alt+U	사용자 입력 노드 작성
Ctrl+Alt+C	캐시 켜기/끄기 토글
Ctrl+Alt+F	캐시 비우기
Ctrl+Alt+X	수퍼노드 확장
Ctrl+Alt+Z	확대/축소
Delete	노드나 연결 삭제

(4) 인쇄

다음 오브젝트를 IBM® SPSS® Modeler에서 인쇄할 수 있습니다.

- 스트림 다이어그램
- 그래프
- 테이블
- 보고서(보고서 노드 및 프로젝트 보고서에서)
- 스크립트(스트림 특성, 독립형 스크립트 또는 수퍼노드 스크립트 대화 상자에서)
- 모델(모델 브라우저, 현재 초점이 있는 대화 상자 탭, 트리 뷰어)
- 주석(출력을 위해 주석 탭 사용)

오브젝트 인쇄

- 미리보지 않고 인쇄하려면 도구 모음에서 인쇄 단추를 클릭하십시오.
- 인쇄하기 전에 페이지를 설정하려면 파일 메뉴에서 **페이지 설정**을 선택하십시오.
- 인쇄 전에 미리 보려면 파일 메뉴에서 **인쇄 미리보기**를 선택하십시오.
- 프린터 선택을 위한 옵션이 있는 표준 인쇄 대화 상자를 보려면 파일 메뉴에서 **인쇄**를 선택하십시오.

(5) IBM SPSS Modeler 자동화

고급 데이터 마이닝은 복잡하고 때로는 시간이 긴 프로세스일 수 있으므로 IBM® SPSS® Modeler에는 몇몇 코딩 유형 및 자동화 지원이 포함됩니다.

- **CLEM(**Control Language for Expression Manipulation)은 IBM SPSS Modeler 스트림을 따라 플로우하는 데이터를 분석하고 조작하기 위한 언어입니다. 데이터 마이너는 비용과 수입 데이터에서 수익을 파생하는 것과 같은 단순한 작업 또는 웹 로그 데이터를 유용한 정보가 있는 필드와 레코드 세트로 변환하는 것과 같은 복잡한 작업을 수행하기 위해 스트림 작업에서 CLEM을 집중적으로 사용합니다.
- **스크립팅**은 사용자 인터페이스에서 프로세스를 자동화하기 위한 강력한 도구입니다. 스크립트는 사용자가 마우스나 키보드를 사용하여 수행하는 것과 동일한 유형의 조치를 수행합니다. 또는 출력을 지정하고 생성된 모델을 조작할 수도 있습니다.

3) 데이터 마이닝 이해

(1) 모델 유형

IBM® SPSS® Modeler는 기계 학습, 인공지능 및 통계로부터 취한 다양한 모델링 방법을 제공합니다. 모델링 팔레트에서 사용할 수 있는 이러한 방법을 통해 데이터로부터 새로운 정보를 얻어서 예측 모형을 개발할 수 있습니다. 각각의 방법은 그것만의 장점이 있으며 특정한 문제점 유형에 가장 적합합니다.

*IBM SPSS Modeler 애플리케이션 안내서*에서는 모델링 프로세스에 대한 일반적인 소개와 함께 이러한 여러 방법의 예제를 제공합니다. 이 안내서는 온라인 자습서로 사용 가능합니다. 자세한 정보.

모델링 방법은 다음 범주로 나뉩니다.

- 감독
- 연관
- 세분화

감독 모델

*감독 모델*은 하나 이상의 출력 또는 **목표** 필드를 예측하기 위해 하나 이상의 **입력** 필드 값을 사용합니다. 이러한 기술의 일부 예는 다음과 같습니다. 의사결정 트리(C&R 트리, QUEST, CHAID 및 C5.0 알고리즘), 회귀분석(1차, 로지스틱, 일반화 선형 및 Cox 회귀 알고리즘), 신경망, 지원 벡터 머신 및 베이지안 네트워크입니다.

감독 모델은 조직이 예를 들어, 고객이 구매할지 또는 떠날지 여부 또는 트랜잭션이 알려진 사기 패턴과 매치하는지 여부 등과 같이 알려진 결과를 예측하는 데 도움을 줍니다. 모델링 기법은 시스템 학습, 규칙 귀납, 하위 그룹 식별, 통계 방법 및 다중 모델 생성을 포함합니다.

감독 노드

	자동 분류자 노드는 이분형 결과(예 또는 아니오, 이탈 또는 이탈 안함 등)에 대해 다수의 여러 모델을 작성하고 비교하여 주어진 분석을 위한 최상의 접근 방식을 선택할 수 있게 합니다. 많은 모델링 알고리즘이 지원되어 사용할 방법, 각각에 대한 특정 옵션, 결과 비교 기준을 선택할 수 있습니다. 이 노드는 지정된 옵션을 기반으로 모델 세트를 생성하고 사용자가 지정하는 기준에 따라 최상의 후보를 순위화합니다.
	자동 수치 노드는 수많은 방법을 사용하여 연속적 수치 범위 결과의 모델을 추정하고 비교합니다. 이 노드는 자동 분류자 노드에서와 같은 방식으로 작동하므로 사용할 알고리즘을 선택하고 단일 모델링 전달에서 여러 옵션의 조합을 실험할 수 있습니다. 지원되는 알고리즘에는 신경망, C&R 트리, CHAID, 선형 회귀, 일반화 선형 회귀 및 지원 벡터 머신(SVM)이 있습니다. 모델은 상관관계, 상대 오차 또는 사용된 변수의 수를 기반으로 비교할 수 있습니다.
	분류 및 회귀(C&R) 트리 노드는 추가 관측값을 예측하거나 분류할 수 있게 하는 의사결정 트리를 생성합니다. 이 방법은 재귀적 파티셔닝을 사용하여 각 단계마다 불순도를 최소화하여 학습 레코드를 세그먼트로 분할합니다. 여기서 트리의 노드는 노드의 케이스의 100%가 대상 필드의 특정 범주에 속하면 "순수"로 간주됩니다. 목표 및 입력 필드는 숫자 범위 또는 범주형(명목형, 순서형 또는 플래그)입니다. 모든 분할은 이분형입니다(오직 두 개의 부집단).
	QUEST 노드는 의사결정 트리를 작성하기 위한 이분형 분류 방법을 제공하며, 대형 C&R 트리 분석에 필요한 처리 시간을 줄이는 동시에 분류 트리 방법에서 찾은 경향을 줄여 더 많은 분할을 허용하는 입력을 선호하도록 설계되었습니다. 입력 필드는 숫자 범위(연속)일 수 있지만 대상 필드는 범주형이어야 합니다. 모든 분할은 이분형입니다.
	CHAID 노드는 최적 분할을 식별하기 위해 카이제곱 통계량을 사용하여 의사결정 트리를 생성합니다. C&R 트리 및 QUEST 노드와 달리, CHAID는 비이분형 트리를 생성할 수 있으며 이는 일부 분할에 둘 이상의 분기가 있음을 의미합니다. 목표 및 입력 필드는 숫자 범위(연속형) 또는 범주형입니다. Exhaustive CHAID는 가능한 모든 분할을 탐색하는 보다 철저한 작업을 수행하지만 계산하는 데 시간이 더 걸리는 변형 CHAID입니다.

C5.0 노드는 의사결정 트리 또는 규칙 세트를 작성합니다. 모델은 각 수준에서 최대 정보 이익을 제공하는 필드를 기반으로 샘플을 분할하여 작동합니다. 대상 필드는 범주형이어야 합니다. 세 개 이상의 부집단으로의 다중 분할이 허용됩니다.

의사결정 목록 노드는 전체 채우기에 상대적인 주어진 이분형 결과의 상위 또는 하위 우도를 표시하는 부집단 또는 세그먼트를 식별합니다. 예를 들어, 캠페인을 이탈할 가능성이 없거나 우호적으로 응답할 가능성이 가장 많은 고객을 찾고 있습니다. 자체 사용자 정의 세그먼트를 추가하고 대체 모델을 나란히 미리보기하여 결과를 비교함으로써 비즈니스 지식을 모델에 통합할 수 있습니다. 의사결정 목록 모델은 각 규칙에 조건과 결과가 있는 규칙 목록으로 구성됩니다. 규칙은 순서대로 적용되며 매치하는 첫 번째 규칙이 결과를 결정합니다.

선형 회귀 모형은 목표와 하나 이상의 예측변수 간의 선형 관계를 기반으로 연속형 목표를 예측합니다.

PCA/요인 노드에서는 강력한 데이터 축소 기법을 제공하여 데이터의 복잡도를 줄입니다. 비선형 주성분분석(PCA)은 구성요소가 서로 직각(수직)인 전체 필드 세트에서 변동을 캡처하는 입력 필드의 선형 조합을 찾습니다. 요인 분석은 관측된 필드 세트 내에서 상관관계 패턴을 설명하는 기본 요인을 식별하려고 시도합니다. 두 접근 방식 모두 목표는 원래 필드 세트의 정보를 효과적으로 요약하는 적은 수의 파생 필드를 찾는 것입니다.

필드선택 노드는 기준(예: 결측값의 퍼센트) 세트를 기반으로 제거용 입력 필드를 차단합니다. 그런 다음 지정된 대상에 상대적인 남아 있는 입력의 중요도에 대해 순위를 매깁니다. 예를 들어, 수백 개의 잠재 입력이 있는 데이터 세트가 있다면 환자 결과 모델링 시 어느 것이 가장 유용합니까?

판별 분석은 로지스틱 회귀분석보다 엄격한 가정을 하지만 해당 가정이 충족되면 로지스틱 회귀 분석의 귀중한 대안 또는 보조물이 될 수 있습니다.

로지스틱 회귀분석은 입력 필드 값을 기반으로 레코드를 분류하는 통계 기법입니다. 선형 회귀와 유사하지만 숫자 범위 대신 범주형 대상 필드를 사용합니다.

일반화 선형 모델은 종속변수가 요인과 선형적으로 관련되고 지정된 연결함수를 통해 공변되도록 일반 선형 모델을 확장합니다. 더욱이 모델을 사용하면 종속변수가 비정규 분포를 가질 수 있습니다. 선형 회귀, 로지스틱 회귀분석, 카운트 데이터에 대한 로그선형 모델, 간격 중도절단 생존 모델을 포함하여 상당수 통계 모델의 기능을 포함합니다.

일반화 선형 혼합 모델(GLMM)은 목표가 비정규 분포를 가질 수 있고 지정된 연결함수를 통해 요인 및 공변량과 선형적으로 관련되며 관측값을 상관시킬 수 있도록 선형 모델을 확장합니다. 일반화 선형 혼합 모델은 단순 선형 회귀에서 비정규 장기적인 데이터에 대한 복합 다중 수준 모델에 이르기까지 다양한 모델을 포함합니다.

	Cox 회귀 노드를 통해 중도절단된 레코드가 있는 데서 시간 대 이벤트 데이터에 대한 생존 모델을 작성할 수 있습니다. 이 모델은 주어진 입력 변수 값에 대해 주어진 시간(t)에 흥미있는 이벤트가 발생한 확률을 예측하는 생존함수를 생성합니다.
	지원 벡터 머신(SVM) 노드를 사용하면 데이터를 과적합 없이 두 개의 그룹 중 하나로 분류할 수 있습니다. SVM은 다수의 입력 필드가 있는 데이터 세트 등과 같은 광범위한 데이터 세트와 잘 작동합니다.
	베이지안 네트워크 노드를 통해 관측 및 레코드된 증거를 실세계 지식과 조합하여 발생 우도를 확립함으로써 확률 모델을 작성할 수 있습니다. 이 노드는 주로 분류에 사용하는 TAN(Tree Augmented Naïve Bayes) 및 Markov Blanket 네트워크에 초점을 맞춥니다.
	SLRM(Self-Learning Response Model) 노드를 사용하면 하나의 새 케이스 또는 소수의 새 케이스를 사용하여 모든 데이터를 사용하는 모델을 다시 학습시킬 필요 없이 모델을 재평가할 수 있는 모델을 작성할 수 있습니다.
	시계열 노드는 시계열 데이터에 대한 지수평활, 일변량 자기회귀 통합 이동 평균(ARIMA), 다변량 ARIMA(또는 전이 함수) 모델을 추정하고 미래 성능을 위한 예측값을 생성합니다. 이 시계열 노드는 SPSS Modeler 버전 18에서 더 이상 사용되지 않는 이전의 시계열 노드와 유사합니다. 그러나 이 새 시계열 노드는 IBM SPSS Analytic Server의 기능을 이용하여 빅 데이터를 처리해서 SPSS Modeler 버전 17에 추가된 출력 뷰어에 결과 모델을 표시하도록 설계되었습니다.
	KNN(k-Nearest Neighbor) 노드는 새 케이스를 k가 정수인 예측자 공간에서 가장 가까이에 있는 k 오브젝트의 범주 또는 값과 연관시킵니다. 유사한 케이스는 서로 가까이에 있고 유사하지 않은 케이스는 서로 멀리 떨어져 있습니다.
	STP(Spatio-Temporal Prediction) 노드는 위치 데이터, 예측(예측자)을 위한 입력 필드, 시간 필드 및 대상 필드를 포함하는 데이터를 사용합니다. 각 위치에는 각 측정 시간에 각 예측변수의 값을 나타내는 데이터에 여러 행이 있습니다. 데이터가 분석된 후에는 분석에 사용된 모양 데이터 내에서 어떤 위치에서든 목표 값을 예측하는 데 사용할 수 있습니다.

연관 모델

*연관 모델*은 하나 이상의 엔티티(예: 이벤트, 구매 또는 속성)가 하나 이상의 다른 엔티티와 연관되어 있는 데이터에서 패턴을 발견합니다. 모델은 이러한 관계를 정의하는 규칙 세트를 구성합니다. 여기에서 데이터 내의 필드는 입력과 목표 둘 모두의 역할을 할 수 있습니다. 이러한 연관을 수동으로 찾을 수 있지만 연관 규칙 알고리즘은 이를 보다 신속하게 수행하므로 더 복잡한 패턴을 탐색할 수 있습니다. Apriori 및 Carma 모델은 이러한 알고리즘 사용의 예입니다. 연관 모델의 또 다른 유형은 순차 발견 모델이며 이는 시간 구조 데이터에서 순차 패턴을 발견합니다.

연관 모델은 다중 결과를 예측할 때 가장 유용합니다(예: 제품 X를 구매한 고객이 Y와 Z도 구매

함). 연관 모델은 특정 결론(예: 구매 결정)을 조건 세트와 연관시킵니다. 다른 표준 의사결정 트리 알고리즘(C5.0 및 C&RT)에 비해 연관 규칙 알고리즘의 장점은 어떤 속성 사이에도 연관이 있을 수 있다는 점입니다. 의사결정 트리 알고리즘은 단일 결론만 포함하는 규칙을 작성하지만, 연관 알고리즘은 각각 다른 결론을 보유할 수 있는 많은 규칙을 찾으려고 합니다.

연관 노드

	Apriori 노드는 데이터에서 규칙 세트를 추출하고 정보 내용이 가장 많은 규칙을 꺼냅니다. Apriori는 규칙을 선택하는 5개의 서로 다른 방법을 제공하며 정교한 색인화 스킴을 사용하여 대형 데이터 세트를 효율적으로 처리합니다. 큰 문제점의 경우, Apriori는 일반적으로 학습 속도가 빠릅니다. 보유할 수 있는 규칙 수에 임의 제한이 없으며 최대 32개의 전제조건을 가진 규칙을 처리할 수 있습니다. Apriori에서는 입력 및 출력 필드가 모두 범주형이어야 하지만 이런 유형의 데이터에 최적화되어 있기 때문에 우수한 성능을 제공합니다.
	CARMA 모델은 입력 또는 대상 필드를 지정하지 않아도 데이터에서 규칙 세트를 추출합니다. Apriori와 대조적으로 CARMA 노드는 단지 전항 지원이 아니라 규칙 지원(전항 및 후항 둘 다에 대한 지원)을 위한 작성 설정을 제공합니다. 이는 생성된 규칙을 보다 다양한 애플리케이션에 사용하여, 예를 들어 후항이 이번 휴가철에 홍보할 항목인 제품 또는 서비스 목록을 찾을 수 있음을 의미합니다.
	순차규칙 노드는 순차 또는 시간 지향 데이터에서 연관 규칙을 발견합니다. 순차규칙은 예측 가능한 순서로 발생하는 경향이 있는 항목 세트 목록입니다. 예를 들어, 면도기와 애프터쉐이브 로션을 구매하는 고객은 다음 번 구매 시에 면도용 크림을 구매할 수 있습니다. 순차규칙 노드는 순차규칙을 찾는 데 효율적인 2패스 방법을 사용하는 CARMA 연관 규칙 알고리즘을 기반으로 합니다.
	연관 규칙 노드는 Apriori 노드와 유사합니다. 그러나 Apriori와는 달리, 연관 규칙 노드는 목록 데이터를 처리할 수 있습니다. 또한, 연관 규칙 노드는 빅 데이터를 처리하고 더 빠른 병렬 처리를 사용하기 위해 IBM SPSS Analytic Server와 함께 사용할 수 있습니다.

세분화 모델

*세분화 모델*은 데이터를 입력 필드의 패턴이 유사한 레코드의 세그먼트 또는 군집으로 나눕니다. 이들은 입력 필드에만 관심이 있으므로 세분화 모델에는 출력이나 목표 필드와 같은 개념이 없습니다. 세분화 모델의 예는 코호넨 네트워크, K-평균 군집, 이단계 군집 및 이상 항목 발견입니다.

세분화 모델("군집 모델"로도 알려짐)은 특정 결과가 알려지지 않은 경우에 유용합니다. 예를 들어, 새로운 사기 패턴을 식별할 때나, 고객 기반에 있는 관심 그룹을 식별할 때입니다. 군집 모델은 유사한 레코드 그룹을 식별하고 레코드에 이들이 속하는 그룹에 따라 레이블을 붙이는 데 초점을 둡니다. 이는 그룹과 그룹의 특성에 대한 사전 지식 없이도 수행되고, 예측할 모델에 대한 사전에 정의된 출력이나 목표 필드가 없다는 면에서 군집 모델을 다른 모델링 기법과 구별해 줍니다. 이러한 모델에는 옳고 그른 응답이 없습니다. 이들 값은 데이터에서 관심 그룹을 캡처하

는 기능에 의해 결정되고 이러한 그룹에 대한 유용한 설명을 제공합니다. 군집 모델은 종종 군집이나 세그먼트를 작성하는 데 사용하고 이러한 군집이나 세그먼트는 이후의 분석에서 입력으로 사용합니다(예: 잠재 고객을 동종 하위그룹으로 분할하는 방법으로).

세분화 노드

	자동 군집 노드는 유사한 특성을 가진 레코드 그룹을 식별하는 군집 모델을 추정하고 비교합니다. 이 노드는 다른 자동 모델링 노드와 동일한 방법으로 작동하여 단일 모델링 패스에서 다중 옵션 조합을 실험할 수 있습니다. 군집 모델의 유용성을 필터링하고 순위화하며 특정 필드의 중요성을 기반으로 측도를 제공하려고 시도하는 기본 측도를 사용하여 모델을 비교할 수 있습니다.
	K-평균 노드는 데이터 세트를 고유 그룹(또는 군집)으로 군집화합니다. 이 방법은 고정된 수의 군집을 정의하고 반복적으로 레코드를 군집에 지정하며, 추가 세분화가 더 이상 모델을 향상시킬 수 없을 때까지 군집중심을 조정합니다. 결과를 예상하는 대신 k-평균은 자율 학습으로 알려진 프로세스를 사용하여 입력 필드 세트의 패턴을 찾아냅니다.
	코호넨 노드는 데이터 세트를 고유 그룹으로 군집화하는 데 사용할 수 있는 신경망 유형을 생성합니다. 네트워크가 완전히 숙달되면, 유사 레코드는 출력 맵 가까이 있지만, 다른 레코드는 멀리 떨어져 있을 것입니다. 모델 너깃에서 각 단위별로 캡처된 관측값을 살펴 강한 단위를 식별할 수 있습니다. 이것은 적당한 군집 수에 대한 감각을 제공할 것입니다.
	이단계 노드는 2단계 군집방법을 사용합니다. 첫 번째 단계는 원시 입력 데이터를 관리 가능한 하위 군집 세트로 압축하기 위해 데이터를 통한 단일 전달을 수행합니다. 두 번째 단계는 계층적 군집 방법을 사용하여 하위 군집을 점점 더 큰 군집으로 계속해서 병합하는 것입니다. 이단계는 학습 데이터에 대한 최적 군집 수를 자동으로 평가하는 장점이 있습니다. 혼합 필드 유형과 대형 데이터 세트를 효율적으로 처리할 수 있습니다.
	이상 항목 발견 노드는 "정상" 데이터 패턴을 따르지 않는 특이 케이스 또는 이상값을 식별합니다. 이 노드를 사용하면 이전에 알려진 패턴에 적합하지 않고, 찾고 있는 패턴을 정확하게 모르더라도 이상값을 식별할 수 있습니다.

In-Database 마이닝 모델

IBM SPSS Modeler에서 Oracle Data Miner 및 Microsoft Analysis Services를 포함하여 데이터베이스 벤더에서 제공하는 데이터 마이닝과 모델링 도구의 통합을 지원합니다. 모두 IBM SPSS Modeler 애플리케이션 내에서 시작하여 데이터베이스 내에 모델을 작성하고, 스코어링하고, 저장할 수 있습니다. 자세한 정보는 *IBM SPSS Modeler In-Database 마이닝 안내서*를 참조하십시오.

(2) 데이터 마이닝 예제

실제로 데이터 마이닝에 대해 배우는 최상의 방법은 예제로 시작하는 것입니다. 특정 모델링 방법과 기술을 중점적으로 간략히 소개하는 *IBM® SPSS® Modeler 애플리케이션 안내서*에는 다양한 애플리케이션 예제가 포함되어 있습니다. 자세한 정보는 애플리케이션 예제의 내용을 참조하십시오.

4) 스트림 작성

(1) 스트림 작성 개요

IBM® SPSS® Modeler를 사용하는 데이터 마이닝은 **스트림**이라고 하는 일련의 노드를 통해 데이터를 실행하는 프로세스에 초점을 맞춥니다. 이 일련의 노드는 데이터에 대해 수행될 작업을 나타내는 반면, 노드 간의 링크는 데이터 플로우의 방향을 표시합니다. 일반적으로, 데이터 스트림을 사용하여 IBM SPSS Modeler로 데이터를 읽어들이고 일련의 조작을 통해 실행한 후 테이블이나 뷰어와 같은 대상으로 보냅니다.

예를 들어, 데이터 소스를 열고 새 필드를 추가하며, 새 필드의 값을 기반으로 레코드를 선택한 후 결과를 테이블에 표시하려 한다고 가정하십시오. 이 경우, 데이터 스트림은 다음 4개의 노드로 구성됩니다.

	가변파일 노드 - 데이터 소스에서 데이터를 읽도록 설정합니다.
	파생 노드 - 데이터 세트에 계산된 새 필드를 추가하는 데 사용합니다.
	선택 노드 - 데이터 스트림에서 레코드를 제외할 선택기준을 설정하는 데 사용합니다.
	테이블 노드 - 조작 결과를 화면에 표시하는 데 사용합니다.

(2) 데이터 스트림 작성

고유한 SPSS® Modeler 인터페이스를 사용하여 데이터 스트림의 다이어그램에 대해 작업함으로써 데이터를 시각적으로 마이닝할 수 있습니다. 가장 기본적인 수준에서 다음 단계를 사용하여 데이터 스트림을 작성할 수 있습니다.

- 노드를 스트림 캔버스에 추가하십시오.
- 스트림을 형성하도록 노드를 연결하십시오.
- 특정 노드 또는 스트림 옵션을 지정하십시오.
- 스트림을 실행하십시오.

이 절에는 노드를 사용하여 작업함으로써 보다 복잡한 데이터 스트림을 작성하는 것에 대한 자세한 정보가 포함되어 있습니다. 또한 노드 및 스트림에 대한 옵션 및 설정에 대해 논의합니다. 프로그램 설치의 Demos 폴더에) SPSS Modeler와 함께 제공되는 데이터를 사용한 스트림 작성의 단계별 예는 애플리케이션 예제의 내용을 참조하십시오.

① 노드에 대한 작업

노드는 IBM® SPSS® Modeler에서 데이터를 탐색하는 데 도움이 됩니다. 작업공간의 다양한 노드는 서로 다른 오브젝트와 조치를 나타냅니다. IBM SPSS Modeler 창의 맨 아래에 있는 팔레트에는 스트림 작성에 사용하는 가능한 모든 노드가 포함됩니다.

몇몇 유형의 노드가 있습니다. **소스 노드**는 데이터를 스트림으로 가져오고 노드 팔레트의 소스 탭에 있습니다. **프로세스 노드**는 개별 데이터 레코드와 필드에서 조작을 수행하고 팔레트의 레코드 작업과 필드 작업 탭에서 발견할 수 있습니다. **출력 노드**는 데이터, 차트 및 모델 결과의 다양한 출력을 생성하고 노드 팔레트의 그래프, 출력 및 내보내기 탭에 포함됩니다. **모델링 노드**는 통계적 알고리즘을 사용하여 모델 너깃을 작성하고 노드 팔레트의 모델링 탭 및 데이터베이스 모델링 탭(있는 경우)에 있습니다. 자세한 정보는 노드 팔레트의 내용을 참조하십시오.

노드를 연결하여 스트림을 형성할 수 있습니다. 스트림은 실행되면 관계를 시각화하고 결론을 그릴 수 있게 해줍니다. 스트림은 스크립트와 같아서 이를 저장하고 다른 데이터 파일과 함께 재사용할 수 있습니다.

스트림 데이터를 처리하는 실행 가능한 노드는 **터미널 노드**로 알려져 있습니다. 모델링 또는 출력 노드는 스트림이나 스트림 분기의 끝에 있으면 터미널 노드입니다. 터미널 노드에 추가로 노드를 연결할 수 없습니다.

> ⓘ **참고:** 노드 팔레트를 사용자 정의할 수 있습니다. 자세한 정보는 노드 팔레트 사용자 정의의 내용을 참조하십시오.

가. 스트림에 노드 추가

노드 팔레트에서 스트림에 노드를 추가하는 몇 가지 방법이 있습니다.

- 팔레트에서 노드를 두 번 클릭하십시오. *참고*: 노드를 두 번 클릭하면 현재 스트림에 노드가 자동으로 연결됩니다. 자세한 정보는 스트림의 노드 연결의 내용을 참조하십시오.

- 팔레트에서 스트림 캔버스로 노드를 끌어서 놓으십시오.
- 팔레트에서 노드를 클릭한 후 스트림 캔버스를 클릭하십시오.
- IBM® SPSS® Modeler의 삽입 메뉴에서 해당 옵션을 선택하십시오.

스트림 캔버스에 노드를 추가했으면 노드를 두 번 클릭하여 해당 대화 상자를 표시하십시오. 사용 가능한 옵션은 추가하고 있는 노드 유형에 따라 다릅니다. 대화 상자 내 특정 제어에 대한 정보는 해당 **도움말** 단추를 클릭하십시오.

노드 제거

데이터 스트림에서 노드를 제거하려면 노드를 클릭하고 Delete 키를 누르거나 마우스 오른쪽 단추를 클릭하고 메뉴에서 **삭제**를 선택하십시오.

나. 스트림의 노드 연결

스트림 캔버스에 추가된 노드는 연결될 때까지 데이터 스트림을 형성하지 않습니다. 노드 간의 연결은 하나의 작업에서 다른 작업으로 데이터가 플로우될 때 데이터의 방향을 표시합니다. 노드를 연결하여 스트림을 형성하는 많은 방법이 있는데, 두 번 클릭, 마우스 가운데 단추 사용 또는 수동으로 이를 수행할 수 있습니다.

두 번 클릭으로 노드 추가 및 연결 방법

스트림을 형성하는 가장 간단한 방법은 팔레트에서 노드를 두 번 클릭하는 것입니다. 이 방법은 새 노드를 스트림 캔버스에서 선택된 노드에 자동으로 연결합니다. 예를 들어, 캔버스에 데이터베이스 노드가 포함된 경우 이 노드를 선택한 후 팔레트에서 다음 노드(예: 파생 노드)를 두 번 클릭할 수 있습니다. 이 동작은 파생 노드를 기존 데이터베이스 노드에 자동으로 연결합니다. 마지막 비터미널 노드 업스트림에 새 노드가 연결되는 지점인 터미널 노드(예: 히스토그램 또는 테이블 노드)에 도달할 때까지 이 프로세스를 반복할 수 있습니다.

마우스 가운데 단추를 사용하여 노드를 연결하는 방법

스트림 캔버스에서 마우스 가운데 단추를 사용하여 한 노드를 다른 노드로 클릭하여 끌 수 있습니다(마우스에 가운데 단추가 없으면 한 노드에서 다른 노드로 마우스를 끄는 동안 Alt 키를 눌러 이를 시뮬레이션할 수 있음).

노드를 수동으로 연결하는 방법

마우스 가운데 단추가 없고 노드를 수동으로 연결하려면 노드 팝업 메뉴를 사용하여 노드를 캔버스에 이미 있는 다른 노드에 연결할 수 있습니다.

1. 연결을 시작할 노드를 마우스 오른쪽 단추로 클릭하십시오. 이렇게 하면 노드 메뉴가 열립니다.
2. 메뉴에서 **연결**을 클릭하십시오.
3. 시작 노드와 커서 둘 다에 연결 아이콘이 표시됩니다. 캔버스에서 또 하나의 노드를 클릭하여 두 노드를 연결하십시오.

노드 연결 시 따라야 할 몇 가지 지침이 있습니다. 다음 유형의 연결을 수행하려고 시도하면 오류 메시지를 수신합니다.

- 소스 노드로 이어지는 연결
- 터미널 노드에서 이어지는 연결
- 최대 입력 연결 수보다 많은 노드
- 이미 연결된 두 개의 노드 연결
- 순환(데이터가 이미 플로우된 노드에 리턴됨)

다. 스트림의 노드 우회

데이터 스트림의 노드를 우회하는 경우 모든 입력 및 출력 연결이 입력 노드에서 출력 노드로 직접 이끄는 연결로 대체됩니다. 노드에 입력 연결과 출력 연결이 둘 다는 없는 경우 재라우팅되지 않고 해당 연결이 모두 삭제됩니다.

예를 들어, 새 필드를 파생시키고 필드를 필터링한 후 히스토그램과 테이블에서 결과를 탐색하는 스트림이 있습니다. 필드가 필터링되기 전에 데이터에 대해서도 동일한 그래프와 테이블을 보려는 경우 새 히스토그램 및 테이블 노드를 스트림에 추가하거나 필터 노드를 우회할 수 있습니다. 필터 노드 우회 시 그래프 및 테이블로의 연결은 파생 노드에서 직접 전달됩니다. 필터 노드는 스트림에서 연결 해제됩니다.

노드 우회 방법

1. 스트림 캔버스에서 가운데 마우스 단추를 사용하여 우회할 노드를 두 번 클릭하십시오. 또는 Alt+두 번 클릭을 사용할 수 있습니다.

참고: 편집 메뉴에서 실행 취소를 클릭하거나 Ctrl+Z를 눌러 이 조치를 실행 취소할 수 있습니다.

라. 스트림의 노드 사용 안함

스트림 내에서 단일 입력이 있는 처리 노드를 사용 안함으로 설정할 수 있는데, 그 결과 스트림 실행 중 노드를 무시합니다. 이로 인해 노드를 제거하거나 우회하지 않아도 되며, 나머지 노드에

계속 연결할 수 있음을 의미합니다. 여전히 노드 설정을 열고 편집할 수 있습니다. 그러나 노드를 다시 사용 가능하게 해야 변경사항이 적용됩니다.

예를 들어, 몇 개의 필드를 필터링한 후 줄어든 데이터 세트로 모델을 작성하는 스트림이 있습니다. 필드를 필터링하지 않고 같은 모델을 작성하여 모델 결과가 향상되는지 여부를 확인하려고 할 경우 필터 노드를 사용 안함으로 설정할 수 있습니다. 필터 노드를 사용 안함으로 설정하면 모델링 노드와의 연결은 파생 노드에서 유형 노드로 직접 전달됩니다.

노드를 사용 안함으로 설정하는 방법

1. 스트림 캔버스에서 사용 안함으로 설정할 노드를 마우스 오른쪽 단추로 클릭하십시오.
2. 도구 메뉴에서 **노드 사용 안함**을 클릭하십시오.

또는 편집 메뉴에서 **노드 > 노드 사용 안함**을 클릭할 수 있습니다. 스트림에 노드를 다시 포함시키려면 같은 방법으로 **노드 사용**을 클릭하십시오.

참고: 편집 메뉴에서 **실행 취소**를 클릭하거나 Ctrl+Z를 눌러 이 조치를 실행 취소할 수 있습니다.

마. 기존 연결의 노드 추가

두 개의 노드를 연결하는 화살표를 끌어서 연결된 두 개의 노드 사이에 새 노드를 추가할 수 있습니다.

1. 가운데 마우스 단추를 사용하여 노드를 삽입할 위치로 연결 화살표를 클릭하여 끄십시오. 또는 클릭하여 끄는 동안 Alt 키를 눌러 가운데 마우스 단추를 시뮬레이션할 수 있습니다.
2. 포함시킬 노드로 연결을 끌고 마우스 단추를 놓으십시오.

참고: 노드에서 새 연결을 제거하고 노드를 **우회**하여 원래값을 복원할 수 있습니다.

바. 노드 간의 연결 삭제

두 노드 간의 연결을 삭제하려면 다음을 수행하십시오.

1. 연결 화살표를 마우스 오른쪽 단추로 클릭하십시오.
2. 메뉴에서 **연결 삭제**를 클릭하십시오.

노드 간의 연결을 모두 삭제하려면 다음 중 하나를 수행하십시오.

- 노드를 선택하고 F3을 누르십시오.
- 기본 메뉴에서 노드를 선택하고 다음을 클릭하십시오.

편집 > 노드 > 연결 해제

사. 노드 옵션 설정

노드를 작성하고 연결한 후, 노드를 사용자 정의하는 몇 가지 옵션이 있습니다. 노드를 마우스 오른쪽 단추로 클릭하고 메뉴 옵션 중 하나를 선택하십시오.

- **편집**을 클릭하여 선택한 노드의 대화 상자를 여십시오.
- **연결**을 클릭하여 한 노드를 다른 노드에 수동으로 연결하십시오.
- **연결 해제**를 클릭하여 노드에서 모든 링크를 삭제하십시오.
- **이름 변경 및 주석달기**를 클릭하여 편집 대화 상자의 주석(Annotation) 탭을 여십시오.
- **새 주석**을 클릭하여 노드와 관련된 주석을 추가하십시오. 자세한 정보는 노드와 스트림에 설명과 주석 추가 주제를 참조하십시오.
- **노드 사용 안함**을 클릭하여 처리 중 노드를 "숨기십시오". 노드를 처리를 위해 다시 표시하려면 노드 사용을 클릭하십시오. 자세한 정보는 스트림의 **노드 사용** 안함 주제를 참조하십시오.
- **잘라내기** 또는 **삭제**를 클릭하여 스트림에서 선택한 노드를 제거하십시오. *참고*: **잘라내기**를 클릭하면 노드를 붙여넣을 수 있지만 **삭제**는 그렇게 할 수 없습니다.
- **노드 복사**를 클릭하여 연결 없이 노드의 사본을 작성하십시오. 이는 신규 또는 기존 스트림에 추가할 수 있습니다.
- **노드 로드**를 클릭하여 이전에 저장된 노드를 열고 해당 옵션을 현재 선택된 노드로 로드하십시오. 노드는 동일한 유형이어야 합니다.
- **노드 저장**을 클릭하여 노드의 세부사항을 파일에 저장하십시오. 노드 세부사항을 동일 유형의 또 다른 노드로 로드할 수 있습니다.
- **캐시**를 클릭하여 선택한 노드를 캐싱하기 위한 옵션과 함께 메뉴를 펼치십시오.
- **데이터 맵핑**을 클릭하여 데이터를 새 소스에 맵핑하고 필수 필드를 지정하기 위한 옵션과 함께 메뉴를 펼치십시오.
- **수퍼노드 작성**을 클릭하여 현재 스트림에서 수퍼노드를 작성하기 위한 옵션과 함께 메뉴를 펼치십시오.
- **사용자 입력 노드 생성**을 클릭하여 선택한 노드를 대체하십시오. 이 노드에 의해 생성된 예제에는 현재 노드와 동일한 필드가 있습니다.
- **여기부터 실행**을 클릭하여 선택한 노드로부터 모든 터미널 노드 다운스트림을 실행하십시오.

아. 노드 캐싱 옵션

실행 중인 스트림을 최적화하기 위해 비터미널 노드에서 캐시를 설정할 수 있습니다. 노드에서 캐시를 설정하면 다음 번에 데이터 스트림을 실행할 때 노드를 통과하는 데이터로 캐시가 채워집니다. 그때부터는 데이터 소스가 아닌 캐시(디스크의 임시 디렉토리에 저장됨)에서 데이터를 읽습니다.

캐싱은 정렬, 병합 또는 통합과 같이 시간이 걸리는 작업 다음으로 가장 유용합니다. 예를 들어, 데이터베이스에서 판매 데이터를 읽는 소스 노드 세트와 지역별로 판매액을 요약하는 통합 노드가 있다고 가정하십시오. 전체 데이터 세트가 아니라 통합된 데이터를 저장할 캐시를 원하기 때문에 소스 노드가 아닌 통합 노드에서 캐시를 설정할 수 있습니다.

> **참고:** 소스 노드에서 캐싱하면(IBM® SPSS® Modeler로 데이터를 읽을 때 원 데이터 세트의 사본만 저장함) 대부분의 상황에서 성능이 향상되지는 않습니다.

캐싱을 사용하는 노드에는 오른쪽 상단 모서리에 작은 문서 아이콘이 표시됩니다. 노드에서 데이터가 캐싱될 때 문서 아이콘은 녹색입니다.

캐시 사용 방법

1. 스트림 캔버스에서 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 **캐시**를 클릭하십시오.
2. 캐싱 하위 메뉴에서 **사용**을 클릭하십시오.
3. 노드를 마우스 오른쪽 단추로 클릭하고 캐싱 하위 메뉴에서 **사용 안함**을 클릭하여 캐시를 설정 해제할 수 있습니다.

데이터베이스의 노드 캐싱

데이터베이스에서 실행되는 스트림의 경우, 스트림 중간에 파일 시스템이 아닌 데이터베이스의 임시 테이블에 데이터를 캐싱할 수 있습니다. SQL 최적화와 결합되면 이로 인해 성능이 상당히 향상됩니다. 예를 들어, 다중 테이블을 병합하여 데이터 마이닝 보기를 작성하는 스트림의 출력을 캐싱하고 필요에 따라 재사용할 수 있습니다. 모든 다운스트림 노드에 대해 SQL을 자동으로 생성하면 성능이 추가로 향상될 수 있습니다.

데이터베이스 캐싱을 이용하려면 SQL 최적화와 데이터베이스 캐싱 둘 다를 사용으로 설정해야 합니다. 서버 최적화 설정은 클라이언트에서 해당 설정을 대체합니다. 자세한 정보는 스트림의 최적화 옵션 설정의 내용을 참조하십시오.

데이터베이스 캐싱을 사용으로 설정한 상태에서는 비터미널 노드를 마우스 오른쪽 단추로 클릭하여 해당 지점에서 데이터를 캐싱하면 되며, 다음 번에 스트림이 실행될 때 데이터베이스에 자동으로 캐시가 직접 작성됩니다. 데이터베이스 캐싱 또는 SQL 최적화를 사용할 수 없으면 캐시는 파일 시스템에 대신 기록됩니다.

> **참고:** 캐싱 목적으로 임시 테이블을 지원하는 데이터베이스는 Db2, Oracle, SQL Server, Teradata입니다. Netezza 등의 기타 데이터베이스는 데이터베이스 캐싱에 정규 테이블을 사용합니다. 특정 데이터베이스에 대해 SQL 코드를 사용자 정의할 수 있습니다. 지원을 받으려면 서비스팀에 문의하십시오.

캐시 비우기 방법

노드의 흰색 문서 아이콘은 해당 캐시가 비어 있음을 표시합니다. **캐시**가 가득 차면 문서 아이콘이 녹색 단색이 됩니다. 캐시 컨텐츠를 대체하려면 먼저 캐시를 비운 후 데이터 스트림을 재실행하여 다시 채우십시오.

1. 스트림 캔버스에서 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 캐시를 클릭하십시오.
2. 캐싱 하위 메뉴에서 **비우기**를 클릭하십시오.

캐시 저장 방법

캐시 컨텐츠를 IBM SPSS Statistics 데이터 파일(*.sav)로 저장할 수 있습니다. 그런 다음 파일을 캐시로 재로드하거나 캐시 파일을 데이터 소스로 사용하는 노드를 설정할 수 있습니다. 또한 다른 프로젝트에서 저장한 캐시를 로드할 수 있습니다.

1. 스트림 캔버스에서 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 캐시를 클릭하십시오.
2. 캐싱 하위 메뉴에서 **캐시 저장**을 클릭하십시오.
3. 캐시 저장 대화 상자에서 캐시 파일을 저장할 위치를 찾아보십시오.
4. 파일 이름 텍스트 상자에 이름을 입력하십시오.
5. 파일 유형 목록에서 ***.sav**가 선택되어 있는지 확인하고 **저장**을 클릭하십시오.

캐시 로드 방법

노드에서 제거하기 전에 **캐시** 파일을 저장했으면 재로드할 수 있습니다.

스트림 캔버스에서 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 캐시를 클릭하십시오.
캐싱 하위 메뉴에서 **캐시 로드**를 클릭하십시오.
캐시 로드 대화 상자에서 캐시 파일 위치를 찾아 선택한 후 **로드**를 클릭하십시오.

자. 노드에서 데이터 미리보기

스트림을 작성할 때 예상한 방법으로 데이터가 변경 중인지 확인하려면 중요 단계마다 테이블 노드를 통해 데이터를 실행하면 됩니다. 그렇게 하지 않으려면 각 노드에서 작성되는 데이터의 샘플을 표시하는 미리보기를 생성하면 됩니다. 그러면 각 노드를 작성하는 시간이 줄어듭니다.

모델 너깃의 노드 업스트림의 경우 미리보기는 입력 필드를 표시합니다. 모델 너깃이나 너깃의 노드 다운스트림의 경우(터미널 노드는 제외) 미리보기는 입력 및 생성된 필드를 표시합니다.

표시되는 기본 행 수는 10입니다. 그러나 스트림 특성에서 이를 변경할 수 있습니다. 자세한 정보는 스트림의 일반 옵션 설정 주제를 참조하십시오.
생성 메뉴에서 몇몇 노드 유형을 작성할 수 있습니다.

> **참고:** 이 노드에 의해 생성된 데이터를 미리 보면 모든 특성 변경사항이 이 노드에 적용되며 취소할 수 없습니다(**적용**을 클릭하는 것과 동일한 작동입니다).

차. 노드 잠금

다른 사용자가 스트림에서 하나 이상의 노드 설정을 수정하지 못하도록 하려면 수퍼노드라는 특수 노드 유형으로 노드를 캡슐화한 다음 비밀번호 보호를 적용하여 수퍼노드를 잠그면 됩니다.

② 스트림에 대한 작업

스트림 캔버스에서 소스, 프로세스 및 터미널 노드를 연결하면 스트림이 작성됩니다. 노드의 콜렉션으로서 스트림은 저장되고, 주석(Annotation)이 작성되고 프로젝트에 추가됩니다. 최적화, 날짜 및 시간 설정, 모수 및 스크립트 등과 같이 스트림에 여러 옵션을 설정할 수도 있습니다. 이러한 특성은 다음과 같은 주제에서 설명됩니다.

IBM® SPSS® Modeler에서, 동일한 IBM SPSS Modeler 세션에서 둘 이상의 데이터 스트림을 사용하고 수정할 수 있습니다. 기본 창의 오른쪽에는 관리자 분할창이 있고 이는 현재 열려 있는 스트림, 출력 및 모델을 탐색하는 데 도움을 줍니다. 관리자 분할창이 보이지 않으면 보기 메뉴에서 관리자를 클릭한 다음 스트림 탭을 클릭하십시오.
이 탭에서 다음을 수행할 수 있습니다.

- 스트림에 액세스합니다.
- 스트림을 저장합니다.
- 스트림을 현재 프로젝트에 저장합니다.
- 스트림을 닫습니다.
- 새 스트림을 엽니다.
- 스트림을 저장하고 IBM SPSS Collaboration and Deployment Services 리포지토리(사용자의 사이트에서 사용 가능한 경우)에서 스트림을 검색합니다. 자세한 정보는 정보의 내용을 참조하십시오.

스트림 탭에서 스트림을 마우스 오른쪽 단추로 클릭하고 이러한 옵션에 액세스합니다.

가. 스트림의 옵션 설정

현재 스트림에 적용하기 위해 수많은 옵션을 지정할 수 있습니다. 이러한 옵션을 모든 스트림에 적용하기 위한 기본값으로 저장할 수도 있습니다. 옵션은 다음과 같습니다.

- **일반.** 스트림에 사용할 기호 및 텍스트 인코딩과 같은 기타 옵션. 자세한 정보는 스트림의 일반 옵션 설정의 내용을 참조하십시오.
- **날짜/시간.** 날짜 및 시간 표현식의 형식과 관련된 옵션. 자세한 정보는 스트림의 날짜 및 시간 옵션 설정의 내용을 참조하십시오.

- **숫자형식.** 숫자표현식의 형식을 제어하는 옵션. 자세한 정보는 스트림의 숫자 형식 옵션 설정의 내용을 참조하십시오.
- **최적화.** 스트림 성능을 최적화하기 위한 옵션. 자세한 정보는 스트림의 최적화 옵션 설정의 내용을 참조하십시오.
- **로깅 및 상태.** SQL 로깅 및 레코드 상태를 제어하는 옵션. 자세한 정보는 스트림의 SQL 로깅 및 레코드 상태 옵션 설정의 내용을 참조하십시오.
- **레이아웃.** 캔버스에서 스트림의 레이아웃과 관련된 옵션. 자세한 정보는 스트림의 레이아웃 옵션 설정를₩ 참조하십시오.
- **지리 공간적.** 스트림에 사용하기 위한 지리 공간적 데이터 형식과 관련된 옵션. 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오.

스트림 옵션 설정

1. 파일 메뉴에서 **스트림 특성**을 클릭하십시오. 또는 관리자 분할창의 스트림 탭에서 스트림을 선택하고 마우스 오른쪽 단추를 클릭한 다음 팝업 메뉴에서 **스트림 특성**을 클릭하십시오.
2. 옵션 탭을 클릭하십시오.

또는 도구 메뉴에서 다음을 클릭하십시오.

스트림 특성 > 옵션

ㄱ. 스트림의 일반 옵션 설정

일반 옵션은 현재 스트림의 다양한 측면에 적용도는 기타 옵션 세트입니다.

기본 섹션에는 다음 기본 옵션이 포함됩니다.
- **소수점 기호.** 쉼표(,) 또는 마침표(.)를 소수점 구분자로 선택하십시오.
- **그룹 기호.** 숫자 표시 형식의 경우 값을 그룹화하는 데 사용하는 기호를 선택하십시오(예: 3,000.00에서 쉼표). 옵션에는 없음, 마침표, 쉼표, 공백 및 정의된 로케일이 포함됩니다(이 경우 현재 로케일의 기본값을 사용함).
- **인코딩.** 텍스트 인코딩을 위한 스트림 기본 방법을 지정하십시오. (참고: 변수에 적용됩니다. 파일 소스 노드와 플랫 파일 내보내기 노드에만 해당합니다. 다른 노드는 이 설정을 사용하지 않습니다. 대부분의 데이터 파일에 임베드된 인코딩 정보가 있습니다.) 시스템 기본값 또는 UTF-8을 선택할 수 있습니다. 시스템 기본값은 Windows 제어판에 지정되어 있거나 분산 모드에서 실행 중인 경우 서버 컴퓨터에 지정되어 있습니다. 자세한 정보는 IBM SPSS Modeler에서 유니코드 지원의 내용을 참조하십시오.
- **규칙 세트 평가.** 규칙 세트 모델이 평가되는 방법을 결정합니다. 기본적으로 규칙 세트는 **투표**를 사용하여 개별 규칙으로부터 예측을 조합하고 최종 예측을 결정합니다. 해당 규칙 세트가 기본적으로 **처음 적중**하는 규칙을 사용하도록 하려면 처음 적중을 선택하십시오. 이 옵션은

항상 알고리즘이 정의한 첫 번째 적중을 사용하는 의사결정 목록 모델에는 적용되지 않음을 유의하십시오.

데이터 미리보기에 표시할 최대 행 수. 노드에 대해 데이터 미리보기가 요청될 때 표시될 행 수를 지정합니다. 자세한 정보는 노드에서 데이터 미리보기의 내용을 참조하십시오.

명목 필드의 최대 수. 필드의 데이터 유형이 **유형 없음** 상태가 되기 전까지의 명목 (세트) 필드의 최대 멤버 수를 지정하려면 선택하십시오. 이 옵션은 대규모의 명목 필드에서 작업할 때 유용합니다. 참고: 필드의 측정 수준이 **유형 없음**으로 설정되면 해당 역할은 자동으로 **없음**으로 설정됩니다. 즉 필드를 모델링에 사용할 수 없음을 의미합니다.

코호넨, K-평균 모델링의 변수군 크기 제한. 코호넨 넷과 K-평균 모델링에 사용된 명목 필드의 최대 멤버 수를 지정하려면 선택하십시오. 기본 변수군 크기는 20이고, 20이 넘으면 필드는 무시되고 경고가 발행되어 문제의 필드에 대한 정보를 제공합니다.

호환성을 위해서 이 옵션은 또한 IBM® SPSS® Modeler의 14 버전에 대체된 이전 신경망 노드에도 적용됩니다. 몇몇 레거시 스트림에는 여전히 이 노드가 포함될 수 있습니다.

실행 시 소스 노드 새로 고침. 현재 스트림을 실행할 때 모든 소스 노드를 자동으로 새로 고치려면 선택하십시오. 이 동작은 소스 노드에서 **새로 고침** 단추를 클릭하는 것과 유사합니다. 단지 이 옵션은 현재 스트림의 모든 소스 노드(사용자 입력 노드는 제외)를 자동으로 새로 고친다는 점이 다릅니다.

> **참고:** 이 옵션을 선택하면 데이터가 변경되지 않았더라도 다운스트림 노드의 캐시가 비워집니다. 도구 모음에서 **현재 스트림 실행**을 실행할 경우, 비우기는 스트림 실행 중 한 번만 발생하지만, 이는 다운스트림 캐시를 여전히 단일 실행 시 임시 저장 공간으로서 사용할 수 있음을 의미합니다. 예를 들어, 복합 파생 작업 후에 캐시 미드스트림을 설정했고 이 파생 노드의 다운스트림에 몇몇 그래프와 보고서가 첨부되었다고 가정해 보십시오. 스트림을 실행할 때 파생 노드에 있는 캐시는 첫 번째 그래프 또는 보고서에 대해서만 비워지고 다시 채워집니다. 후속 터미널 노드는 파생 노드 캐시로부터 데이터를 읽습니다. (2개 이상의 터미널 노드가 있을 때) **현재 스트림 실행**을 사용하는 대신 각각의 터미널 노드를 개별적으로 실행하도록 선택하는 경우, 캐시 비우기는 사용자가 터미널 노드를 실행할 때마다 발생한다는 점을 참고하십시오.

출력에 필드와 값 레이블 표시. 테이블, 차트 및 기타 출력에 필드와 값 레이블을 표시합니다. 레이블이 없는 경우에는 필드 이름과 데이터 값이 대신에 표시됩니다. 레이블은 기본적으로 꺼져 있습니다. 그러나 IBM SPSS Modeler의 다른 곳에서 개별적으로 레이블을 토글할 수 있습니다. 도구 모음에서 사용 가능한 토글 단추를 사용하여 출력 레이블을 표시하도록 선택할 수도 있습니다.

그림 1. 필드와 값 레이블을 토글하는 데 사용하는 도구 모음 아이콘

실행 시간 표시. 스트림이 실행된 후 실행 시간 탭에서 스트림 노드의 개별 실행 시간을 표시합니다. 자세한 정보는 실행 시간 보기의 내용을 참조하십시오.

자동 노드 작성 섹션에는 개별 스트림에서 노드를 자동으로 작성하기 위해 다음 옵션을 포함합니다. 이러한 옵션은 새 너깃을 생성할 때 스트림 캔버스에 모델링 너깃을 삽입하는지 여부를 제어합니다. 기본적으로 이러한 옵션은 버전 16 이상에서 작성된 스트림에만 적용됩니다. IBM SPSS Modeler 16 이상에서 버전 15 이전에서 작성된 스트림을 열고 모델링 노드를 실행하면 너깃은 이전 릴리스에서 사용된 것처럼 스트림 캔버스에 배치되지 않습니다. IBM SPSS Modeler 16 이상을 사용하여 새 스트림을 작성하고 모델링 노드를 실행하는 경우 생성된 너깃이 스트림 캔버스에 배치됩니다. 이는 예를 들어, **새 모델 출력의 모델 적용 노드 작성** 옵션이 IBM SPSS Collaboration and Deployment Services 및 IBM SPSS Modeler Server 클라이언트 사용자 인터페이스가 없는 다른 환경에서 배치로 실행된 앞의 16개 스트림을 구분하기 때문입니다.

- **새 모델 출력의 모델 적용 노드 작성.** 새 모델 출력의 모델 적용 노드를 자동으로 작성합니다. 이 옵션을 선택하면 링크를 사용, 사용 안함 또는 작성하지 않음으로 설정하는지 여부와 관계없이 **모델 업데이트 링크 작성**에서 선택할 수도 있습니다.
  새 모델 적용자 또는 소스 노드가 작성되면 드롭 다운 제어의 링크 옵션은 작성자 노드와 새 노드 간의 업데이트 링크가 작성되는지 여부를 제어하고, 작성되는 경우 어떤 모드에 있는지를 제어합니다. 링크가 작성되면 이를 사용 가능하게 만들 수 있지만 이러한 옵션은 사용자에게 완전한 제어를 제공합니다.

- **소스 작성기에서 소스 노드 작성.** 소스 작성기에서 소스 노드를 자동으로 작성합니다. 이전 옵션과 유사하게, 이 옵션을 선택하면 새로 고침 링크가 사용, 사용 안함으로 설정되는지 또는 이를 작성하지 않는지 여부와 관계없이 **소스 새로 고침 링크 작성** 드롭 다운에서 선택할 수도 있습니다.

기본값으로 저장. 지정된 옵션은 현재 스트림에만 적용됩니다. 이러한 옵션을 모든 스트림의 기본값으로 설정하려면 이 단추를 클릭하십시오.

ㄴ. 스트림의 날짜 및 시간 옵션 설정

이러한 옵션은 현재 스트림에서 다양한 날짜 및 시간 표현식에 사용하기 위한 형식을 지정합니다.

다른 이름으로 날짜/시간 가져오기 - 날짜/시간 필드에 날짜/시간 저장 공간을 사용할지 또는 이를 문자열 변수로 가져올지 여부를 선택하십시오.

날짜 형식 - 날짜 저장 공간 필드에 사용하거나 문자열이 CLEM 날짜 함수를 통해 날짜로 해석될 때 사용할 날짜 형식을 선택하십시오.

시간 형식 - 시간 저장 공간 필드에 사용하거나 문자열이 CLEM 시간 함수를 통해 시간으로 해석될 때 사용할 시간 형식을 선택하십시오.

롤오버 일/분 - 시간 형식에서 음의 시간 차이가 이전 일 또는 시간을 가리키는 것으로 해석할지 여부를 선택하십시오.

날짜 기준선(1월 1일) - 단일 날짜를 처리하는 CLEM 날짜 함수가 사용할 기준선 연도(항상 1월 1일)를 선택하십시오.

두 자리 시작 날짜 - 두 자릿수로 표시되는 연도에 세기 자릿수를 추가하기 위한 분리점 연도를 지정하십시오. 예를 들어, 1930을 분리점 연도로 지정하면 05/11/02는 2002년에 있다고 가정합니다. 이와 같은 설정은 30년 후의 날짜에도 20세기를 사용합니다. 따라서 05/11/73은 1973년으로 가정됩니다.

시간대 - datetime_now CLEM 표현식에서 사용할 시간대의 선택 방법을 선택하십시오.
- **서버**를 선택할 경우 시간대가 다음 항목에 따라 달라집니다.
  - 현재 스트림이 Analytic Server 데이터 소스를 사용할 경우 datetime_now 표현식에는 Analytic Server의 시간이 사용됩니다. 기본적으로 서버는 협정 세계시(UTC)를 사용합니다.
  - 현재 스트림이 데이터베이스 소스 노드를 사용할 경우 지원되는 데이터베이스에는 SQL 푸시백이 사용되고, datetime_now 표현식에는 데이터베이스의 시간이 사용됩니다.
  - 기타 모든 스트림의 경우 시간대가 SPSS® Modeler Server의 시간을 사용합니다.

- **Modeler 클라이언트**를 선택할 경우 SPSS Modeler가 설치된 시스템의 시간대 세부사항이 시간대에 적용됩니다.
- 또는 시간대에 대해 협정 세계시(UTC) 값을 선택할 수 있습니다.

기본값으로 저장 - 현재 스트림에만 적용되는 옵션입니다. 이러한 옵션을 모든 스트림의 기본값으로 설정하려면 이 단추를 클릭하십시오.

ㄷ. 스트림의 숫자 형식 옵션 설정

이러한 옵션은 현재 스트림에서 다양한 숫자 표현식에 사용하기 위한 형식을 지정합니다.

숫자 표시 형식. 표준(####.###), 지수표기(#.###E+##) 또는 통화 표시 형식($###.##)에서 선택할 수 있습니다.

소수점이하자리수(표준, 지수표기, 통화). 숫자 표시 형식의 경우 실수를 표시하거나 인쇄할 때 사용할 소수점이하자리수를 지정합니다. 이 옵션은 각 표시 형식마다 별도로 지정됩니다.

계산 위치. 삼각법 CLEM 표현식에 측정 단위로 사용할 **라디안** 또는 **도**를 선택하십시오. 자세한 정보는 삼각 함수 주제를 참조하십시오.

기본값으로 저장. 지정된 옵션은 현재 스트림에만 적용됩니다. 이러한 옵션을 모든 스트림의 기본값으로 설정하려면 이 단추를 클릭하십시오.

ㄹ. 스트림의 최적화 옵션 설정

최적화 설정을 사용하여 스트림 성능을 최적화할 수 있습니다. IBM® SPSS® Modeler Server (사용하는 경우)의 성능 및 최적화 설정은 클라이언트의 해당 설정을 대체합니다. 이러한 설정을 서버에서 사용할 수 없는 경우 클라이언트는 이를 사용할 수 없습니다. 그러나 서버에서 사용으로 설정된 경우에는 클라이언트가 사용 안함으로 설정하도록 선택할 수 있습니다.

> **참고:** 데이터베이스 모델링 및 SQL 최적화를 위해서는 IBM SPSS Modeler 컴퓨터에서 IBM SPSS Modeler Server 연결성이 설정되어 있어야 합니다. 이 설정이 사용 가능하면 데이터베이스 알고리즘에 액세스하고, SQL을 IBM SPSS Modeler에서 직접 푸시백하고, IBM SPSS Modeler Server에 액세스할 수 있습니다. 현재 라이센스 상태를 검증하려면 IBM SPSS Modeler 메뉴에서 다음을 선택하십시오.

도움말 > 정보 > 추가 세부사항

연결성이 사용 가능으로 설정되면 라이센스 상태 탭에 **서버 사용 가능** 옵션이 표시됩니다.

> **참고:** SQL 푸시백 및 최적화 지원 여부는 사용 중인 데이터베이스 유형에 따라 다릅니다. IBM SPSS Modeler에서 사용할 수 있도록 검정되고 지원되는 데이터베이스 및 ODBC 드라이버에 대한 최신 정보는 회사 지원 사이트(http://www.ibm.com/support)를 참조하십시오.

스트림 재작성 사용. IBM SPSS Modeler에서 스트림 재작성을 사용하려면 이 옵션을 선택하십시오. 4가지 유형의 재작성을 사용할 수 있으며 이 중 하나 이상을 선택할 수 있습니다. 스트림 재작성은 스트림 시맨틱을 변경하지 않고 보다 효율적인 작업을 위해 막후에서 스트림의 노드를 재정렬합니다.

- **SQL 생성 최적화.** 이 옵션을 사용하면 데이터베이스에서의 실행에 SQL 생성을 사용하여 더 많은 작업을 푸시백할 수 있도록 스트림 내에서 노드를 재정렬할 수 있습니다. SQL로 렌더링할 수 없는 노드를 발견하면 최적화 프로그램이 SQL로 렌더링할 수 있고 스트림 시맨틱에 영향을 주지 않고 문제점 노드 앞에 안전하게 이동할 수 있는 다운스트림 노드가 있는지 여부를 미리 확인합니다. 데이터베이스가 IBM SPSS Modeler 보다 효율적으로 작업을 수행할 수 있

을 뿐만 아니라 이러한 푸시백은 처리를 위해 IBM SPSS Modeler로 리턴되는 데이터의 크기를 줄입니다. 이는 결국 네트워크 트래픽을 줄이고 스트림 작업 속도를 높일 수 있습니다. SQL 최적화가 효과를 보려면 **SQL 생성** 선택란을 선택해야 합니다.

- **CLEM 표현식 최적화.** 이 옵션을 통해 최적화 프로그램은 처리 속도를 높이기 위해 스트림이 실행되기 전에 사전 처리할 수 있는 CLEM 표현식을 검색할 수 있습니다. 단순한 예로, *log(salary)*와 같은 표현식이 있으면 최적화 프로그램이 실제 급여 값을 계산하고 처리를 위해 이를 전달합니다. 이를 사용하여 SQL 푸시백과 IBM SPSS Modeler Server 성능을 모두 향상시킬 수 있습니다.
- **구문 실행 최적화.** 이러한 스트림 재작성 방법은 IBM SPSS Statistics 구문이 포함된 두 개 이상의 노드를 통합하는 작업의 효율성을 증대합니다. 구문 명령문을 각각 별도의 작업으로 실행하는 대신 단일 작업에 결합하여 최적화를 달성합니다.
- **기타 실행 최적화.** 이러한 스트림 재작성 방법은 데이터베이스에 위임할 수 없는 작업의 효율성을 증대합니다. 스트림의 데이터 양을 가능한 조기에 줄여 최적화를 달성합니다. 데이터 무결성을 유지하는 동안 스트림이 재작성되어 데이터 소스 가까이에 작업을 푸시하므로 결합과 같이 비용이 많이 드는 작업에 대한 데이터 다운스트림을 줄입니다.

병렬 처리 사용. 다중 프로세서를 가진 컴퓨터에서 실행할 때 이 옵션을 사용하면 시스템이 해당 프로세서 전반에서 로드 균형을 맞출 수 있으므로 성능이 향상됩니다. 다중 노드를 사용하거나 개별 노드(C5.0, 병합(키별), 정렬, 구간(순위 및 분위수 방법), 통합(하나 이상의 키 필드 사용))를 사용하면 병렬 처리를 통해 이익을 얻을 수 있습니다.

SQL 생성. 실행 프로세스를 생성하는 데 SQL 코드를 사용하여 데이터베이스에 스트림 작업을 푸시백(이렇게 하면 성능이 향상됨)할 수 있도록 SQL 생성을 사용하려면 이 옵션을 선택하십시오. 성능을 추가로 개선하려면 **SQL 생성 최적화**도 선택하여 데이터베이스에 푸시백되는 작업 수를 최대화할 수 있습니다. 노드에 대한 작업이 데이터베이스에 푸시백된 경우, 스트림이 실행될 때 노드가 보라색으로 강조 표시됩니다.

- **데이터베이스 캐싱.** 데이터베이스에서 실행될 SQL을 생성하는 스트림의 경우, 스트림 중간에 파일 시스템이 아닌 데이터베이스의 임시 테이블에 데이터를 캐싱할 수 있습니다. SQL 최적화와 결합되면 이로 인해 성능이 상당히 향상됩니다. 예를 들어, 다중 테이블을 병합하여 데이터 마이닝 보기를 작성하는 스트림의 출력을 캐싱하고 필요에 따라 재사용할 수 있습니다. 데이터베이스 캐싱을 사용으로 설정한 상태에서는 비터미널 노드를 마우스 오른쪽 단추로 클릭하여 해당 지점에서 데이터를 캐싱하면 되며, 다음 번에 스트림이 실행될 때 데이터베이스에 자동으로 캐시가 직접 작성됩니다. 따라서 다운스트림 노드에 대해 SQL이 생성되며 성능이 추가로 향상됩니다. 또는 필요하면 이 옵션을 사용 안 함으로 설정할 수 있습니다(예: 정책 또는 권한으로 인해 데이터베이스에 데이터를 기록하지 못하는 경우). 데이터베이스 캐싱 또는 SQL 최적화를 사용할 수 없으면 캐시는 파일 시스템에 대신 기록됩니다. 자세한 정보는 노드 캐싱 옵션 주제를 참조하십시오.
- **완화된 변환 사용.** 이 옵션을 사용하면 적절한 형식으로 저장된 경우 문자열에서 숫자로 또는 숫자에서 문자열로의 데이터 변환이 가능합니다. 예를 들어, 데이터가 데이터베이스에 문자열로 유지되지만 실제로는 의미있는 숫자를 포함하는 경우 푸시백이 발생할 때 사용하도록 데이

터를 변환할 수 있습니다.

> **참고:** SQL 구현에서의 사소한 차이 때문에 데이터베이스에서 실행되는 스트림이 IBM SPSS Modeler에서 실행될 때 리턴되는 것과 약간 다른 결과를 리턴할 수 있습니다. 비슷한 이유로 이러한 차이는 데이터베이스 벤더에 따라서도 다릅니다.

기본값으로 저장. 지정된 옵션은 현재 스트림에만 적용됩니다. 이러한 옵션을 모든 스트림의 기본값으로 설정하려면 이 단추를 클릭하십시오.

ㅁ. 스트림의 SQL 로깅 및 레코드 상태 옵션 설정

이러한 설정에는 스트림이 생성한 SQL문의 표시와 스트림이 처리한 레코드 수의 표시를 제어하는 다양한 옵션을 포함합니다.

SQL을 스트림 실행 동안에 메시지 로그에 표시. 스트림을 실행하는 동안 생성된 SQL이 메시지 로그로 전달되는지 여부를 지정합니다.

스트림 전파 동안에 메시지 로그에 SQL 생성 세부사항 표시. 스트림 미리보기 중에 생성된 SQL의 미리보기가 메시지 로그에 전달되는지 여부를 지정합니다.

SQL 표시. 로그에 표시되는 SQL에 SPSS® Modeler이 생성하는 기본 SQL 함수 또는 {fn FUNC(…)} 양식의 표준 ODBC 함수가 포함되는지 여부를 지정합니다. 전자는 구현되지 않을 수도 있는 ODBC 드라이버 기능에 의존합니다.

가독성 향상을 위해 SQL 다시 형식화. 로그에 표시된 SQL이 가독성을 위해 형식화되어야 하는지 여부를 지정합니다.

레코드의 상태 표시. 레코드가 터미널 노드에 도착할 때 언제 레코드를 보고해야 하는지를 지정합니다. 상태를 N개의 레코드마다 업데이트하는 데 사용하는 숫자를 지정합니다.

기본값으로 저장. 지정된 옵션은 현재 스트림에만 적용됩니다. 이러한 옵션을 모든 스트림의 기본값으로 설정하려면 이 단추를 클릭하십시오.

ㅂ. 스트림의 레이아웃 옵션 설정

이러한 설정은 스트림 캔버스의 표시 및 사용과 관련된 수많은 옵션을 제공합니다.

최소 스트림 캔버스 너비. 스트림 캔버스의 최소 너비를 픽셀 단위로 지정합니다.

최소 스트림 캔버스 높이. 스트림 캔버스의 최소 높이를 픽셀 단위로 지정합니다.

스트림 스크롤링 속도. 노드를 캔버스에서 한 위치에서 다른 위치로 끌 때 스트림 캔버스 분할 창이 스크롤링되는 속도를 제어하기 위한 스트림 캔버스의 스크롤링 비율을 지정합니다. 숫자가 높을수록 더 빠른 스크롤링 속도를 지정합니다.

아이콘 이름 최대값. 스트림 캔버스에서 노드 이름의 문자 수 제한을 지정합니다.

아이콘 크기. 전체 스트림 보기를 표준 아이콘 크기의 8%와 200% 사이의 크기의 숫자 중 하나로 배율 조정하려면 옵션을 선택하십시오.

눈금 셀 크기. 목록에서 눈금 셀 크기를 선택합니다. 이 숫자는 개별 눈금을 사용하여 스트림 캔버스에서 노드를 정렬하는 데 사용합니다. 기본 눈금 셀 크기는 0.25입니다.

눈금에 정렬. 아이콘을 보이지 않는 눈금 패턴으로 맞추려면 선택합니다(기본적으로 선택됨).

생성된 아이콘 배치. 캔버스에서, 모델 너깃에서 생성된 노드의 아이콘을 배치할 위치를 선택합니다. 기본값은 왼쪽 상단입니다.

기본값으로 저장. 지정된 옵션은 현재 스트림에만 적용됩니다. 이러한 옵션을 모든 스트림의 기본값으로 설정하려면 이 단추를 클릭하십시오.

ㅅ. 스트림의 지리 공간적 옵션 설정

모양, 좌표 또는 단일 축 값(x 또는 y, 또는 위도와 경도)인지 관계없이 모든 지리 공간적 필드에는 연관된 좌표계가 있습니다. 이 좌표계는 속성을 원점(0,0) 및 값과 연관된 단위로 설정합니다.

수많은 좌표계가 있고 유형에는 지리적 및 투영의 두 가지가 있습니다. SPSS® Modeler에서 모든 공간 함수는 투영된 좌표계와 함께만 사용할 수 있습니다.

좌표계의 특성으로 인해, 두 개의 별개의 지리 공간적 데이터 소스에서 데이터를 병합하거나 붙여쓰려면 소스가 동일한 좌표계를 사용해야 합니다. 이로 인해 스트림에 사용된 지리 공간적 데이터에 좌표 설정을 지정해야 합니다.

데이터는 자동으로 다음 상황에서 선택된 스트림 좌표계를 사용하도록 다시 투영됩니다.

- 공간 함수(예: 영역, closeto, within)의 경우, 함수에 전달된 모수는 자동으로 다시 투영됩니다. 그러나 원본 행 데이터는 변경되지 않은 상태로 남아 있습니다.
- STP(Spatio-Temporal Prediction)에서 작성 또는 스코어링 (너깃) 노드를 사용할 때 위치 필드는 자동으로 재투영됩니다. 스코어링할 때 너깃에서 나오는 위치는 원본 위치입니다.
- 맵 시각화 노드를 사용할 때.

스트림 좌표계. 선택란을 선택하는 경우에만 사용 가능합니다. 사용 가능한 투영된 좌표계 목록을 표시하고 현재 스트림에 사용하려는 좌표계를 선택하려면 **변경**을 클릭합니다.

기본값으로 저장. 선택하는 좌표계는 현재 스트림에만 적용됩니다. 스트림을 모든 스트림의 기본값으로 선택하려면 이 단추를 클릭하십시오.

- **지리 공간적 좌표계 선택**

SPSS® Modeler에서 모든 공간 함수는 투영된 좌표계와만 함께 사용할 수 있습니다.

스트림 좌표계 선택 대화 상자에는 스트림에 사용하는 지리 공간적 데이터를 위해 선택할 수 있는 모든 투영된 좌표계 목록이 포함됩니다.

다음 정보는 각 좌표계에 나열되어 있습니다.

- **WKID** 각 좌표계에 고유한 잘 알려진 ID.
- **이름** 좌표계의 이름.
- **단위** 좌표계와 연관된 측정 단위.

모든 좌표계의 목록에 추가로 대화 상자에는 **필터링** 제어가 있습니다. 필요한 좌표계 **이름**의 모두 또는 일부를 알고 있는 경우에는 이를 대화 상자 아래쪽의 이름 필드에 입력하십시오. 선택할 수 있는 좌표계 목록은 입력한 텍스트를 포함하는 이름이 있는 시스템만을 표시하기 위해 자동으로 필터링됩니다.

나. 스트림 작업 메시지 보기

모델 작성과 평가 실행, 최적화 및 경과 시간 등과 같이 스트림 작업과 관련된 메시지는 스트림 특성 대화 상자의 메시지 탭을 사용하여 쉽게 볼 수 있습니다. 오류 메시지는 또한 이 테이블에 보고됩니다.

스트림 메시지 보기

1. 파일 메뉴에서 **스트림 특성**을 클릭하십시오. 또는 관리자 분할창의 스트림 탭에서 스트림을 선택하고 마우스 오른쪽 단추를 클릭한 다음 팝업 메뉴에서 **스트림 특성**을 클릭하십시오.
2. **메시지** 탭을 클릭하십시오.

또는 도구 메뉴에서 다음을 클릭하십시오.
스트림 특성 > 메시지

스트림 작업에 관한 메시지에 추가로 오류 메시지도 여기에 보고됩니다. 오류로 인해 스트림 실행이 종료되면 이 대화 상자 오류 메시지가 표시된 상태로 메시지 탭으로 열립니다. 또한 오류가 있는 노드는 스트림 캔버스에서 빨간색으로 강조 표시됩니다.

SQL 최적화와 로깅 옵션이 사용자 옵션 대화 상자에서 사용 가능한 경우에는 생성된 SQL에 대한 정보 또한 표시됩니다. 자세한 정보는 스트림의 최적화 옵션 설정의 내용을 참조하십시오.

저장 단추 드롭 다운 목록에서(메시지 탭 바로 아래 왼쪽에서) **메시지 저장**을 클릭하여 스트림에 대해 보고된 메시지를 여기에 저장할 수 있습니다.

드롭 다운에서 **모든 메시지 지우기**를 선택하여 지정된 스트림에 대한 모든 메시지를 지울 수 있습니다.

CPU 시간은 서버 프로세스가 CPU를 이용하는 시간임을 유의하십시오. 경과 시간은 실행 시작과 실행 종료 사이의 총 시간이므로 파일 전송과 출력 렌더링 등을 포함합니다. CPU 시간은 스트림이 다중 CPU(병렬 실행)를 사용할 때는 경과 시간보다 길 수 있습니다. 스트림이 데이터 소스로 사용 중인 데이터베이스로 완전히 푸시백되면 CPU 시간은 0이 됩니다.

다. 실행 시간 보기

메시지 탭에서 IBM® SPSS® Modeler Server에서 실행되는 스트림의 모든 노드에 대한 개별 실행 시간을 볼 수 있는 실행 시간을 표시하도록 선택할 수 있습니다. 시간이 R 또는 Analytic Server 등과 같은 다른 영역에서 스트림 실행이 정확하지 않을 수도 있음을 유의하십시오. 또한 일부 노드의 실행 시간은 계산할 수 없습니다.

> **참고:** 이 기능이 작동하려면 **옵션** 탭의 **일반** 설정에서 **실행 시간 표시** 옵션이 선택되어야 합니다.

노드 실행 시간의 테이블에서 열은 다음과 같습니다. 열 머리말을 클릭하여 항목을 오름차순이나 내림차순으로 정렬하십시오(예를 들어, 어떤 노드가 실행 시간이 가장 긴지 확인하기 위해).

터미널 노드. 노드가 속하는 분기의 ID. ID는 분기의 끝에 있는 터미널 노드의 이름입니다.

노드 레이블. 실행 시간이 참조하는 노드의 이름입니다.

노드 ID. 실행 시간이 참조하는 노드의 고유 ID입니다. 이 ID는 노드가 작성될 때 시스템에 의해 생성됩니다.
실행 시간. 이 노드를 실행하는 데 걸리는 시간(초). 스트림을 실행할 때 데이터를 준비하고 출력에서 데이터를 검색하는 데 시간이 걸리고 이 유형의 시간은 계산할 수 없으므로 실행 시간은 일반 메시지에 표시되는 시간과 다를 수 있습니다.

라. 스트림과 세션 모수 설정

CLEM 표현식과 스크립팅에서 사용할 모수를 정의할 수 있습니다. 사실상 모수는 현재 스트림, 세션 또는 수퍼노드와 함께 저장되고 지속되는 사용자 정의 변수이며 스크립팅을 통해서는 물론 사용자 인터페이스에서도 액세스할 수 있습니다. 예를 들어, 스트림을 저장하면 해당 스트림에 대해 설정된 모수도 저장됩니다(이는 선언된 스크립트에서만 사용할 수 있는 로컬 스크립트 변수와 모수를 구별함). 모수는 스크립트에 하드 코드화하지 않아도 되는 필드 및 값에 대한 정보를 제공하여 스크립트팅에서 스크립트 동작을 제어하는 데 사용하는 경우가 많습니다.

모수 범위는 모수가 설정된 위치에 따라 다릅니다.

- 스트림 모수는 스트림 스크립트 또는 스트림 특성 대화 상자에서 설정할 수 있으며 스트림의 모든 노드에 사용할 수 있습니다. 표현식 작성기의 모수 목록에 표시됩니다.
- 세션 모수는 독립형 스크립트 또는 세션 모수 대화 상자에서 설정할 수 있습니다. 현재 세션에서 사용하는 모든 스트림(관리자 분할창의 스트림 탭에 나열된 모든 스트림)에 사용할 수 있습니다.

수퍼노드에 대해서도 모수를 설정할 수 있는데, 이 경우 해당 수퍼노드 내에 캡슐화된 노드에만 표시됩니다.

사용자 인터페이스를 통해 스트림 및 세션 모수 설정

1. 스트림 모수를 설정하려면 기본 메뉴에서 다음을 클릭하십시오.
   도구 > 스트림 특성 > 모수

2. 세션 모수를 설정하려면 도구 메뉴에서 **세션 모수 설정**을 클릭하십시오.

프롬프트 표시 여부. 사용자가 런타임 시에 이 모수의 값을 입력하도록 프롬프트하려면 이 상자를 선택하십시오.

이름 모수 이름이 여기 나열됩니다. 이 필드에 이름을 입력하여 새 모수를 작성할 수 있습니다. 예를 들어, 최저 기온에 대한 모수를 작성하려면 minvalue를 입력할 수 있습니다. CLEM 표현식에서 모수를 나타내는 $P- 접두문자를 포함시키지 마십시오. 이 이름은 CLEM 표현식 작성기에 표시하는 데도 사용합니다.

긴 이름. 작성된 각 모수에 대한 설명 이름을 나열합니다.

저장 공간 목록에서 저장 유형을 선택하십시오. 저장 공간은 모수에 데이터 값이 저장되는 방법을 표시합니다. 예를 들어, 유지할 선행 0이 포함된 값(예: 008)에 대한 작업 시 저장 유형으로 **문자열**을 선택해야 합니다. 그렇지 않으면, 값에서 0이 제거됩니다. 사용 가능한 저장 유형은 문

자열, 정수, 실수, 시간, 날짜, 시간소인입니다. 날짜 모수의 경우, 다음 단락에 표시된 대로 ISO 표준 표기법을 사용하여 값을 지정해야 합니다.

값 각 모수의 현재 값을 나열합니다. 필요에 따라 모수를 조정하십시오. 데이터 모수의 경우, ISO 표준 표기법(즉, YYYY-MM-DD)으로 값을 지정해야 합니다. 다른 형식으로 지정된 날짜는 허용되지 않습니다.

유형(선택사항). 외부 애플리케이션에 스트림을 배포할 계획이면 목록에서 측정 수준을 선택하십시오. 그렇지 않으면 유형 열을 있는 그대로 두는 것이 바람직합니다. 모수의 값 제한조건(예: 숫자 범위의 상한 및 하한)을 지정하려면 목록에서 **지정**을 선택하십시오.

사용자 인터페이스를 통해서만 모수에 대해 긴 이름, 저장 공간, 유형 옵션을 설정할 수 있습니다. 이러한 옵션은 스크립트를 사용하여 설정할 수 없습니다.

선택된 모수를 사용 가능한 모수 목록 위, 아래로 추가로 이동하려면 오른쪽에 있는 화살표를 선택하십시오. 선택된 모수를 제거하려면 삭제 단추(X로 표시)를 사용하십시오.

마. 모수 값의 런타임 프롬프트 지정

서로 다른 경우에 동일한 모수에 대해 서로 다른 값을 입력해야 하는 스트림이 있는 경우에는 하나 이상의 스트림 또는 세션 모수 값에 런타임 프롬프트를 지정할 수 있습니다.

모수. (선택적) 모수 값을 입력하거나 기본값이 있으면 그대로 두십시오.

이러한 프롬프트 끄기. 스트림을 실행할 때 이러한 프롬프트가 표시되는 것을 원치 않으면 이 상자를 선택하십시오. 모수가 정의된 스트림 특성이나 세션 특성 대화 상자에서 **프롬프트?** 선택란을 선택하여 이를 다시 표시되게 할 수 있습니다. 자세한 정보는 스트림과 세션 모수 설정 주제를 참조하십시오.

바. 모수 유형의 값 제한조건 지정

데이터 모델링 스트림을 읽는 외부 애플리케이션에 스트림 배포 동안에 사용 가능한 모수의 값 제한조건을 만들 수 있습니다. 이 대화 상자를 사용하면 스트림을 실행하는 외부 사용자가 사용 가능한 값을 지정할 수 있습니다. 데이터 유형에 따라서 값 제한조건은 대화 상자에서 동적으로 변화합니다. 여기에 표시된 옵션은 유형 노드의 값에 사용 가능한 옵션과 동일합니다.

유형. 현재 선택된 측정 수준을 표시합니다. IBM® SPSS® Modeler에서 모수를 사용하려는 방법을 반영하여 이 값을 변경할 수 있습니다.

저장 공간. 알려진 경우 저장 유형을 표시합니다. 저장 유형은 사용자가 IBM SPSS Modeler에서 작업하기 위해 선택하는 측정 수준(연속, 명목 또는 플래그)에 영향을 받지 않습니다. 기본 모수 탭에서 저장 유형을 변경할 수 있습니다.

대화 상자의 하반부는 **유형** 필드에 선택된 측정 수준에 따라서 동적으로 변경됩니다.

연속 측정 수준

하한. 모수 값의 하한을 지정합니다.

상한. 모수 값의 상한을 지정합니다.

레이블. 범위 필드의 값에 레이블을 지정할 수 있습니다. 레이블 단추를 클릭하여 값 레이블을 지정하기 위한 별도의 대화 상자를 여십시오.

명목 측정 수준

값. 이 옵션을 사용하면 명목 필드로 사용할 모수의 값을 지정할 수 있습니다. 값은 IBM SPSS Modeler 스트림에서 강요되지 않지만 외부 배포 애플리케이션의 드롭 다운 목록에 사용합니다. 화살표와 삭제 단추를 사용하면 기존 값을 수정할 뿐만 아니라 값을 다시 정렬하고 삭제할 수 있습니다.

플래그 측정 수준

참. 조건이 충족될 때 모수의 플래그 값을 지정합니다.

거짓. 조건이 충족되지 않을 때 모수의 플래그 값을 지정합니다.

레이블. 플래그 필드의 값에 레이블을 지정할 수 있습니다.

사. 스트림의 실행 반복

스트림 특성 대화 상자에서 실행 탭을 사용하면 현재 스트림에서 반복 작업을 자동화하도록 반복 조건을 설정할 수 있습니다.

이러한 조건을 설정한 후에는 이를 스크립팅에 소개로 사용하여 스크립트 창을 스트림의 기본 스크립팅으로 사용할 수 있습니다. 그런 다음 더 나은 스크립트를 작성하기 위한 기본으로 사용하기 위해 이를 수정할 수 있습니다. 자세한 정보는 글로벌 함수 주제를 참조하십시오.

스트림의 루핑 설정

1. 파일 메뉴에서 **스트림 특성**을 클릭하십시오. 또는 관리자 분할창의 스트림 탭에서 스트림을 선택하고 마우스 오른쪽 단추를 클릭한 다음 팝업 메뉴에서 스트림 특성을 클릭하십시오.
2. **실행** 탭을 클릭하십시오.

3. **루핑/조건부 실행** 실행 모드를 선택하십시오.
4. **루핑** 탭을 클릭하십시오.

또는 도구 메뉴에서 다음을 클릭하십시오.

스트림 특성 > 실행

또 다른 대안으로서 노드를 마우스 오른쪽 단추로 클릭하고 컨텍스트 메뉴에서 다음을 클릭하십시오.

루핑/조건부 실행 > 루핑 설정 편집

반복. 이 행 번호 값을 편집할 수 없지만 테이블 오른쪽의 단추를 사용하여 반복을 추가, 삭제 또는 위아래로 이동할 수 있습니다.

테이블 헤더. 이는 루프를 설정할 때 작성되는 반복 키와 반복 변수를 반영합니다.

아. 스트림의 글로벌 값 보기

스트림 특성 대화 상자의 글로벌 탭을 사용하여 현재 스트림에 설정된 글로벌 값을 볼 수 있습니다. 글로벌 값은 선택한 필드의 평균, 합계 또는 표준 편차 등과 같은 통계를 판별하기 위해 전역값 설정 노드를 사용하여 작성됩니다.

전역값 설정 노드가 실행되면 이러한 값은 스트림 작업에서 다양한 용도로 사용 가능합니다. 자세한 정보는 글로벌 함수의 내용을 참조하십시오.

스트림의 글로벌 값 보기

1. 파일 메뉴에서 **스트림 특성**을 클릭하십시오. 또는 관리자 분할창의 스트림 탭에서 스트림을 선택하고 마우스 오른쪽 단추를 클릭한 다음 팝업 메뉴에서 **스트림 특성**을 클릭하십시오.
2. 전역값 탭을 클릭하십시오.

또는 도구 메뉴에서 다음을 클릭하십시오.

스트림 특성 > 전역값

전역값 사용가능. 사용 가능한 전역값이 이 테이블에 나열되어 있습니다. 여기에서는 글로벌 값을 편집할 수 없지만 테이블 오른쪽의 모두 선택 취소 단추를 사용하여 스트림의 모든 글로벌 값을 지울 수 있습니다.

자. 스트림에서 노드 검색

노드 이름, 범주 및 식별자 등과 같은 여러 검색 기준을 지정하여 스트림에서 노드를 검색할 수 있습니다. 이 기능은 다수의 노드를 포함하는 복합 스트림에서 특히 유용할 수 있습니다.

스트림에서 노드 검색

1. 파일 메뉴에서 **스트림 특성**을 클릭하십시오. 또는 관리자 분할창의 스트림 탭에서 스트림을 선택하고 마우스 오른쪽 단추를 클릭한 다음 팝업 메뉴에서 **스트림 특성**을 클릭하십시오.
2. **검색** 탭을 클릭하십시오.

또는 도구 메뉴에서 다음을 클릭하십시오.

스트림 특성 > 검색

둘 이상의 옵션을 지정하여 검색을 제한할 수 있습니다. 그러나 노드 ID로 검색하면(**ID 동일** 필드 사용) 다른 옵션은 제외된다는 점은 예외입니다.

노드 레이블 포함. 특정 노드를 검색하려면 이 상자를 선택하고 노드 레이블의 모두나 일부를 입력하십시오. 검색은 대소문자를 구분하지 않고 다중 단어는 단일 텍스트 조각으로 처리됩니다.

노드 범주. 특정 유형의 노드를 검색하려면 이 상자를 선택하고 목록에서 범주를 클릭하십시오. **프로세스 노드**는 노드 팔레트의 레코드 작업 또는 필드 작업 탭의 노드를 의미합니다. **모델 노드 적용**은 모델 너깃을 가리킵니다.

키워드 포함. 노드 대화 상자의 주석(Annotation) 탭에서 키워드 필드에 해당 텍스트가 있는 노드를 검색하려면 이 상자를 선택하고 하나 이상의 완전한 키워드를 입력하십시오. 입력하는 키워드 텍스트는 정확히 매치해야 합니다. 대안을 검색하려면 여러 키워드를 세미콜론으로 구분하십시오. 예를 들어, proton;neutron을 입력하면 이러한 키워드 중 하나가 있는 모든 노드를 찾습니다. 자세한 정보는 주석(Annotation)의 내용을 참조하십시오.

주석(Annotation) 포함. 노드 대화 상자의 주석(Annotation) 탭에서 기본 텍스트 영역에 이 텍스트를 포함하는 노드를 검색하려면 이 상자를 선택하고 하나 이상의 단어를 입력하십시오. 검색은 대소문자를 구분하지 않고 다중 단어는 단일 텍스트 조각으로 처리됩니다. 자세한 정보는 주석(Annotation)의 내용을 참조하십시오.

호출된 필드 생성. 이 상자를 선택하고 생성된 필드의 이름을 입력하십시오(예: $C-Drug). 이 옵션을 사용하여 특정 필드를 생성하는 모델링 노드를 검색할 수 있습니다. 정확히 매치해야 하는 단 하나의 필드 이름을 입력하십시오.

ID 동일. 해당 ID가 있는 특정 노드를 검색하려면 이 상자를 선택하고 노드 ID를 입력하십시오. 노드 ID는 노드가 작성될 때 시스템이 지정하고 스크립팅이나 자동화를 위해 노드를 참조할 때

사용할 수 있습니다. 정확히 매치해야 하는 단 하나의 노드 ID만을 입력하십시오. 자세한 정보는 주석(Annotation)의 내용을 참조하십시오.

수퍼노드에서 검색. 이 상자는 기본적으로 선택되어 있습니다. 즉 수퍼노드의 내부와 외부 둘 모두의 노드에서 검색이 수행됨을 의미합니다. 스트림의 맨 위 수준에서 수퍼노드의 외부에 있는 노드에서만 검색을 수행하려면 이 상자를 지우십시오.

찾기. 원하는 모든 옵션을 지정한 후에는 이 단추를 클릭하여 검색을 시작하십시오.

지정된 옵션과 매치하는 노드가 대화 상자의 하부에 나열됩니다. 목록에서 노드를 선택하여 이를 스트림 캔버스에서 강조 표시하십시오.

차. 스트림 이름 변경

스트림 특성 대화 상자에서 주석(Annotation) 탭을 사용하면 스트림의 설명 주석(Annotation)을 추가하고 스트림의 사용자 정의 이름을 작성할 수 있습니다. 이러한 옵션은 프로젝트 분할창에 추가된 스트림에 대한 보고서를 생성할 때 특히 유용합니다. 자세한 정보는 주석(Annotation)의 내용을 참조하십시오.

③ 스트림 설명

작성하는 각 스트림에 대해 IBM® SPSS® Modeler는 스트림의 내용에 대한 정보를 포함하는 스트림 설명을 생성합니다. 스트림의 수행 작업은 보려고 하지만 IBM SPSS Modeler가 설치되지 않은 경우를 통해 스트림에 액세스하는 경우) 유용할 수 있습니다.

스트림 설명은 여러 섹션으로 구성되어 HTML 문서의 양식으로 표시됩니다.

일반 스트림 정보

이 섹션에는 스트림 작성 시간과 마지막 저장 시간에 대한 세부사항과 함께 스트림 이름이 포함됩니다.

설명과 주석
이 섹션에는 다음이 포함됩니다.

- 스트림 주석(Annotation)(주석(Annotation) 참조)
- 특정 노드에 연결되지 않은 주석
- 스트림의 모델링 및 스코어링 분기 둘 모두에서 노드에 연결된 주석

스코어링 정보

이 섹션에는 스트림의 스코어링 분기와 관련하여 다양한 머리말 아래에 정보를 포함합니다.

- **주석** 스코어링 분기에서 노드에 링크된 주석만을 포함합니다.
- **입력** 입력 필드를 저장 유형과 함께 나열합니다(예: 문자열, 정수, 실수 등).
- **출력** 모델링 노드에 의해 생성된 추가 필드를 포함하여 출력 필드를 저장 유형과 함께 나열합니다.
- **모수** 스트림의 스코어링 분기와 관련되었고 모델이 스코어링될 때마다 보거나 편집할 수 있는 모수를 나열합니다. 이러한 모수는 스트림 특성 대화 상자의 **배포** 탭에서 **스코어링 모수** 단추를 클릭할 때 식별됩니다.
- **모델 노드** 모델 이름과 유형을 표시합니다(예: 신경망, C&R 트리 등). 이는 스트림 특성 대화 상자의 **배포** 탭에서 **모델 노드** 필드에 대해 선택된 모델 너깃입니다.
- **모델 세부사항** 식별된 모델 너깃의 세부사항을 이전 머리말 아래에 표시합니다. 가능한 경우 모델의 예측자 중요도와 평가 차트가 포함됩니다.

모델링 정보

스트림의 모델링 분기와 관련된 정보를 포함합니다.

- **주석** 모델링 분기에서 노드에 연결된 주석 또는 주석(Annotation)을 나열합니다.
- **입력** 입력 필드를 모델링 분기에서의 역할과 함께 나열합니다(예를 들어, 입력, 목표, 분할 등의 필드 역할 값의 양식으로).
- **모수** 스트림의 모델링 분기와 관련되었고 모델이 업데이트될 때마다 보거나 편집할 수 있는 모수를 나열합니다. 이러한 모수는 스트림 특성 대화 상자의 **배포** 탭에서 **모델 작성 모수** 단추를 클릭할 때 식별됩니다.
- **모델링 노드** 모델을 생성하거나 업데이트하는 데 사용하는 모델링 노드의 이름과 유형을 표시합니다.

가. 스트림 설명 미리보기

스트림 특성 대화 상자에서 옵션을 클릭하여 스트림 설명의 내용을 웹 브라우저에서 볼 수 있습니다. 설명의 내용은 대화 상자의 배포 탭에 지정하는 옵션에 따라 다릅니다. 자세한 정보는 스트림 배포 옵션의 내용을 참조하십시오.

스트림 설명 보기

1. 기본 IBM® SPSS® Modeler 메뉴에서 다음을 클릭하십시오.
도구 > 스트림 특성 > 배포

2. 배포 유형, 전용 스코어링 노드 및 모든 스코어링 모수를 설정하십시오.
3. 배포 유형이 모델 새로 고침이면 선택적으로 다음을 선택할 수 있습니다.
   - 모델링 노드 및 모든 모델 작성 모수
   - 스트림의 스코어링 분기 위의 모델 너깃
4. **스트림 설명 미리보기** 단추를 클릭하십시오.

나. 스트림 설명 내보내기

스트림 설명의 내용을 HTML 파일로 내보낼 수 있습니다.

스트림 설명 내보내기

1. 기본 메뉴에서 다음을 클릭하십시오.
   파일 > 스트림 설명 내보내기
2. HTML 파일의 이름을 입력하고 **저장**을 클릭하십시오.

④ 스트림 실행

필수 스트림 옵션을 지정했고 필수 노드를 연결했으면 스트림의 노드를 통해 데이터를 실행하여 스트림을 실행할 수 있습니다. IBM® SPSS® Modeler에서 스트림을 실행하는 몇 가지 방법이 있습니다. 다음을 수행할 수 있습니다.

- 도구 메뉴에서 **실행**을 클릭하십시오.
- 도구 모음에서 **실행...** 단추 중 하나를 클릭하십시오. 이러한 단추를 사용하여 전체 스트림을 실행하거나 선택된 터미널 노드만 실행할 수 있습니다. 자세한 정보는 IBM SPSS Modeler 도구 모음의 내용을 참조하십시오.
- 터미널 노드를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **실행**을 클릭하여 단일 데이터 스트림을 실행하십시오.
- 비터미널 노드를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **여기에서 실행**을 클릭하여 데이터 스트림의 일부를 실행하십시오. 이렇게 하면 선택된 노드 뒤의 해당 작업만 수행됩니다.

진행 중인 스트림의 실행을 정지하려면 도구 모음에서 빨간색 **중지** 단추를 클릭하거나 도구 메뉴에서 **실행 중지**를 선택하십시오. 중지 단추를 한 번 클릭하면 Modeler Server의 실행이 중지됩니다. 어떤 경우에는 실행이 즉시 중지되지만 다른 경우에는 전체 실행을 중지하기 전에 현재 단계를 완료해야 합니다. 따라서 시간이 달라질 수 있습니다. 이 단추를 두 번 클릭하면 서버 연결이 끊기고 새 연결이 설정됩니다. 대부분의 경우 이렇게 하면 서버 프로세스가 모두 닫힙니다(시간이 걸릴 수 있음). 그러나 어떤 경우에는 일부 서버 프로세스가 중지되지 않습니다.

스트림을 실행하는 데 3초가 넘으면, 실행 피드백 대화 상자가 표시되어 진행률을 나타냅니다.

일부 노드에는 스트림 실행에 대한 추가 정보를 제공하는 추가 표시가 있습니다. 대화 상자에서 해당 행을 선택하면 이러한 표시를 볼 수 있습니다. 첫째 행이 자동으로 선택됩니다.

⑤ 모델에 대한 작업

스트림이 모델링 노드(즉, 노드 팔레트의 모델링 또는 데이터베이스 모델링 탭의 노드)를 포함하는 경우 스트림이 실행될 때 모델 너깃이 작성됩니다. 모델 너깃은 모델 컨테이너 즉, 소스 데이터에 대해 예측을 생성할 수 있게 하고 예측 분석의 심장에 놓여 있는 규칙 세트, 수식 또는 등식입니다.

그림 1. 모델 너깃

모델링 노드를 실행할 때 해당 모델 너깃이 스트림 캔버스에 배치되는데, 여기서 황금색 마름모꼴 아이콘으로 표시되기 때문에 이름이 "너깃"입니다. 너깃을 열고 내용을 찾아 모델에 대한 세부사항을 볼 수 있습니다. 예측을 보려면 하나 이상의 터미널 노드에 접속하여 예측을 읽을 수 있는 양식으로 제공하는 출력을 실행하십시오.

일반적인 모델링 스트림은 2개의 분기로 구성됩니다. **모델링 분기**는 모델링 노드를 포함하며, 모델링 노드 앞에 소스 및 처리 노드가 있습니다. **스코어링 분기**는 모델링 노드를 실행할 때 작성되며, 예측을 보는 데 사용하는 모델 너깃과 터미널 노드를 포함합니다.

자세한 정보는 *IBM® SPSS® Modeler 모델링 노드 안내서*를 참조하십시오.

⑥ 노드와 스트림에 설명과 주석 추가

사용자의 조직에서 다른 사람에게 스트림을 설명해야 할 수도 있습니다. 이를 위해서 스트림, 노드 및 모델 너깃에 설명을 첨부할 수 있습니다.

그러면 다른 사람은 화면에서 이러한 설명을 보거나 설명을 포함하는 스트림의 이미지를 출력할 수 있습니다.

스트림이나 수퍼노드의 모든 설명을 나열하고, 목록에서 설명의 순서를 변경하고, 주석 텍스트를 편집하고, 설명의 전경 또는 배경 색상을 변경할 수 있습니다. 자세한 정보는 스트림 주석 나열의 내용을 참조하십시오.

스트림 특성 대화 상자, 노드 대화 상자 또는 모델 너깃 창의 주석 탭을 사용하여 노트를 텍스트 주석의 양식으로 스트림, 노드 및 너깃에 추가할 수도 있습니다. 이러한 노트는 주석 탭이 열려 있는 경우에만 표시되지만 스트림 주석은 화면 설명으로도 표시 가능합니다. 자세한 정보는 주석(Annotation)의 내용을 참조하십시오.

가. 설명

주석은 텍스트를 원하는 양만큼 입력할 수 있고 원하는 수만큼의 주석을 추가할 수 있는 텍스트 상자의 양식으로 되어 있습니다. 주석은 독립적이거나(스트림 오브젝트에 첨부되지 않음) 스트림에서 하나 이상의 노드나 모델 너깃에 연결될 수 있습니다. 독립적인 주석은 일반적으로 스트림의 전체 목적을 설명하는 데 사용합니다. 연결된 주석은 이들이 첨부된 노드나 너깃을 설명합니다. 노드와 너깃에는 둘 이상의 주석이 첨부될 수 있으며 스트림에는 독립적인 주석을 원하는 수만큼 가질 수 있습니다.

참고: 스트림 주석(Annotation)을 화면 주석으로 표시할 수도 있지만 이들은 노드나 너깃에는 첨부할 수 없습니다. 자세한 정보는 주석(Annotation)을 주석(Comment)으로 변환의 내용을 참조하십시오.

텍스트 상자의 모양은 다음 테이블에 표시된 대로 주석의 현재 모드(또는 주석으로 표시된 주석(Annotation))를 나타내기 위해 변경됩니다.

표 1. 주석과 주석(Annotation) 텍스트 상자 모드

주석 텍스트 상자	주석(Annotation) 텍스트 상자	모드	표시	획득 방법..
		편집	주석이 편집을 위해 열려 있습니다.	새 주석 또는 주석(Annotation)을 작성하거나 기존의 것을 두 번 클릭합니다.
		마지막 선택	주석은 이동하거나, 크기를 조정하거나 삭제할 수 있습니다.	기존 주석이나 주석(Annotation)을 편집하거나 한 번 클릭한 후에 스트림 배경을 클릭.
		보기	편집이 완료되었습니다.	편집 후에 또 다른 노드, 주석 또는 주석(Annotation) 클릭.

새로운 독립된 주석을 작성할 때 처음에는 스트림 캔버스의 왼쪽 상단 구석에 표시됩니다.

노드나 너깃에 주석을 첨부하는 경우에는 주석은 처음에는 주석이 첨부된 스트림 오브젝트 위에 표시됩니다.

텍스트 상자는 텍스트를 입력할 수 있다는 점을 표시하기 위해 흰색으로 표시되어 있습니다. 텍스트를 입력한 후에는 텍스트 상자 밖을 클릭하십시오. 텍스트 입력이 완료되었음을 보여주기 위해 주석 배경이 노란색으로 표시됩니다. 주석이 선택된 상태로 남아 있어서 이를 이동, 크기 조정 또는 삭제할 수 있습니다.

다시 클릭하면 경계는 편집이 완료되었음을 보여주기 위해 실선으로 변경됩니다.

주석을 두 번 클릭하면 텍스트 상자가 편집 모드로 변경됩니다. 배경이 흰색으로 변경되고 주석 텍스트를 편집할 수 있습니다.

수퍼노드에 주석을 첨부할 수도 있습니다.

ㄱ. 주석 관련 작업

주석에 대해 다수의 조작을 수행할 수 있습니다. 다음을 수행할 수 있습니다.

- 독립된 주석 추가
- 노드 또는 너깃에 주석 첨부
- 주석 편집
- 주석 크기 조정
- 주석 이동
- 주석 연결 해제
- 주석 삭제
- 스트림에 대한 주석 모두 표시 또는 숨기기

독립된 주석 추가 방법

1. 스트림에서 아무것도 선택되어 있지 않은지 확인하십시오.
2. 다음 중 하나를 수행하십시오.
   - 기본 메뉴에서 다음을 클릭하십시오.
     삽입 > 새 주석
   - 스트림 배경을 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 새 주석을 클릭하십시오.
   - 도구 모음에서 **새 주석** 단추를 클릭하십시오.
3. 주석 텍스트를 입력하거나 클립보드에서 텍스트를 붙여넣으십시오.
4. 스트림에서 노드를 클릭하여 주석을 저장하십시오.

노드 또는 너깃에 주석 첨부 방법

1. 스트림 캔버스에서 하나 이상의 노드 또는 너깃을 선택하십시오.
2. 다음 중 하나를 수행하십시오.
   - 기본 메뉴에서 다음을 클릭하십시오.
     삽입 > 새 주석
   - 스트림 배경을 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **새 주석**을 클릭하십시오.
   - 도구 모음에서 **새 주석** 단추를 클릭하십시오.
3. 주석 텍스트를 입력하십시오.

4. 스트림에서 다른 노드를 클릭하여 주석을 저장하십시오.
   또는 다음을 수행할 수 있습니다.

5. 독립된 주석을 삽입하십시오(이전 절 참조).
6. 다음 중 하나를 수행하십시오.

- 주석을 선택하고 F2를 누른 후 노드 또는 너깃을 선택하십시오.
- 노드 또는 너깃을 선택하고 F2를 누른 후 주석을 선택하십시오.
- (3단추 마우스만) 마우스 포인터를 주석 위로 이동하고 가운데 단추를 누른 상태에서 노드나 너깃 위로 마우스 포인터를 끈 후 마우스 단추를 놓으십시오.

추가 노드 또는 너깃에 주석 첨부 방법

주석이 노드나 너깃에 이미 첨부되어 있거나 현재 스트림 수준에 있는데 추가 노드나 너깃에 첨부하려면 다음 중 하나를 수행하십시오.

- 주석을 선택하고 F2를 누른 후 노드 또는 너깃을 선택하십시오.
- 노드 또는 너깃을 선택하고 F2를 누른 후 주석을 선택하십시오.
- (3단추 마우스만) 마우스 포인터를 주석 위로 이동하고 가운데 단추를 누른 상태에서 노드나 너깃 위로 마우스 포인터를 끈 후 마우스 단추를 놓으십시오.

기존 주석 편집 방법

1. 다음 중 하나를 수행하십시오.
   - 주석 텍스트 상자를 두 번 클릭하십시오.
   - 텍스트 상자를 선택하고 Enter를 누르십시오.
   - 텍스트 상자를 마우스 오른쪽 단추로 클릭하여 메뉴를 표시한 후 편집을 클릭하십시오.

2. 주석 텍스트를 편집하십시오. 편집 시 표준 Windows 단축키를 사용할 수 있습니다(예: 복사의 경우 Ctrl+C). 편집 중에 기타 옵션이 주석 팝업 메뉴에 나열됩니다.
3. 텍스트 상자 바깥쪽을 한 번 클릭하여 크기 조정 제어를 표시한 후 다시 주석을 완료하십시오.

주석 텍스트 상자 크기 조정 방법

1. 크기 조정 제어를 표시할 주석을 선택하십시오.
2. 제어를 클릭하여 끌어서 상자 크기를 조정하십시오.
3. 텍스트 상자 바깥쪽을 클릭하여 변경사항을 저장하십시오.

기존 주석 이동 방법

첨부된 오브젝트(있는 경우)는 이동하지 않고 주석을 이동하려면 다음 중 하나를 수행하십시오.

- 마우스 포인터를 주석 위로 이동하고 마우스 왼쪽 단추를 누른 상태에서 주석을 새 위치로 끄십시오.
- 주석을 선택하고, Alt 키를 누른 상태에서 화살표 키를 사용하여 주석을 이동하십시오.

주석이 첨부된 노드나 너깃과 함께 주석을 이동하려면 다음을 수행하십시오.

1. 이동할 오브젝트를 모두 선택하십시오.
2. 다음 중 하나를 수행하십시오.

- 마우스 포인터를 오브젝트 중 하나로 이동하고 마우스 왼쪽 단추를 누른 상태에서 오브젝트를 새 위치로 끄십시오.
- 오브젝트 중 하나를 선택하고, Alt 키를 누른 상태에서 화살표 키를 사용하여 오브젝트를 이동하십시오.

노드 또는 너깃에서 주석 연결 해제 방법

1. 연결을 해제할 하나 이상의 주석을 선택하십시오.
2. 다음 중 하나를 수행하십시오.

- F3을 누르십시오.
- 선택된 주석을 마우스 오른쪽 단추로 클릭하고 해당 메뉴에서 연결 해제를 클릭하십시오.

주석 삭제 방법

1. 삭제할 하나 이상의 주석을 선택하십시오.
2. 다음 중 하나를 수행하십시오.

- Delete 키를 누르십시오.
- 선택된 주석을 마우스 오른쪽 단추로 클릭하고 해당 메뉴에서 삭제를 클릭하십시오.

주석이 노드나 너깃에 첨부된 경우 연결 행도 삭제됩니다.

주석이 원래 독립된 주석으로 변환된 스트림이나 수퍼노드였으면 캔버스에서 주석이 삭제되지만 텍스트는 스트림 또는 수퍼노드의 주석(Annotation) 탭에 유지됩니다.

스트림에 대한 주석 표시 또는 숨기기 방법

1. 다음 중 하나를 수행하십시오.
- 기본 메뉴에서 다음을 클릭하십시오.
  보기 > 주석

- 도구 모음에서 **주석** 표시/**숨기기** 단추를 클릭하십시오.

ㄴ. 스트림 주석 나열

특정 스트림 또는 수퍼노드에 대해 작성된 모든 주석 목록을 볼 수 있습니다.

이 목록에서 다음을 수행할 수 있습니다.

- 주석 순서 변경
- 주석 텍스트 편집
- 주석의 전경 또는 배경색 변경

주석 나열

스트림에 대해 작성된 주석을 나열하려면 다음 중 하나를 수행하십시오.

- 기본 메뉴에서 다음을 클릭하십시오.
  도구 > 스트림 특성 > 주석

- 관리자 분할창에서 스트림을 마우스 오른쪽 단추로 클릭하고 **스트림 특성**, **주석**을 차례로 클릭하십시오.
- 캔버스에서 스트림 배경을 마우스 오른쪽 단추로 클릭하고 **스트림 특성**, **주석**을 차례로 클릭하십시오.

텍스트. 주석 텍스트입니다. 필드를 편집 가능한 텍스트 상자로 변경하려면 텍스트를 두 번 클릭하십시오.

링크. 주석을 첨부할 노드 이름입니다. 이 필드가 비어 있으면 스트림에 주석이 적용됩니다.

위치지정 단추. 선택된 주석을 목록에서 위, 아래로 이동합니다.

주석 색상. 주석의 전경 또는 배경색을 변경하려면 주석을 선택하고 **사용자 정의 색상** 확인 상자를 선택한 후 **배경** 또는 **전경** 목록(또는 둘 다)에서 색상을 선택하십시오. **적용**을 클릭한 후 스트림 배경을 클릭하여 변경 효과를 확인하십시오. **확인**을 클릭하여 변경사항을 저장하십시오.

ㄷ. 주석(Annotation)을 주석(Comment)으로 변환

스트림이나 수퍼노드에 작성된 주석(Annotation)을 주석(Comment)으로 변환할 수 있습니다.

스트림의 경우, 주석(Annotation)은 스트림 캔버스에서 독립된 주석(Comment)(즉, 노드에 첨부되지 않음)으로 변환됩니다.

수퍼노드 주석(Annotation)이 주석(Comment)으로 변환될 때 주석(Comment)은 스트림 캔버스에서 수퍼노드에 첨부되지 않지만 수퍼노드를 확대하면 표시됩니다.

스트림 주석(Annotation)을 주석(Comment)으로 변환하는 방법

1. 도구 메뉴에서 **스트림 특성**을 클릭하십시오. (또는 관리자 분할창에서 스트림을 마우스 오른쪽 단추로 클릭하고 스트림 특성을 클릭하십시오.)
2. **주석(Annotation)** 탭을 클릭하십시오.
3. **주석(Annotation)을 주석(Comment)으로 표시** 선택란을 선택하십시오.
4. **확인**을 클릭하십시오.

수퍼노드 주석(Annotation)을 주석(Comment)으로 변환하는 방법

1. 캔버스에서 수퍼노드 아이콘을 두 번 클릭하십시오.
2. **주석(Annotation)** 탭을 클릭하십시오.
3. **주석(Annotation)을 주석(Comment)으로 표시** 선택란을 선택하십시오.
4. **확인**을 클릭하십시오.

나. 주석(Annotation)

노드, 스트림 및 모델은 여러 가지 방법으로 주석(Annotation)을 달 수 있습니다. 설명 주석(Annotation)을 추가하고 사용자 정의 이름을 추가할 수 있습니다. 이러한 옵션은 프로젝트 분할창에 추가된 스트림에 대한 보고서를 생성할 때 특히 유용합니다. 노드와 모델 너깃의 경우 스트림 캔버스에서 유사한 노드 간에 구별할 수 있도록 도구 팁 텍스트를 추가할 수도 있습니다.

주석 추가

노드나 모델 너깃을 편집하면 다양한 주석(Annotation) 옵션을 설정하는 데 사용하는 주석(Annotation) 탭을 포함하는 탭이 있는 대화 상자가 열립니다. 주석(Annotation) 탭을 직접 열 수도 있습니다.

1. 노드나 너깃에 주석(Annotation)을 달려면 스트림 캔버스에서 노드나 너깃을 마우스 오른쪽 단추로 클릭하고 이름 변경 및 주석달기를 클릭하십시오. 편집 대화 상자는 주석(Annotation) 탭이 표시된 상태로 열립니다.
2. 스트림에 주석(Annotation)을 달려면 도구 메뉴에서 스트림 특성을 클릭하십시오. (또는 관리자 분할창에서 스트림을 마우스 오른쪽 단추로 클릭하고 스트림 특성을 클릭하십시오.) 주석(Annotation) 탭을 클릭하십시오.

이름 사용자 정의를 선택하여 자동 생성된 이름을 조정하거나 스트림 캔버스에 표시되는 노드의 고유 이름을 작성하십시오.

도구 팁 텍스트. (노드 및 모델 너깃에만 해당) 스트림 캔버스에서 도구 팁으로 사용하는 텍스트를 입력합니다. 이는 특히 대량의 유사한 노드에 대해 작업할 때 유용합니다.
키워드. 프로젝트 보고서에 사용하고, 스트림에서 노드를 검색할 때 또는 리포지토리에 저장된 오브젝트를 추적할 때 사용할 키워드를 지정하십시오(정보 참조). 여러 개의 키워드는 세미콜론으로 구분할 수 있습니다(예: income; crop type; claim value). 각 키워드의 시작부와 끝부분에 있는 공백은 잘립니다(예: income ; crop type은 income;crop type과 같은 결과임). (그러나, 키워드 내의 공백은 잘리지 않습니다. 예를 들어, 하나의 공백이 있는 crop type과 두 개의 공백이 있는 crop type은 서로 다릅니다.)

기본 텍스트 영역은 노드의 작업이나 노드에서 결정된 의사결정과 관계없이 긴 주석(Annotation)을 입력하는 데 사용할 수 있습니다. 예를 들어, 스트림을 공유하고 재사용할 때는 필터 노드를 사용하여 여러 공백이 있는 필드를 삭제하는 등 의사결정에 대해 기록을 남겨두는 것이 유용합니다. 노드에 주석을 작성하면 이 정보가 노드와 함께 저장됩니다. 이러한 주석(Annotation)을 프로젝트 분할창에서 작성된 프로젝트 보고서에 포함하도록 선택할 수도 있습니다. 자세한 정보는 프로젝트에 대한 소개의 내용을 참조하십시오.

주석(Annotation)을 주석으로 표시. (스트림과 수퍼노드 주석(Annotation)에만 해당) 주석(Annotation)을 스트림 캔버스에 표시되는 독립된 주석으로 변환하려면 이 상자를 선택하십시오. 자세한 정보는 노드와 스트림에 설명과 주석 추가 주제를 참조하십시오.

ID. 스크립팅이나 자동화를 위해 노드를 참조하는 데 사용할 수 있는 고유한 ID를 표시합니다. 이 값은 노드가 작성될 때 자동으로 생성되고 변경되지 않습니다. 문자 "0"과의 혼동을 피하기 위해 0값은 노드 ID에는 사용하지 않음을 유의하십시오. 오른쪽의 복사 단추를 사용하여 ID를 복사하여 스크립트 또는 다른 필요한 곳에 붙여 넣으십시오.

⑦ 데이터 스트림 저장

스트림을 작성했으면 나중에 사용하기 위해 저장할 수 있습니다.

스트림 저장

1. 파일 메뉴에서 **스트림 저장** 또는 **다른 이름으로 스트림 저장**을 클릭하십시오.
2. 저장 대화 상자에서 스트림 파일을 저장할 폴더를 찾아보십시오.
3. 파일 이름 텍스트 상자에 스트림 이름을 입력하십시오.
4. 저장된 스트림을 현재 프로젝트에 추가하려면 **프로젝트에 추가**를 선택하십시오.

저장을 클릭하면 지정된 디렉토리에 *.str 확장자로 스트림이 저장됩니다.

자동 백업 파일. 스트림이 저장될 때마다 이전에 저장된 파일 버전이 백업으로 자동으로 보존되는데, 파일 이름에 하이픈이 추가됩니다(예: mystream.str-). 백업된 버전을 복원하려면 하이픈을 삭제하고 파일을 다시 열면 됩니다.

가. 상태 저장

스트림에 추가로, 현재 표시된 스트림 다이어그램과 작성한 모든 모델 너깃(관리자 분할창의 모델에 나열됨)을 포함하는 **상태**를 저장할 수 있습니다.

상태 저장

1. 파일 메뉴에서 다음을 클릭하십시오.
   상태 > 상태 저장 또는 다른 이름으로 상태 저장

2. 저장 대화 상자에서 상태 파일을 저장하려는 폴더로 이동하십시오.

저장을 클릭하면 확장자가 *.cst인 상태가 지정된 디렉토리에 저장됩니다.

나. 노드 저장

스트림 캔버스에서 노드를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **노드 저장**을 클릭하여 개별 노드를 저장할 수도 있습니다. 파일 확장자 **.nod*를 사용하십시오.

다. 여러 스트림 오브젝트 저장

스트림, 프로젝트 또는 모델 너깃과 같이 저장되지 않은 여러 오브젝트가 있는 상태로 IBM® SPSS® Modeler를 종료하면 소프트웨어를 완전히 닫기 전에 저장하라는 프롬프트가 표시됩니다. 항목을 저장하기로 선택하면 대화 상자가 각 오브젝트를 저장하기 위한 옵션을 표시합니다.

1. 단순히 저장하려는 오브젝트의 선택란을 선택하십시오.
2. **확인**을 클릭하여 각 오브젝트를 필수 위치에 저장하십시오.

그런 다음 각 오브젝트의 표준 저장 대화 상자가 표시됩니다. 저장을 완료하고 나면 애플리케이션이 닫힙니다.

라. 출력 저장

IBM® SPSS® Modeler 출력 노드에서 생성된 테이블, 그래프 및 보고서는 출력 오브젝트 *(*.cou)* 형식으로 저장할 수 있습니다.

1. 저장하려는 출력을 볼 때 출력 창 메뉴에서 다음을 클릭하십시오.
   파일 > 저장

2. 출력 파일의 이름과 위치를 지정하십시오.
3. 또는 저장 대화 상자에서 **파일을 프로젝트에 추가**를 선택하여 파일을 현재 프로젝트에 포함시키십시오. 자세한 정보는 프로젝트에 대한 소개의 내용을 참조하십시오.

또는 관리자 분할창에 나열된 출력 오브젝트를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **저장**을 선택하십시오.

마. 정보 암호화 및 복호화

스트림, 노드, 프로젝트, 출력 파일 또는 모델 너깃을 저장할 때 암호화하여 무단 사용을 방지할 수 있습니다. 이를 수행하려면 저장 시 추가 옵션을 선택하고 저장되는 항목에 비밀번호를 추가하십시오. 이 암호화는 사용자가 저장하는 항목에 대해 설정할 수 있으며 항목에 추가 보안을 추가합니다. IBM® SPSS® Modeler와 IBM SPSS Modeler Server 간에 파일을 전달하는 경우 사용하는 SSL 암호화와 같지 않습니다.

암호화된 항목을 열려고 하면 비밀번호를 입력하도록 프롬프트가 표시됩니다. 올바른 비밀번호를 입력한 후에 항목이 복호화되어 평소처럼 열립니다.

항목 암호화 방법

1. 저장 대화 상자에서 암호화할 항목에 대해 **옵션**을 클릭하십시오. 암호화 옵션 대화 상자가 열립니다.
2. **이 파일 암호화**를 선택하십시오.
3. 선택적으로, 추가 보안을 위해 **비밀번호 마스크**를 선택하십시오. 이렇게 하면 입력하는 것이 일련의 점으로 표시됩니다.
4. 비밀번호를 입력하십시오. 경고: 비밀번호를 잊어버리면 파일 또는 모델을 열 수 없습니다.
5. **비밀번호 마스크**를 선택했으면 비밀번호를 다시 입력하여 올바로 입력했는지 확인하십시오.
6. **확인**을 클릭하여 저장 대화 상자로 돌아가십시오.

참고: 암호화 보호 항목 사본을 저장하는 경우, 새 항목은 암호화 옵션 대화 상자에서 설정을 변경하지 않는 한 원래 비밀번호를 사용하여 암호화된 형식으로 자동으로 저장됩니다.

⑧ 파일 로드

IBM® SPSS® Modeler에 저장된 다수의 오브젝트를 재로드할 수 있습니다.

- 스트림(.str)
- 상태(.cst)
- 모델(.gm)
- 모델 팔레트(.gen)
- 노드(.nod)
- 출력(.cou)
- 프로젝트(.cpj)

새 파일 열기

파일 메뉴에서 직접 스트림을 로드할 수 있습니다.

- 파일 메뉴에서 **스트림 열기**를 클릭하십시오.

파일 메뉴에서 사용 가능한 하위 메뉴 항목을 사용하여 기타 모든 파일 유형을 열 수 있습니다. 예를 들어, 모델을 로드하려면 파일 메뉴에서 다음을 클릭하십시오.

모델 > 모델 열기 또는 모델 팔레트 로드

최근에 사용한 파일 열기

최근에 사용한 파일을 빨리 로드하기 위해 파일 메뉴 맨 아래에 있는 옵션을 사용할 수 있습니다.

최근에 사용한 파일 목록을 펼치려면 **최근 스트림**, **최근 프로젝트** 또는 **최근 상태**를 선택하십시오.

⑨ 데이터 스트림 맵핑

맵핑 도구를 사용하면 새 데이터 소스를 기존 스트림에 연결할 수 있습니다. 맵핑 도구는 연결을 설정할 뿐만 아니라 새 소스의 필드가 기존 스트림의 필드를 대체하는 방법을 지정하는 데에도 도움을 줍니다. 새 데이터 소스의 전체 데이터를 다시 작성하는 대신에, 간단하게 기존 스트림에 연결할 수 있습니다.

데이터 맵핑 도구를 사용하면 두 개의 스트림 단편을 함께 결합하고 모든 본질적 필드 이름이 적절하게 매치하는지 확인할 수 있습니다. 기본적으로 데이터 맵핑 결과로 단순히 새 필터 노드가 작성되고 이는 이름을 바꿔서 적합한 필드와 매치시킵니다.

데이터를 맵핑하는 데에는 두 개의 동등한 방법이 있습니다.

대체 노드 선택. 이 방법은 대체할 노드로 시작합니다. 먼저 대체할 노드를 마우스 오른쪽 단추로 클릭한 다음 팝업 메뉴에서 **데이터 맵핑** > **대체 노드 선택** 옵션을 사용하여 이를 대체할 노드를 선택하십시오.

맵핑 대상. 이 방법은 스트림에 도입될 노드로 시작합니다. 먼저 도입할 노드를 마우스 오른쪽 단추로 클릭한 다음 팝업 메뉴에서 **데이터 맵핑** > **맵핑 대상** 옵션을 사용하여 결합할 노드를 선택하십시오. 이 방법은 특히 터미널 노드에 맵핑할 때 유용합니다. 참고: 병합 또는 붙여쓰기 노드에는 맵핑할 수 없습니다. 대신 간단하게 스트림을 정상적인 방법으로 병합 노드에 연결해야 합니다.

데이터 맵핑은 스트림 작성에 엄격하게 통합됩니다. 이미 연결이 있는 노드에 연결하려고 시도하면 해당 노드에 대한 연결 또는 맵핑을 대체하는 옵션이 제공됩니다.

가. 데이터를 템플리트에 맵핑

템플리트 스트림의 데이터 소스를 새 소스 노드로 대체하여 사용자 고유의 데이터를 IBM® SPSS® Modeler으로 가져오려면 데이터 맵핑 팝업 메뉴에서 **대체 노드 선택** 옵션을 사용해야 합니다. 이 옵션은 병합, 통합 및 모든 터미널 노드를 제외한 모든 노드에서 사용 가능합니다. 데이터 맵핑 도구를 사용하여 이 조치를 수행하면 기존 스트림 작업과 새 데이터 소스 간에 필드가 제대로 매치하는지 확인하는 데 도움이 됩니다. 다음 단계는 데이터 맵핑 프로세스의 개요를 제공합니다.

1단계: 원본 소스 노드에서 본질적 필드를 지정하십시오. 스트림 작업이 제대로 실행되려면 본질적 필드가 지정되어야 합니다. 자세한 정보는 본질적 필드 지정의 내용을 참조하십시오.

2단계: 새 데이터 소스를 스트림 캔버스에 추가하십시오. 소스 노드 중 하나를 사용하여 새 대체 데이터를 가져오십시오.

3단계: 템플리트 소스 노드를 대체하십시오. 템플리트 소스 노드의 팝업 메뉴에서 데이터 맵핑 옵션을 사용하여 **대체 노드 선택**을 클릭한 다음 대체 데이터의 소스 노드를 선택하십시오.

4단계: 맵핑된 필드를 확인하십시오. 대화 상자가 열리면 소프트웨어가 대체 데이터 소스에서 스트림으로 필드를 제대로 맵핑하 확인하십시오. 맵핑되지 않은 모든 본질적 필드가 빨간색으로 표시됩니다. 이러한 필드는 스트림 작업에 사용하고 다운스트림 작업이 제대로 작동하기 위해서는

새 데이터 소스의 유사한 필드로 대체되어야 합니다. 자세한 정보는 맵핑된 필드 탐색의 내용을 참조하십시오.

대화 상자를 사용하여 모든 본질적 필드가 제대로 맵핑되었는지 확인한 후에는 이전 데이터 소스가 연결이 해제되고 새 데이터 소스가 맵이라는 필터 노드를 사용하여 스트림에 연결됩니다. 이 필터 노드는 스트림에서 필드의 실제 맵핑을 지시합니다. 맵핑 해제 필터 노드는 또한 스트림 캔버스에 포함됩니다. *맵핑 해제* 필터 노드는 필드 이름 맵핑을 스트림에 추가하여 이를 역전시키는 데 사용할 수 있습니다. 이는 맵핑된 필드를 실행 취소하지만 필드와 오버레이를 다시 선택하기 위해 모든 다운스트림 터미널 노드를 편집해야 함을 유의하십시오.

나. 스트림 사이에 맵핑

노드 연결과 유사하게, 이 데이터 맵핑 방법에서는 본질적 필드를 미리 설정할 필요가 없습니다. 이 방법을 사용하면 데이터 맵핑 팝업 메뉴에서 **맵핑 대상**을 사용하여 한 스트림에서 다른 스트림으로 간단하게 연결할 수 있습니다. 이 데이터 맵핑 유형은 터미널 노드에 맵핑하고 스트림 간에 복사하고 붙여넣는 데 유용합니다. *참고*: **맵핑 대상** 옵션을 사용하면 병합, 붙여쓰기 및 모든 소스 노드 유형으로 맵핑할 수 없습니다.

스트림 간에 데이터 맵핑

1. 새 스트림에 연결하기 위해 사용하려는 노드를 마우스 오른쪽 단추로 클릭하십시오.
2. 메뉴에서 다음을 클릭하십시오.
   데이터 맵핑 > 맵핑 대상

3. 목표 스트림에서 커서를 사용하여 대상 노드를 선택하십시오.
4. 대화 상자가 열리면 해당 필드가 적절하게 매치하는지 확인하고 **확인**을 클릭하십시오.

다. 본질적 필드 지정

기존 스트림에 맵핑할 때 본질적 필드는 일반적으로 스트림 작성자가 지정합니다. 이러한 본질적 필드는 특정 필드가 다운스트림 작업에 사용하는지 여부를 표시합니다. 예를 들어, 기존 스트림은 *Churn*이라는 필드를 사용하는 모델을 작성할 수 있습니다. 이 스트림에서 *Churn*은 이 필드 없이는 모델을 작성할 수 없으므로 본질적 필드입니다. 마찬가지로, 파생 노드처럼 조작 노드에 사용하는 필드는 새 필드를 파생하는 데 필요합니다. 이러한 필드를 본질적인 것으로 명시적으로 설정하면 새 소스 노드에서 적합한 필드가 여기로 맵핑됩니다. 필수 필드가 맵핑되지 않으면 오류 메시지를 받게 됩니다. 특정 조작 또는 출력 노드가 불필요하다고 결정하는 경우에는 스트림에서 노드를 삭제하고 본질적 필드 목록에서 적합한 필드를 제거할 수 있습니다.

본질적 필드 설정

1. 대체될 템플리트 스트림의 소스 노드를 마우스 오른쪽 단추로 클릭하십시오.
2. 메뉴에서 다음을 클릭하십시오.
   데이터 맵핑 > 본질적 필드 지정

3. 필드 선택기를 사용하여 목록에서 필드를 추가하거나 제거할 수 있습니다. 필드 선택기를 열려면 필드 목록 오른쪽의 아이콘을 클릭하십시오.

라. 맵핑된 필드 탐색

하나의 데이터 스트림 또는 데이터 소스가 또 다른 데이터 스트림이나 데이터 소스로 맵핑되는 위치를 선택하면 맵핑을 위한 필드를 선택하거나 시스템 기본값 맵핑이 올바른지 확인하기 위한 대화 상자가 표시됩니다. 스트림이나 데이터 소스에 기본 필드가 설정되었고 이들이 매치하지 않은 경우에는 이러한 필드는 빨간색으로 표시됩니다. 데이터 소스에서 맵핑되지 않은 필드는 변경되지 않은 필터 노드를 통과하지만 비필수 필드 또한 맵핑할 수 있음을 유의하십시오.

원래값. 템플리트의 모든 필드 또는 기존 스트림의 현재 추가적 다운스트림인 모든 필드를 나열합니다. 새 데이터 소스의 필드는 이러한 필드로 맵핑됩니다.

맵핑됨. 템플리트 필드에 맵핑하기 위해 선택된 필드를 나열합니다. 이들은 이름을 스트림 작업에 사용된 원래 필드와 매치하도록 변경해야 할 수도 있는 필드입니다. 필드가 사용 가능한 필드 목록을 활성화하기 위해 테이블 셀을 클릭하십시오.

어떤 필드를 맵핑할지 확실하지 않으면 맵핑 전에 소스 데이터를 자세히 검사하는 것이 유용할 수 있습니다. 예를 들어, 소스 노드에서 유형 탭을 사용하여 소스 데이터의 요약을 검토할 수 있습니다.

(3) 팁 및 단축키

다음 단축 아이콘과 팁에 익숙해지면 빠르고 쉽게 작업할 수 있습니다.

- **두 번 클릭하여 스트림을 빠르게 작성.** 팔레트에서 노드를 두 번 클릭하기만 하면 현재 스트림에 추가하고 연결할 수 있습니다.
- **다운스트림 노드를 선택하기 위해 키 조합 사용.** Ctrl+Q 및 Ctrl+W를 눌러 모든 노드 다운스트림 선택을 토글할 수 있습니다.
- **단축키를 사용하여 노드에 연결 및 연결 해제.** 캔버스에서 노드가 선택되었을 때 연결을 시작하려면 F2를 누르고, 필수 노드로 이동하려면 탭을 누르고, 연결을 완료하려면 Shift+스페이스바를 누르십시오. 선택한 노드에 대한 모든 입력과 출력을 연결 해제하려면 F3을 누르십시오.

- **노드 팔레트 탭을 즐겨찾는 노드로 사용자 정의.** 도구 메뉴에서 **팔레트 관리**를 클릭하여 노드 팔레트에 표시된 노드를 추가, 제거 또는 이동하기 위한 대화 상자를 여십시오.
- **노드 이름 변경 및 도구 팁 추가.** 각 노드 대화 상자에는 캔버스에서 노드에 사용자 정의 이름을 지정할 수 있는 주석(Annotation) 탭뿐만 아니라 스트림을 구성하는 데 도움을 주는 도구 팁이 포함됩니다. 진행률을 추적하고, 프로세스 세부사항을 저장하고, 비즈니스 의사결정을 필수 또는 달성으로 표시하기 위해 긴 주석(Annotation)을 포함할 수도 있습니다.
- **CLEM 표현식에 자동으로 값 삽입.** 다양한 대화 상자(예: 파생 및 채움 노드의 대화 상자)에서 액세스 가능한 표현식 작성기를 사용하여 필드 값을 CLEM 표현식에 자동으로 삽입할 수 있습니다. 표현식 작성기의 값 단추를 클릭하여 기존 필드 값에서 선택하십시오.

  그림 1. 값 단추

- **신속하게 파일 찾아보기.** 열기 대화 상자에서 파일을 찾아볼 때, 파일 목록(찾아보기 필드 옆의 대화 상자 맨 위에 있는 노란색 다이아몬드 단추 클릭)을 사용하여 이전에 사용된 디렉토리뿐만 아니라 IBM® SPSS® Modeler 기본 디렉토리에 액세스하십시오. 전진 및 이전 단추를 사용하여 액세스된 디렉토리를 통해 스크롤하십시오.
- **출력 창 클러터 최소화.** 모든 출력 창 오른쪽 상단 구석에 있는 빨간색 X 단추를 사용하여 출력을 빠르게 닫고 삭제할 수 있습니다. 그러면 관리자 분할창의 출력 탭에 유망하거나 흥미로운 결과만을 유지할 수 있습니다.

키보드 단축키의 전체 범위를 소프트웨어에 사용 가능합니다. 자세한 정보는 내게 필요한 옵션의 키보드 기능의 내용을 참조하십시오.

다음을 수행할 수 있습니다.

- 마우스를 사용하여 스트림 캔버스에서 노드 그룹을 끌고 선택합니다.
- 하나의 스트림에서 다른 스트림으로 노드를 복사하고 붙여넣을 수 있습니다.
- 모든 대화 상자와 출력 창에서 도움말에 액세스합니다.
- CRISP-DM(Cross-Industry Standard Process for Data Mining)에 대한 도움말을 얻습니다. (도움말 메뉴에서 **CRISP-DM 도움말**을 클릭하십시오.)

5) 데이터에 대한 작업

데이터 노드를 마우스 오른쪽 단추로 클릭한 후 **데이터 보기**를 선택하여 스트림 데이터를 탐색할 수 있습니다. 열리는 창에서 다음 작업을 수행할 수 있습니다.
- **차트** 탭에서는 고급 데이터 시각화를 작성하여 다른 퍼스펙티브에서 데이터를 탐색하고 데이터 내에서 패턴, 연결 및 관계를 식별할 수 있습니다. 전체 세부사항은 차트 작성의 내용을 참조하십시오.
- **스프레드시트** 탭에서는 데이터의 읽기 전용 보기를 표 형식으로 표시할 수 있습니다.

- **데이터 검토** 탭에는 각 열에 대한 빈도와 통계가 표시됩니다.
- **대시보드** 탭에서 **레이아웃 디자인 시작**을 클릭하고 여러 개의 차트를 단일 페이지에서 보기 위한 레이아웃을 작성할 수 있습니다. 레이아웃을 템플리트로 저장하고 저장된 차트를 레이아웃의 위치로 끌어다 놓을 수 있습니다.
- **기본 설정** 탭에서는 언어 및 모양과 느낌 등의 사용자 인터페이스 기본 설정을 지정할 수 있습니다.

이 기능은 기본적으로 포트 28900을 사용합니다. 다른 포트를 사용해야 하는 경우 options.cfg 파일에서 data_view_port_number 구성 설정의 값을 변경하십시오.

> **참고:** 이 **데이터 보기** 기능은 현재 SPSS Modeler의 Subscription 버전에서 사용할 수 없습니다.

(1) 차트 작성

차트 탭에서는 미리 정의된 갤러리 차트 또는 개별 부분(예: 축, 막대)을 통해 차트를 작성할 수 있습니다. 차트를 작성하려면 제공된 차트 유형 옵션에서 갤러리 차트 유형 또는 기본 요소를 선택하십시오. 설명은 레이아웃 및 용어의 내용을 참조하십시오.

차트를 작성할 때 해당 차트의 미리보기가 캔버스에 표시됩니다. 미리보기에는 실제 데이터를 대표하는 실제 변수 레이블과 측정 수준이 사용됩니다.

초보자라면 차트를 작성할 때 갤러리를 사용하는 것이 좋습니다. 갤러리 사용에 대한 자세한 정보는 차트 유형 갤러리에서 차트 작성의 내용을 참조하십시오.

차트 작성기 시작

- 작업하려는 데이터 노드를 마우스 오른쪽 단추로 클릭한 후 **데이터 보기**를 클릭하십시오.

① 차트 카탈로그

차트 카탈로그는 차트 범주를 탐색하거나 키워드 검색을 사용하거나 차트 특성으로 필터링하여 적합한 차트를 찾는 옵션을 제공합니다. 차트 카탈로그는 활성 데이터 세트 탭에서 액세스할 수 있습니다.

활성 데이터 세트 탭에서 도구 모음에서 사용 가능한 **차트 작성기** 컨트롤을 클릭하십시오. 또는 **시각화 > 차트 작성기...**를 통해 카탈로그에 액세스하십시오. 다양한 방법으로 적합한 차트 유형을 검색할 수 있습니다.

- **차트 유형 찾기 필드**에서는 특정 차트와 키워드를 검색할 수 있습니다. 차트 이름 또는 키워드를 입력하면 사용 가능한 차트 타일 수가 필터링되어 지정된 차트 이름 또는 키워드에 대한 결과만 제공됩니다.

> ⓘ **참고: 차트 유형 찾기 필드** 옆에 있는 타일 보기 및 목록 보기 아이콘을 사용하여 차트 유형이 차트 카탈로그에 표시되는 방식을 제어할 수 있습니다.

- 차트 카탈로그 타일은 해당 범주에 따라 그룹화됩니다. 원하는 차트 유형을 찾을 때까지 사용 가능한 타일을 스크롤할 수 있습니다.
- **필터** 분할창은 해당 범주, 목적, 종속 변수 및 통계에 따라 차트를 필터링하는 옵션을 제공합니다.

> ⓘ **참고:**
> - 각 **필터** 분할창 머리말 옆에 있는 숫자는 각 섹션에 대해 현재 선택된 필터의 수를 나타냅니다.
> - 필터 분할창 설정을 지우려면 필터 재설정을 클릭하십시오.

② 레이아웃 및 용어

시작 화면

차트 작성기를 실행하면 차트 유형을 선택하거나 활성 데이터 세트에서 열을 선택하는 옵션이 표시됩니다. 시각화할 열을 추가하면 차트 유형 옵션이 업데이트되어 선택한 열에 권장되는 차트 유형이 표시됩니다.

캔버스

캔버스는 차트를 작성하는 데 사용되는 차트 작성기 대화 상자의 영역입니다.

도표 유형

사용 가능한 도표 유형이 나열됩니다. 그래픽 요소는 데이터를 나타내는 차트 내 항목입니다. 이러한 항목으로는 막대, 점, 선 등이 있습니다.

세부사항 분할창

세부사항 분할창은 기본 차트 작성 블록을 제공합니다.

차트 설정

차트를 작성하는 데 사용되는 변수, 분포 방법, 제목 및 부제목 필드 등을 선택하는 옵션을 제공합니다. 선택한 차트 유형에 따라 세부사항 분할창에 제공된 옵션이 달라집니다. 차트 유형별 사용 가능한 옵션에 대한 자세한 정보는 차트 유형의 내용을 참조하십시오.

조치

차트 구성 파일 다운로드, 차트를 이미지 파일로 다운로드, 차트 재설정 및 글로벌 기본 설정 지정에 대한 옵션을 제공합니다.

③ 차트 유형 갤러리에서 차트 작성

차트를 작성하는 가장 쉬운 방법은 차트 유형 갤러리를 사용하는 것입니다. 다음은 갤러리를 사용하여 차트를 작성하는 일반적인 단계입니다. 다른 차트 유형에 대한 내용은 차트 유형의 내용을 참조하십시오.

1. **차트 유형** 섹션에서 차트 범주를 선택하십시오. 선택한 차트 유형의 미리보기 버전이 차트 캔버스에 표시됩니다.
2. 캔버스에 이미 차트가 표시된 경우 새 차트가 차트의 축 세트와 그래픽 요소를 대체합니다.
   a. 선택한 차트 유형에 따라 사용 가능한 변수는 세부사항 분할창에서 서로 다른 여러 머리말(예: 막대형 차트의 경우 **범주**, 선형 차트의 경우 **X축** 및 **Y축**) 아래에 표시됩니다. 선택한 차트 유형에 적합한 변수를 선택하십시오.

④ 차트 유형

갤러리에는 가장 일반적으로 사용되는 차트의 콜렉션이 포함되어 있습니다. 다음과 같은 차트가 있습니다.

3차원 차트
단순. 자세한 정보는 3차원 차트의 내용을 참조하십시오.
막대형 차트
단순, 수직누적, 수평배열. 자세한 정보는 막대형 차트의 내용을 참조하십시오.
상자도표
단순 및 수평배열. 자세한 정보는 상자도표의 내용을 참조하십시오.
거품형
단순. 자세한 정보는 거품형 차트의 내용을 참조하십시오.
원통형 차트
단순. 자세한 정보는 원통형 차트의 내용을 참조하십시오.
원 채우기
단순. 자세한 정보는 원 채우기 차트의 내용을 참조하십시오.
사용자 정의 차트
자세한 정보는 사용자 정의 차트의 내용을 참조하십시오.
이중 Y축 차트
단순. 자세한 정보는 이중 Y축 차트의 내용을 참조하십시오.
오차 막대 차트
단순. 자세한 정보는 오차 막대 차트의 내용을 참조하십시오.
평가 차트
단순. 자세한 정보는 평가 차트의 내용을 참조하십시오.
히트 맵
단순. 자세한 정보는 히트 맵 차트의 내용을 참조하십시오.

히스토그램

단순, 수직누적, 빈도 다각형. 자세한 정보는 히스토그램 차트의 내용을 참조하십시오.

선형 차트

단순. 자세한 정보는 선형 차트의 내용을 참조하십시오.

맵 차트

단순. 자세한 정보는 맵 차트의 내용을 참조하십시오.

수학 곡선 차트

단순. 자세한 정보는 수학 곡선 차트의 내용을 참조하십시오.

다중 차트

단순. 자세한 정보는 다중 차트의 내용을 참조하십시오.

다중 계열 차트

단순. 자세한 정보는 다중 계열 차트의 내용을 참조하십시오.

평형 차트

단순. 자세한 정보는 평형 차트의 내용을 참조하십시오.

원형 차트

단순. 자세한 정보는 원형 차트의 내용을 참조하십시오.

인구 피라미드 차트

단순. 자세한 정보는 인구 피라미드 차트의 내용을 참조하십시오.

Q-Q 도표

단순. 자세한 정보는 Q-Q 도표의 내용을 참조하십시오.

방사형 차트

단순. 자세한 정보는 방사형 차트의 내용을 참조하십시오.

관계 차트

단순. 자세한 정보는 관계 차트의 내용을 참조하십시오.

산점도 및 점도표

1차원, 단순, 그룹, 오버레이 및 3차원 산점도(요약 점도표, 1차원 점도표, 하락-선 차트). 자세한 정보는 산점도 및 점도표의 내용을 참조하십시오.

산점도 행렬 차트

단순. 자세한 정보는 산점도 행렬 차트의 내용을 참조하십시오.

햇살 차트

단순. 자세한 정보는 햇살 차트의 내용을 참조하십시오.

시간 도표

단순. 자세한 정보는 시간 도표의 내용을 참조하십시오.

테마 리버

단순. 자세한 정보는 테마 리버 차트의 내용을 참조하십시오.

트리 차트

단순. 자세한 정보는 트리 차트의 내용을 참조하십시오.

트리 맵 차트

단순 및 햇살. 자세한 정보는 트리 맵 차트의 내용을 참조하십시오.

t-SNE 차트

단순. 자세한 정보는 t-SNE 차트의 내용을 참조하십시오.

단어 클라우드

단순. 자세한 정보는 단어 클라우드 차트의 내용을 참조하십시오.

가. 3차원 차트

3차원 차트는 일반적으로 다중 변수 함수를 나타내는 데 사용되며, X축 및 Y축 변수의 함수인 Z축 변수를 포함합니다.

단순 3차원 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **3차원** 아이콘을 클릭하십시오.

캔버스가 업데이트되어 3차원 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 차트 **유형**을 선택하십시오.
3. 드롭 다운 목록에서 **X축** 변수를 선택하십시오.
4. 드롭 다운 목록에서 **Y축** 변수를 선택하십시오.
5. 드롭 다운 목록에서 **Z축** 변수를 선택하십시오.

추가 기능

유형

데이터를 표시하는 데 사용할 수 있는 차트 유형이 나열됩니다.

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Y축

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Z축

차트의 z축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

도구 팁 정보

데이터 점 위로 마우스를 이동할 때 표시되는 도구 팁 정보를 생성하는 데 사용할 수 있는 변수가 나열됩니다.

색상 맵

사용 가능한 색상 맵 변수가 나열됩니다. 이 변수는 색상 진행(지정된 열의 값 범위를 기준으로 함)을 사용하여 플롯 점에 자신을 표시합니다. 색상 맵은 코로플레스 맵이라고도 합니다.

크기 맵

사용 가능한 크기 맵 변수가 나열됩니다. 이 변수는 서로 다른 크기를 사용하여 플롯 점에 자신을 표시합니다.

Z 비율

x축과 Y축에 상대적인 z축 데이터 값의 척도를 설정합니다.

회전

차트 회전을 사용 및 사용 안함으로 설정합니다.

데이터 점 도구 팁

데이터 점 도구 팁이 표시되는 위치(데이터 점의 오른쪽, 차트의 오른쪽 상단, 숨기기)를 제어합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

나. 막대형 차트

막대형 차트는 범주형 변수 요약 시 유용합니다. 예를 들어 막대형 차트를 사용하여 설문조사에 참여한 남자 및 여자 인원 수를 표시하거나, 남자의 평균 급여 및 여자의 평균 급여를 표시할 수 있습니다.

단순 막대형 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **막대** 아이콘을 클릭하십시오.

   Bar •

   캔버스가 업데이트되어 막대형 차트 템플리트가 표시됩니다.

2. 범주형(명목 또는 순서) 변수를 **범주** 변수로 선택하십시오. 척도변수를 사용할 수는 있지만 소수의 특별한 케이스에서만 결과가 유용합니다. 막대형 차트는 제한된 수의 고별된 값이 있는 경우에 가장 적절하게 표시됩니다. 척도를 **범주** 축으로 사용하여 막대형 차트를 작성할 경우 각 막대가 정확한 값에서 그려지고 막대가 다른 연속 값과 겹칠 수 없기 때문에 막대가 매우 얇습니다.
3. **요약** 목록에서 통계를 선택하십시오. 통계 결과에 따라 막대 높이가 결정됩니다. 원하는 통계가 **요약** 목록에 나타나지 않을 경우 변수가 필요할 수 있습니다. **값** 목록에서 변수를 선택하고 통계가 지금 사용 가능한지 확인하십시오. 기타 차트 유형 제한사항이 있을 수 있습니다. 예를 들어 오차 막대 차트는 특정 통계에 대해서만 계산할 수 있습니다.

추가 기능

범주

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

순서 기준

변수 내의 범주에 대한 정렬 옵션을 선택하십시오.

범주 이름

변수의 범주를 정렬하는 데 범주 레이블을 사용합니다. 이는 차트에 표시되는 레이블로, 일반적으로 눈금 또는 범례 레이블입니다.

범주 값

변수의 범주를 정렬하는 데 데이터 세트에 저장된 값을 사용합니다. 범주 값은 데이터 세트에서 범주를 식별하는 데 사용됩니다. 일반적으로 이 값은 레이블과는 다르며 구체적인 설명을 제공하지 않아도 됩니다. 예를 들어, 값은 1 같은 숫자가 될 수 있는 반면, 레이블은 여성과 같이 범주에 대한 텍스트 설명입니다.

범주 순서

변수 범주의 정렬 순서를 선택하십시오.

읽는 대로

변수 범주가 데이터 세트에 나타나는 대로 표시됩니다.

오름차순

변수 범주가 오름차순으로 정렬됩니다.

내림차순

변수 범주가 내림차순으로 정렬됩니다.

요약

그래픽 요소에 대한 통계 요약함수를 선택하십시오. 통계 결과에 따라 y축에서 그래픽 요소의 위치가 결정됩니다. 2차원 차트의 경우 축의 각 x값에 대해 통계가 계산됩니다. 3차원 차트의 경우 x축 및 z축의 교차 값에 대해 통계가 계산됩니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 값으로 사용할 변수를 선택하십시오.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오. 그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

분할 유형

분할 기준 변수를 선택한 경우 생성되는 범주 막대를 수직누적 또는 수평배열로 표시하도록 선택할 수 있습니다. 군집 및 누적이 차트 내에서 차원으로 추가됩니다. 군집은 한 막대를 여러 막대로 분할하고, 누적은 각 막대에 여러 세그먼트를 작성합니다. 누적의 경우 정확한 통계를 선택해야 합니다. 값을 함께 추가하면(누적) 적절한 결과가 제공되어야 합니다. 예를 들어 평균값을 더하거나 누적하는 것은 일반적으로 의미가 없습니다.

막대 유형

제공된 옵션에서 막대형 차트 유형을 선택하십시오.

- X축
- Y축
- X축 반전
- Y축 반전
- 극좌표 각도 축
- 극반경 축
- 극도표-레인보우

레이블 위치

드롭 다운 메뉴에서 차트의 레이블 위치를 선택하십시오.

- 없음
- 위쪽
- 왼쪽
- 오른쪽
- 아래쪽
- 내부
- insideLeft
- insideRight
- insideTop
- insideBottom
- insideTopLeft
- insideBottomLeft
- insideTopRight
- insideBottomRight

참조선 표시

이 전환 컨트롤은 차트에 참조선 표시를 사용/사용 안함으로 설정합니다. 사용가능한 옵션은 **최소값**, **최대값** 및 **평균**이며, 차트의 최소값, 최대값 및 평균값에 각각 참조선을 표시합니다.

참조선 값 입력

참조선 표시가 사용으로 설정된 경우 이 설정을 사용하여 특정 참조선 값을 지정할 수 있습니다. 추가 참조선 값을 지정하려면 **다른 열 추가**를 클릭하십시오.

전치

사용으로 설정할 경우 차트의 x축과 y축이 전치됩니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

다. 상자도표

상자도표는 5개의 통계(최소값, 첫 번째 사분위수, 중위수, 세 번째 사분위수 및 최대값)를 표시합니다. 척도변수의 분포를 표시하고 이상값을 정확히 표시하는 데 유용합니다.

단순 상자도표 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **상자도표** 아이콘을 클릭하십시오.

   Box plot

   캔버스가 업데이트되어 상자도표 차트 템플리트가 표시됩니다.

2. 하나 이상의 척도변수를 **열** 변수로 선택하십시오.

> ⓘ **참고**: 점도표에 대한 통계는 상자도표입니다. 이는 변경할 수 없습니다.

추가 기능

열

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다. 열을 더 추가하려면 **다른 열 추가**를 클릭하십시오.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오. 그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

범주 순서

변수 범주의 정렬 순서를 선택하십시오.

읽는 대로

변수 범주가 데이터 세트에 나타나는 대로 표시됩니다.

오름차순

변수 범주가 오름차순으로 정렬됩니다.

내림차순

변수 범주가 내림차순으로 정렬됩니다.

데이터 표준화

사용으로 설정할 경우 이 설정은 다중 데이터 세트 또는 다중 열의 데이터를 비교할 수 있도록 데이터를 정규 분포로 변환합니다. 이 설정은 빈도 및 변환 통계에 대한 100% 수직누적을 퍼센트로 작성합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

라. 거품형 차트

거품형 차트에서는 그룹의 범주를 계층 구조가 아닌 원으로 채워 표시합니다. 각 원(거품)의 크기는 해당 값에 비례합니다. 거품형 차트는 데이터의 관계를 비교하는 데 유용합니다.

단순 거품형 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **거품형** 아이콘을 클릭하십시오.

   Bubble

   캔버스가 업데이트되어 거품형 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 **열** 변수를 선택하십시오.

> ⓘ **참고**: 열 변수를 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

추가 기능

그룹 색상
그룹 색상을 켜거나 끕니다.
기본 제목
차트 제목입니다.
부제목
차트 제목 바로 아래에 표시되는 차트 부제목입니다.
꼬리말
차트 아래에 표시되는 차트 꼬리말입니다.

마. 원통형 차트

원통형 차트는 재무 차트의 한 스타일로, 유가증권, 파생상품 또는 통화의 가격 변화를 설명하는 데 사용됩니다. 각 원통형 요소는 일반적으로 1일을 나타냅니다. 따라서 한 달 차트에는 거래일 20일이 20개의 원통형 요소로 표시될 수 있습니다. 원통형 차트는 주식 및 통화 가격 패턴 분석에 가장 자주 사용되며 상자도표와 유사합니다.

원통형 차트를 작성하는 데 사용되는 데이터 세트에는 표시하려는 각 기간에 대한 고가, 저가 및 종가가 포함되어 있어야 합니다.

단순 원통형 차트 작성

1. 도표 작성기의 **도표 유형** 섹션에서 **원통형 차트** 아이콘을 클릭하십시오.
   Candlestick

   캔버스가 업데이트되어 원통형 차트 템플리트가 표시됩니다.

2. 변수를 **X축** 변수로 선택하십시오.
3. 변수를 **고가** 변수로 선택하십시오.
4. 변수를 **저가** 변수로 선택하십시오.

추가 기능

X축
차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.
고가
차트의 고가 값에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

상한 필드 요약

선택한 상한 변수에 대한 통계 요약함수를 선택하십시오.

저가

차트의 저가 값에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

시가

차트의 시가 값에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

종가

차트의 종가 값에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

볼륨

차트의 볼륨 막대에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

범주 순서

변수 범주의 정렬 순서를 선택하십시오.

읽는 대로

변수 범주가 데이터 세트에 나타나는 대로 표시됩니다.

오름차순

변수 범주가 오름차순으로 정렬됩니다.

내림차순

변수 범주가 내림차순으로 정렬됩니다.

원통형

차트 데이터를 원통형 또는 선형으로 표시하도록 토글합니다.

이동 평균

이동 평균을 차트에 표시하기 위한 옵션을 제공합니다. 사용 가능한 옵션은 5, 10, 20, 30, 60 및 120입니다.

색상 전환

이 전환 컨트롤은 증가 색상과 감소 색상을 전환합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

바. 원 채우기 차트

원 채우기 차트는 계층 구조 데이터를 중첩 영역 세트로 표시하므로, 대량의 계층 구조 데이터를 시각화할 수 있습니다. 이 도표는 나무 지도와 비슷하지만 직사각형 대신 원을 사용합니다. 원 채우기 차트에서는 포함(중첩)를 사용하여 계층 구조 데이터를 표시합니다.

단순 원 채우기 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **시간 도표** 아이콘을 클릭하십시오.

   Circle packing

   캔버스가 업데이트되어 원 채우기 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 **열** 변수를 선택하십시오.

> ⓘ **참고:** 열 변수를 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

추가 기능

그룹 색상

그룹 색상을 켜거나 끕니다.

요약

통계 요약함수를 선택하십시오. 각 범주를 요약하는 데 사용되는 방법입니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 값 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 값으로 사용할 변수를 선택하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

사. 사용자 정의 차트

사용자 정의 차트 옵션으로 JSON 코드를 붙여넣거나 편집하여 원하는 차트를 작성할 수 있습니다.

사용자 정의 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **사용자 정의** 아이콘을 클릭하십시오.
   Customized

2. 차트 지정 사항이 포함된 JSON 코드를 세부사항 분할창의 제공된 필드에 붙여넣으십시오.
3. **차트 생성**을 클릭하십시오.

아. 이중 Y축 차트

이중 *y*축 차트를 사용하여 서로 다른 도메인을 사용하는 두 개의 *y*축 변수를 요약하거나 차트로 작성할 수 있습니다. 예를 들어 케이스 수는 특정 축에, 평균 급여는 다른 축에 도표로 작성할 수 있습니다. 또한 이 차트는 이전에 설명한 다양한 차트 유형이 이중 *y*축 차트에 포함되도록 다양한 그래픽 요소가 혼합된 것일 수 있습니다. 개수를 선으로 표시하고 각 범주의 평균을 막대로 표시할 수 있습니다.

단순 이중 Y축 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **이중 Y축** 아이콘을 클릭하십시오.
   Dual Y-axes

   캔버스가 업데이트되어 이중 Y축 차트 템플리트가 표시됩니다.
2. 변수를 **X축** 변수로 선택하십시오.
3. 첫 번째 **Y축** 변수에 대한 변수를 선택한 후 변수를 나타내는 차트 유형(**막대**, **선**, **산점도**)을 선택하십시오.
4. 두 번째 **Y축** 변수에 대한 변수를 선택한 후 변수를 나타내는 차트 유형(**막대**, **선**, **산점도**)을 선택하십시오.

> ⓘ **참고**: 위로 및 아래로 화살표 컨트롤을 사용하여 y축 순서를 변경할 수 있습니다.

추가 기능

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Y축

차트의 이중 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

요약

사용으로 설정할 경우 각 범주의 요약 방법을 선택할 수 있는 옵션이 표시됩니다.

왼쪽 Y축 요약

차트의 왼쪽에 표시될 Y축에 대한 요약 방법을 설정합니다. 옵션은 **합계**, **평균**, **최대값**, **최소값**입니다.

오른쪽 Y축 요약

차트의 오른쪽에 표시될 Y축에 대한 요약 방법을 설정합니다. 옵션은 **합계, 평균, 최대값, 최소값**입니다.

데이터 표준화

사용으로 설정할 경우 데이터가 정규 분포로 변환됩니다. 이 경우 여러 데이터 세트 또는 열의 데이터를 쉽게 비교할 수 있습니다.

두 번째 축선

사용으로 설정할 경우 차트의 두 번째 축선이 표시됩니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

자. 오차 막대 차트

오차 막대 차트는 데이터의 변동을 나타내며 보고된 측정치로 오차(불확실성)를 표시합니다. 오차 막대로 차이가 통계적으로 유의적인지 판별할 수 있습니다. 오차 막대는 제공된 함수의 적합도를 제시할 수도 있습니다.

단순 오차 막대 차트 작성

1. 도표 작성기의 **도표 유형** 섹션에서 **오차 막대** 아이콘을 클릭하십시오.

   Error bar

   캔버스가 업데이트되어 오차 막대 차트 템플리트가 표시됩니다.

2. 척도변수를 **범주** 변수로 선택하십시오. 이 변수의 데이터는 x축에 표시됩니다.
3. 변수를 **Y축** 변수로 선택하십시오. 이 변수의 데이터는 y축에 표시됩니다.

추가 기능

범주

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Y축

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

범주 순서

변수 범주의 정렬 순서를 선택하십시오.

읽는 대로

변수 범주가 데이터 세트에 나타나는 대로 표시됩니다.

오름차순

변수 범주가 오름차순으로 정렬됩니다.

내림차순

변수 범주가 내림차순으로 정렬됩니다.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오. 그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

참조선

사용으로 설정할 경우 차트에 참조선이 표시됩니다. 참조선은 선택한 **통계 방법**과 상관됩니다.

오차 막대

사용으로 설정할 경우 오차 범위를 나타내는 선이 차트에 표시됩니다.

측도

오차 막대로 표시할 측도 유형을 선택하십시오.

신뢰구간

선택한 변수에 대한 신뢰구간을 설정합니다. 기본값은 0.95(95%)이며, **표시 값** 필드에 적용된 것과 같습니다.

표준 오차

선택한 변수의 표준 오차를 측정합니다.

표준 편차

선택한 변수의 표준 편차를 측정합니다.

신뢰수준

이 값은 선택한 **측도**에 대한 신뢰구간을 나타냅니다. 기본값은 0.95(95%)입니다.

통계 방법

중심 경향 설명 방법을 선택하십시오.

평균

비율을 합산한 후 그 결과를 총 비율 수로 나눈 결과입니다.

중앙값

이 값보다 작은 비율의 수와 이 값보다 큰 비율의 수가 같은 값입니다.

표시 모드

통계 방법 선택사항의 표시 방법(막대, 선, 원)을 선택하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

차. 평가 차트

평가 차트는 히스토그램 또는 콜렉션 그래프와 비슷합니다. 평가 차트는 특정 결과 예측 시 모델이 작동하는 방식을 보여줍니다. 평가 차트는 예측의 신뢰도 및 예측값을 기반으로 레코드를 정렬하고 레코드를 동일한 크기의 그룹(**분위수**)으로 분할한 후 각 분위수에 대한 기준의 값을 내림차순으로 도표로 작성하여 작동합니다. 다중 모델이 도표에 선구분 변수로 표시됩니다.

결과는 특정 값 또는 값 범위를 **적중**으로 정의하여 처리됩니다. 적중은 일반적으로 관심 있는 이벤트(예: 특정 의료 진단) 또는 일부 정렬(예: 고객에 대한 판매)의 성공을 표시합니다.

플래그

출력 필드는 직설적이어서 적중은 *참* 값에 해당합니다.

명목

명목 출력 필드의 경우 세트의 첫 번째 값이 적중을 정의합니다.

연속

연속 출력 필드의 경우 적중은 필드 범위의 중심점보다 큰 값과 동일합니다.

각각의 점이 해당 분위수에 대한 값에 더 높은 모든 분위수를 더한 값과 동일하도록 평가 차트는 누적일 수도 있습니다. 누적 차트가 일반적으로 모델의 전체 성능을 더 잘 전달하지만 비누적 차트가 모델에 대한 특정 문제점 영역 표시에서 뛰어날 수도 있습니다.

단순 평가 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **평가** 아이콘을 클릭하십시오.

   Evaluation

   캔버스가 업데이트되어 평가 차트 템플리트가 표시됩니다.

2. **목표 필드**, **예측 필드** 및 **신뢰도 필드** 변수를 설정하십시오. 목표 필드는 둘 이상의 값을 가진 명목 필드 또는 인스턴스화된 플래그일 수 있습니다. 예측 필드는 예측값으로 사용되는 변수를 정의합니다. 신뢰도 필드는 예측의 신뢰도를 설정하는 데 사용되는 변수를 정의합니다.

   > **참고**: **예측 필드** 변수 유형은 **목표 필드**에 대해 선택된 변수 유형과 일치해야 합니다.

3. **사용자 정의 적중**을 표시하는 데 사용하는 사용자 정의 조건을 지정하십시오. 이 옵션은 값의 순서 및 대상 필드의 유형에서 관심 있는 결과를 추론할 때보다 관심 있는 결과를 정의하는 경우에 유용합니다.
   적중 조건에 대해 CLEM 표현식을 지정해야 합니다. 예를 들어, @TARGET = "YES"는 목표 필드에 대한 Yes 값이 평가에서 적중으로 계수됨을 나타내는 유효한 조건입니다. 지정된 조건은 모든 대상 필드에 사용됩니다.

추가 기능

누적 도표

사용으로 설정된 경우 누적 차트를 작성합니다. 누적 차트에 각 분위수 및 더 높은 분위수에 대해 값이 표시됩니다.

표시 모드

이 설정은 미리보기 모드 및 출력에 표시되는 차트를 제어합니다.

클래식 모드

선택할 경우 모델 분류 조정, 분리점, 행렬 막대, ROC, 이득, ROI 및 이익 차트가 미리보기 모드와 출력에 표시됩니다.

단일 모드

선택할 경우 모델 분류 조정 차트가 미리보기 모드와 출력에 표시되는 차트에만 있습니다.

전체 모드

선택할 경우 모델 분류 조정, 분리점, 행렬 막대, ROC, 이득, ROI, 이익, GINI, 리프트 및 반응 차트가 미리보기 모드와 출력에 표시됩니다.

평가 차트

분리점

분리점 차트는 지정된 분리점 값의 선택된 변수에 대한 예측 값과 실제 값을 보여줍니다.

행렬 막대

행렬 막대 차트는 다중 변수 간에 선형 상관관계가 있는지 판별하는 좋은 방법입니다.

ROC

ROC(Receiver Operating Characteristic)는 개체를 분류하는 두 개의 범주가 하나의 변수에 있는 분류 방법의 성능을 평가합니다.

이득

이득은 각 분위수에서 발생하는 적중 총계의 비율로 정의됩니다. 이득은 (분위수의 적중 수 / 적중 수 총계) × 100%로 계산됩니다.

ROI

ROI(Return On Investment)는 수입 및 비용 정의를 포함한다는 점에서 이익과 비슷합니다. ROI는 분위수에 대한 비용과 이익을 비교합니다. ROI는 (분위수에 대한 이익 / 분위수에 대한 비용) × 100%로 계산됩니다.

이익

이익은 각 레코드의 수입에서 레코드의 비용을 뺀 값입니다. 분위수의 이익은 단순히 분위수의 전체 레코드 이익 합계입니다. 수입은 적중에만 적용되는 것으로 가정되지만 비용은 모든 레코드에 적용됩니다. 이익 및 비용은 고정이거나 데이터의 필드에 의해 정의될 수 있습니다. 이익은 (분위수의 레코드에 대한 수입의 합계 - 분위수의 레코드에 대한 비용의 합계)로 계산됩니다.

Kolmogorov-Smirnov

한 변수에 대해 관측된 누적 분포 함수를 지정된 이론적 분포와 비교합니다. 이때 이론적 분포는 정규 분포, 균일 분포, 지수 분포, 포아송 분포일 수 있습니다.

GINI

GINI는 통계적 산포도를 측정하고 수입 또는 재산 분포를 나타냅니다. 불균등을 측정하는 데 가장 일반적으로 사용됩니다.

리프트

리프트는 적중인 각 분위수의 레코드 백분율을 훈련 데이터의 전체 적중 백분율과 비교합니다. 리프트는 (분위수의 적중 수 / 분위수의 레코드 수) / (적중 총계 / 레코드 총계)로 계산됩니다.

응답

반응은 단순히 적중인 분위수의 레코드 백분율입니다. 반응 은 (분위수의 적중 수 / 분위수의 레코드 수) × 100%로 계산됩니다.

평가 차트 설정

다음 설정은 이익과 ROI 차트에만 적용됩니다.

비용

각 레코드와 연관된 고정 비용을 지정합니다.

수입

적중을 나타내는 각 레코드와 연관된 고정수입을 지정합니다.

가중치

데이터의 레코드가 둘 이상의 단위를 표시하는 경우 빈도 가중치를 사용하여 결과를 조정할 수 있습니다. 각 레코드와 연관된 고정 가중치를 지정하십시오.

카. 히트 맵 차트

히트 맵 차트는 행렬에 포함된 개별 값이 색상으로 표시되는 데이터를 제공합니다.

단순 히트 맵 차트 작성

1. 차트 작성기의 차트 유형 섹션에서 히트 맵 아이콘을 클릭하십시오.

   Heat map

   캔버스가 업데이트되어 히트 맵 차트 템플리트가 표시됩니다.

2. 변수를 **열** 변수로 선택하십시오. 각 변수 범주가 개별 차트 열로 표시됩니다.
3. 변수를 **행** 변수로 선택하십시오. 각 변수 범주가 개별 차트 행으로 표시됩니다.

추가 기능

Column

차트 열에 사용할 수 있는 데이터 세트 변수가 나열됩니다. 각 변수 범주가 개별 차트 열로 표시됩니다.

행

차트 행에 사용할 수 있는 데이터 세트 변수가 나열됩니다. 각 변수 범주가 개별 차트 행으로 표시됩니다.

범주 순서

변수 범주의 정렬 순서를 선택하십시오.

읽는 대로

변수 범주가 데이터 세트에 나타나는 대로 표시됩니다.

오름차순

변수 범주가 오름차순으로 정렬됩니다.

내림차순

변수 범주가 내림차순으로 정렬됩니다.

요약

그래픽 요소에 대한 통계 요약함수를 선택하십시오. 통계 결과에 따라 y축에서 그래픽 요소의 위치가 결정됩니다. 2차원 차트의 경우 축의 각 x값에 대해 통계가 계산됩니다. 3차원 차트의 경우 x축 및 z축의 교차 값에 대해 통계가 계산됩니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 값으로 사용할 변수를 선택하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

타. 히스토그램 차트

히스토그램은 막대형 차트와 모양이 유사하지만, 범주를 비교하거나 시간 경과에 따른 추세를 찾을 수 있습니다. 각 막대는 데이터가 단일 범주에서 어떻게 분포되어 있는지를 나타냅니다. 각 막대는 특정 데이터 점에 대해 연속 범위의 데이터 또는 빈도 수를 나타냅니다.

히스토그램은 단일 척도변수의 분포를 표시하는 데 유용합니다. 데이터는 개수 또는 퍼센트 통계를 사용하여 구간화되고 요약됩니다. 빈도 다각형은 히스토그램의 변동이며 막대 그래픽 요소 대신 영역 그래픽 요소가 사용되는 것을 제외하고 일반적인 히스토그램과 같습니다.

다른 히스토그램의 변동은 인구 피라미드입니다. 이 이름은 가장 일반적인 용도(인구 데이터 요약)에서 파생되었습니다. 인구 데이터와 함께 사용하는 경우 성별로 분할되어 나이 데이터에 대해 두 개의 백 투 백, 수평 히스토그램을 제공합니다. 젊은이 인구가 있는 국가의 경우 결과 그래프의 형태가 피라미드와 유사합니다.

히스토그램 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **히스토그램** 아이콘을 클릭하십시오.

Histogram •

캔버스가 업데이트되어 히스토그램 차트 템플리트가 표시됩니다.

2. 척도변수를 **X축** 변수로 선택하십시오.

> **참고**: 히스토그램에 대한 통계는 히스토그램 또는 히스토그램 퍼센트입니다. 이러한 통계는 데이터를 구간화하고 각 구간에 대한 개수를 계산합니다.

추가 기능

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오. 그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

KDE 곡선 표시

사용으로 설정할 경우, 커널 밀도 추정값 곡선이 차트에 표시됩니다.

분포 곡선 표시

사용으로 설정할 경우 분포 적합 곡선이 차트에 표시됩니다.

분포

드롭 다운 목록은 다음 분포 옵션을 제공합니다.

분포 자동 적합

기본 설정입니다.

베타

지정한 형태 모수를 사용하여 베타 분포의 값을 반환합니다.

지수

지수 분포에서 값을 리턴합니다.

감마

지정한 형태와 척도 모수를 사용하여 감마 분포의 값을 반환합니다.

로그정규

지정한 모수를 사용하여 로그-정규 분포의 값을 반환합니다.

정규

지정한 평균과 표준 편차를 사용하여 정규 분포의 값을 반환합니다.

삼각

지정한 모수를 사용하여 삼각 분포의 값을 반환합니다.

균일

최소값과 최대값 간 균일 분포의 값을 반환합니다.

와이블

지정한 모수를 사용하여 Weibull 분포의 값을 반환합니다.

구간 너비

데이터를 그룹으로 분할하는 데 사용되는 간격의 크기를 슬라이더로 제어할 수 있습니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

파. 선형 차트

선형 차트는 일련의 데이터 점을 그래프에 도표화한 후 이를 선으로 연결합니다. 선형 차트는 미묘한 차이가 있는 추세선을 표시하거나 서로 교차하는 데이터 선이 있는 경우에 특히 유용합니다. 선형 차트를 사용하여 범주형 변수를 요약할 수 있으며 이 경우 막대형 차트와 비슷합니다(막대형 차트 참조). 선형 차트는 시계열 데이터에도 유용합니다.

단순 시계열 선형 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **선** 아이콘을 클릭하십시오.

   Line •

   캔버스가 업데이트되어 선형 차트 템플리트가 표시됩니다.

2. 날짜변수를 **X축** 변수로 선택하십시오.
3. 척도변수를 **Y축** 변수로 선택하십시오. 이 변수는 값이 시간 경과에 따라 기록된 변수입니다.

추가 기능

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Y축

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오. 그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

영역

사용으로 설정할 경우 선 아래의 영역이 다른 색상으로 채워집니다.

평활

사용으로 설정할 경우 차트에 평활 곡선이 표시됩니다.

데이터 점 표시

사용으로 설정할 경우 차트에 데이터 점이 표시됩니다.

재정렬

이 전환 컨트롤은 X축 및 Y축 값을 기준으로 데이터를 다시 정렬합니다.

참조선 표시

사용으로 설정할 경우, X축 및 Y축 값을 기준으로 하는 차트에 참조선이 표시됩니다.

X축에 참조선 값 입력

참조선 표시가 사용으로 설정된 경우 이 설정을 사용하여 X축에 대한 특정 참조선 값을 지정할 수 있습니다. 추가 참조선 값을 지정하려면 **다른 열 추가**를 클릭하십시오.

Y축에 참조선 값 입력

참조선 표시가 사용으로 설정된 경우 이 설정을 사용하여 Y축에 대한 특정 참조선 값을 지정할 수 있습니다. 추가 참조선 값을 지정하려면 **다른 열 추가**를 클릭하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

하. 맵 차트

맵 차트는 일반적으로 값을 비교하고 지역 전체의 범주를 표시하는 데 사용되며, 데이터에 지리 정보(국가, 지역, 시/도, County, 우편번호 등)가 포함되어 있는 경우에 특히 유용합니다.

단순 맵 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **맵** 아이콘을 클릭하십시오.

   Map

   캔버스가 업데이트되어 맵 차트 템플리트가 표시됩니다.

2. **맵 서비스** 드롭 다운 목록에서 맵 이미지를 제공하는 데 사용할 서비스를 선택하십시오. 이 목록은 특정 세계 지역을 포함하는 옵션을 제공합니다.
3. 드롭 다운 메뉴에서 차트 **유형**을 선택하십시오. 선택한 차트 유형에 따라 다음 옵션이 사용가능합니다.

   경도

   드롭 다운 목록에서 경도 값으로 사용할 변수를 선택하십시오.

   위도

   드롭 다운 목록에서 위도 값으로 사용할 변수를 선택하십시오.

   그룹

   드롭 다운 메뉴에서 데이터 점 위치를 그룹화하는 변수를 선택하십시오.

   범주

   시각화할 열 변수를 선택하십시오.

   요약

   통계 요약함수를 선택하십시오. 각 범주를 요약하는 데 사용되는 방법입니다.

   두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

   - **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
   - **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

   값

   이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 값으로 사용할 변수를 선택하십시오.

추가 기능

맵 서비스

맵 이미지를 제공하는 데 사용할 수 있는 서비스가 나열됩니다. 맵 서비스 옵션의 내용을 참조하십시오.

유형

데이터를 표시하는 데 사용할 수 있는 차트 유형이 나열됩니다.

경도

경도 값으로 사용할 수 있는 변수가 나열됩니다.

위도

위도 값으로 사용할 수 있는 변수가 나열됩니다.

그룹

데이터 점 위치를 그룹화하는 데 사용할 수 있는 변수를 나열합니다.

범주

열 변수를 나열합니다.

요약

통계 요약함수를 선택하십시오. 각 범주를 요약하는 데 사용되는 방법입니다.

두 가지 유형의 통계 요약함수가 있습니다. 값 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 요약 함수를 선택한 경우에 표시됩니다. 값으로 사용할 변수를 선택하십시오.

도구 팁 정보

데이터 점 위로 마우스를 이동할 때 표시되는 도구 팁 정보를 생성하는 데 사용할 수 있는 변수가 나열됩니다.

크기 맵

사용 가능한 크기 맵 변수가 나열됩니다. 이 변수는 서로 다른 크기를 사용하여 플롯 점에 자신을 표시합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

ㄱ. 맵 서비스 옵션

맵 차트의 세부사항 분할창에서 맵 이미지를 제공하는 데 사용할 맵 서비스를 선택할 수 있습니다.

맵 서비스 옵션 구성

1. IBM® SPSS® Modeler Server를 중지하십시오.
2. geojson 형식의 로컬 맵 파일을 적용하려면 <Modeler_installation_directory>/dataview/conf/public/mapfiles 디렉토리에 두십시오. 디렉토리가 없으면 작성하십시오.
3. 텍스트 편집기에서 <Modeler_installation_directory>/dataview/conf/application.conf 파일을 열고 다음 옵션을 map.resources 섹션에 추가하십시오. 이 예에서는 World - Local Map과 World - Online Map이라는 두 개의 새로운 맵이 추가됩니다.

```
map.resources {
  "mapservices" :
    [

      {
        "type"        : "geojson",
        "name"        : "World (Bulit-in)",
        "location"  : "built_in_world",
        "id"          : "built_in_world"
      },
      {
        "type"        : "geojson",
        "name"        : "Blank (Built-in)",
        "location"  : "built_in_blank",
        "id"          : "built_in_blank"
      },
      {
        "type"        : "geojson",
        "name"        : "World - Local Map",
        "location"  : "/mapfiles/world.json",
        "useProxy"  : "false",
        "id"          : "world_local"
      },
      {
        "type"        : "geojson",
        "name"        : "World - Online Map",
        "location"  : "http://map.example.com/map/world.json",
        "useProxy"  : "true",
        "id"          : "world_online"
      }    ]
}
```

다음 모수를 맵 옵션에 사용할 수 있습니다.

표 1. 맵 서비스 모수

옵션	값	유형	필요 여부	설명
type	geojson	문자열	예	현재 이 옵션에는 geojson 값만 지원됩니다.

옵션	값	유형	필요 여부	설명
이름	맵 이름	문자열	예	맵의 이름을 입력하십시오. 이 이름은 맵 차트의 세부사항 분할창에 있는 맵 서비스 드롭다운에 표시됩니다.
location	맵 위치	문자열	예	맵의 URL 또는 로컬 파일 위치를 입력하십시오. URL의 경우, http:// 또는 https://로 시작해야 합니다. 로컬 파일의 경우, 경로 /mapfiles/로 시작해야 합니다.
useProxy	true 또는 false	문자열	예	맵 위치가 URL일 경우 true를 사용하고, 맵 위치가 로컬 파일일 경우 false를 사용하십시오.
id	맵 ID	문자열	예	맵마다 고유한 ID를 사용하십시오.

참고:

- 기본 **World(기본 제공)** 및 **Blank(기본 제공)** 옵션을 제거하지 마십시오.
- 예에 표시된 대로 내용 블록을 쉼표로 구분하고 새 map.resources 섹션이 유효한 JSON 형식의 내용인지 확인하십시오.
- IBM SPSS Modeler Server(또는 서버가 없는 독립형 SPSS Modeler 클라이언트)에서 맵 지도에 액세스할 수 있어야 합니다.

4. IBM SPSS Modeler Server를 시작하십시오.

이제 맵 차트에 대해 작업하는 동안 **세부사항** 분할창에 사용자가 추가한 새로운 맵 서비스가 표시됩니다.

거. 수학 곡선 차트

수학 곡선 차트는 사용자가 입력한 표현식에 따라 수학 방정식 곡선을 도표화합니다.

단순 수학 곡선 차트 작성

1. 차트 작성기의 차트 유형 섹션에서 수학 곡선 아이콘을 클릭하십시오.

   Math curve

   캔버스가 업데이트되어 수학 곡선 차트 템플리트가 표시됩니다.

2. **X 값 시작 문자** 필드에 X축의 시작값을 입력합니다.
3. **X 값 끝 문자** 필드에 X축의 종료값을 입력합니다.
4. 방정식 필드에 그래프 곡선을 도표화하는 방정식을 입력합니다. 추가 방정식을 포함시키려면 다른 열 추가를 클릭하십시오.
   각 방정식은 x를 독립변수로 처리합니다. 방정식은 다음을 허용합니다.
   - +, -, *, /, % 및 ^
   - abs(x)
   - ceil(x)
   - floor(x)
   - log(x)
   - max(a,b,c...)
   - min(a,b,c...)
   - random()
   - round(x)
   - sqrt(x)
   - sin
   - cos
   - exp
   - tan
   - atan
   - atan2
   - asin
   - acos

추가 기능

X 값 시작 문자
X축 시작값입니다.
X 값 끝 문자
X축 종료값입니다.
방정식
그래프 곡선을 도표화하는 사용자 입력 방정식입니다.
기본 제목
차트 제목입니다.
부제목
차트 제목 바로 아래에 표시되는 차트 부제목입니다.
꼬리말
차트 아래에 표시되는 차트 꼬리말입니다.

너. 다중 차트

다중 차트는 여러 도표를 작성하는 옵션을 제공합니다. 차트는 동일하거나 서로 다른 유형일 수 있으며 동일한 데이터 세트의 여러 변수를 포함할 수 있습니다.

단순 다중 차트 작성

1. 차트 작성기의 차트 유형 섹션에서 다중 차트 아이콘을 클릭하십시오.
   Multi-chart

   캔버스가 업데이트되어 다중 차트 템플리트가 표시됩니다.

2. 서브차트를 추가하려면 **다른 서브차트 추가**를 클릭하십시오.
3. **유형** 드롭 다운 목록에서 차트 유형을 선택하십시오.
4. 선택한 차트 유형에 따라 다음 옵션이 사용 가능합니다.
   막대 및 원형 차트
      드롭 다운 목록에서 **범주** 변수를 선택하십시오.
   선 및 산점도 차트
      드롭 다운 목록에서 **X축** 변수를 선택하십시오.
      드롭 다운 목록에서 **Y축** 변수를 선택하십시오.

5. 추가 차트 유형을 포함시키려면 **다른 서브차트 추가**를 클릭하십시오.

추가 기능

제목
   서브차트 제목입니다.
기본 제목
   차트 제목입니다.
부제목
   차트 제목 바로 아래에 표시되는 차트 부제목입니다.
꼬리말
   차트 아래에 표시되는 차트 꼬리말입니다.

더. 다중 계열 차트

다중 계열 차트는 선형 차트와 유사하지만, y축에 변수를 여러 개 작성할 수 있다는 점이 다릅니다.

단순 다중 계열 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **다중 계열** 아이콘을 클릭하십시오.

   Multi-series

   캔버스가 업데이트되어 다중 계열 차트 템플리트가 표시됩니다.

2. 변수를 **X축** 변수로 선택하십시오.
3. 최소 두 개의 척도변수를 **Y축** 변수로 선택하십시오.

추가 기능

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Y축

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

드롭 다운 목록에서 차트 유형(선, 막대 또는 산점도)을 선택하십시오.

차트에 열을 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

계열 유형

Y축 방향을 정의하기 위한 옵션을 제공합니다. 사용 가능한 옵션은 다음과 같습니다.

- 기본값
- Y축 분리
- 보조 Y축 표시
- 이중 Y축

데이터 표준화

사용으로 설정할 경우 이 설정은 다중 데이터 세트 또는 다중 열의 데이터를 비교할 수 있도록 데이터를 정규 분포로 변환합니다. 이 설정은 빈도 및 변환 통계에 대한 100% 수직누적을 퍼센트로 작성합니다.

보조 Y축

차트의 보조 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

드롭 다운 목록에서 차트 유형(선, 막대 또는 산점도)을 선택하십시오.

차트에 열을 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

러. 평형 차트

평형 차트는 고차원 기하학을 시각화하고 다변량 데이터를 분석할 때 유용합니다. 평형 차트는 시계열 데이터에 대한 선형 차트와 유사하지만, 축이 시점에 해당하지 않습니다(자연 순서가 없음).

단순 평형 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **평형** 아이콘을 클릭하십시오.

   Parallel

   캔버스가 업데이트되어 평형 차트 템플리트가 표시됩니다.

2. 최소 두 개의 변수를 **열** 변수로 선택하십시오. 각 필드는 차트의 평형 수직축을 나타냅니다.

> **참고**: 열 순서는 기능을 찾는 데 중요합니다. 일반 데이터 분석에서는 열을 여러 번 다시 정렬해야 할 수 있습니다.

추가 기능

열

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.
열을 더 추가하려면 **다른 열 추가**를 클릭하십시오.

색상 맵

사용 가능한 색상 맵 변수가 나열됩니다. 이 변수는 색상 진행(지정된 열의 값 범위를 기준으로 함)을 사용하여 플롯 점에 자신을 표시합니다. 색상 맵은 코로플레스 맵이라고도 합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

머. 원형 차트

원형 차트는 비율을 비교하는 데 유용합니다. 예를 들어 원형 차트를 사용하여 높은 비율의 여성이 특정 클래스에 등록했음을 설명할 수 있습니다.

단순 원형 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **원형 차트** 아이콘을 클릭하십시오.

   Pie •

   캔버스가 업데이트되어 원형 차트 템플리트가 표시됩니다.

2. **범주** 목록에서 범주형(명목 또는 순서) 변수를 선택하십시오. 이 변수의 범주에 따라 원형 차트의 조각 수가 결정됩니다.
3. 그래픽 요소에 대한 통계 요약함수를 선택하십시오. 원형 차트의 경우 일반적으로 개수 기반 통계 또는 합계를 원할 때 사용됩니다. 통계의 결과에 따라 각 조각의 크기가 결정됩니다.

추가 기능

범주

원형 차트의 조각 수를 결정하는 범주형(명목 또는 순서) 변수를 선택하십시오.

요약

그래픽 요소에 대한 통계 요약함수를 선택하십시오. 원형 차트의 경우 일반적으로 개수 기반 통계 또는 합계를 원할 때 사용됩니다. 통계의 결과에 따라 각 조각의 크기가 결정됩니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 척도변수가 필요한 요약 함수를 선택한 경우에 표시됩니다. 척도변수로 사용할 변수를 선택하십시오.

원 유형

사용 가능한 유형은 다음과 같습니다.

정규

원 세그먼트가 정규 조각으로 표시됩니다.

링

원 세그먼트가 링으로 표시됩니다. 이 유형은 도넛 차트라고도 합니다.

Rose

일반 반경을 사용하는 정규 원형 차트와 달리, 원 세그먼트의 크기가 값에 따라 달라집니다.

장미 영역

일반 반경을 사용하는 정규 원형 차트와 달리, 원 세그먼트의 크기가 영역에 따라 달라집니다.

장미 고리

일반 반경을 사용하는 정규 원형 차트와 달리, 원 세그먼트의 크기가 링에 따라 달라집니다.

반 장미

도표가 원도표의 절반으로 표시되는 점을 제외하고, 장미와 동일합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

버. 인구 피라미드 차트

인구 피라미드 차트("연령-성별 피라미드"라고도 함)는 일반적으로 연령과 성별을 기준으로 인구 정보를 표시하고 분석하는 데 사용됩니다.

단순 인구 피라미드 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **인구 피라미드** 아이콘을 클릭하십시오.

   Population ...

   캔버스가 업데이트되어 인구 피라미드 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 **Y축** 변수를 선택하십시오.
3. 드롭 다운 목록에서 **분할 기준** 변수를 선택하십시오.

추가 기능

Y축

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오.

그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

구간 너비

데이터를 그룹으로 분할하는 데 사용되는 간격의 크기를 슬라이더로 제어할 수 있습니다.

분포 곡선 표시

사용으로 설정할 경우 분포 적합 곡선이 차트에 표시됩니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

서. Q-Q 도표

Q-Q(quantile-quantile) 도표는 두 확률 분포에 대해 분위수를 도표화하여 두 확률 분포를 비교합니다. Q-Q 도표는 분포 형태를 비교하는 데 사용되며, 두 개의 분포에서 위치, 척도, 왜도 등의 특성이 얼마나 유사한지 또는 다른지를 그래픽으로 보여줍니다.

단순 Q-Q 도표 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **Q-Q 도표** 아이콘을 클릭하십시오.

   Q-Q plot

   캔버스가 업데이트되어 Q-Q 도표 차트 템플리트가 표시됩니다.

2. 변수를 **X축** 변수로 선택하십시오.

추가 기능

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

분포

드롭 다운 목록에 사용 가능한 모든 분포 방법이 나열됩니다.

분포 자동 적합

기본 설정입니다.

베타

지정한 형태 모수를 사용하여 베타 분포의 값을 반환합니다.

지수

지수 분포에서 값을 리턴합니다.

감마

지정한 형태와 척도 모수를 사용하여 감마 분포의 값을 반환합니다.

로그정규

지정한 모수를 사용하여 로그-정규 분포의 값을 반환합니다.

정규

지정한 평균과 표준 편차를 사용하여 정규 분포의 값을 반환합니다.

균일

최소값과 최대값 간 균일 분포의 값을 반환합니다.

스튜던트 T

지정한 자유도를 사용하여 스튜던트 T 분포의 값을 반환합니다.

도표 유형

Q-Q(quantile-quantile) 도표 또는 P-P(percent-percent) 도표를 선택하십시오.

자동 적합

사용으로 설정할 경우 선택한 **분포**에 대한 데이터 모수가 추정됩니다. 사용으로 설정할 경우 **모양** 및 **척도** 분포 값이 표시됩니다.

형태1

베타 분포의 형태1 값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **베타**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

형태2

베타 분포의 형태2 값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **베타**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

형태

선택한 분포의 형태 값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **감마** 또는 **로그정규**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

척도

선택한 분포의 척도 값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **지수**, **감마** 또는 **로그정규**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

평균

정규 분포의 평균 값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **정규**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

표준 편차

정규 분포의 표준 편차 값을 설정합니다. 이 설정은 **정규**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

최소값

균일 분포의 최소값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **균일**이 **분포**로 선택된 경우에만 사용할 수 있습니다.

최대값

균일 분포의 최대값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **균일**이 **분포**로 선택된 경우에만 사용할 수 있습니다.

자유도(df)

스튜던트 t 분포에 대한 자유도 값을 설정합니다. 이 설정은 **자동 적합**이 사용으로 설정되어 있지 않고 **스튜던트 T**가 **분포**로 선택된 경우에만 사용할 수 있습니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

어. 방사형 차트

방사형 차트는 여러 양적 변수를 비교하고, 값이 비슷한 변수를 시각화하거나 변수 간에 이상치가 있는지 표시하는 데 유용합니다. 방사형 차트는 일련의 스포크로 구성되며, 각 스포크는 단일 변수를 나타냅니다. 방사형 차트는 데이터 세트에서 점수가 높고 낮은 변수를 판별하는 데도 유용합니다.

단순 방사형 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **방사형** 아이콘을 클릭하십시오.

   Radar

   캔버스가 업데이트되어 방사형 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 **열** 변수를 선택하십시오.

   > ⓘ **참고:** 추가 열을 포함시키려면 **다른 열 추가**를 클릭하십시오. 세 개 이상의 열 변수를 정의해야 합니다.

추가 기능

범주

범주형(명목 또는 순서) 변수를 선택하십시오. 범주 값으로 **없음**을 선택하면 모든 값이 개별적으로 표시되며 요약 방법도 적용되지 않습니다.

요약

범주형 변수를 선택하면, 다음 요약 통계량이 사용 가능합니다.

수

총 케이스 수입니다.

합계

값의 합계입니다.

평균

합계를 케이스 수로 나눈 산술 평균 값입니다.

최대값

가장 큰(가장 높은) 값입니다.

최소값

가장 작은(가장 낮은) 값입니다.

방사형 레이아웃

방사형 차트의 배경 이미지 레이아웃을 판별합니다.

원형

선택할 경우 방사형 차트가 원형 레이아웃 위에 표시됩니다.

다각형

선택할 경우 방사형 차트가 다각형 레이아웃 위에 표시됩니다.

분할 기준

분할 기준 변수에 범주별 셀이 있는 차트 테이블을 작성하는 범주형 변수를 선택하십시오. 그룹화와 마찬가지로, 분할 기준 변수는 기본적으로 변수 범주 각각에 대한 정보를 표시하여 차트에 차원을 더 추가합니다. 자세한 정보는 분할 기준 변수 추가 주제를 참조하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

저. 관계 차트

관계 차트는 변수의 상관 정도를 판별하는 데 유용합니다.

단순 관계 차트 작성

1. 차트 작성기의 차트 유형 섹션에서 관계 아이콘을 클릭하십시오.

   Relationship

   캔버스가 업데이트되어 관계 차트 템플리트가 표시됩니다.

2. 최소 두 개의 변수를 **열** 변수로 선택하십시오.

   ⓘ **참고:** 열 변수를 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

추가 기능

열

사용 가능한 데이터 세트 변수가 나열됩니다.
열을 더 추가하려면 **다른 열 추가**를 클릭하십시오.

선 유형

관련 데이터 점을 서로 연결하는 선 유형을 제어합니다.

곡선

선택할 경우 관련 데이터 점이 곡선으로 연결됩니다.

직선

선택할 경우 관련 데이터 포인트가 직선으로 연결됩니다.

레이블 임계값

정의된 값을 초과하는 값을 갖는 데이터 점의 레이블을 표시합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

처. 산점도 및 점도표

포인트 그래픽 요소로 작성된 여러 광범위한 범주의 차트가 있습니다.

- **산점도**: 다변량 데이터를 도표화하는 데 유용하며, 척도변수 간에 잠재적 관계를 판별하는 데

도움이 됩니다. 단순 산점도는 2차원 좌표계를 사용하여 두 개의 변수를 도표화합니다. 3차원 산점도는 3차원 좌표계를 사용하여 세 개의 변수를 도표화합니다. 추가 변수를 도표화해야 하는 경우, 오버레이 산점도 및 산점도 행렬(SPLOM)을 시도할 수 있습니다. 오버레이 산점도는 각 대응 변수를 색상 또는 형태별로 구분하여 x-y 변수의 오버레이 대응을 표시합니다. SPLOM은 각 변수를 SPLOM의 다른 각 변수에 대해 도표화한 2차원 산점도 행렬을 작성합니다.

- **점도표**: 히스토그램과 마찬가지로, 단일 척도변수의 분포를 표시하는 데 유용합니다. 데이터가 구간화되지만, 구간당 하나의 값(예: 개수) 대신 각 구간의 모든 포인트가 표시되고 누적됩니다. 이 그래프는 밀도 도표라고도 합니다.
- **요약 포인트 도표**: 이 차트는 포인트가 막대 맨 위에 표시되는 것을 제외하고 막대형 차트와 거의 비슷합니다. 요약 포인트 도표는 막대형 차트와 매우 비슷하므로, 이 도표 작성에 대한 정보는 막대형 차트의 내용을 참조하십시오.
- **하락-선 차트**: 이 차트는 특별한 유형의 요약 포인트 도표입니다. 포인트는 그룹화되고 선은 각 범주 내 포인트를 통해 작성됩니다. 하락-선 차트는 범주형 변수에 대한 통계를 비교하는 데 유용합니다.

단순 산점도 작성

1. 차트 작성기의 차트 유형 섹션에서 산점도 아이콘을 클릭하십시오.

   Scatter plot

   캔버스가 업데이트되어 산점도 차트 템플리트가 표시됩니다.

2. 척도변수를 **X축** 변수로 선택하십시오.
3. 척도변수를 **Y축** 변수로 선택하십시오. 산점도는 일반적으로 원시 값을 표시하기 때문에 통계를 지정할 필요가 없습니다.

추가 기능

X축

차트의 x축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

Y축

차트의 y축에 사용할 수 있는 데이터 세트 변수가 나열됩니다.

색상 맵

사용 가능한 색상 맵 변수가 나열됩니다. 이 변수는 색상 진행(지정된 열의 값 범위를 기준으로 함)을 사용하여 플롯 점에 자신을 표시합니다. 색상 맵은 코로플레스 맵이라고도 합니다.

크기 맵

사용 가능한 크기 맵 변수가 나열됩니다. 이 변수는 서로 다른 크기를 사용하여 플롯 점에 자신을 표시합니다.

모양 맵

사용 가능한 모양 맵 변수가 나열됩니다. 이 변수는 서로 다른 모양을 사용하여 플롯 점에 자신을 표시합니다.

회귀선 적합

회귀선 적합에서는 데이터 점이 하나의 선에 맞춰지는데 이 선이 데이터 점을 모두 통과하는 것은 아닙니다. 회귀선 적합은 데이터의 추세를 나타냅니다. 일부 회귀선 적합은 회귀분석을 기반으로 하고 다른 회귀선 적합은 반복 가중치를 부여한 최소제곱을 기반으로 합니다. 드롭다운 목록에서 회귀선 적합 옵션을 선택하십시오.

기울기 거품

이 전환 컨트롤은 거품형 차트에서 색상 기울기 및 3D 효과를 사용/사용 안함으로 설정합니다. **색상 맵** 변수를 선택한 경우에는 이 설정이 사용 안 함으로 설정됩니다.

최소 거품 크기

최소 거품 크기를 설정합니다. 5와 20 사이의 값을 입력합니다.

최대 거품 크기

최대 거품 크기를 설정합니다. 20과 80 사이의 값을 입력합니다.

참조선 표시

사용으로 설정할 경우, X축 및 Y축 값을 기준으로 하는 차트에 참조선이 표시됩니다.

X축에 참조선 값 입력

참조선 표시가 사용으로 설정된 경우 이 설정을 사용하여 X축에 대한 특정 참조선 값을 지정할 수 있습니다. 추가 참조선 값을 지정하려면 **다른 열 추가**를 클릭하십시오.

Y축에 참조선 값 입력

참조선 표시가 사용으로 설정된 경우 이 설정을 사용하여 Y축에 대한 특정 참조선 값을 지정할 수 있습니다. 추가 참조선 값을 지정하려면 **다른 열 추가**를 클릭하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

커. 산점도 행렬 차트

산점도 행렬은 다중 변수 간에 선형 상관관계가 있는지 판별하는 좋은 방법입니다.

산점도 행렬 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **산점도 행렬** 아이콘을 클릭하십시오.

   Scatterplot matrix

   캔버스가 업데이트되어 산점도 행렬 차트 템플리트가 표시됩니다.

2. 최소 두 개 이상의 **열** 척도변수를 선택하십시오.

   > ⓘ **참고:** 열 변수를 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

선택한 각 변수가 다른 변수 각각에 대해 도표화되어 개별 산점도 행렬이 작성됩니다.

추가 기능

열

행렬 변수를 최소 두 개 선택하십시오. 변수는 숫자여야 합니다(날짜 형식 제외).
열을 더 추가하려면 **다른 열 추가**를 클릭하십시오.

색상 맵

사용 가능한 색상 맵 변수가 나열됩니다. 이 변수는 색상 진행(지정된 열의 값 범위를 기준으로 함)을 사용하여 플롯 점에 자신을 표시합니다. 색상 맵은 코로플레스 맵이라고도 합니다.

상관관계

사용으로 설정할 경우 선택한 변수에 대해 선형 상관관계 정보(강함, 중간, 약함)가 표시됩니다.

KDE 곡선 표시

사용으로 설정할 경우, 커널 밀도 추정값 곡선이 차트에 표시됩니다.

기울기 거품

이 전환 컨트롤은 거품형 차트에서 색상 기울기 및 3D 효과를 사용/사용 안함으로 설정합니다. **색상 맵** 변수를 선택한 경우에는 이 설정이 사용 안 함으로 설정됩니다.

최소 거품 크기

최소 거품 크기를 설정합니다. 5와 20 사이의 값을 입력합니다.

최대 거품 크기

최대 거품 크기를 설정합니다. 20과 80 사이의 값을 입력합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

터. 햇살 차트

햇살 차트는 계층형 데이터 구조를 시각화할 때 유용합니다. 햇살 차트는 내부 원과 이를 둘러싸고 있는 더 깊은 계층 수준 고리로 구성됩니다. 각 세그먼트의 각도는 값에 비례하거나 내부 세그먼트 아래에서 동일하게 나뉩니다. 차트 세그먼트는 속해 있는 범주 또는 계층 구조 수준에 따라 색상이 지정됩니다.

단순 햇살 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **햇살** 아이콘을 클릭하십시오.
   Sunburst

   캔버스가 업데이트되어 햇살 차트 템플리트가 표시됩니다.

2. **열** 목록에서 범주형(명목 또는 순서) 변수를 선택하십시오. 이 변수의 범주에 따라 차트의 세그먼트 수가 결정됩니다.
3. **다른 열 추가**를 클릭하고 **열** 목록에서 범주형(명목 또는 순서) 변수를 선택하십시오. 이 변수의 범주에 따라 차트 두 번째 고리의 세그먼트 수가 결정되며, 이는 더 높은 계층 구조 수준을 나타냅니다.
4. 그래픽 요소(개수 기반 통계 또는 합계)에 대한 통계 요약함수를 선택하십시오. 통계의 결과에 따라 각 세그먼트의 크기가 결정됩니다. **합계**를 선택한 경우, 요약할 데이터 세트의 값을 나타낼 척도변수를 **값** 목록에서 선택하십시오.
5. **햇살 레이아웃** 옵션을 **기존** 또는 **분산형** 중에서 선택하십시오.

추가 기능

열

차트의 세그먼트 수를 결정하는 범주형(명목 또는 순서) 변수를 선택하십시오.

요약

그래픽 요소(개수 기반 통계 또는 합계)에 대한 통계 요약함수를 선택하십시오. 통계의 결과에 따라 각 조각의 크기가 결정됩니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 이 함수는 **값** 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어, 합계 함수의 경우 합계를 계산할 때 변수가 필요합니다.

값

이 필드는 척도변수가 필요한 요약 함수를 선택한 경우에 표시됩니다. 척도변수로 사용할 변수를 선택하십시오.

햇살 레이아웃

사용 가능한 옵션은 기존 및 분기입니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

퍼. 시간 도표

시간 도표에서는 연속된 시간 구간으로 된 데이터 점을 보여줍니다. 구성하는 시계열은 숫자 값을 포함해야 하며 주기가 균일한 시간 범위에서 발생한다고 가정합니다. 시간 도표에서는 기본 통계량과 검정에서 시계열 데이터의 공정특성 변수를 예비 분석하므로, 모델링하기 전에 데이터에 관한 유용한 정보가 생성됩니다. 시간 도표에는 분해, ADF(Augemented Dickey-Fuller) 검정, 상관계수(ACF/PACF) 및 스펙트럼 분석과 같은 분석 방법이 포함됩니다.

단순 시간 도표 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **시간 도표** 아이콘을 클릭하십시오.

   Time plot

   캔버스가 업데이트되어 시간 도표 차트 템플리트가 표시됩니다.

2. **값** 드롭 다운에서 y축 값을 선택하십시오.

추가 기능

날짜

해당되는 경우, 드롭 다운 목록에서 날짜를 선택하십시오. 각 관측값은 기본적으로 동일한 시간 구간으로 구분됩니다. 날짜 변수를 선택하는 경우 표본 재추출 옵션이 표시됩니다. 이 옵션을 사용하면 지정된 구간과 일치하는 값 필드를 통합할 수 있습니다.

시간 도표 알고리즘

시계열 데이터를 분석하는 데 사용하는 시간 도표 알고리즘입니다.

- **분해.** 시계열을 세 개의 구성요소(추세-순환, 계절 및 비정규)로 분해합니다. 분해는 가법 방식으로 수행됩니다.
- **ADF 검정.** ADF(Augmented Dickey-Fuller)는 계열에 단위 근이 있으며 계열이 정상이 아니라는 영가설을 검정합니다. 검정 결과에서 영가설을 거부하면 계열이 정상이거나 차분 모델을 사용하여 정상으로 표시될 수 있음을 나타냅니다.
- **ACF/PACF.** 계열의 상관관계입니다.
- S**스펙트럼 분석.** 빈도 도메인의 분석 도구입니다. 최고 빈도는 다이아몬드로 표시됩니다.

도표 위치 바꾸기

평면 측정 직사각형 좌표계 및 극 좌표계에서 도표의 위치를 반전시킵니다.

전환점 표시

도표에서 전환점을 표시하거나 숨깁니다. 시계열 분해의 추세-순환 구성요소를 기반으로 시계열에 전체 추세가 있는지 아니면 추세 패턴의 방향을 변경하는 전환점이 있는지 탐색하십시오.

이상치 표시

이상치를 표시하거나 숨깁니다. 시계열의 이상치는 시계열 분해의 비정규 구성요소에서 분석됩니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

허. 테마 리버 차트

테마 리버는 시간 경과에 따른 변화를 보여주는 특수화된 플로우 그래프입니다.

단순 테마 리버 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **테마 리버** 아이콘을 클릭하십시오.

   Theme River

   캔버스가 업데이트되어 테마 리버 차트 템플리트가 표시됩니다.

2. **X축** 변수를 선택하십시오.
3. **범주** 변수를 선택하십시오.

⊘ **제한사항:** 범주 필드에 50개가 넘는 개별 범주가 있는 경우 상위 50개의 최대 범주만 차트에서 이벤트로 사용됩니다.

추가 기능

순서 기준

선택한 X축 값에 따라, 범주 순서가 범주 이름 또는 범주 값을 기준으로 하는지 여부를 지정할 수 있습니다.

범주 순서

선택한 X축 값에 따라, 범주 순서가 오름차순 또는 내림차순인지 아니면 데이터 세트에서 읽을 것인지 여부를 지정할 수 있습니다.

요약

통계 요약함수를 선택하십시오. 각 범주를 요약하는 데 사용되는 방법입니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 값으로 사용할 변수를 선택하십시오.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

고. 트리 차트

트리 차트는 트리와 유사한 구조로 계층을 표시합니다. 트리 차트의 구조는 루트 노드(상위 노드 없음), 분기라고 하는 선 연결(멤버 간의 관계와 연결을 나타냄), 리프 노드(하위 노드 없음)로 구성됩니다.

단순 트리 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **트리** 아이콘을 클릭하십시오.

   Tree

   캔버스가 업데이트되어 트리 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 **열** 변수를 선택하십시오.

   > ⓘ **참고:** 열 변수를 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

추가 기능

열

차트 열을 표시하는 데 사용할 수 있는 데이터 세트 변수가 나열됩니다.

요약

그래픽 요소에 대한 통계 요약함수를 선택하십시오. 통계의 결과에 따라 그래픽 요소의 위치가 결정됩니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 값 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 요약값의 기준으로 사용할 변수를 선택하십시오.

트리 레이아웃

왼쪽에서 오른쪽

루트 노드가 왼쪽에 표시되고 리프 노드는 오른쪽에 표시됩니다.

오른쪽에서 왼쪽

루트 노드가 오른쪽에 표시되고 리프 노드는 왼쪽에 표시됩니다.

위쪽에서 아래쪽

루트 노드가 위쪽에 표시되고 리프 노드는 아래쪽에 표시됩니다.

아래쪽에서 위쪽

루트 노드가 아래쪽에 표시되고 리프 노드는 위쪽에 표시됩니다.

방사형

루트 노드가 가운에 표시되고 리프 노드는 루트에서 방사형으로 뻗어나갑니다.

리프 깊이

리프 노드에 대한 드릴다운 수준 값을 설정합니다.

리프 레이블 표시

사용으로 설정할 경우 레이블이 리프 노드마다 표시됩니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

노. 트리 맵 차트

트리 맵 차트는 트리 다이어그램의 계층 구조를 시각화하는 대체 방법이며, 범주별로 수량도 표시합니다. 트리 맵 차트는 특히 데이터의 패턴을 식별하는 데 유용합니다. 트리 분기는 직사각형으로 표시되며, 각 하위 분기는 더 작은 직사각형으로 표시됩니다.

단순 트리 맵 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **트리 맵** 아이콘을 클릭하십시오.

   Treemap

   캔버스가 업데이트되어 트리 맵 차트 템플리트가 표시됩니다.

2. 드롭 다운 목록에서 **열** 변수를 선택하십시오.

   > **참고:** 열 변수를 더 포함시키려면 **다른 열 추가**를 클릭하십시오.

추가 기능

열

차트 열을 표시하는 데 사용할 수 있는 데이터 세트 변수가 나열됩니다.

요약

그래픽 요소에 대한 통계 요약함수를 선택하십시오. 통계의 결과에 따라 그래픽 요소의 위치가 결정됩니다.

두 가지 유형의 통계 요약함수가 있습니다. **값** 변수를 지정해야 하는지 여부가 판별되기 때문에 구분하는 것이 중요합니다.

- **값 변수가 필요하지 않은 함수**: 값 변수가 필요하지 않은 함수입니다. 모든 개수 및 퍼센트 통계가 이 범주에 속합니다. 이 통계는 정의된 **값** 변수가 없을 때 사용할 수 있습니다.
- **값 변수가 필요한 함수**: 이 함수는 **값** 변수가 필요한 함수입니다. 예를 들어 평균 함수의 경우 평균을 계산할 때 변수가 필요합니다. 이 통계는 정의된 **값** 변수가 있을 때 사용할 수 있습니다.

값

이 필드는 값 변수가 필요한 **요약** 함수를 선택한 경우에 표시됩니다. 요약값의 기준으로 사용할 변수를 선택하십시오.

리프 깊이

리프 노드에 대한 드릴다운 수준 값을 설정합니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

도. t-SNE 차트

t-SNE(T-distributed Stochastic Neighbor Embedding)는 시각화를 위한 머신 학습 알고리즘입니다. t-SNE 차트는 유사한 오브젝트는 근접 지점 옆에 모델링하고 유사하지 않은 오브젝트는 높은 확률의 먼 지점에 모델링하는 방식으로, 고차원 오브젝트 각각을 2차원 또는 3차원 지점 옆에 모델링합니다.

단순 t-SNE 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **t-SNE** 아이콘을 클릭하십시오.

   t-SNE

   캔버스가 업데이트되어 t-SNE 차트 템플리트가 표시됩니다.

2. **혼란**, **학습률**, **최대 반복** 값을 설정하십시오.
3. 선택적으로 **색상 맵** 변수를 선택하십시오.

추가 기능

혼란

각 데이터 점의 근접 이웃 수에 대해 경험적 추측을 구축할 숫자를 설정합니다. 이는 데이터의 로컬 측면과 글로벌 측면의 균형을 맞추기 위한 것입니다.

학습률

이 값은 반복 시마다 가중치 크기 변경사항을 지정하여 학습 속도에 영향을 미칩니다.

최대반복계산

수행할 최대 반복 수입니다.

색상 맵

사용 가능한 색상 맵 변수가 나열됩니다. 이 변수는 색상 진행(지정된 열의 값 범위를 기준으로 함)을 사용하여 플롯 점에 자신을 표시합니다. 색상 맵은 코로플레스 맵이라고도 합니다.

기본 제목

차트 제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

로. 단어 클라우드 차트

단어 클라우드 차트는 데이터를 단어로 표시하며, 개별 단어의 크기와 배치는 가중값에 따라 달라집니다.

단순 단어 클라우드 차트 작성

1. 차트 작성기의 **차트 유형** 섹션에서 **단어 클라우드** 아이콘을 클릭하십시오.

   Word cloud

   캔버스가 업데이트되어 단어 클라우드 차트 템플리트가 표시됩니다.

2. 변수를 **소스** 변수로 선택하십시오. 각 변수 범주가 가중값을 기준으로 차트에 표시됩니다.
3. 차트의 **형태** 값을 선택하십시오. 생성되는 차트 데이터가 선택한 형태로 표시됩니다.

추가 기능

소스

차트의 소스로 사용할 수 있는 데이터 세트 변수가 나열됩니다. 각 변수 범주가 가중값을 기준으로 차트에 표시됩니다.

형태

사용 가능한 차트 형태가 나열됩니다. 생성되는 차트 데이터가 선택한 형태로 표시됩니다.

기본 제목

차트 제목입니다.

부제목

차트 제목 바로 아래에 표시되는 차트 부제목입니다.

꼬리말

차트 아래에 표시되는 차트 꼬리말입니다.

(2) 대시보드

여러 차트를 동시에 볼 수 있는 차트 대시보드 레이아웃을 작성할 수 있습니다. 레이아웃을 템플리트로 저장한 다음 저장한 차트를 레이아웃의 위치로 끌어다 놓으십시오.

1. 데이터 노드를 마우스 오른쪽 단추로 클릭하고 데이터 보기를 선택한 후 동작 섹션에서 대시보드 제어를 클릭하십시오.

   그림 1. 대시보드 제어

   대시보드는 다음과 같은 설정을 표시하고 제공합니다.

   템플리트

   이 섹션에서는 레이아웃 템플리트를 새로 작성하거나 사전 정의된 템플리트 또는 저장된 템플리트 중에서 선택할 수 있습니다.

   레이아웃 템플리트 선택

   사용 가능한 레이아웃 템플리트 중에서 선택하거나 새 템플리트를 사용하여 시작하십시오.

   조치

   선택한 레이아웃 템플리트를 편집하려면 **레이아웃 편집** 아이콘을 클릭하십시오. 또한 대시보드 파일을 가져올 수도 있습니다. 완료되면, 템플리트를 저장하십시오.

   완료된 경우 **레이아웃 편집 모드 나가기** 아이콘을 클릭하십시오.

   컨텐츠

   이 섹션을 사용하여 항목을 대시보드 레이아웃으로 끌어다 놓을 수 있습니다.

   저장된 차트 선택

   이 섹션은 저장된 차트 목록을 제공합니다. 대시보드 레이아웃에서 원하는 위치로 차트를 끌어다 놓으십시오.

   오브젝트 선택

   HTML 텍스트 또는 이미지를 대시보드 레이아웃으로 끌어올 수도 있습니다.

(3) 글로벌 시각화 기본 설정

제목, 범위 슬라이더, 격자선 및 마우스 추적에 대한 기본 설정을 대체할 수 있습니다. 다른 색 구성표 템플리트를 지정할 수도 있습니다.

1. 데이터 노드를 마우스 오른쪽 단추로 클릭하고 **데이터 보기**를 선택한 후 **동작** 섹션에서 글로벌 시각화 기본 설정 제어를 클릭하십시오.

   그림 1. 글로벌 시각화 기본 설정 제어

   글로벌 시각화 기본 설정 대화 상자는 다음과 같은 설정을 표시하고 제공합니다.

   제목

   이 섹션은 글로벌 차트 제목 설정을 제공합니다.

   글로벌 제목

   모든 차트에 대해 글로벌 제목을 사용 또는 사용 안함으로 설정합니다.

   글로벌 1차 제목

   글로벌 1차 차트 제목의 표시를 사용 또는 사용 안함으로 설정합니다. 사용으로 설정할 경우 여기에 입력한 최상위 레벨 차트 제목이 모든 차트에 적용되며, 결과적으로 각 차트의 개별 **1차 제목** 설정을 대체합니다.

   글로벌 부제목

   글로벌 차트 부제목의 표시를 사용 또는 사용 안함으로 설정합니다. 사용으로 설정할 경우 여기에 입력한 차트 부제목이 모든 차트에 적용되며, 결과적으로 각 차트의 개별 **부제목** 설정을 대체합니다.

   기본 제목

   모든 차트에 대해 기본 제목을 사용 또는 사용 안함으로 설정합니다.

   도구

   이 섹션은 차트 동작을 제어하는 옵션을 제공합니다.

   범위 슬라이더

   각 차트에 대해 범위 슬라이더를 사용 또는 사용 안함으로 설정합니다. 사용으로 설정할 경우 각 차트의 아래에 제공되는 범위 슬라이더를 통해 표시되는 차트 데이터의 양을 제어할 수 있습니다.

   격자선

   X축(세로) 및 Y축(가로) 격자선의 표시를 제어합니다.

   마우스 추적

   사용으로 설정할 경우 커서를 차트 위의 원하는 위치에 두면 차트 데이터와 관련하여 마우스 커서 위치가 표시됩니다.

도구 상자

각 차트에 대해 도구 상자를 사용 또는 사용 안함으로 설정합니다. 차트 유형에 따라 화면의 오른쪽에 있는 도구 상자는 확대/축소, 이미지로 저장, 복원, 데이터 선택, 선택사항 지우기 등의 도구를 제공합니다.

ARIA

사용으로 설정할 경우, 장애가 있는 사용자가 웹 컨텐츠와 웹 애플리케이션에 보다 쉽게 액세스할 수 있습니다.

널값 필터링

널 차트 데이터 필터링을 사용 또는 사용 안함으로 설정합니다.

X축을 원점에

사용으로 설정할 경우 X축이 다른 원점 위치에 있습니다. 사용 안함으로 설정할 경우 X축이 항상 0에서 시작합니다.

Y축을 원점에

사용으로 설정할 경우 Y축이 다른 원점 위치에 있습니다. 사용 안함으로 설정할 경우 Y축이 항상 0에서 시작합니다.

X축 레이블 표시

X축 레이블을 사용 또는 사용 안함으로 설정합니다.

Y축 레이블 표시

Y축 레이블을 사용 또는 사용 안함으로 설정합니다.

X축 선 표시

X축 선을 사용 또는 사용 안함으로 설정합니다.

Y축 선 표시

Y축 선을 사용 또는 사용 안함으로 설정합니다.

테마

그룹화 또는 스택 변수가 있는 차트에서 사용되는 색상을 변경하려면 템플리트를 선택하십시오. 선택된 템플리트 파일에 정의된 요소 속성은 해당 요소 속성의 기본 템플리트 설정을 대체합니다.

2. 설정을 저장하려면 **적용**을 클릭하고, 변경을 무시하려면 **취소**를 클릭하십시오.

6) 출력 작업

일부 스트림을 실행할 때 모델 너깃 노드의 모델 탭 또는 **고급** 탭을 통해 뷰어에서 결과를 사용할 수 있습니다. 뷰어에서 보려는 출력으로 쉽게 이동할 수 있습니다. 출력을 조작하고 원하는 출력을 정확히 포함하는 문서를 작성할 수도 있습니다. 또한 일부 그래프 출력의 경우 뷰어가 사용됩니다.

뷰어는 IBM® SPSS® Modeler에서 다음 출력에 사용됩니다.

- TCM 모델 너깃
- STP 모델 너깃
- TwoStep-AS 군집 모델 너깃
- GSAR 모델 너깃
- 맵 시각화 그래프 노드

(1) 뷰어

결과는 뷰어에 표시됩니다. 뷰어를 사용하여 다음을 수행할 수 있습니다.

- 결과 찾아보기
- 선택한 테이블과 차트를 표시하거나 숨깁니다.
- 선택한 항목을 이동하여 결과의 표시 순서를 변경합니다.
- 항목을 뷰어와 다른 애플리케이션 간에 이동합니다.

뷰어는 두 개의 분할창으로 나뉘어져 있습니다.

- 왼쪽 분할창에는 내용의 개요 보기가 있습니다.
- 오른쪽 분할창에는 통계표, 차트, 텍스트 출력이 들어 있습니다.

개요에서 항목을 눌러 해당하는 테이블이나 차트로 바로 이동할 수 있습니다. 개요 분할창의 오른쪽 경계선을 끌어 개요 분할창의 너비를 변경합니다.

① 결과 표시 및 숨김

뷰어에서 전체 프로시저의 개별 테이블이나 결과를 선택적으로 표시하거나 숨길 수 있습니다. 이 방법은 컨텐츠 창에서 볼 수 있는 출력 양을 줄이려고 할 때 유용합니다.

가. 테이블 및 차트 숨김

1. 뷰어의 개요 분할창에서 항목의 책 모양 아이콘을 두 번 클릭하십시오.
   또는
2. 항목을 클릭하여 선택하십시오.
3. 메뉴에서 다음을 선택하십시오.
   보기 > 숨김

   또는
4. 개요 도구 모음에서 닫힌 책 모양(숨김) 아이콘을 클릭하십시오.

열린 책 모양(표시) 아이콘은 활성 아이콘이 되며 해당 항목이 현재 숨겨져 있음을 나타냅니다.

나. 프로시저 결과 숨김

1. 개요 분할창에서 프로시저 이름 왼쪽의 상자를 클릭하십시오.

프로시저의 결과를 모두 숨기고 개요 보기를 접습니다.

② 출력 이동, 삭제, 복사

출력을 항목이나 항목 그룹의 복사, 이동, 삭제 등의 방법으로 재정렬합니다.

가. 뷰어에서 출력 이동

1. 개요 또는 내용 창에서 항목을 선택하십시오.
2. 선택한 항목을 다른 위치로 끌어다 놓습니다.

1단계에서 잘라낸 항목이 선택한 위치로 이동합니다.

나. 뷰어에서 출력 삭제

1. 개요 또는 내용 창에서 항목을 선택하십시오.
2. **Delete** 키를 누르십시오.
   또는

3. 메뉴에서 다음을 선택하십시오.
   편집 > 삭제

③ 초기 맞춤 변경

기본적으로 모든 결과는 처음에는 왼쪽 맞춤으로 표시됩니다. 새로운 출력 항목의 초기 맞춤 변경

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션

2. **뷰어** 탭을 클릭하십시오.
3. 초기 출력결과 상태 그룹에서 항목 유형(예: 피벗 테이블. 차트, 텍스트 출력)을 선택합니다.
4. 원하는 맞춤 옵션을 선택하십시오.

④ 출력 항목의 맞추기 변경

1. 개요 또는 컨텐츠 창에서 맞추려는 항목을 선택합니다.
2. 메뉴에서 다음을 선택하십시오.
   형식 > 왼쪽 맞춤

   또는

   형식 > 가운데 맞춤

   또는

   형식 > 오른쪽 맞춤

⑤ 뷰어 개요

개요 분할창은 뷰어 문서의 목차를 제공합니다. 개요 분할창을 사용하여 결과를 탐색하고 표시를 제어할 수 있습니다. 개요 분할창의 동작은 대부분 내용 창에 영향을 줍니다.

- 개요 분할창에서 항목을 선택하면 내용 창에서 해당하는 항목도 표시됩니다.
- 개요 분할창에서 항목을 이동하면 내용 창에서 해당하는 항목도 이동됩니다.
- 개요 보기를 접으면 접혀진 수준의 모든 항목으로부터 결과를 숨깁니다.

개요 표시 제어

개요 표시 형식을 제어하려면 다음을 수행할 수 있습니다.

- 개요 보기를 펼치거나 접습니다.
- 선택한 항목에 대한 개요 수준을 바꿉니다.
- 개요 표시에서 항목 크기를 바꿉니다.
- 개요 표시에 사용하는 글꼴을 바꿉니다.

가. 개요 보기 접기 및 펼치기

1. 접거나 펼치려는 개요 항목의 왼쪽에 있는 상자를 클릭하십시오.
   또는
2. 개요에 있는 항목을 클릭하십시오.
3. 메뉴에서 다음을 선택하십시오.
   보기 > 접기

   또는 **보기 > 펼치기**

나. 개요 수준 변경

1. 개요 분할창에 있는 항목을 클릭하십시오.
2. 메뉴에서 다음을 선택하십시오.
   편집 > 개요 > 올리기

   또는

   편집 > 개요 > 내리기

다. 개요 항목의 크기 변경

1. 메뉴에서 다음을 선택하십시오.
   보기 > 개요 크기

2. 개요 크기(**작게**, **중간** 또는 **크게**)를 선택하십시오.

라. 개요에서 글꼴 변경

1. 메뉴에서 다음을 선택하십시오.
   보기 > 개요 글꼴...

2. 글꼴을 선택하십시오.

⑥ 항목 편집 및 뷰어에 추가

뷰어에서 제목, 새로운 텍스트, 차트 또는 다른 애플리케이션의 자료 같은 항목을 편집하고 추가할 수 있습니다.

가. 제목 또는 텍스트 추가

테이블이나 차트에 연결되어 있지 않은 텍스트 항목을 뷰어에 추가할 수 있습니다.

1. 제목이나 텍스트 앞에 올 테이블, 차트, 기타 오브젝트를 클릭하십시오.
2. 메뉴에서 다음을 선택하십시오.
   삽입 > 새 제목

   또는

   삽입 > 새 텍스트

3. 새 오브젝트를 두 번 클릭하십시오.
4. 텍스트를 입력하십시오.

나. 텍스트 파일 추가

1. 뷰어의 개요 영역이나 내용 창에서 텍스트 앞에 올 테이블, 차트 또는 기타 오브젝트를 클릭하십시오.
2. 메뉴에서 다음을 선택하십시오.
   삽입 > 텍스트 파일...
3. 텍스트 파일을 선택하십시오.

텍스트를 편집하려면 이를 두 번 클릭하십시오.

다. 뷰어에 오브젝트 붙여넣기

다른 애플리케이션의 오브젝트를 뷰어에 붙여넣을 수 있습니다. **뒤에 붙여넣기** 또는 **선택하여 붙여넣기**를 사용할 수 있습니다. 붙여넣기 유형은 뷰어에서 현재 선택된 오브젝트 뒤에 새 오브젝트를 붙여 넣습니다. 붙여넣은 오브젝트의 형식을 선택하려면 **선택하여 붙여넣기**를 사용하십시오.

⑦ 뷰어에서 찾기 및 바꾸기

1. 뷰어에서 정보를 찾거나 바꾸려면 메뉴에서 다음을 선택합니다.
   편집 > 찾기

   또는

   편집 > 바꾸기

찾기 및 바꾸기를 사용하여 다음 작업을 수행할 수 있습니다.

- 전체 문서 또는 선택한 항목만 검색합니다.
- 현재 위치에서 위나 아래로 검색합니다.
- 두 영역을 모두 검색하거나 내용 또는 개요 분할창으로 검색을 제한합니다.
- 숨겨진 항목을 검색합니다. 여기에는 내용 창의 숨겨진 항목(예: 기본적으로 숨겨지는 참고 테이블) 및 피벗 테이블의 숨겨진 행과 열이 포함됩니다.
- 대소문자가 매치하도록 검색 기준을 제한합니다.
- 전체 셀 내용이 매치하도록 피벗 테이블의 검색 기준을 제한합니다.
- 피벗 테이블의 검색 기준을 꼬리말 표식으로만 제한합니다. 뷰어의 선택에 피벗 테이블이 아닌 다른 것이 포함될 경우 이 옵션을 사용할 수 없습니다.

숨겨진 항목 및 피벗 테이블 레이어

- 다차원 피벗 테이블의 현재 표시된 레이어 아래에 있는 레이어는 숨겨진 것으로 간주되지 않으며 숨겨진 항목이 검색에 포함되어 있지 않을 때도 검색 영역에 포함됩니다.
- 숨겨진 항목은 내용 창의 숨겨진 항목(내용 창에서 닫힌 책 모양 아이콘이 있는 항목 또는 개요 분할창의 접힌 블록 내에 포함된 항목) 및 기본적으로 숨겨졌거나(예: 기본적으로 숨겨진 빈 열과 행) 테이블을 편집하고 특정 행이나 열을 선택적으로 숨겨 수동으로 숨겨진 행과 열을 포함합니다. 숨겨진 항목은 **숨겨진 항목 포함**을 명시적으로 선택하는 경우에만 검색에 포함됩니다.
- 두 경우 모두 검색 텍스트나 값이 포함되어 있는 숨겨졌거나 표시되지 않은 요소는 발견되면 표시되지만 그 후에 항목은 원래 상태로 돌아갑니다.

피벗 테이블의 값 범위 검색

피벗 테이블의 특정 값 범위에 속하는 값을 찾으려면 다음을 수행합니다.

1. 피벗 테이블을 활성화하거나 뷰어에서 하나 이상의 피벗 테이블을 선택합니다. 피벗 테이블만 선택했는지 확인합니다. 다른 오브젝트가 선택된 경우 범위 옵션을 사용할 수 없습니다.
2. 메뉴에서 다음을 선택하십시오.
   편집 > 찾기

3. **범위** 탭을 클릭하십시오.
4. 범위 유형: 사이, 이상 또는 이하에서 선택합니다.
5. 값을 선택하거나 범위를 정의하는 값을 선택합니다.

- 어느 한쪽 값에 숫자가 아닌 문자가 포함될 경우 두 값 모두 문자열로 처리됩니다.
- 양쪽 값이 숫자일 경우 숫자 값만 검색됩니다.
- 범위 탭을 사용하여 값을 바꿀 수 없습니다.

이 기능은 레거시 테이블에 사용할 수 없습니다. 자세한 정보는 레거시 테이블의 내용을 참조하십시오.

(2) 다른 애플리케이션으로 출력 복사

출력 오브젝트를 다른 워드프로세서나 스프레드시트 같은 애플리케이션으로 복사하고 붙여넣을 수 있습니다. 여러 가지 형식으로 출력을 붙여 넣을 수 있습니다. 대상 애플리케이션 및 선택한 출력 오브젝트에 따라 다음 형식의 일부 또는 모두를 사용할 수 있습니다.

메타파일. WMF 및 EMF 메타파일 형식. 이러한 형식은 Windows 운영 체제에서만 사용할 수 있습니다.

RTF(서식있는 텍스트). 선택한 다중 오브젝트, 텍스트 출력 및 피벗 표는 RTF 형식으로 복사하고 붙여넣을 수 있습니다. 피벗 표의 경우 대부분의 애플리케이션에서 테이블을 다른 애플리케이션에서 편집 가능한 테이블로 붙여 넣을 수 있음을 의미합니다. 문서 너비보다 너무 큰 피벗 표는 피벗 표 옵션 설정에 따라 문서 너비에 맞게 줄 바꿈 또는 축소되거나 변경되지 않고 그대로 있게 됩니다. 자세한 정보는 피벗 테이블 옵션 주제를 참조하십시오.

> **참고**: 매우 넓은 테이블은 Microsoft Word에서 올바르게 표시되지 않을 수 있습니다.

이미지. JPG 및 PNG 이미지 형식.

BIFF. 피벗 표 및 텍스트 출력은 BIFF 형식으로 스프레드시트에 붙여 넣을 수 있습니다. 피벗 표의 숫자는 숫자 정밀도를 유지합니다. 이 형식은 Windows 운영 체제에서만 사용할 수 있습니다.

텍스트. 피벗 표 및 텍스트 출력은 텍스트로 복사하고 붙여 넣을 수 있습니다. 이 프로세스는 텍스트만 허용하거나 전송할 수 있는 이메일과 같은 애플리케이션에서 유용합니다.

Microsoft Office Graphic Object. 이 형식을 지원하는 차트를 Microsoft Office 애플리케이션에 복사하여 해당 애플리케이션에서 기본 Microsoft Office 차트로 편집할 수 있습니다. SPSS® Statistics/SPSS Modeler 차트와 Microsoft Office 차트 사이의 차이점으로 인해 SPSS Statistics/SPSS Modeler 차트의 일부 기능은 복사된 버전에 유지되지 않습니다. Microsoft Office Graphic Object 형식으로 선택한 여러 차트를 복사하는 기능은 지원되지 않습니다.

대상 애플리케이션이 여러 사용 가능한 형식을 지원하는 경우, 선택하여 붙여넣기 메뉴 항목을 갖고 있어 형식을 선택할 수 있도록 하거나 사용 가능한 형식 목록을 자동으로 표시할 수 있습니다.

> **참고**: 상자도표 및 히스토그램을 복사하여 붙여넣을 때 Microsoft Office 버전 16(이상)이 필요합니다.

여러 출력 오브젝트를 복사하여 붙여넣기

다중 출력 오브젝트를 다른 애플리케이션으로 붙여넣을 경우 다음 제한이 적용됩니다.

- **RTF 형식.** 대부분의 애플리케이션에서 피벗 표는 해당 애플리케이션에서 편집할 수 있는 테이블로 붙여 넣어집니다. 차트, 트리 및 모델 보기는 이미지로 붙여 넣어집니다.
- **메타파일 및 이미지 형식.** 선택한 모든 출력 오브젝트는 다른 애플리케이션에 하나의 오브젝트로 붙여 넣어집니다.

- **BIFF 형식.** 차트, 트리 및 모델 보기가 제외됩니다.

선택하여 복사

대량의 출력을 복사 및 붙여넣을 때 특히 매우 큰 피벗 표는 편집 > 선택하여 복사를 사용하여 클립보드로 복사되는 형식 수를 제한하여 작업 속도를 개선할 수 있습니다.

선택한 형식을 클립보드에 복사할 기본 형식 세트로 저장할 수도 있습니다. 이 설정은 세션 전체에 지속됩니다.

복사 유형

출력 뷰어에서 마우스 오른쪽 단추로 선택된 오브젝트를 클릭하고 **편집** > **복사 유형**을 선택하여 가장 대중적인 형식(예: **모두**, **이미지** 또는 **Microsoft Office 그래픽 오브젝트**)으로 복사하십시오. **편집** > **복사**를 선택하면 **모두** 복사됩니다. 선택된 오브젝트에 대해 **복사 유형**이 회색으로 표시되거나 표시되지 않으면 해당되는 특정 오브젝트에 대해 이 복사 형식을 사용할 수 없습니다.

(3) 대화형 출력

대화형 출력 오브젝트는 여러 개의 관련 출력 오브젝트를 포함합니다. 한 오브젝트에서의 선택사항은 다른 오브젝트에 표시되거나 강조 표시되는 내용을 변경할 수 있습니다. 예를 들어, 테이블에서 행을 선택하면 맵의 영역이 강조 표시되거나 다른 범주의 차트가 표시될 수도 있습니다.

대화형 출력 오브젝트는 텍스트, 색상, 글꼴 또는 테이블 경계선 변경 등과 같은 편집 기능을 지원하지 않습니다. 개별 오브젝트는 대화형 오브젝트에서 뷰어로 복사될 수 있습니다. 대화형 출력에서 복사된 테이블은 피벗 테이블 편집기에서 편집할 수 있습니다.

대화형 출력으로부터 오브젝트 복사

파일>뷰어로 복사는 개별 출력 오브젝트를 뷰어 창으로 복사합니다.

- 사용 가능한 옵션은 대화형 출력의 내용에 따라 다릅니다.
- **차트** 및 **맵**은 차트 오브젝트를 작성합니다.
- **테이블**은 피벗 테이블 편집기에서 편집할 수 있는 피벗 테이블을 작성합니다.
- **스냅샷**은 현재 보기의 이미지를 작성합니다.
- **모델**은 현재 대화형 출력 오브젝트의 사본을 작성합니다.

편집>오브젝트 복사는 개별 출력 오브젝트를 클립보드로 복사합니다.

- 복사된 오브젝트를 뷰어에 붙여넣는 것은 **파일>뷰어로 복사**에 해당합니다.
- 오브젝트를 다른 애플리케이션에 붙여넣으면 오브젝트를 이미지로 붙여넣습니다.

확대/축소 및 팬

맵의 경우 **보기>확대/축소**를 사용하여 맵의 보기를 확대/축소할 수 있습니다. 확대/축소된 맵 보기 내에서 **보기>팬**을 사용하여 보기를 이동할 수 있습니다.

인쇄 설정

파일>인쇄 설정은 대화형 오브젝트를 인쇄하는 방법을 제어합니다.

- **볼 수 있는 보기만 인쇄.** 현재 표시되는 보기만을 인쇄합니다. 이 옵션은 기본 설정입니다.
- **모든 보기 인쇄.** 대화형 출력에 포함된 모든 보기를 인쇄합니다.
- 선택한 옵션은 또한 출력 오브젝트를 내보내기 위한 기본 동작을 결정합니다.

(4) 출력 내보내기

내보내기 출력결과는 뷰어 결과를 HTML, 텍스트, Work/RTF, Excel, PowerPoint(PowerPoint 97 이상 필요) 또는 PDF 형식으로 저장합니다. 여러 가지 서로 다른 그래픽 형식으로 차트를 내보낼 수도 있습니다.

ⓘ **참고**: PowerPoint로 내보내기는 Windows 운영 체제에서만 사용할 수 있습니다.

출력 내보내기

1. 뷰어를 활성 대화 상자로 설정하십시오(대화 상자에서 임의 위치 클릭).
2. 도구 모음에서 **내보내기** 단추를 클릭하거나 출력 창을 마우스 오른쪽 단추로 클릭하고 **내보내기**를 선택하십시오.
3. 파일 이름(또는 차트의 접두문자)을 입력하고 내보내기 형식을 선택하십시오.

내보낼 오브젝트. 뷰어의 모든 오브젝트, 보이는 모든 오브젝트, 또는 선택한 오브젝트만 내보낼 수 있습니다.

문서 유형. 다음 옵션을 사용할 수 있습니다.

- **Word/RTF(*.doc)**: 피벗 테이블은 모든 형식 속성이 온전한 상태로 Word 테이블로 내보냅니다(예: 셀 경계선, 글꼴 유형, 배경색). 텍스트 출력은 서식있는 RTF로 내보내집니다. 차트, 트리 다이어그램 및 모델 보기는 PNG 형식에 포함됩니다. Microsoft Word는 너무 넓은 테이블을 제대로 표시하지 않을 수도 있습니다.
- **Excel 97-2004(*.xls)/Excel 2007 이상(*.xlsx)**: 피벗 테이블 행, 열 및 셀은 모든 형식 속성이 온전한 상태로 Excel 행, 열 및 셀로 내보냅니다(예: 셀 경계, 글꼴 유형 및 배경색). 텍스트 출력은 모든 글꼴 속성을 그대로 유지한 채 내보냅니다. 텍스트 출력의 각 줄은 줄의 전체 내용이 한 셀이 있는 Excel 파일의 행입니다. 차트, 트리 다이어그램 및 모델 보기는 PNG 형식에 포함됩니다. 출력은 Excel 97-2004 또는 Excel 2007 이상으로 내보낼 수 있습니다.
- **HTML(*.htm)**: 피벗 테이블을 HTML 테이블로 내보냅니다. 텍스트 출력은 형식이 미리 설정된 HTML로 내보냅니다. 차트, 트리 다이어그램 및 모델 보기는 선택된 그래픽 형식으로 문서에 임베드됩니다. HTML 5와 호환 가능한 브라우저는 HTML 형식으로 내보내는 출력을 보는 데 필요합니다.
- **포터블 문서 형식(*.pdf)**: 모든 출력이 인쇄 미리보기에 표시되는 그대로 내보내지고 모든 형식 속성이 유지됩니다.
- **텍스트 - 일반/UTF8/UTF16 (*.txt)**: 텍스트 출력 형식은 일반 텍스트, UTF-8 및 UTF-16을 포함합니다. 피벗 테이블은 탭으로 구분된 형식이나 공백으로 구분된 형식으로 내보낼 수 있습니다. 모든 텍스트 출력은 공백-구분 형식으로 내보냅니다. 차트, 트리 다이어그램 및 모델 보기의 경우 줄은 각 그래픽마다 텍스트 파일에 삽입되어 이미지 파일 이름을 표시합니다.

- **없음(그래픽만)** 사용 가능한 내보내기 형식에는 EPS, JPEG, TIFF, PNG 및 BMP가 포함됩니다. Windows 운영 체제에서는 EMF(확장 메타파일) 형식도 사용할 수 있습니다.

포함 폴더 열기. 내보내기가 작성한 파일을 포함하는 폴더를 엽니다.

① HTML 옵션

HTML 내보내기에는 HTML 5와 호환 가능한 브라우저가 필요합니다.

출력을 HTML 형식으로 내보낼 때 사용할 수 있는 옵션은 다음과 같습니다.

피벗 표의 레이어. 기본적으로 피벗 표 레이어의 포함 또는 제외는 각 피벗 표의 테이블 특성으로 제어됩니다. 이 설정을 무시하고 모든 레이어를 포함하거나 현재 표시된 레이어를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.

대화형으로 레이어된 테이블 내보내기. 레이어된 테이블은 뷰어에 나타날 때 표시되고 표시된 레이어를 브라우저에서 대화식으로 변경할 수 있습니다. 이 옵션이 선택되지 않으면 각 테이블 레이어는 별도의 테이블로 표시됩니다.

테이블을 HTML로. 내보낸 피벗 테이블에 포함된 유형 정보를 제어합니다.

- **유형 및 고정 열 너비로 내보내기**. 모든 피벗 테이블 유형 정보(글꼴 유형, 배경색 등) 및 열 너비가 보존됩니다.
- **유형 없이 내보내기**. 피벗 테이블이 기본 HTML 테이블로 변환됩니다. 유형 속성이 유지되지 않습니다. 열 너비가 자동으로 결정됩니다.

꼬리말과 캡션을 포함합니다. 모든 피벗 테이블 꼬리말과 캡션의 포함과 제외를 제어합니다.

모델 보기. 기본적으로 모델 보기 포함 또는 제외 여부는 각 모델의 모델 특성에 의해 제어됩니다. 이 설정을 무시하고 모든 보기를 포함하거나 현재 표시된 보기를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 모델 특성의 내용을 참조하십시오. (참고: 테이블을 포함하여 모든 모델 보기는 그래픽으로 내보냅니다.)

참고: HTML의 경우 내보낸 차트의 이미지 파일 형식을 제어할 수도 있습니다. 자세한 정보는 그래픽 형식 옵션의 내용을 참조하십시오.

HTML 내보내기 옵션 설정

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...

3. 내보내기 형식으로 **HTML**을 선택하십시오.
4. **옵션 변경**을 클릭하십시오.

② 웹 보고서 옵션

웹 보고서는 대부분의 브라우저와 호환 가능한 대화식 문서입니다. 뷰어에서 사용 가능한 피벗 표의 대화식 기능의 대부분은 웹 보고서에서도 사용 가능합니다.

보고서 제목. 보고서의 헤더에 표시되는 제목. 기본적으로, 파일 이름을 사용합니다. 파일 이름 대신에 사용할 사용자 정의 제목을 지정할 수 있습니다.

형식. 보고서 형식에는 두 개의 옵션이 있습니다.

- **SPSS 웹 보고서(HTML 5)**. 이 형식은 HTML 5와 호환 가능한 브라우저가 필요합니다.
- **Cognos Active Report(mht)**. 이 형식에는 MHT 형식 파일 또는 Cognos Active Report 애플리케이션을 지원하는 브라우저가 필요합니다.

오브젝트 제외. 선택한 오브젝트 유형을 보고서에서 제외할 수 있습니다.

- **텍스트**. 로그가 아닌 텍스트 오브젝트입니다. 이 옵션에는 활성 데이터 세트에 대한 정보를 포함하는 텍스트 오브젝트가 포함됩니다.
- **로그**. 실행된 명령 구문의 목록을 포함하는 텍스트 오브젝트입니다. 로그 항목에는 뷰어 출력을 생성하지 않는 명령이 마주치는 경고 및 오류 메시지를 포함합니다.
- **참고 테이블**. 통계 및 차트 프로시저의 출력에는 참고 테이블이 포함됩니다. 이 테이블에는 사용된 데이터 세트에 대한 정보, 결측값 및 프로시저를 실행하는 데 사용된 명령 구문이 포함됩니다.
- **경고 및 오류 메시지**. 통계 및 차트 프로시저로부터의 경고 및 오류 메시지

웹 보고서와 매치하도록 테이블 및 차트를 재유형화. 이 옵션은 표준 웹 보고서 양식을 모든 테이블과 차트에 적용합니다. 이 설정은 뷰어에 표시된 대로 출력에 있는 모든 글꼴, 색상 또는 다른 유형을 대체합니다. 표준 웹 보고서 유형을 수정할 수 없습니다.

웹 서버 연결. IBM® SPSS® Statistics Web Report Application Server를 실행 중인 하나 이상의 애플리케이션 서버의 URL 위치를 포함할 수 있습니다. 웹 애플리케이션 서버는 테이블을 피벗하고, 차트를 편집하고 수정된 웹 보고서를 저장하기 위한 기능을 제공합니다.

- 웹 보고서에 포함하려는 각 애플리케이션 서버마다 **사용**을 선택하십시오.
- 웹 보고서에 URL 스펙이 포함된 경우, 웹 보고서는 추가 편집 기능을 제공하기 위해 해당 애플리케이션 서버에 연결합니다.
- 여러 URL을 지정하는 경우 웹 보고서는 서버가 지정된 순서대로 각 서버에 연결하려고 시도합니다.

IBM SPSS Statistics Web Report Application Server는
http://www.ibm.com/developerworks/spssdevcentral에서 다운로드할 수 있습니다.

③ Word 옵션

결과를 Word 형식으로 내보낼 때 사용할 수 있는 옵션은 다음과 같습니다.

구분점 및 집단 보존구분점을 정의한 경우에는 이러한 설정이 Word 테이블에 보존됩니다.

꼬리말과 캡션 포함. 모든 피벗 테이블 꼬리말과 캡션의 포함 또는 제외를 제어합니다.

모델 보기. 기본적으로 모델 보기 포함 또는 제외 여부는 각 모델의 모델 특성에 의해 제어됩니다. 이 설정을 무시하고 모든 보기를 포함하거나 현재 표시된 보기를 제외한 모든 레이어를 제외할 수 있습니다. 자세한 정보는 모델 특성의 내용을 참조하십시오. (참고: 테이블을 포함하여 모든 모델 보기는 그래픽으로 내보냅니다.)

내보내기를 위한 페이지 설정. 내보낼 문서의 페이지 크기 및 여백을 정의할 수 있는 대화 상자

를 엽니다. 줄바꿈이나 축소 동작 결정에 사용하는 문서 너비는 페이지 너비에서 오른쪽 및 왼쪽 여백을 뺀 값입니다.

Word 내보내기 옵션 설정

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...

3. 내보내기 형식으로 **Word/RTF**를 선택합니다.
4. **옵션 변경**을 클릭하십시오.

④ Excel 옵션

결과를 Excel 형식으로 내보낼 때 사용할 수 있는 옵션은 다음과 같습니다.

워크시트 또는 워크북을 작성하거나 기존 워크시트를 수정. 기본적으로 새 워크북이 작성됩니다. 특정 이름이 이미 존재할 경우에는 이 이름을 덮어쓰게 됩니다. 옵션을 선택하여 워크시트를 생성할 때 지정된 이름을 가진 워크시트가 지정된 파일에 이미 존재할 경우에는 기존 이름을 덮어쓰게 됩니다. 옵션을 선택하여 기존 워크시트를 수정하려면 워크시트 이름도 지정해야 합니다. (이는 워크시트를 작성하기 위한 선택사항입니다.) 워크시트 이름은 31자를 초과할 수 없으며 슬래시(/) 또는 백슬래시(₩), 대괄호, 물음표, 별표 등이 포함될 수 없습니다.

Excel 97-2004로 내보낼 때 기존의 워크시트를 수정할 경우 차트, 모델 보기 및 트리 다이어그램은 내보낸 결과에 포함되지 않습니다.

워크시트에서의 위치. 내보낸 파일에 대한 워크시트 내의 위치를 제어합니다. 기본적으로 내보낸 파일은 첫 번째 행에서 시작하여 기존 내용을 수정하지 않고 모든 내용이 들어 있는 마지막 열 다음에 추가됩니다. 기존의 워크시트에 새 열을 추가하는 방식이 좋습니다. 마지막 열 다음에 내보낸 결과를 추가하는 것은 기존의 워크시트에 새 행을 추가하는 데 유용한 방식입니다. 특정 셀의 위치에서부터 내보내는 출력을 추가하면 내보낸 파일이 추가되는 영역에 있는 기존의 모든 내용을 덮어쓰게 됩니다.

피벗 표의 레이어. 기본적으로 피벗 표 레이어의 포함 또는 제외는 각 피벗 표의 테이블 특성으로 제어됩니다. 이 설정을 무시하고 모든 레이어를 포함하거나 현재 표시된 레이어를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.

꼬리말과 캡션 포함. 모든 피벗 테이블 꼬리말과 캡션의 포함 또는 제외를 제어합니다.

모델 보기. 기본적으로 모델 보기 포함 또는 제외 여부는 각 모델의 모델 특성에 의해 제어됩니

다. 이 설정을 무시하고 모든 보기를 포함하거나 현재 표시된 보기를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 모델 특성의 내용을 참조하십시오. (참고: 테이블을 포함하여 모든 모델 보기는 그래픽으로 내보냅니다.)

Excel 내보내기 옵션 설정

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...

3. 내보내기 형식으로 **Excel**을 선택하십시오.
4. **옵션 변경**을 클릭하십시오.

⑤ PowerPoint 옵션

다음 옵션을 PowerPoint에 사용할 수 있습니다.

피벗 표의 레이어. 기본적으로 피벗 표 레이어의 포함 또는 제외는 각 피벗 표의 테이블 특성으로 제어됩니다. 이 설정을 무시하고 모든 레이어를 포함하거나 현재 표시된 레이어를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.

넓은 피벗 표. 정의된 문서 너비에 비해 너무 넓은 테이블 처리를 제어합니다. 기본적으로 테이블이 너비에 맞게 줄바꿈됩니다. 테이블이 여러 섹션으로 나뉘며 행 레이블이 테이블의 각 섹션에 대해 반복됩니다. 또는 넓은 테이블을 축소하거나 그대로 둘 수 있으며 정의된 문서의 너비보다 크게 확장할 수 있습니다.

꼬리말과 캡션을 포함합니다. 모든 피벗 표 꼬리말과 캡션의 포함과 제외를 제어합니다.

뷰어 개요 항목을 슬라이드 제목으로 사용. 내보내기로 작성한 각 슬라이드에 제목을 포함합니다. 각 슬라이드에는 뷰어에서 내보내는 항목이 하나씩 포함됩니다. 제목은 뷰어의 개요 분할창에 있는 항목에 대한 개요 항목을 통해 구성됩니다.

모델 보기. 기본적으로 모델 보기 포함 또는 제외 여부는 각 모델의 모델 특성에 의해 제어됩니다. 이 설정을 무시하고 모든 보기를 포함하거나 현재 표시된 보기를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 모델 특성의 내용을 참조하십시오. (참고: 테이블을 포함하여 모든 모델 보기는 그래픽으로 내보냅니다.)

내보내기를 위한 페이지 설정. 내보낼 문서의 페이지 크기 및 여백을 정의할 수 있는 대화 상자를 엽니다. 줄바꿈이나 축소 동작 결정에 사용하는 문서 너비는 페이지 너비에서 오른쪽 및 왼쪽 여백을 뺀 값입니다.

PowerPoint 내보내기 옵션 설정

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...

3. 내보내기 형식으로 **PowerPoint**를 선택합니다.
4. **옵션 변경**을 클릭하십시오.

참고: PowerPoint로 내보내기는 Windows 운영 체제에서만 사용 가능합니다.

⑥ PDF 옵션

다음 옵션을 PDF에 사용할 수 있습니다.

책갈피 포함. 이 옵션을 선택하면 뷰어 개요 항목에 해당하는 책갈피가 PDF 문서에 포함됩니다. 뷰어 개요 분할창에서와 마찬가지로 책갈피를 사용하면 출력 오브젝트 수가 매우 많은 문서도 훨씬 쉽게 탐색할 수 있습니다.

글꼴 포함. 어떤 컴퓨터에서건 PDF 문서가 동일하게 표시되도록 글꼴을 삽입합니다. 이 옵션을 사용하지 않는 경우 PDF 문서를 보거나 인쇄하는 데 사용하는 컴퓨터에 이 문서를 작성할 때 사용했던 글꼴 중 일부가 설치되어 있지 않으면 대체 글꼴을 사용해야 하므로 최상의 결과물을 얻지 못할 수도 있습니다.

피벗 표의 레이어. 기본적으로 피벗 표 레이어의 포함 또는 제외는 각 피벗 표의 테이블 특성으로 제어됩니다. 이 설정을 무시하고 모든 레이어를 포함하거나 현재 표시된 레이어를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.

모델 보기. 기본적으로 모델 보기 포함 또는 제외 여부는 각 모델의 모델 특성에 의해 제어됩니다. 이 설정을 무시하고 모든 보기를 포함하거나 현재 표시된 보기를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 모델 특성의 내용을 참조하십시오. (참고: 테이블을 포함하여 모든 모델 보기는 그래픽으로 내보냅니다.)

PDF 내보내기 옵션 설정

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...

3. 내보내기 형식으로 **포터블 문서 형식**을 선택합니다.

4. **옵션 변경**을 클릭하십시오.

PDF 출력에 영향을 주는 기타 설정

페이지 설정/페이지 속성. PDF 문서의 페이지 크기, 방향, 여백, 페이지 머리말과 꼬리말의 컨텐츠 및 출력, 인쇄된 문자 크기는 페이지 설정과 페이지 속성 옵션으로 제어합니다.

테이블 특성/테이블모양. 가로로 길거나 세로로 긴 테이블의 척도 및 테이블 레이어의 인쇄는 각 테이블에 대한 테이블 특성을 통해 제어합니다. 이러한 특성은 테이블모양에도 저장할 수 있습니다.

기본값/현재 프린터. PDF 문서의 해상도(DPI)는 기본값 프린터 또는 현재 선택된 프린터의 현재 해상도 설정입니다. 이는 페이지 설정을 사용하여 변경할 수 있습니다.최대 해상도는 1200DPI입니다.프린터 설정이 이보다 높더라도 PDF 문서는 1200DPI로 설정됩니다.

참고: 고해상도 문서를 저해상도 프린터에서 인쇄하면 인쇄 품질이 떨어질 수 있습니다.

⑦ 텍스트 옵션

텍스트 출력 옵션 설정

다음 옵션을 텍스트 내보내기에 사용할 수 있습니다.

피벗 표 형식. 피벗 테이블은 탭으로 구분된 형식이나 공백으로 구분된 형식으로 내보낼 수 있습니다. 공백으로 구분된 형식의 경우 다음을 제어할 수도 있습니다.

- **열 너비**. **자동맞춤**은 열 컨텐츠를 줄바꿈하지 않으며 각 열의 너비는 해당 열에서 가장 넓은 레이블이나 값 너비와 동일합니다. **사용자 정의** 설정은 테이블에 있는 모든 열에 적용되는 최대 열 너비를 설정하며 이 너비를 초과하는 값은 해당 열에서 다음 줄로 줄바꿈됩니다.

- **행/열 경계 문자**. 행과 열 경계를 작성하는 데 사용하는 문자를 제어합니다. 행과 열 경계를 표시하지 않으려면 값으로 빈 공백을 입력하십시오.

피벗 표의 레이어. 기본적으로 피벗 표 레이어의 포함 또는 제외는 각 피벗 표의 테이블 특성으로 제어됩니다. 이 설정을 무시하고 모든 레이어를 포함하거나 현재 표시된 레이어를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.

꼬리말과 캡션을 포함합니다. 모든 피벗 테이블 꼬리말과 캡션의 포함과 제외를 제어합니다.

모델 보기. 기본적으로 모델 보기 포함 또는 제외 여부는 각 모델의 모델 특성에 의해 제어됩니다. 이 설정을 무시하고 모든 보기를 포함하거나 현재 표시된 보기를 제외한 모든 레이어를 제외할 수 있습니다.자세한 정보는 모델 특성의 내용을 참조하십시오. (참고: 테이블을 포함하여 모든 모델 보기는 그래픽으로 내보냅니다.)

텍스트 내보내기 옵션 설정

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...

3. 내보내기 형식으로 **텍스트**를 선택합니다.
4. **옵션 변경**을 클릭하십시오.

⑧ 이미지만 옵션

다음 옵션을 이미지만 내보내기에 사용할 수 있습니다.

1. 탭의 아무 곳이나 클릭하여 출력 탭을 활성 탭으로 설정하십시오.
2. 내보낼 차트 항목을 선택하고 수직 생략 기호 컨트롤을 클릭한 후 메뉴에서 **내보내기**를 선택하십시오. 또한 도구 모음에서 **출력 내보내기** 단추를 클릭할 수도 있습니다. 내보내기 출력결과 대화 상자가 표시됩니다.
   문서 설정

   유형
   문서 **유형**으로 **이미지만**을 선택하십시오.
   이미지 형식
   내보내는 이미지의 이미지 형식을 선택하십시오.

   PNG
   선택할 경우 이미지를 Portable Network Graphics(*.png) 형식으로 내보냅니다. 이 설정은 이미지 **척도(%)** 및 **색 농도**를 선택하는 옵션을 제공합니다.
   JPG
   선택할 경우 이미지를 Joint Photographic Experts Group(*.jpg) 형식으로 내보냅니다. 이 설정은 이미지 **척도(%)** 및 **그레이스케일로 변환**을 선택하는 옵션을 제공합니다.
   BMP
   선택할 경우 이미지를 비트맵(*.bmp) 형식으로 내보냅니다. 이 설정은 이미지 **척도(%)** 및 **파일 크기를 줄이기 위해 이미지 압축**을 선택하는 옵션을 제공합니다.

포함 폴더 열기

(선택사항) 내보내기 후 내보내기 저장 파일 위치 폴더를 열려면 이 설정을 선택하십시오.

내보낼 오브젝트

포함될 오브젝트

모든, **표시된** 또는 **선택된**(기본 설정) 출력 오브젝트를 내보내려면 선택하십시오.

3. 문서 설정을 지정한 후 **내보내기**를 클릭하십시오.

⑨ 그래픽 형식 옵션

HTML 및 텍스트 문서에서 차트 내보내기만 하는 경우에는 그래픽 형식을 선택할 수 있고 각 그래픽 형식에 대해서는 다양한 옵션 설정을 제어할 수 있습니다.

내보낸 차트의 그래픽 형식 및 옵션을 선택

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 내보내기...
3. **HTML**, **텍스트** 또는 **없음(그래픽만 해당)**을 문서 유형으로 선택하십시오.
4. 드롭 다운 목록에서 그래픽 파일 형식을 선택하십시오.
5. **옵션 변경**을 클릭하여 선택한 그래픽 파일 형식의 옵션을 변경하십시오.

가. JPEG 차트 내보내기 옵션

이미지 크기(%)

최대 200 퍼센트의 원래 차트 크기 백분율.

그레이스케일로 변환

색상을 회색 음영으로 변환합니다.

나. BMP 차트 내보내기 옵션

이미지 크기(%)

최대 200 퍼센트의 원래 차트 크기 백분율.

파일 크기를 줄이기 위해 이미지 압축

이미지 품질에 영향을 주지 않고 작은 파일을 작성하는 무손실 압축 기법입니다.

다. PNG 차트 내보내기 옵션

이미지 크기(%)

최대 200 퍼센트의 원래 차트 크기 백분율.

색 농도

내보낸 차트에서 사용하는 색상 수를 결정합니다. 임의의 깊이로 저장된 차트에는 실제로 사용된 최소 색상 수와 해당 깊이가 허용하는 최대 색상 수가 포함됩니다. 예를 들어, 차트를 세 가지 색(빨강, 흰색, 검정)을 사용하여 작성하고 16색으로 저장하면 차트는 세 가지 색상으로 표시됩니다.

> ⓘ **참고**: 차트에서 사용한 색상 수가 해당 깊이에 대한 색상 수를 초과하는 경우 색상이 혼합되어 차트의 색상을 복제합니다.

라. EMF 및 TIFF 차트 내보내기 옵션

이미지 크기. 최대 200 퍼센트의 원래 차트 크기 백분율.

참고: EMF(향상된 메타파일) 형식은 Windows 운영 체제에서만 사용 가능합니다.

마. EPS 차트 내보내기 옵션

이미지 크기. 원래 이미지 크기(최대 200 퍼센트)의 백분율로 크기를 지정하거나 이미지 너비를 픽셀(너비 값과 가로 세로 비율로 결정되는 높이)로 지정할 수 있습니다. 내보낸 이미지는 항상 원본에 비례합니다.

TIFF 미리보기 이미지 포함. EPS 이미지를 화면에 표시할 수 없는 애플리케이션에서 EPS 이미지를 볼 수 있도록 이를 TIFF 형식의 미리보기와 함께 저장합니다.

글꼴. EPS 이미지에서의 글꼴 처리 방법을 제어합니다.

- **글꼴 참조 사용**. 차트에 사용된 글꼴을 출력 장치에서도 사용할 수 있으면 해당 글꼴을 사용합니다. 그렇지 않으면 출력 장치에서 대체 글꼴을 사용합니다.
- **글꼴을 곡선으로 바꾸기**. 글꼴을 포스트스크립트 곡선 데이터로 변경합니다. 텍스트 자체는 EPS 그래픽을 편집할 수 있는 애플리케이션의 텍스트로는 더 이상 편집할 수 없습니다. 이 옵션은 차트를 작성하는 데 사용한 글꼴을 출력 장치에서 사용할 수 없는 경우에 유용합니다.

(5) 뷰어 인쇄

뷰어 창의 컨텐츠를 인쇄하는 옵션은 두 가지가 있습니다.

모든 가시출력결과. 컨텐츠 창에 현재 표시되어 있는 항목만 인쇄합니다. 숨겨진 항목(개요 분할 창에 닫힌 책 모양 아이콘이 있는 항목 또는 접힌 개요 레이어에 숨겨진 항목)은 인쇄되지 않습니다.

선택. 개요 및/또는 컨텐츠 창에서 현재 선택된 항목만 인쇄합니다.

① 출력과 차트 인쇄

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 인쇄...

3. 원하는 인쇄 설정을 선택합니다.
4. **확인**을 클릭하여 인쇄합니다.

② 인쇄 미리보기

인쇄 미리보기로 뷰어 문서의 각 페이지에 인쇄할 내용을 볼 수 있습니다. 인쇄 미리보기로 다음을 포함하여 뷰어의 컨텐츠 창에서 찾아 볼 수 없는 항목을 볼 수 있으므로 뷰어 문서를 실제로 인쇄하기 전에 인쇄 미리보기로 확인하는 것이 좋습니다.

- 페이지 구분
- 피벗 표의 은닉층
- 넓은 테이블의 구분
- 각 페이지에 인쇄되는 헤더 및 꼬리말

어떤 출력이 뷰어에 현재 선택되어 있으면 미리보기는 선택한 출력만 나타냅니다. 모든 출력에 대한 미리보기를 실행하려면 뷰어에 선택된 출력이 없는지 확인하십시오.

③ 페이지 속성: 머리말 및 꼬리말

머리말과 꼬리말은 각 페이지의 상단과 하단에 인쇄되는 정보입니다. 머리말 및 꼬리말로 사용할 텍스트를 입력할 수 있습니다. 삽입할 대화 상자 가운데에 있는 도구 모음을 사용하여 다음을 입력할 수도 있습니다.

- 날짜 및 시간
- 페이지 번호
- 뷰어 파일 이름

- 개요 머리말 레이블
- 페이지 제목 및 부제목
- **기본값 설정**은 새로운 뷰어 문서의 기본 설정으로 여기에 지정한 설정을 사용합니다. (참고: 그러면 머리말/꼬리말 탭과 기본 설정의 옵션 탭 둘 모두에서 현재 설정이 됩니다.
- 개요 머리말 레이블은 각 페이지에 있는 첫 번째 항목에 대한 첫 번째, 두 번째, 세 번째 또는 네 번째 수준의 개요 머리말을 나타냅니다.
- 페이지 제목과 부제목은 현재 페이지 제목과 부제목을 인쇄합니다. 이것은 뷰어 삽입 메뉴의 새 페이지 제목 또는 TITLE 및 SUBTITLE 명령을 사용하여 작성할 수 있습니다. 페이지 제목이나 부제목을 지정하지 않았다면 이 설정은 무시됩니다.
  참고: 새 페이지 제목과 부제목의 글꼴 특성은 옵션 대화 상자의 뷰어 탭에서 제어됩니다(편집 메뉴에서 옵션을 선택하여 액세스). 기존의 페이지 제목과 부제목에 대한 글꼴 특성을 변경하려면 뷰어에서 제목을 편집하면 됩니다.

인쇄된 페이지에서 머리말과 꼬리말이 어떻게 표시되는지 확인하려면 파일 메뉴에서 인쇄 미리보기를 선택합니다.

가. 페이지 헤더 및 꼬리말 삽입

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 헤더 및 꼬리말...

3. 각 페이지에 표시할 헤더 및/또는 꼬리말을 입력합니다.

④ 페이지 속성: 옵션

이 대화 상자는 인쇄된 차트 크기, 인쇄된 출력 항목 사이의 공간 및 페이지 번호 매김을 제어합니다.

- **차트 인쇄 크기**. 정의된 페이지 크기에 상대적인 차트 인쇄 크기를 제어합니다. 차트의 가로 세로 비율(너비 대 높이 비율)은 차트 인쇄 크기의 영향을 받지 않습니다. 차트의 전체 인쇄 크기는 높이와 너비로 제한됩니다. 차트 외부의 경계가 페이지의 왼쪽과 오른쪽 경계선에 도달하면 차트 크기가 페이지 높이를 더 채우기 위해서 더 이상 커지지 않습니다.
- **항목 간 간격**. 인쇄된 항목 간의 공백을 제어합니다. 각 피벗 표, 차트 및 텍스트 오브젝트는 개별 항목입니다. 이 설정은 뷰어의 항목 표시에 영향을 미치지 않습니다.
- **페이지 번호 시작**. 지정한 번호부터 순차적으로 페이지 번호를 시작합니다.
- **기본값 지정**. 이 옵션은 새 뷰어 문서의 기본 설정으로 여기에 지정되는 설정을 사용합니다. (이 옵션은 현재 머리말/꼬리말 설정 및 옵션 설정을 기본값으로 만드는 점을 유의하십시오.)

가. 차트 인쇄 크기, 페이지 매김, 인쇄된 항목 간 공백 변경 방법

1. 뷰어를 활성 창으로 만드십시오(창의 임의의 위치 클릭).
2. 메뉴에서 다음을 선택하십시오.
   파일 > 페이지 속성...

3. **옵션** 탭을 클릭하십시오.
4. 설정을 변경하고 **확인**을 클릭하십시오.

(6) 출력 저장

뷰어의 내용은 저장할 수 있습니다.

- **출력 오브젝트(*.cou)**. 이 형식은 그래프, 탭, 주석(Annotation) 등을 포함하여 전체 출력 컨테이너를 저장합니다. 이 형식은 IBM® SPSS® Modeler에서 열리고 표시되고, 프로젝트에 추가되고, IBM SPSS Collaboration and Deployment Services Repository를 사용하여 게시되고 추적됩니다. 이 형식은 IBM SPSS Statistics과 호환되지 않습니다.
- **뷰어 파일(*.spv)**. 뷰어 창에 파일을 표시하는 데 사용되는 형식입니다. IBM SPSS Modeler에서 모델 너깃으로부터 이 형식을 저장하면 **모델** 탭의 뷰어 내용만 저장됩니다.

웹 보고서를 저장하기 위한 옵션을 제어하거나 결과를 다른 형식으로 저장하려면(예: 텍스트, Word, Excel), **파일** 메뉴에서 **내보내기**를 사용하십시오.

① 뷰어 문서 저장

1. 뷰어 창에서 다음을 선택합니다.
   파일 > 저장

2. 문서 이름을 입력하고 **저장**을 클릭하십시오.
   선택적으로 다음을 수행할 수 있습니다.

   파일을 잠가 IBM® SPSS® Smartreader에서의 편집 방지.
   뷰어 문서가 잠기면 피벗 테이블을 조작할 수 있습니다(행과 열 바꾸기, 표시된 레이어 변경 등). 그러나 IBM SPSS Smartreader에서 출력을 편집하거나 변경사항을 뷰어 문서에 저장할 수 없습니다(뷰어 문서에서 작업하기 위한 별도의 제품). 이 설정은 IBM SPSS Statistics 또는 IBM SPSS Modeler에서 열린 뷰어 문서에는 영향을 미치지 않습니다.
   파일을 비밀번호로 암호화
   비밀번호로 문서를 암호화하여 뷰어 문서에 저장된 기밀 정보를 보호할 수 있습니다. 암호화되면 비밀번호를 제공해야만 문서를 열 수 있습니다. IBM SPSS Smartreader 사용자

또한 파일을 열기 위해서는 비밀번호를 제공해야 합니다.
뷰어 문서를 암호화는 방법은 다음과 같습니다.

a. 다른 이름으로 출력 저장 대화 상자에서 **비밀번호로 파일 암호화**를 선택합니다.
b. **저장**을 클릭합니다.
c. 파일 암호화 대화 상자에서 비밀번호를 제공하고 비밀번호 확인 텍스트 상자에 비밀번호를 다시 입력합니다. 비밀번호는 10자로 제한되며 대소문자를 구분합니다.

경고: 비밀번호는 분실되면 복구할 수 없습니다. 비밀번호를 분실하면 파일을 열 수 없습니다.

강력한 비밀번호 작성

- 8자 이상을 사용하십시오.
- 비밀번호에 숫자, 기호 및 문장 부호까지 포함하십시오.
- "123" 및 "abc"처럼 연속된 숫자나 문자를 피하며 "111aaa"처럼 반복하지 않습니다.
- 생일 또는 별명과 같은 개인 정보를 사용하여 비밀번호를 만들지 마십시오.
- 비밀번호를 주기적으로 변경하십시오.

참고: 암호화된 파일을 IBM SPSS Collaboration and Deployment Services Repository에 저장하는 것은 지원되지 않습니다.

암호화된 파일 수정

- 암호화된 파일을 열어 파일을 수정하고 파일 > 저장을 선택할 경우 수정된 파일은 같은 비밀번호로 저장됩니다.
- 파일을 열고 암호화 단계를 반복한 후 파일 암호화 대화 상자에 다른 비밀번호를 지정하여 비밀번호를 변경할 수 있습니다.
- 파일을 열고 파일 > 다른 이름으로 저장을 선택한 후 다른 이름으로 출력 저장 대화 상자에서 **비밀번호로 파일 암호화**를 선택 해제하여 암호화된 파일의 암호화되지 않은 버전을 저장할 수 있습니다.

참고: 암호화된 데이터 파일 및 출력 문서는 버전 21 이전의 IBM SPSS Statistics 버전에서는 열 수 없습니다. 암호화된 구문 파일은 버전 22 이전에서는 열 수 없습니다.

필수 모델 정보를 출력 문서와 함께 저장

이 옵션은 일부 대화식 기능을 사용하기 위해서는 보조 정보가 필요한 출력 문서에 모델 뷰어 항목이 있는 경우에만 적용됩니다. 보조 정보를 요구하는 대화식 기능 및 이러한 모델 뷰어 항목 목록을 표시하려면 **자세한 정보**를 클릭하십시오. 이 정보를 출력 문서와 함께 저장하면 문서 크기가 대폭 커집니다. 이 정보를 저장하지 않기로 선택한 경우에도 이러한 출력 항목을 열 수는 있지만 지정된 대화식 기능은 사용할 수 없습니다.

(7) 피벗 테이블

① 피벗 테이블

여러 결과는 대화형으로 피벗할 수 있는 테이블로 나타낼 수 있습니다. 즉, 행, 열 및 레이어를 다시 배열할 수 있습니다.

② 피벗 표 조작

피벗 표 조작에는 다음과 같은 옵션이 있습니다.

- 행과 열 전치
- 행과 열 이동
- 다차원 레이어 작성
- 행과 열의 그룹 설정 및 해제
- 행, 열 및 기타 정보 표시 및 숨김
- 행과 열 레이블 회전
- 항에 대한 정의 찾기

가. 피벗 테이블 활성화

피벗 테이블을 조작하거나 수정하기 전에 테이블을 **활성화**해야 합니다. 테이블 활성화 방법

1. 테이블을 두 번 클릭하십시오.
   또는

2. 테이블을 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **내용 편집**을 선택하십시오.
3. 하위 메뉴에서 **뷰어에서** 또는 **별도의 창에서**를 선택하십시오.

나. 테이블 피벗

테이블에는 행, 열 및 레이어의 3차원이 있습니다. 차원은 여러 요소를 포함할 수 있습니다(또는 어떠한 요소도 포함하지 않을 수 있음). 차원 간에 또는 차원 내에서 요소를 이동하여 테이블의 구성을 변경할 수 있습니다. 요소를 이동하려면 끌어다 원하는 곳에 놓습니다.

다. 차원 내에서 요소의 표시 순서 변경

테이블 차원(행, 열 또는 레이어) 내에서 요소의 표시 순서 변경

1. 피벗 표를 활성화합니다.
2. 피벗 트레이가 설정되지 않은 경우에는 피벗 표 메뉴에서 다음을 선택합니다.
   피벗 > 피벗 트레이

3. 피벗 트레이의 차원 내에 요소를 끌어다 놓습니다.

라. 차원 요소 내에서 열 및 행 이동

1. 피벗 표를 활성화합니다.
2. 테이블 자체에서(피벗 트레이가 아님) 이동하려는 행이나 열의 레이블을 클릭하십시오.
3. 레이블을 새 위치에 끌어다 놓으십시오.

마. 행과 열 전치

1. 피벗 테이블을 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   피벗 > 행과 열 전치

열 차원으로 행 요소를 모두 끌 때와 행 요소로 열 요소를 모두 끌 때 같은 효과가 나타납니다.

바. 행 또는 열 그룹화

1. 피벗 표를 활성화합니다.
2. 그룹화할 행이나 열에 대해 레이블을 선택합니다. 마우스로 끌거나 Shift와 함께 마우스로 눌러 여러 레이블을 선택합니다.
3. 메뉴에서 다음을 선택하십시오.
   편집 > 그룹

그룹 레이블이 자동으로 삽입됩니다. 레이블 텍스트를 편집하려면 해당 그룹 레이블을 두 번 클릭합니다.

참고: 기존 그룹에 행이나 열을 추가하려면 먼저 현재 그룹에 있는 항목을 그룹 해제해야 합니다. 그런 다음 추가 항목을 포함하는 새 그룹을 작성할 수 있습니다.

사. 행이나 열 그룹 해제

그룹 해제를 사용하면 그룹 레이블이 자동으로 삭제됩니다.

아. 행이나 열 레이블 회전

테이블에서 가장 안쪽의 열 레이블과 가장 바깥쪽의 행 레이블에 대해 수평 및 수직 출력 간에 레이블을 회전할 수 있습니다.

1. 피벗 테이블을 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   형식 > 내부 열 레이블 회전

   또는

   형식 > 외부 행 레이블 회전

가장 안쪽의 열 레이블과 가장 바깥쪽의 행 레이블만 회전할 수 있습니다.

자. 행 정렬

피벗 표의 행을 정렬하려면 다음을 수행합니다.

1. 테이블을 활성화합니다.
2. 정렬을 사용하려는 열에서 셀을 선택합니다. 선택한 행 그룹만 정렬하려면 정렬을 사용하려는 열에서 두 개 이상의 연속 셀을 선택합니다.
3. 메뉴에서 다음을 선택하십시오.
   편집 > 행 정렬

4. 하위 메뉴에서 **오름차순** 또는 **내림차순**을 선택합니다.

- 행 차원에 그룹이 포함된 경우, 정렬은 선택을 포함하는 그룹에만 영향을 미칩니다.
- 그룹 경계를 걸쳐 정렬할 수 없습니다.
- 행 차원에 항목이 둘 이상인 테이블은 정렬할 수 없습니다.

차. 행 및 열 삽입

피벗 표에 행 또는 열을 삽입하려면 다음을 수행하십시오.

1. 테이블을 활성화합니다.
2. 테이블에서 셀을 선택합니다.
3. 메뉴에서 다음을 선택하십시오.

앞에 삽입 또는 **뒤에 삽입**

하위 메뉴에서 다음을 선택하십시오.

행 또는 **열**

- 셀이 비어있기 때문에 자동으로 숨겨지는 것을 방지하기 위해 새 행이나 열의 각 셀에 더하기 부호(+)를 삽입합니다.
- 중첩되거나 레이어 차원인 테이블에서는 모든 해당하는 차원 수준에 열이나 행이 삽입됩니다.

카. 변수 및 변수값 레이블 표시 제어

변수에 구체적인 변수 또는 변수값 레이블이 들어 있을 경우, 변수 이름과 레이블 및 피벗 표의 데이터 값과 변수값 레이블 표시를 제어할 수 있습니다.

1. 피벗 표를 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   보기 > 변수 레이블

   또는

   보기 > 값 레이블

3. 하위 메뉴에서 다음 옵션 중 하나를 선택합니다.

- **이름** 또는 **값**. 변수 이름(또는 값)만이 표시됩니다. 설명 레이블은 표시되지 않습니다.
- **레이블**. 설명 레이블만 표시됩니다. 변수 이름(또는 값)은 표시되지 않습니다.
- **모두**. 이름(또는 값) 및 설명 레이블 둘 모두가 표시됩니다.

타. 출력 언어 변경

피벗 표의 출력 언어를 변경하려면 다음을 수행하십시오.

1. 테이블을 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   보기 > 언어

3. 사용 가능한 언어 중 하나를 선택합니다.

언어를 변경하면 테이블 제목, 행과 열 레이블 및 꼬리말 텍스트 등과 같이 애플리케이션이 생성한 텍스트에만 영향을 미칩니다. 변수 이름과 설명 변수 및 값 레이블은 영향을 받지 않습니다.

파. 큰 테이블 탐색

탐색 창을 사용하여 큰 테이블을 탐색하려면 다음을 수행하십시오.

1. 테이블을 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   보기 > 탐색

하. 변경 실행 취소

활성화된 피벗 표에 작성한 마지막 변경 또는 모든 변경을 실행 취소할 수 있습니다. 두 동작은 모두 테이블의 가장 최근 활성화 이후에 작성된 변경에만 적용됩니다.

마지막 변경 실행 취소:

1. 메뉴에서 다음을 선택하십시오.
   편집 > 실행 취소

   모든 변경 실행 취소:

2. 메뉴에서 다음을 선택하십시오.
   편집 > 복원

③ 레이어 작업

각 범주나 범주 조합에 대해 별도의 2차원 테이블을 표시할 수 있습니다. 이러한 테이블은 레이어에서 누적된 것으로 간주되며 맨 위 레이어만 볼 수 있습니다.

가. 레이어 작성 및 표시

레이어 작성

1. 피벗 트레이가 설정되지 않은 경우에는 **피벗 표** 메뉴에서 다음을 선택합니다.
   피벗 > 피벗 트레이

2. 행이나 열 차원의 요소를 레이어 차원으로 끕니다.

요소를 레이어 차원으로 이동하면 다차원 테이블이 작성되지만 단 하나의 2차원 "조각"만이 표시됩니다. 표시되는 테이블은 위쪽 레이어에 대한 테이블입니다. 예를 들어, 예/아니오 범주형 변수가 레이어 차원에 있는 경우에는 다차원 테이블에는 2개의 레이어가 있습니다. 하나는 "예" 범주이고 하나는 "아니오" 범주입니다.

표시된 레이어 변경

1. 레이어의 드롭 다운 목록에서 범주를 선택합니다(피벗 트레이가 아닌 피벗 표 자체에 있음).
   또는
2. 메뉴에서 다음을 선택하십시오.

피벗 > 레이어로 이동...

나. 레이어 범주로 이동

레이어 범주로 이동을 사용하여 피벗 테이블에서 레이어를 바꿀 수 있습니다. 이 대화 상자는 많은 레이어가 있거나 선택한 레이어에 많은 범주가 있을 때 특히 유용합니다.

④ 항목 표시 및 숨김

다음을 비롯한 많은 셀 유형을 숨길 수 있습니다.

- 차원 레이블
- 행이나 열에서 레이블 셀과 데이터 셀을 포함하는 범주
- 범주 레이블(데이터 셀을 숨김없이)
- 꼬리말, 제목, 캡션

가. 테이블에서 행과 열을 숨김

나. 테이블에서 숨겨진 행과 열을 표시

1. 피벗 표를 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   보기 > 모든 범주 표시

테이블에 숨겨진 모든 행과 열을 표시합니다. (현재 테이블의 테이블 특성에서 **비어있는 행과 열 숨김**을 선택하는 경우 빈 행 또는 빈 열 전체가 숨겨집니다.)

다. 차원 레이블 숨김 및 표시

1. 피벗 표를 활성화합니다.
2. 차원 레이블이나 차원 내 범주 레이블을 선택합니다.
3. 보기 메뉴나 팝업 메뉴에서 **차원 레이블 숨김** 또는 **차원 레이블 표시**를 선택합니다.

라. 테이블 제목 숨김 및 표시

제목 숨김

1. 피벗 테이블을 활성화합니다.
2. 제목을 선택합니다.
3. 보기 메뉴에서 **숨김**을 선택합니다.

   숨겨진 제목 표시:

4. 보기 메뉴에서 **모두 표시**를 선택합니다.

⑤ 테이블모양

테이블모양은 테이블의 외형을 정의하는 일련의 특성입니다. 이전에 정의된 테이블모양을 선택하거나 사용자만의 테이블모양을 작성할 수 있습니다.

- 테이블모양 적용 전과 후에, 셀 특성을 사용하여 개별 셀이나 셀 그룹의 셀 형식을 변경할 수 있습니다. 편집된 셀 형식은 새 테이블모양을 적용할 때에도 온전한 상태로 남아 있습니다.
- 또는 모든 셀을 현재 테이블모양으로 정의되는 셀 형식으로 재설정할 수 있습니다. 이 옵션은 편집된 모든 셀을 재설정합니다. **표시된 대로**가 테이블모양 파일 목록에서 선택된 경우 편집된 셀은 현재 테이블 특성으로 재설정됩니다.
- 테이블 특성 대화 상자에 정의된 테이블 특성만 테이블모양에 저장됩니다. 테이블모양은 개별 셀 수정을 포함하지 않습니다.

가. 테이블모양 적용

1. 피벗 표를 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블모양...
3. 파일 목록에서 테이블모양을 선택합니다. 다른 디렉토리에서 파일을 선택하려면 **찾아보기**를 클릭하십시오.
4. **확인**을 클릭하여 선택한 피벗 표에 테이블모양을 적용합니다.

테이블모양을 편집하여 새 테이블모양을 작성할 수도 있습니다.

나. 테이블모양 편집 또는 작성

1. 피벗 표를 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블모양...

1. 테이블모양 대화 상자의 파일 목록에서 테이블모양을 선택합니다.
2. **모양 편집**을 클릭합니다.
3. 원하는 속성에 대한 테이블 특성을 조정하고 **확인**을 클릭합니다.
4. **모양 저장**을 클릭하여 편집한 테이블모양을 저장하거나 **다른 이름으로 저장**을 클릭하여 새 테이블모양으로 저장합니다.

- 테이블모양을 편집하면 선택한 피벗 표만 영향을 받습니다. 편집된 테이블모양은 테이블을 선택하고 테이블모양을 다시 적용하지 않는 한 해당 테이블모양을 사용하는 다른 테이블에 적용되지 않습니다. 모든 새 테이블에 적용된 기본 테이블모양 설정에 대한 자세한 정보는 새 피벗 테이블의 테이블모양 설정의 내용을 참조하십시오.
- 테이블 특성 대화 상자에 정의된 테이블 특성만 테이블모양에 저장됩니다. 테이블모양은 개별 셀 수정을 포함하지 않습니다.

⑥ 테이블 특성

테이블 특성을 사용하면 테이블의 일반 특성을 설정하고, 및 테이블의 여러 부분에 대한 셀 유형을 설정하고 할 수 있습니다. 다음을 수행할 수 있습니다.

- 빈 행이나 열 숨김과 인쇄 특성 조정 같은 일반 특성을 제어합니다.
- 꼬리말 표식 형식과 위치를 제어합니다.
- 행과 열 레이블 및 테이블의 다른 영역에 대해 데이터 영역에서 셀의 특정 형식을 결정합니다.
- 테이블의 각 영역 경계를 형성하는 선의 너비와 색상을 제어합니다.

가. 피벗 표 특성 변경

1. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블 특성...

2. 탭(**일반**, **꼬리말**, **셀 형식**, **경계선** 또는 **인쇄**)을 선택하십시오.
3. 원하는 옵션을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

새 특성은 선택한 피벗 표에 적용됩니다.

나. 테이블 특성: 일반

여러 특성이 테이블에 전체적으로 적용됩니다. 다음을 수행할 수 있습니다.

- 빈 행과 열을 표시하거나 숨깁니다. (빈 행이나 열에는 데이터 셀에 아무 것도 없습니다.)
- 왼쪽 위 코너에 배치하거나 중첩할 수 있는 행 레이블의 위치를 제어합니다.
- 열 너비 최대값과 최소값을 제어합니다. 포인트로 표시됩니다.

일반 테이블 특성 변경

1. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블 특성...

2. **일반** 탭을 클릭하십시오.
3. 원하는 옵션을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

새 특성은 선택한 피벗 표에 적용됩니다.

ㄱ. 표시할 행 수 설정

ⓘ **참고:** 이 기능은 레거시 테이블에만 적용됩니다.

기본적으로, 행 수가 많은 테이블은 100개 행 섹션으로 표시됩니다. 테이블에 표시되는 행 수를 제어하려면 다음을 수행합니다.

1. 테이블을 활성화합니다.
2. 메뉴에서 다음을 선택하십시오.
   테이블 특성 > 일반

3. **행별 테이블 표시**를 선택하십시오.
4. **표시할 행 수 설정**을 클릭하십시오.
   또는

5. 활성화된 피벗 표의 보기 메뉴에서 **행별 테이블 표시** 및 **표시할 행 수 설정**을 선택합니다.

표시할 행 수. 한 번에 표시되는 최대 행 수를 제어합니다. 탐색 제어를 사용하여 테이블의 여러 섹션으로 이동할 수 있습니다. 최소 값은 10입니다. 기본값은 100입니다.

과부/고아 줄 공차. 테이블에서 가장 안쪽에 있는 행 차원의 최대 행 수를 제어하여 표시되는 테이블 보기 전체에 분할합니다. 예를 들어, 가장 안쪽 행 차원의 각 그룹에 6개의 범주가 있는 경우 값을 6으로 지정하면 표시된 보기에서 그룹이 분할되지 않습니다. 이렇게 설정하면 표시된 보기의 총 행 수가 표시되도록 지정된 최대 행 수를 초과합니다.

다. 테이블 특성: 참고

테이블 특성 대화 상자의 참고 탭은 꼬리말 형식과 테이블 주석 텍스트를 제어합니다.

꼬리말. 꼬리말 표식 특성에는 텍스트와 관련된 유형과 위치가 포함됩니다.

- 꼬리말 표식의 유형은 숫자(1, 2, 3, ...) 또는 문자(a, b, c, ...)입니다.
- 꼬리말 표식은 위첨자나 아래첨자로 텍스트에 첨부됩니다.

주석 텍스트. 각 테이블에 주석 텍스트를 추가할 수 있습니다.

- 뷰어에서 마우스를 테이블 위에 올리면 주석 텍스트가 도구 팁에 표시됩니다.
- 테이블에 초점이 맞으면 스크린 리더가 주석 텍스트를 읽습니다.
- 뷰어의 도구 팁은 주석의 처음 200문자만을 표시하지만 스크린 리더는 전체 텍스트를 읽습니다.
- 출력을 HTML로 내보낼 때 주석 텍스트는 alt 텍스트로 사용합니다.

새 특성은 선택한 피벗 테이블에 적용됩니다. 선택한 테이블 대신 새로운 테이블 특성을 테이블 모양에 적용하려면 테이블모양을 편집합니다(형식 메뉴, 테이블모양).

라. 테이블 특성: 셀 형식

형식의 경우 테이블은 제목, 레이어, 코너 레이블, 행 레이블, 열 레이블, 데이터, 캡션 및 꼬리말의 영역으로 구분됩니다. 각 테이블 영역에서는 관련된 셀 형식을 수정할 수 있습니다. 셀 형식에는 텍스트 특성(글꼴, 크기, 색상, 유형 등), 수평 및 수직맞춤, 배경색 및 내부 셀 여백 등이 포함됩니다.

셀 형식은 영역(정보의 범주)에 적용됩니다. 이들은 개별 셀의 특성이 아닙니다. 이러한 차이는 테이블을 피벗할 때 중요한 사항이 됩니다.

예를 들어,

- 열 레이블의 셀 형식으로 굵은 글꼴을 지정하는 경우 열 레이블은 열 차원에 현재 표시된 정보에 관계 없이 굵게 나타납니다. 열 차원에서 다른 차원으로 항목을 이동하는 경우 열 레이블의 굵은 특성을 유지하지 않습니다.

- 현재 피벗 표의 셀을 선택하고 도구 모음에서 굵게 단추를 눌러 열 레이블을 굵게 표시하면 셀 컨텐츠는 이동하는 차원에 관계 없이 계속 굵게 표시되지만, 열 차원으로 옮긴 다른 항목에 대해서는 열 레이블이 굵게 유지되지 않습니다.

셀 형식 변경

1. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블 특성...

2. **셀 형식** 탭을 선택합니다.
3. 드롭 다운 목록에서 영역을 선택하거나 보기 영역을 클릭하십시오.
4. 영역 특성을 선택합니다. 선택은 보기에 반영됩니다.
5. **확인** 또는 **적용**을 클릭하십시오.

새 특성은 선택한 피벗 표에 적용됩니다.

행 색상 변경

테이블의 데이터 영역에 있는 행의 배경 및/또는 텍스트 색상을 변경하려면 다음을 수행합니다.

1. 영역 드롭 다운 목록에서 **데이터**를 선택합니다.
2. 배경색 그룹에서 **행 색상 변경**을 선택합니다.
3. 행의 배경 및 텍스트를 변경하는 데 사용할 색상을 선택합니다.

행 색상 변경은 테이블의 데이터 영역에만 적용됩니다. 행 또는 열 레이블 영역에는 적용되지 않습니다.

마. 테이블 특성: 경계선

테이블의 각 경계선 위치에서 선 유형과 색상을 선택할 수 있습니다. 유형을 **지정않음**으로 선택하면 선택한 위치에서 선이 보이지 않게 됩니다.

테이블 경계선 변경

1. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블 특성...

2. **경계선** 탭을 클릭하십시오.
3. 목록에서 해당 이름을 누르거나 보기 영역에서 선을 눌러 경계 위치를 선택하십시오.
4. 선 유형을 선택하거나 **지정않음**을 선택하십시오.

5. 색상을 선택하십시오.
6. **확인** 또는 **적용**을 클릭하십시오.

새 특성은 선택한 피벗 표에 적용됩니다.

바. 테이블 특성: 인쇄

인쇄할 피벗 표에서 다음 특성을 제어할 수 있습니다.

- 모든 레이어나 테이블의 위쪽 레이어만 인쇄하고 각 레이어를 별도의 페이지에 인쇄합니다.
- 테이블을 인쇄할 수 있도록 가로 또는 세로로 크기를 줄여 용지 크기에 맞춥니다.
- 정의된 페이지 크기에 비해 테이블이 너무 넓거나 너무 길면 테이블의 인쇄 영역에 포함할 행과 열의 최소값을 제어하여 과부/고아 줄을 조정합니다.
  참고: 위에 다른 출력이 있으므로 테이블이 너무 길어서 현재 페이지에 맞출 수 없지만 정의된 페이지 길이에는 맞는 경우에는 테이블은 과부/고아 설정과 관계없이 새 페이지에 자동으로 인쇄됩니다.
- 한 페이지에 들어가지 못하는 테이블에는 '계속' 표시를 합니다. 각 페이지의 위와 아래에 '계속' 표시를 할 수 있습니다. 어떤 옵션도 선택하지 않으면 '계속' 표시는 나타나지 않습니다.

피벗 표 인쇄 특성 제어

1. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블 특성...

2. **인쇄** 탭을 클릭하십시오.
3. 원하는 인쇄 옵션을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

새 특성은 선택한 피벗 표에 적용됩니다.

⑦ 셀 특성

셀 특성은 선택한 셀에 적용됩니다. 글꼴, 값 형식, 맞춤, 여백 및 색상을 변경할 수 있습니다. 셀 특성은 테이블 특성을 무시하므로 테이블 특성을 변경하는 경우 개별적으로 적용된 셀 특성은 변경되지 않습니다.

셀 특성 변경:

1. 테이블에서 셀을 선택하십시오.
2. 형식 메뉴 또는 팝업 메뉴에서 **셀 특성**을 선택하십시오.

가. 셀 특성: 글꼴 및 배경

글꼴 및 배경 탭은 테이블에서 선택한 셀에 대한 글꼴 유형과 색상 및 배경 색상을 제어합니다.

피벗 표에서 선택한 셀에 대한 글꼴 및 배경 설정 바꾸기

나. 셀 특성: 형식 값

형식 값 탭은 선택한 셀에 대한 값 형식을 제어합니다. 숫자, 날짜, 시간, 통화 등에 대한 형식을 선택하고 표시할 소수점 자릿수를 조정할 수 있습니다.

피벗 테이블에서 선택한 셀에 대한 값 형식 변경:

다. 셀 특성: 맞춤 및 여백

맞춤 및 여백 탭은 값의 수평과 수직 맞춤 및 선택한 셀에 대한 위, 아래, 왼쪽, 오른쪽 여백을 제어합니다. **혼합** 수평 맞춤은 유형에 따라 각 셀의 컨텐츠를 맞춥니다. 예를 들어, 날짜는 오른쪽으로 맞추고 텍스트 값은 왼쪽으로 맞춥니다.

피벗 표에서 선택한 셀에 대한 맞춤 또는 여백 변경:

⑧ 꼬리말 및 캡션

테이블에 꼬리말과 캡션을 추가할 수 있습니다. 꼬리말이나 캡션을 숨기고, 꼬리말 표식을 변경하고, 꼬리말 번호를 재지정할 수도 있습니다.

가. 꼬리말 및 캡션 추가

테이블에 캡션 추가

1. 삽입 메뉴에서 **캡션**을 선택합니다.

테이블의 항목에 꼬리말을 첨부할 수 있습니다. 꼬리말 추가

1. 활성화된 피벗 표 내에서 제목, 셀 또는 캡션을 클릭하십시오.
2. 삽입 메뉴에서 **꼬리말**을 선택하십시오.
3. 제공된 영역에 꼬리말 텍스트를 삽입하십시오.

나. 캡션 숨김 또는 표시

캡션 숨김

1. 피벗 테이블을 활성화합니다.
2. 캡션을 선택합니다.
3. 보기 메뉴에서 **숨김**을 선택합니다.

숨겨진 캡션 표시

1. 보기 메뉴에서 **모두 표시**를 선택합니다.

다. 테이블에서 꼬리말 숨김 또는 표시

꼬리말 숨김

1. 피벗 테이블을 활성화합니다.
2. 꼬리말 참조를 포함하는 셀을 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **꼬리말 숨김**을 선택하십시오.
   또는

3. 테이블의 꼬리말 영역에서 꼬리말을 선택하고 팝업 메뉴에서 **숨김**을 선택하십시오.
   참고: 레거시 테이블의 경우 테이블의 꼬리말 영역을 선택하고, 팝업 메뉴에서 **꼬리말 편집**을 선택한 다음 숨기려는 꼬리말의 표시 특성을 선택 취소(지우기)하십시오.

   셀에 여러 꼬리말이 포함된 경우에는 후자의 방법을 사용하여 꼬리말을 선택적으로 숨기십시오.

테이블에서 모든 꼬리말 숨김

1. 테이블의 꼬리말 영역에서 모든 꼬리말을 선택하고(클릭과 끌기 또는 Shift+클릭을 사용하여 꼬리말 선택) 보기 메뉴에서 **숨김**을 선택하십시오.

숨겨진 꼬리말 표시

1. 보기 메뉴에서 **모든 꼬리말 표시**를 선택합니다.

라. 꼬리말 표식

꼬리말 표식은 꼬리말을 표시할 때 사용할 수 있는 문자를 변경합니다. 기본적으로 표준 꼬리말 표식은 테이블 특성 설정에 따라 순차 문자 또는 숫자입니다. 특수 표식을 할당할 수도 있습니다. 특수 표식은 꼬리말 번호를 재지정하거나 표준 표식의 숫자와 문자를 전환할 경우 영향을 받지 않습니다. 표준 표식의 번호 또는 문자 표시와 꼬리말 표식의 아래첨자 또는 위첨자 위치는 테이블 특성 대화 상자의 꼬리말 탭으로 제어됩니다.

꼬리말 표식 변경

1. 꼬리말을 선택합니다.
2. **형식** 메뉴에서 **꼬리말 표식**을 선택하십시오.

특수 표식은 2문자로 제한됩니다. 특수 표식이 있는 꼬리말은 테이블의 꼬리말 영역에서 순차 문자나 숫자가 있는 꼬리말보다 앞섭니다. 따라서 특수 표식으로 변경하면 꼬리말 목록을 다시 정렬할 수 있습니다.

마. 꼬리말에 다시 번호 매기기

행, 열, 레이어를 전환하여 테이블을 피벗하면 꼬리말 순서가 달라질 수 있습니다. 꼬리말 번호를 재지정하려면 다음을 수행합니다.

1. 형식 메뉴에서 **꼬리말에 다시 번호 매기기**를 선택하십시오.

바. 레거시 테이블의 꼬리말 편집

레거시 테이블의 경우, 꼬리말 편집 대화 상자를 사용하여 꼬리말 텍스트와 글꼴 설정을 입력 및 수정하고 꼬리말 표식을 변경하며 선택적으로 꼬리말을 숨기거나 삭제할 수 있습니다.

레거시 테이블에 새 꼬리말을 삽입할 경우 꼬리말 편집 대화 상자가 자동으로 열립니다. 꼬리말 편집 대화 상자를 사용하여 기존 꼬리말 편집(새 꼬리말 작성하지 않음)

표식. 기본적으로 표준 꼬리말 표식은 테이블 특성 설정에 따라 순차 문자 또는 숫자입니다. 특수 표식을 지정하려면 표식 열에 새 표식 값을 입력합니다. 특수 표식은 꼬리말 번호를 재지정하거나 표준 표식의 숫자와 문자를 전환할 경우 영향을 받지 않습니다. 표준 표식의 번호 또는 문자 표시와 꼬리말 표식의 아래첨자 또는 위첨자 위치는 테이블 특성 대화 상자의 꼬리말 탭으로 제어됩니다. 자세한 정보는 테이블 특성: 참고의 내용을 참조하십시오.

특수 표식을 다시 표준 표식으로 변경하려면 꼬리말 편집 대화 상자에서 표식을 마우스 오른쪽 단추로 클릭하고, 팝업 메뉴에서 **꼬리말 표식**을 선택하고 꼬리말 표식 대화 상자에서 표준 표식을 선택하십시오.

꼬리말. 꼬리말의 컨텐츠입니다. 표시는 현재 글꼴 및 배경 설정을 반영합니다. 글꼴 설정은 형식 하위 대화 상자를 사용하여 개별 꼬리말에 대해 변경할 수 있습니다. 자세한 정보는 꼬리말 글꼴 및 색상 설정 주제를 참조하십시오. 하나의 배경 색상이 모든 꼬리말에 적용되고 셀 특성 대화 상자의 글꼴 및 배경 탭에서 변경할 수 있습니다. 자세한 정보는 셀 특성: 글꼴 및 배경 주제를 참조하십시오.

표시. 모든 꼬리말은 기본적으로 표시됩니다. 꼬리말을 숨기려면 표시 선택란을 선택 취소(지우기)하십시오.

ㄱ. 꼬리말 글꼴 및 색상 설정

레거시 테이블의 경우, 형식 대화 상자를 사용하여 하나 이상의 선택한 꼬리말에 대해 글꼴군, 유형, 크기 및 색상을 변경할 수 있습니다.

1. 테이블의 꼬리말 영역을 두 번 클릭하거나 메뉴에서 형식 > 꼬리말 편집을 선택합니다.
2. 꼬리말 편집 대화 상자에서 꼬리말 눈금에 있는 하나 이상의 꼬리말을 선택(클릭)합니다.
3. **형식** 단추를 클릭하십시오.

선택한 글꼴군, 유형, 크기 및 색상이 선택한 모든 꼬리말에 적용됩니다.

배경색, 맞춤 및 여백은 셀 속성 대화 상자에서 설정하여 모든 꼬리말에 적용할 수 있습니다. 이러한 설정은 개별 꼬리말에 대해 변경할 수 없습니다. 자세한 정보는 셀 특성: 글꼴 및 배경의 내용을 참조하십시오.

⑨ 데이터 셀 너비

데이터 셀 너비 설정은 모든 데이터 셀 너비를 동일하게 하는 데 사용합니다.

모든 데이터 셀의 너비 설정

1. 메뉴에서 다음을 선택하십시오.
   형식 > 데이터 셀 너비 설정...

2. 셀 너비 값을 입력합니다.

⑩ 열 너비 변경

1. 열 경계를 클릭하고 끕니다.

피벗 표에서 숨겨진 경계선을 보려면 보기 메뉴에서 **격자선**을 클릭하십시오.

⑪ 피벗 표에서 숨겨진 경계선 표시

경계선이 보이지 않는 테이블에서는 숨겨진 경계선을 표시할 수 있습니다. 이렇게 하면 열 너비 변경 같은 작업을 쉽게 수행할 수 있습니다.

1. 피벗 테이블을 활성화합니다.
2. 보기 메뉴에서 **격자선**을 선택하십시오.

⑫ 피벗 표에서 행, 열과 셀을 선택

전체 행이나 열 또는 지정된 데이터 세트 및 레이블 셀을 선택할 수 있습니다.

여러 셀 선택

선택 > 데이터 및 레이블 셀

⑬ 피벗 테이블 인쇄

여러 요소가 인쇄된 피벗 테이블이 표시되는 방식에 영향을 미칠 수 있으며 이러한 요소는 피벗 테이블 속성을 변경하여 제어할 수 있습니다.

- 다차원 피벗 테이블(레이어가 있는 테이블)의 경우 모든 레이어를 인쇄하거나 맨 위(표시된) 레이어만 인쇄할 수 있습니다. 자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.
- 길거나 넓은 피벗 테이블의 경우 페이지에 맞게 테이블을 자동으로 크기 조정하거나 페이지 구분과 테이블 구분 위치를 제어할 수 있습니다. 자세한 정보는 테이블 특성: 인쇄의 내용을 참조하십시오.

파일 메뉴의 인쇄 미리보기를 사용하여 인쇄된 피벗 테이블의 모습을 확인하십시오.

가. 넓거나 긴 테이블에 대한 테이블 구분 제어

정의된 페이지 크기 내에 인쇄하기에는 너무 넓거나 긴 피벗 표는 자동으로 분할되고 다중 섹션에 인쇄됩니다. 다음을 수행할 수 있습니다.

- 큰 테이블이 분할되는 행과 열 위치를 제어합니다.
- 테이블을 분할할 때 분할되지 않도록 해야 하는 행과 열을 지정합니다.
- 정의된 페이지 크기에 맞게 큰 테이블의 크기를 조정합니다.

인쇄된 피벗 표를 위해 행과 열 구분을 지정합니다.

1. 피벗 테이블을 활성화합니다.
2. 구분을 삽입하려는 위치 왼쪽의 열에서 셀을 클릭하거나 구분을 삽입하려는 행 앞의 행의 셀을 클릭하십시오.
3. 메뉴에서 다음을 선택하십시오.
   형식 > 중단점 > 수직 중단점

   또는

   형식 > 중단점 > 수평 중단점

1. 피벗 테이블을 활성화합니다.
2. 구분을 삽입하려는 위치 왼쪽의 열에서 셀을 클릭하거나 구분을 삽입하려는 행 앞의 행의 셀을 클릭하십시오.
3. 메뉴에서 다음을 선택하십시오.
   형식 > 중단점 > 수직 중단점

   또는

   형식 > 중단점 > 수평 중단점

연결할 행이나 열을 지정:

1. 연결 할 행 레이블이나 열 레이블을 선택합니다. 여러 행이나 열 레이블을 선택하려면 클릭한 후 끌거나 Shift+클릭하십시오.
2. 메뉴에서 다음을 선택하십시오.
   형식 > 중단점 > 연결

중단점을 보고 그룹을 연결:

1. 메뉴에서 다음을 선택하십시오.

형식 > 중단점 > 중단점 표시

중단점은 수직선 또는 수평선으로 나타납니다. 그룹 연결은 어두운 경계선으로 둘러싸인 회색 직사각형 영역으로 나타납니다.

참고: 중단점 표시와 그룹 연결은 레거시 테이블에서는 지원되지 않습니다.

중단점 및 연결 그룹 지우기

중단점 지우기

1. 수직 중단점 왼쪽 열의 셀을 클릭하거나 수평 중단점 위 행의 셀을 클릭하십시오.
2. 메뉴에서 다음을 선택하십시오.
   형식 > 중단점 > 중단점 또는 그룹 지우기

   연결 그룹 지우기

3. 그룹을 지정하는 열 레이블 또는 행 레이블을 선택합니다.
4. 메뉴에서 다음을 선택하십시오.
   형식 > 중단점 > 중단점 또는 그룹 지우기

모든 중단점 및 연결 그룹은 행이나 열을 피벗하거나 다시 정렬할 때 자동으로 지워집니다. 이 동작은 레거시 테이블에 적용되지 않습니다.

⑭ 피벗 표에서 차트 작성

1. 피벗 표를 두 번 클릭하여 이를 활성화하십시오.
2. 차트에 표시할 행, 열 또는 셀을 선택하십시오.
3. 선택한 영역의 임의의 위치에 마우스 오른쪽 단추로 클릭하십시오.
4. 팝업 메뉴에서 **그래프 작성**을 선택하고 차트 유형을 선택하십시오.

⑮ 레거시 테이블

테이블을 레거시 테이블(릴리스 19에서 완전히 기능하는 테이블로 알려짐)로서 렌더링하기로 선택할 수 있고 그러면 이는 20 이전의 IBM® SPSS® Statistics 릴리스와 완전히 호환 가능합니다. 레거시 테이블은 느리게 렌더링할 수 있으며 20 이전의 릴리스와 호환성이 필요한 경우에만 권장됩니다. 레거시 테이블을 작성하는 방법에 대한 자세한 정보는 피벗 테이블 옵션의 내용을 참조하십시오.

(8) 옵션

① 일반 옵션

최대 스레드 수

결과를 계산할 때 멀티스레드 프로시저가 사용할 스레드 수입니다. **자동** 설정은 사용 가능한 처리 코어 수를 기준으로 합니다. 멀티스레드 프로시저가 실행되는 동안 다른 애플리케이션에서 더 많은 처리 자원을 사용할 수 있도록 하려면 낮은 값을 지정하십시오. 분산 분석 모드에서는 이 옵션을 사용할 수 없습니다.

출력

소수점 값에 선행 0 표시 - 소수점 부분으로만 구성된 숫자 값에 선행 0을 표시합니다. 예를 들어 선행 0을 표시할 경우 값 .123은 0.123으로 표시됩니다. 통화 또는 퍼센트 형식의 숫자 값에는 이 설정이 적용되지 않습니다. 데이터를 외부 파일에 저장할 때 선행 0은 포함되지 않습니다. 단, 고정 ASCII 파일(*.dat)은 예외입니다.

측정 시스템. 피벗 표 인쇄 작업에서 셀 여백, 셀 너비, 테이블 간 간격 등의 속성을 지정하는데 사용하는 측정 시스템(포인트, 인치, 센티미터)입니다.

가. 일반 옵션 설정

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션...

2. **일반** 탭을 클릭하십시오.
3. 원하는 설정을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

② 뷰어 옵션

뷰어 출력 표시 옵션은 설정을 변경한 후에 생성되는 새 출력에만 영향을 미칩니다. 뷰어에 이미 표시된 출력은 이러한 설정의 변경에 영향을 받지 않습니다.

초기 출력결과 상태

프로시저를 실행할 때마다 어떤 항목이 자동으로 표시되거나 숨겨지는지와 항목이 처음에 정렬되는 방법을 제어합니다. 다음 항목의 표시를 제어할 수 있습니다. 로그, 경고, 참고, 제목, 피벗 테이블, 차트, 트리 다이어그램 및 텍스트 출력입니다. 로그에서 명령의 표시를 켜거나 끌 수도 있습니다. 로그에서 명령 구문을 복사하고 이를 구문 파일에 저장할 수 있습니다.

> **참고**: 모든 결과 항목은 뷰어에서 왼쪽 맞춤으로 표시됩니다. 인쇄된 출력의 정렬만 맞춤 설정으로 영향을 받습니다. 가운데 맞춤과 오른쪽 맞춤 항목은 작은 기호로 식별됩니다.

제목

새 결과 제목의 글꼴 유형, 크기, 색상 등을 제어합니다. 글꼴 **크기** 목록은 일련의 사전 정의된 크기를 제공하지만, 지원되는 다른 크기 값을 수동으로 입력할 수 있습니다.

페이지 제목

새 페이지 제목 및 TITLE 및 SUBTITLE 명령 구문을 통해 생성되거나 **삽입** 메뉴의 **새 페이지 제목**을 통해 작성된 페이지 제목의 글꼴 유형, 크기 및 색상을 제어합니다. 글꼴 **크기** 목록은 일련의 사전 정의된 크기를 제공하지만, 지원되는 다른 크기 값을 수동으로 입력할 수 있습니다.

텍스트 출력

텍스트 출력에 사용되는 글꼴입니다. 텍스트 출력에는 단일자간(고정 피치) 글꼴을 사용하는 것이 적합합니다. 비례 글꼴을 선택하면 테이블 출력은 제대로 정렬되지 않습니다. 글꼴 **크기** 목록은 일련의 사전 정의된 크기를 제공하지만, 지원되는 다른 크기 값을 수동으로 입력할 수 있습니다.

기본 페이지 설정

인쇄의 출력방향과 여백을 위한 기본 옵션을 제어합니다.

가. 뷰어 옵션 설정

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션...

2. **뷰어** 탭을 클릭하십시오.
3. 원하는 설정을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

뷰어 출력 표시 옵션은 설정을 변경한 후에 생성되는 새 출력에만 영향을 미칩니다. 뷰어에 이미 표시된 출력은 이러한 설정의 변경에 영향을 받지 않습니다.

나. 새 출력의 초기 표시 상태 제어

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션…
2. 옵션 대화 상자에서 **뷰어** 탭을 클릭하십시오.
3. 초기 출력 상태 그룹에서 새 출력이 생성될 때마다 표시하거나 숨길 항목 유형(예: 피벗 테이블, 제목, 참고, 로그)의 아이콘을 선택하십시오.
4. **표시됨** 또는 **숨겨짐**을 선택하여 해당 항목의 초기 출력 상태를 지정하십시오.

다. 새 결과의 초기 맞추기 변경

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션…
2. **뷰어** 탭을 선택하고 초기 출력 상태 그룹에서 맞추기(자리 맞춤)를 선택하십시오.

③ 피벗 테이블 옵션

피벗 테이블 옵션은 피벗 테이블의 표시에 대해 다양한 옵션을 설정합니다.

테이블모양

파일 목록에서 테이블모양을 선택하고 **확인** 또는 **적용**을 클릭하십시오. IBM® SPSS® Statistics에서 제공하는 테이블모양 중 하나를 사용하거나, 형식 메뉴의 **테이블모양**을 선택하여 피벗 테이블 편집기에서 테이블모양을 직접 작성할 수 있습니다.

- **찾아보기**. 다른 디렉토리에서 테이블모양을 선택할 수 있습니다.
- **테이블모양 디렉토리 설정**. 기본값 테이블모양 디렉토리를 변경할 수 있습니다. 찾아보기를 사용하여 사용할 디렉토리를 탐색하고 해당 디렉토리에서 테이블모양을 선택한 다음 **테이블모양 디렉토리 설정**을 선택합니다.

참고: IBM SPSS Statistics의 이전 버전에서 작성된 테이블모양은 버전 16.0 이상에서는 사용할 수 없습니다.

열 너비

이러한 옵션은 피벗 테이블에서 열 너비의 자동 조절을 제어합니다.

- **레이블만 수정**. 열 너비를 열 레이블 너비에 맞춥니다. 이 프로시저를 사용하면 간결한 테이블이 생성되지만 레이블보다 넓은 데이터 값은 잘릴 수 있습니다.
- **모든 테이블에 대해 레이블과 데이터 수정**. 열 너비를 열 레이블 또는 가장 큰 데이터 값 중 더 큰 값으로 조정합니다. 이 프로시저를 사용하면 테이블 너비가 더 넓지만 모든 값이 빠짐없이 표시됩니다.

기본 편집 모드

이 옵션은 뷰어 창이나 별도의 창에서 피벗 테이블의 활성화를 제어합니다. 기본적으로 피벗 테이블을 두 번 클릭하면 매우 큰 테이블을 제외하고 뷰어 창에서 테이블이 활성화됩니다. 개별 창에서 피벗 테이블을 활성화할 것을 선택하거나 뷰어 창에서 피벗 테이블을 작게 열거나 개별 창에서 피벗 테이블을 크게 열도록 하는 크기 설정을 선택할 수 있습니다.

넓은 테이블을 서식있는 텍스트 형식으로 클립보드에 복사

피벗 테이블을 Word/RTF 형식으로 붙여넣을 경우 문서 너비보다 큰 테이블은 문서 너비에 맞게 줄바꿈 또는 축소되거나 변경되지 않고 그대로 있게 됩니다.

가. 새 피벗 테이블의 테이블모양 설정

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션...

2. **피벗 테이블** 탭을 클릭하십시오.
3. 원하는 테이블모양을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

> ⓘ **참고**: IBM® SPSS® Statistics의 이전 버전에서 작성된 테이블모양은 버전 16.0 이상에서는 사용할 수 없습니다.

나. 새 피벗 표의 열 너비 제어

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션...

2. **피벗 테이블** 탭을 클릭하십시오.
3. 열 너비에 대해 원하는 옵션을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

- **매우 큰 테이블을 제외하고 레이블과 데이터 조정**. 셀이 1만개를 초과하지 않는 테이블의 경우 열 레이블이나 가장 큰 데이터 값 중 더 큰 값으로 열 너비를 조정하십시오. 셀이 10,000개 이상인 테이블의 경우 열 너비를 열 레이블의 너비로 조정합니다.
- **레이블만 수정**. 열 너비를 열 레이블 너비에 맞춥니다. 이 프로시저를 사용하면 간결한 테이블이 생성되지만 레이블보다 넓은 데이터 값은 잘릴 수 있습니다.
- **레이블 및 데이터에 맞게 조정**. 열 너비를 열 레이블 또는 가장 큰 데이터 값 중 더 큰 값으로 조정합니다. 이 프로시저를 사용하면 테이블 너비가 더 넓지만 모든 값이 빠짐 없이 표시됩니다.

다. 사용자 정의 기본 테이블모양 작성

1. 피벗 테이블을 활성화하십시오(테이블의 임의 위치에서 두 번 클릭).
2. 메뉴에서 다음을 선택하십시오.
   형식 > 테이블모양…
3. 목록에서 테이블모양을 선택하고 **모양 편집**을 클릭하십시오.
4. 원하는 속성에 대한 테이블 특성을 조정하십시오.
5. **모양 저장** 또는 **다른 이름으로 저장**을 클릭하여 테이블모양을 저장하고 **확인**을 클릭하십시오.
6. 피벗 테이블을 선택 취소하십시오.
7. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션…

8. **피벗 테이블** 탭을 클릭하십시오.
9. 목록에서 테이블모양을 선택하고 **확인**을 클릭하십시오.

참고: IBM® SPSS® Statistics의 이전 버전에서 작성된 테이블모양은 버전 16.0 이상에서는 사용할 수 없습니다.

④ 출력 옵션

출력 옵션은 출력 옵션 수의 기본 설정을 제어합니다.

스크린 리더 내게 필요한 옵션. 스크린 리더가 피벗 테이블과 열 레이블을 읽는 방법을 제어합니다. 각 데이터 셀의 전체 행과 열 레이블을 읽거나 테이블에서 데이터 셀 간에 이동할 때 변경하는 레이블만 읽을 수 있습니다.

가. 출력 옵션 설정

1. 메뉴에서 다음을 선택하십시오.
   편집 > 옵션...

2. **출력** 탭을 클릭하십시오.
3. 원하는 설정을 선택하십시오.
4. **확인** 또는 **적용**을 클릭하십시오.

뷰어 출력 표시 옵션은 설정을 변경한 후에 생성되는 새 출력에만 영향을 미칩니다. 뷰어에 이미 표시된 출력은 이러한 설정의 변경에 영향을 받지 않습니다.

7) 결측값 처리

(1) 결측값의 개요

데이터 마이닝의 데이터 준비 단계 동안 종종 데이터에서 **결측값**을 대체하려고 할 수 있습니다. 결측값은 알려지지 않고, 수집되지 않거나 잘못 입력된 데이터 세트의 값입니다. 일반적으로 이러한 값은 필드에 유효하지 않습니다. 예를 들어, *성별* 필드에는 *M* 및 *F* 값이 포함되어야 합니다. 필드에 *Y* 또는 *Z* 값이 있다면 이를 유효하지 않은 값으로 가정하여 공백으로 해석해야 합니다. 마찬가지로 *연령* 필드의 음수값은 의미가 없으므로 공백으로 해석되어야 합니다. 때때로 무응답을 보여주기 위해 질문지에 이렇게 명백하게 잘못된 값을 고의로 입력하거나 필드를 공백으로 두기도 합니다. 때로는 이러한 공백을 자세히 조사하여 누군가가 나이 제공을 거부하는 등 무응답한 것이 특정 결과를 예측하는 데 있어서 요인인지 여부를 판별하려고 할 수 있습니다.

몇몇 모델링 기술은 다른 기술보다 결측 데이터를 보다 잘 처리합니다. 예를 들어, C5.0 및 Apriori는 유형 노드에서 "결측"으로 명시적으로 선언된 값에 잘 대처합니다. 다른 모델링 기법은 결측값을 다루는 데 문제가 있고 학습 시간이 길고 정확성이 떨어지는 모델을 생성합니다.

IBM® SPSS® Modeler에 의해 인지되는 결측값에는 몇 가지 유형이 있습니다.

- **널 또는 시스템 결측값**. 이들은 데이터베이스나 소스 파일에 공백으로 남겨졌고 소스 또는 유형 노드에서 "결측"으로 정의되지 않은 문자열이 아닌 값입니다. 시스템 결측값은 **$null$**로 표시됩니다. 빈 문자열은 특정 데이터베이스에 의해 널로 처리되더라도 IBM SPSS Modeler에서는 널로 간주되지 않음을 유의하십시오.
- **빈 문자열 및 공백**. 빈 문자열 값과 공백(눈에 보이는 문자가 없는 문자열)은 널값과는 별개로 처리됩니다. 빈 문자열은 대부분의 경우에서 공백과 동일하게 처리됩니다. 예를 들어, 소스나 유형 노드에서 공백을 공란으로 처리하는 옵션을 선택한 경우 이 설정은 빈 문자열에도 적용됩니다.
- **공백 또는 사용자 정의 결측값**. 이들은 소스 노드 또는 유형 노드에서 결측으로 명백하게 정의되어 있는 unknown, 99 또는 -1 등과 같은 값입니다. 또는 널과 공백을 공란으로 처리하기로 선택할 수도 있는데 그러면 이들은 특수 처리용으로 플래그가 지정되고 대부분의 계산에서 제외됩니다. 예를 들어, @BLANK 함수를 사용하여 이들 값 및 다른 유형의 결측값을 공란으로 처리할 수 있습니다.

혼합 데이터 읽기. 숫자 저장 공간(정수, 실수, 시간, 시간소인 또는 날짜)이 있는 필드를 읽을 때 숫자가 아닌 값은 널 또는 시스템 결측값으로 설정됨을 유의하십시오. 이는 몇몇 애플리케이션과는 달리 필드 내에서 혼합된 저장 유형을 허용하지 않기 때문입니다. 이를 피하려면 혼합 데이터가 있는 필드는 소스 노드나 외부 애플리케이션에서 필요에 따라 저장 유형을 변경하여 문자열로 읽어야 합니다.

Oracle에서 빈 문자열 읽기. Oracle 데이터베이스에서 읽거나 Oracle 데이터베이스에 쓸 때 IBM SPSS Modeler 및 대부분의 다른 데이터베이스와 달리 Oracle은 빈 문자열 값을 널값과 같다고 간주하고 저장합니다. 이는 Oracle 데이터베이스에서 추출된 동일한 데이터가 파일 또는 다른 데이터베이스에서 추출된 경우와 다르게 동작하고 데이터가 다른 결과를 리턴할 수 있음을 의미합니다.

(2) 결측값 처리

사용자의 비즈니스나 도메인 지식에 비추어 결측값을 처리하는 방법을 결정해야 합니다. 학습 시간을 줄이고 정확도를 높이기 위해 데이터 세트에서 공백을 제거하려고 할 수 있습니다. 반면, 공백값이 있으면 새로운 비즈니스 기회나 추가적인 통찰로 이어질 수도 있습니다. 최상의 기법을 선택하려면 데이터의 다음 측면을 고려해야 합니다.

- 데이터 세트의 크기
- 공백을 포함하는 필드 수
- 결측 정보의 양

일반적으로 두 가지의 방법을 따를 수 있습니다.

- 결측값이 있는 필드나 레코드를 제외할 수 있습니다.
- 다양한 방법을 사용하여 결측값을 대치, 바꾸거나 강요할 수 있습니다.

이러한 두 방법은 모두 데이터 검토 노드를 사용하여 대부분 자동화할 수 있습니다. 예를 들어, 결측값이 너무 많은 필드가 모델링에서 유용하게 되도록 이를 제외하는 필터 노드를 생성하거나 남아 있는 필드의 일부 또는 모두의 결측값을 대치하는 수퍼노드를 생성할 수 있습니다. 이는 감사의 실제 힘이며 이를 통해, 데이터의 현재 상태를 평가할 수 있을 뿐만 아니라 평가를 기반으로 조치를 취할 수도 있습니다.

① 결측값이 있는 레코드 처리

대다수의 결측값이 소수의 레코드에 집중되면 이러한 레코드를 제외하기만 하면 됩니다. 예를 들어, 은행은 일반적으로 대출 고객에 대해서 자세하고 완전한 레코드를 유지합니다. 그러나 은행이 자체 직원에게 대출을 승인할 때는 덜 제한적이라면 직원 대출에 대해 수집한 데이터에는 몇

몇 공백 필드가 있을 수 있습니다. 이런 경우에는 이러한 결측값을 다루기 위한 두 가지 옵션이 있습니다.

- 선택 노드를 사용하여 직원 레코드를 제거할 수 있습니다.
- 데이터 세트가 크면 공백이 있는 모든 레코드를 삭제할 수 있습니다.

② 결측값이 있는 필드 처리

결측값의 대부분이 소스의 필드에 집중되면 이들을 레코드 수준이 아니라 필드 수준에서 해결할 수 있습니다. 이 방법을 사용하면 결측값 처리를 위한 방법을 결정하기 전에 특정 필드의 상대적 중요성을 실험할 수 있습니다. 필드가 모델링에서 중요하지 않은 경우에는 결측값이 얼마나 많은지와는 관계 없이 유지할 가치가 없을 수 있습니다.

예를 들어, 시장 조사 회사는 50개의 질문을 포함하는 일반 질문지에서 데이터를 수집할 수 있습니다. 질문 중 두 개는 연령과 정치적 신념 등 대다수의 사람이 응답을 꺼려하는 정보를 다룹니다. 이 경우, *Age* 및 *Political_persuasion*에는 결측값이 많습니다.

필드 측정 수준
사용할 방법을 결정할 때 결측값이 있는 필드의 측정 수준 또한 고려해야 합니다.

수치 필드. 연속과 같은 수치 필드 유형의 경우에는 수치 필드에 공백이 포함되면 대다수의 모델이 작동하지 않으므로 모델을 작성하기 전에 항상 수치가 아닌 값을 제거해야 합니다.

범주형 필드. 명목 및 플래그 등과 같은 범주형 필드의 경우 결측값을 변경하는 것은 필수는 아니지만 모델의 정확도를 높여줍니다. 예를 들어, 성별 필드를 사용하는 모델은 Y와 Z 같은 의미 없는 값이 있어도 잘 작동하지만 M과 F를 제외한 모든 값을 제거하면 모델의 정확도가 증가합니다.

필드 선별 또는 제거

결측값이 너무 많은 필드를 선별하는 데에는 몇몇 옵션이 있습니다.

- 데이터 검토 노드를 사용하여 품질을 기반으로 필드를 필터링할 수 있습니다.
- 필드선택 노드를 사용하여 지정된 결측값 비율보다 높은 필드를 선별하고 지정된 목표와 관련된 중요도를 기반으로 필드의 등급을 지정할 수 있습니다.
- 필드를 제거하는 대신에 유형 노드를 사용하여 필드 역할을 없음으로 설정할 수 있습니다. 그러면 필드가 데이터 세트에 보관되지만 모델링 프로세스에서는 제외됩니다.

③ 시스템 결측값이 있는 레코드 처리

시스템 결측값이란?

시스템 결측값은 알려지지 않거나 적용되지 않는 데이터 값을 나타냅니다. 데이터베이스에서 이러한 값은 종종 *NULL* 값으로 언급됩니다.

시스템 결측값은 공백값과는 다릅니다. 공백값은 일반적으로 유형 노드에 특정 값 또는 값의 범위로 정의되어 있고 사용자 정의 결측으로 간주될 수 있습니다. 공백값은 모델링 컨텍스트에서 다르게 처리됩니다.

시스템 결측값 구성

시스템 결측값은 데이터 소스에서 읽혀지는 데이터에 있을 수 있습니다(예: 데이터베이스 테이블에 *NULL* 값이 포함될 수 있음).

시스템 결측값은 표현식에서 **undef** 값을 사용하여 구성될 수 있습니다. 예를 들어, 다음 CLEM 표현식은 30 이하의 경우 연령을 리턴하고 30보다 큰 경우 결측값을 리턴합니다.

```
if Age > 30 then undef else Age endif
```

결측값은 또한 외부 결합이 수행될 때, 숫자가 0으로 나눠질 때, 음수의 제곱근이 계산될 때 및 기타 상황에서도 작성될 수 있습니다.

시스템 결측값 표시

시스템 결측값은 테이블과 다른 출력에 $null$로 표시됩니다.

시스템 결측값 검정

특수 함수 **@NULL**을 사용하여 다음과 같이 인수 값이 시스템 결측값인 경우 true를 리턴하십시오.

```
if @NULL(MyFieldName) then 'It is null' else 'It is not null' endif
```

함수에 전달된 시스템 결측값

함수에 전달된 시스템 결측값은 일반적으로 출력에 결측값을 전파합니다. 예를 들어, f1 필드 값

이 특정 행에서 시스템 결측값이면 log(f1) 표현식 또한 해당 행을 시스템 결측값으로 평가합니다. 예외는 **@NULL** 함수입니다.

산술 연산자와 관련된 표현식에서 시스템 결측값

시스템 결측값을 포함하는 값에 산술 연산자를 적용하면 시스템 결측값이 생깁니다. 예를 들어, f1 필드 값이 특정 행에서 시스템 결측값이면 f1 + 10 표현식 또한 해당 행을 시스템 결측값으로 평가합니다.

논리 연산자와 관련된 표현식에서 시스템 결측값

논리 연산자와 관련된 표현식에서 시스템 결측값에 대해 작업할 때 3차 값 논리(*true*, *false* 및 *missing*)의 규칙이 적용되고 진리표에서 설명될 수 있습니다. *not*, *and* 및 *or*의 공통 논리 연산자에 대한 진리표는 다음 테이블에 표시됩니다.

표 1. NOT에 대한 진리표

피연산자	NOT 피연산자
true	false
false	true
결측	결측

표 2. AND에 대한 진리표

Operand1	Operand2	Operand1 AND Operand2
true	true	true
true	false	false
true	결측	결측
false	true	false
false	false	false
false	결측	false
결측	true	결측
결측	false	false
결측	결측	결측

표 3. OR에 대한 진리표

Operand1	Operand2	Operand1 OR Operand2
true	true	true
true	false	true
true	결측	true
false	true	true
false	false	false
false	결측	결측
결측	true	true
결측	false	결측
결측	결측	결측

비교 연산자와 관련된 표현식에서 시스템 결측값

시스템 결측값과 비시스템 결측값을 비교할 때 출력은 참 또는 거짓 결과 대신에 시스템 결측값으로 평가합니다. 시스템 결측값은 서로 비교할 수 있습니다. 두 개의 시스템 결측값은 동일한 것으로 간주됩니다.

if/then/else/endif 표현식에서 시스템 결측값

조건식을 사용할 때 조건식이 시스템 결측값을 리턴하면 else 절의 값이 조건식에서 리턴됩니다.

선택 노드의 시스템 결측값

특정 레코드의 경우 선택 표현식이 결측값으로 평가되면 레코드는 선택 노드의 출력이 아닙니다. 이 조치는 포함 및 삭제 모드 둘 모두에 적용됩니다.

병합 노드의 시스템 결측값

키를 사용하여 병합할 때 키 필드에 시스템 결측값이 있는 모든 레코드는 병합되지 않습니다.

통합의 시스템 결측값

열에서 데이터를 통합할 때 결측값은 계산에 포함되지 않습니다. 예를 들어, 3개의 값 {1, 2 및 undef}가 있는 열에서 열의 값 합계는 3으로 계산되고, 평균 값은 1.5로 계산됩니다.

(3) 결측값 대치 또는 채우기

결측값이 몇 개밖에 없는 경우에는 공백을 대체하기 위해 값을 삽입하는 것이 유용할 수도 있습니다. 이를 데이터 검토 보고서에서 수행할 수 있고 이를 통해 특정 필드에 옵션을 적합하게 지정한 다음 여러 방법을 사용하여 값을 대치하는 SuperNode를 생성할 수 있습니다. 이는 가장 유용한 방법이고 단일 노드에서 대량 필드의 처리를 지정할 수도 있습니다.

결측값을 대치하기 위해 다음 방법을 사용할 수 있습니다.

고정됨 고정된 값을 대체합니다(필드 평균, 범위의 중심점 또는 사용자가 지정하는 상수).

임의 보통 또는 균일 분포를 기반으로 변량 값을 대체합니다.

표현식. 사용자 정의 표현식을 지정할 수 있습니다. 예를 들어, 전역값 설정 노드에 의해 생성된 글로벌 변수로 값을 대체할 수 있습니다.

알고리즘. C&RT 알고리즘을 기반으로 모델에 의해 예측된 값을 대체합니다. 이 방법을 사용하여 대체된 각 필드의 경우, 공백과 널을 모델에 의해 예측된 값으로 대체하는 채움 노드와 함께 별도의 C&RT 모델이 있습니다. 그러면 필터 노드가 모델에 의해 생성된 예측 필드를 제거하는 데 사용합니다.

또는 특정 필드의 값을 **강제 실행**하려면 유형 노드를 사용하여 필드 유형이 합법적인 값만을 다루는지 확인한 다음 공백값을 대체해야 하는 필드의 선택 열을 강제 실행으로 설정하십시오.

(4) CLEM 함수 결측값

결측값을 처리하는 데에는 몇 가지 함수를 사용합니다. 다음 함수는 종종 결측값을 버리거나 채우기 위해 선택 및 채움 노드에 사용합니다.

- count_nulls(LIST)
- @BLANK(FIELD)
- @NULL(FIELD)
- undef

@ 함수는 @FIELD 함수와 함께 사용하여 하나 이상의 필드에서 공백이나 널값을 식별할 수 있습니다. 필드는 공백이나 널값이 있을 때 간단히 플래그를 지정하거나 대체 값으로 채우거나 다양한 다른 작업에 사용할 수 있습니다.

다음과 같이 필드 목록 전체에 걸쳐 널을 계산할 수 있습니다.

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

필드 목록을 입력으로 받아들이는 함수를 사용할 때 다음 예제에 표시된 대로 특수 함수 @FIELDS_BETWEEN 및 @FIELDS_MATCHING을 사용할 수 있습니다.

```
count_nulls(@FIELDS_MATCHING('card*'))
```

undef 함수를 사용하여 **$null$**로 표시되는 시스템 결측값이 있는 필드를 채울 수 있습니다. 예를 들어, 숫자 값을 대체하기 위해 다음과 같이 조건문을 사용할 수 있습니다.

```
if not(Age > 17) or not(Age < 66) then undef else Age endif
```

이는 **$null$**로 표시되는 시스템 결측값이 있는 범위가 아닌 모든 것을 대체합니다. not() 함수를 사용하여 음수를 포함하여 다른 모든 숫자 값을 포착할 수 있습니다. 자세한 정보는 공백 및 널값 처리 함수의 내용을 참조하십시오.

레코드 삭제에 대한 참고

선택 노드를 사용하여 레코드를 제거할 때 구문이 3값 논리를 사용하고 선택 명령문에 널값을 자동으로 포함함을 유의하십시오. 널값(시스템 결측값)을 선택 표현식에서 제외하려면 표현식에서 and not을 사용하여 이를 명시적으로 지정해야 합니다. 예를 들어, 처방약의 유형이 Drug C인 모든 레코드를 선택하고 포함하려면 다음 선택 명령문을 사용합니다.

```
Drug = 'drugC' and not(@NULL(Drug))
```

이러한 상황에서 제외된 널값의 이전 버전

8) CLEM 표현식 작성

(1) CLEM 정보

CLEM(Control Language for Expression Manipulation)은 IBM® SPSS® Modeler 스트림을 따라 플로우하는 데이터를 분석하고 조작하기 위한 강력한 언어입니다. 데이터 마이너는 비용과 수입 데이터에서 수익을 파생하는 것과 같은 단순한 작업 또는 웹 로그 데이터를 유용한 정보가

있는 필드와 레코드 세트로 변환하는 것과 같은 복잡한 작업을 수행하기 위해 스트림 작업에서 CLEM을 집중적으로 사용합니다.

CLEM은 IBM SPSS Modeler 내에서 다음을 위해 사용합니다.

- 레코드 필드에서 조건을 비교하고 평가합니다.
- 새 필드의 값을 파생합니다.
- 기존 필드의 새 값을 파생합니다.
- 레코드 시퀀스에 대한 원인.
- 데이터를 레코드에서 보고서에 삽입합니다.

CLEM 표현식은 IBM SPSS Modeler에서 데이터 준비에는 필수적이며 레코드와 필드 작업(선택, 균형, 채움)부터 차트와 출력(분석, 보고서, 테이블)까지 광범위한 노드에서 사용할 수 있습니다. 예를 들어, 파생 노드에서 CLEM을 사용하여 비율과 같은 수식을 기반으로 새 필드를 작성할 수 있습니다.

CLEM 표현식은 또한 글로벌 검색과 바꾸기 작업에 사용할 수도 있습니다. 예를 들어, @NULL(@FIELD) 표현식은 채움 노드에서 **시스템 결측값**을 정수값 0으로 바꾸는 데 사용할 수 있습니다. (공란이라고도 불리는 **사용자 결측값**을 바꾸려면 @BLANK 함수를 사용하십시오.)

(2) CLEM 예제

CLEM으로 가능한 표현식 유형은 물론 올바른 구문을 설명하기 위해 예제 표현식이 이어집니다.

단순 표현식

수식은 *After* 및 *Before* 필드 값을 기반으로 새 필드를 파생시키는 이 수식만큼 단순할 수 있습니다.

```
(After - Before) / Before * 100.0
```

필드 값을 참조할 때 필드 이름에는 따옴표가 없습니다.

마찬가지로 다음 표현식은 *salary* 필드 각 값의 로그를 리턴하기만 합니다.

```
log(salary)
```

복합 표현식

또한 표현식은 길고 더 복잡할 수 있습니다. 다음 표현식은 두 필드(*$KX-Kohonen* 및 *$KY-Kohonen*) 값이 지정된 범위 내에 있을 경우 참을 리턴합니다. 필드 이름에 특수 문자가

포함되어 있기 때문에 여기서 필드 이름에는 작은따옴표가 있습니다.

```
('$KX-Kohonen' >= -0.2635771036148072 and '$KX-Kohonen' <= 0.3146203637123107
and '$KY-Kohonen' >= -0.18975617885589602 and
'$KY-Kohonen' <= 0.17674794197082522) -> T
```

문자열 함수와 같은 몇 개의 함수에는 올바른 구문을 사용하여 몇 개의 모수를 입력해야 합니다. 다음 예제에서 subscrs 함수는 *produce_ID* 필드의 첫 번째 문자를 리턴하는 데 사용되어 항목이 유기농인지 유전자 변형인지 일반인지 여부를 표시합니다. 표현식 결과는 -> ‘result’로 설명됩니다.

```
subscrs(1,produce_ID) -> ‘c‘
```

마찬가지로 다음 표현식이 있습니다.

```
stripchar(`3`,`123`) -> `12`
```

문자는 단일 역따옴표 안에서 항상 캡슐화됨에 유의하는 것이 중요합니다.

표현식에서 함수 조합

CLEM 표현식은 함수 조합으로 구성되는 경우가 빈번합니다. 다음 표현식은 subscr 및 lowertoupper를 조합하여 *produce_ID*의 첫 번째 문자를 리턴하고 이를 대문자로 변환합니다.

```
lowertoupper(subscr(1,produce_ID)) -> `C`
```

이 동일한 표현식을 간단하게 다음과 같이 작성할 수 있습니다.

```
lowertoupper(produce_ID(1)) -> `C`
```

일반적으로 사용하는 또 다른 함수 조합은 다음과 같습니다.

```
locchar_back(`n`, (length(web_page)), web_page)
```

이 표현식은 필드 값의 마지막 문자부터 거꾸로 읽어 *web_page* 필드 값에 `n` 문자를 위치시킵니다. length 함수도 포함시켜 이 표현식은 7과 같은 정적 숫자(7자 미만인 값의 경우 올바르지 않음)를 사용하지 않고 현재 값의 길이를 동적으로 계산합니다.

특수 함수

수많은 특수 함수(앞에 @ 기호가 붙음)를 사용할 수 있습니다. 일반적으로 사용하는 함수는 다음과 같습니다.

```
@BLANK('referrer ID') -> T
```

특수 함수를 조합에서 사용하는 경우가 빈번한데, 한 번에 두 개 이상의 필드에서 공백에 플래그를 표시하는 일반적으로 사용하는 방법입니다.

```
@BLANK(@FIELD)-> T
```

추가 예제가 CLEM 문서 전반에 설명되어 있습니다. 자세한 정보는 CLEM 참조 개요의 내용을 참조하십시오.

(3) 값 및 데이터 유형

CLEM 표현식은 값, 필드 이름, 연산자 및 함수로부터 구성된 수식과 유사합니다. 가장 단순한 CLEM 표현식은 값 또는 필드 이름입니다. 유효한 값의 예제는 다음과 같습니다.

```
31.79
'banana'
```

필드 이름의 예:

```
Product_ID
'$P-NextField'
```

여기서 *Product*는 장바구니 데이터 세트의 필드 이름이고, *'$P-NextField'*는 모수의 이름이고, 표현식의 값은 이름 지정된 필드의 값입니다. 일반적으로 필드 이름은 문자로 시작하고 숫자와 밑줄(_)을 포함할 수도 있습니다. 이름을 따옴표 내에 놓으면 이러한 규칙을 따르지 않는 이름을 사용할 수 있습니다. CLEM 값은 다음 중 하나일 수 있습니다.

- 문자열--예: "c1", "Type 2", "a piece of free text"
- 정수--예: 12, 0, -189
- 실수--예: 12.34, 0.0, -0.0045
- 날짜/시간 필드--예: 05/12/2002, 12/05/2002, 12/05/02

다음 요소를 사용하는 것도 가능합니다.

- 문자 코드--예: `a` or 3
- 항목 목록--예: [1 2 3], ['Type 1' 'Type 2']

문자 코드와 목록은 보통 필드 값으로서 발생하지 않습니다. 일반적으로 이들은 CLEM 함수의 인수로서 사용합니다.

인용 규칙

CLEM 표현식에서 사용하는 필드, 값, 모수 및 문자열을 결정할 때 소프트웨어는 유연하지만, 다음 일반 규칙은 표현식을 작성할 때 사용할 "우수 사례" 목록을 제공합니다.

- **문자열**—문자열을 쓸 때 항상 큰따옴표를 사용하십시오("Type 2" 또는 "value"). 작은따옴표를 대신 사용할 수 있지만 인용된 필드와 혼동될 위험이 있습니다.
- **문자**—항상 `와 같이 하나의 역따옴표를 사용하십시오. 예를 들어, stripchar(`d`,"drugA") 함수에서 d 문자를 참고하십시오. 이에 대한 유일한 예외는 문자열에서 특정 문자를 가리키기 위해 정수를 사용하는 경우입니다. 예를 들어, lowertoupper("druga"(5)) —> "A" 함수에서 5 문자를 참고하십시오. 참고: 표준 영국 및 미국 키보드에서 역따옴표 문자의 키(억음 부호, 유니코드 0060)는 Esc 키 바로 아래에서 찾을 수 있습니다.
- **필드**—필드는 일반적으로 CLEM 표현식에서 사용할 때 인용을 사용하지 않습니다(subscr(2,arrayID)) —> CHAR). 필요한 경우 작은따옴표를 사용하여 공백이나 기타 특수 문자를 묶을 수 있습니다('Order Number'). 인용되었지만 데이터 세트에 정의되지 않은 필드는 문자열로 잘못 읽힙니다.
- **모수**—항상 작은따옴표('$P-threshold')를 사용합니다.

(4) 표현식 및 조건

CLEM 표현식은 다음과 같은 결과를 리턴할 수 있습니다(예: 새 값을 파생시킬 때).

```
Weight * 2.2
Age + 1
sqrt(Signal-Echo)
```

또는 다음과 같은 *참* 또는 *거짓*을 평가할 수 있습니다(조건을 기준으로 선택할 때).

```
Drug = "drugA"
Age < 16
not(PowerFlux) and Power > 2000
```

다음과 같이 CLEM 표현식에서 연산자와 함수를 임의로 결합할 수 있습니다.

```
sqrt(abs(Signal)) * max(T1, T2) + Baseline
```

대괄호와 연산자 우선순위는 표현식이 평가되는 순서를 결정합니다. 이 예에서는 평가의 순서는 다음과 같습니다.

- abs(Signal)가 평가되고, sqrt가 해당 결과에 적용됩니다.
- max(T1, T2)가 평가됩니다.
- 두 개의 결과를 곱합니다. x가 +보다 우선순위가 높습니다.
- 마지막으로, Baseline이 결과에 추가됩니다.

우선순위의 내림차순(즉, 가장 먼저 수행되는 작업부터 가장 마지막에 수행되는 작업)은 다음과 같습니다.

- 함수 인수
- 함수 호출
- **xx**
- **x / mod div rem**
- + -
- > < >= <= /== == = /=

우선순위를 대체하려거나 평가 순서에 의심이 든다면 다음과 같이 소괄호를 사용하여 이를 명백하게 만들 수 있습니다.

```
sqrt(abs(Signal)) * (max(T1, T2) + Baseline)
```

(5) 스트림, 세션, 수퍼 노드 모수

CLEM 표현식과 스크립팅에서 사용할 모수를 정의할 수 있습니다. 사실상 모수는 현재 스트림, 세션 또는 수퍼노드와 함께 저장되고 지속되는 사용자 정의 변수이며 스크립팅을 통해서는 물론 사용자 인터페이스에서도 액세스할 수 있습니다. 예를 들어, 스트림을 저장하면 해당 스트림에 대해 설정된 모수도 저장됩니다(이는 선언된 스크립트에서만 사용할 수 있는 로컬 스크립트 변수와 모수를 구별함). 모수는 스크립트에 하드 코드화하지 않아도 되는 필드 및 값에 대한 정보를 제공하여 스크립트팅에서 스크립트 동작을 제어하는 데 사용하는 경우가 많습니다.

모수 범위는 모수가 설정된 위치에 따라 다릅니다.

- 스트림 모수는 스트림 스크립트 또는 스트림 특성 대화 상자에서 설정할 수 있으며 스트림의 모든 노드에 사용할 수 있습니다. 표현식 작성기의 모수 목록에 표시됩니다.
- 세션 모수는 독립형 스크립트 또는 세션 모수 대화 상자에서 설정할 수 있습니다. 현재 세션에서 사용하는 모든 스트림(관리자 분할창의 스트림 탭에 나열된 모든 스트림)에 사용할 수 있습니다.

수퍼노드에 대해서도 모수를 설정할 수 있는데, 이 경우 해당 수퍼노드 내에 캡슐화된 노드에만 표시됩니다.

CLEM 표현식에서 모수 사용

모수는 CLEM 표현식에서 $P-pname으로 표시됩니다. 여기서 pname은 모수의 이름입니다. CLEM 표현식에서 사용하면 모수는 작은 따옴표 안에 두어야 합니다(예: '$P-scale').

사용 가능한 모수는 표현식 작성기를 사용하여 쉽게 볼 수 있습니다. 현재 모수 보기

1. CLEM 표현식을 승인하는 대화 상자에서 표현식 작성기 단추를 클릭하십시오.

2. 필드 목록에서 **모수**를 선택하십시오.

CLEM 표현식에 삽입하기 위해 목록에서 모수를 선택할 수 있습니다. 자세한 정보는 필드, 모수 및 글로벌 변수 선택 주제를 참조하십시오.

(6) 문자열에 대한 작업

문자열에 사용 가능한 많은 연산이 있는데, 다음과 같습니다.

- 문자열을 대문자 또는 소문자로 변환 - uppertolower(CHAR)
- 문자열 변수에서 지정된 문자(예: `ID_` 또는 `$`) 제거 - stripchar(CHAR,STRING)
- 문자열 변수의 길이(문자 수) 판별 - length(STRING)
- 문자열 값의 알파벳 순서 검사 - alphabefore(STRING1, STRING2)
- 값에서 선행 또는 후행 공백 제거 - trim(STRING), trim_start(STRING) 또는 trimend(STRING)
- 문자열에서 처음 또는 마지막 n개의 문자 추출 - startstring(LENGTH, STRING) 또는 endstring(LENGTH, STRING). 예를 들어, 제품 이름을 4자리 ID 코드와 결합(ACME CAMERA-D109)하는 *item* 필드가 있다고 가정하십시오. 4자리 코드만 포함하는 새 필드를 작성하려면 파생 노드에 다음 수식을 지정하십시오.

```
endstring(4, item)
```

- 특정 패턴 매치 - STRING matches PATTERN. 예를 들어, 직위 어딘가에 "market"이 포함되어 있는 사용자를 선택하려면 선택 노드에 다음을 지정할 수 있습니다.

```
job_title matches "*market*"
```

- 문자열 내 하위 문자열의 인스턴스 모두 바꾸기 - replace(SUBSTRING, NEWSUBSTRING, STRING). 예를 들어, 텍스트 마이닝 이전에 지원되지 않는 문자(예: 수직 파이프(|))를 세미콜론으로 모두 바꾸려면 채움 노드 replace 함수를 사용하십시오. **필드 채우기**에서 문자가 나타날 수 있는 필드를 모두 선택하십시오. **바꾸기**: 조건에 대해 **항상**을 선택하고 **바꿀 문자열**: 아래에 다음 조건을 지정하십시오.

```
replace('|',';',@FIELD)
```

- 특정 하위 문자열의 존재를 기준으로 플래그 필드 파생. 예를 들어, 파생 노드에서 문자열 함수를 사용하여 다음과 같은 표현식으로 각 반응에 대해 별도의 플래그 필드를 생성할 수 있습니다.

```
hassubstring(museums,"museum_of_design")
```

자세한 정보는 문자열 함수의 내용을 참조하십시오.

(7) 공백 및 결측값 처리

공백 또는 결측값을 대체하는 것은 데이터 마이너를 위한 공통적인 데이터 준비 작업입니다. CLEM는 공백 처리를 자동화하기 위한 많은 도구를 제공합니다. 채움 노드는 공백에 대한 작업을 하는 가장 일반적인 위치입니다. 그러나 CLEM 표현식을 허용하는 모든 노드에서 다음 함수를 사용할 수 있습니다.

- @BLANK(FIELD)를 사용하여 특정 필드(예: *Age*)에 대해 값이 공백인 레코드를 판별할 수 있습니다.
- @NULL(FIELD)을 사용하여 지정된 필드에 대해 시스템 결측값인 레코드를 판별할 수 있습니다. IBM® SPSS® Modeler에서 시스템 결측값은 **$null$** 값으로 표시됩니다.

자세한 정보는 공백 및 널값 처리 함수의 내용을 참조하십시오.

(8) 숫자에 대한 작업

IBM® SPSS® Modeler에서는 숫자 값에 대한 다수의 표준 연산을 사용할 수 있는데, 다음과 같습니다.

- 지정된 각의 사인 계산 — sin(NUM)
- 숫자 필드의 자연로그 계산 — log(NUM)
- 두 숫자의 합계 계산 — NUM1 + NUM2

자세한 정보는 숫자 함수의 내용을 참조하십시오.

(9) 시간 및 날짜에 대한 작업

시간 및 날짜 형식은 데이터 소스와 로케일에 따라 다릅니다. 날짜 및 시간 형식은 각 스트림에 고유하며 스트림 특성 대화 상자에서 설정됩니다. 다음 예제는 날짜/시간 필드에 대한 작업에 일반적으로 사용하는 함수입니다.

경과 시간 계산

다음 함수와 유사한 함수 패밀리를 사용하여 기준 날짜부터 경과한 시간을 쉽게 계산할 수 있습니다. 이 함수는 기준 날짜부터 날짜 문자열 DATE가 나타내는 날짜까지의 시간(월)을 실수로 리턴합니다. 이는 30.0일을 1개월로 한 대략적인 숫자입니다.

```
date_in_months(Date)
```

날짜/시간 값 비교

다음 함수와 유사한 함수를 사용하여 레코드 전반에서 날짜/시간 필드 값을 비교할 수 있습니다. 이 함수는 날짜 문자열 DATE1이 날짜 문자열 DATE2가 나타내는 날짜 이전의 날짜를 나타내는 경우 참 값을 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다.

```
date_before(Date1, Date2)
```

차이 계산

다음과 같은 함수를 사용하여 두 개의 시간과 두 개의 날짜 간의 차이도 계산할 수 있습니다.

```
date_weeks_difference(Date1, Date2)
```

이 함수는 날짜 문자열 DATE1이 나타내는 날짜부터 날짜 문자열 DATE2가 나타내는 날짜까지의 시간(주)을 실수로 리턴합니다. 이는 7.0일을 1주로 합니다. DATE2가 DATE1 이전인 경우 이 함수는 음수를 리턴합니다.

오늘 날짜

@TODAY 함수를 사용하여 데이터 세트에 현재 날짜를 추가할 수 있습니다. 오늘 날짜는 스트림 특성 대화 상자에서 선택된 날짜 형식을 사용하여 지정된 필드 또는 새 필드에 문자열로 추가됩니다. 자세한 정보는 날짜 및 시간 함수의 내용을 참조하십시오.

(10) 다중 필드 요약

CLEM 언어는 다중 필드 전반에서 요약 통계를 리턴하는 많은 함수를 포함합니다. 이러한 함수는 특히 하나의 질문에 대한 다중반응이 다중 필드에 저장될 수 있는 설문조사 데이터 분석 시 유용합니다. 자세한 정보는 다중반응 데이터에 대한 작업의 내용을 참조하십시오.

비교 함수

min_n과 max_n 함수를 사용하여 다중 필드 전반에서 값을 비교할 수 있습니다. 예를 들어, 다음과 같습니다.

```
max_n(['card1fee' 'card2fee''card3fee''card4fee'])
```

또한 다수의 계산 함수를 사용하여 특정 기준을 충족시키는 값 수를 확보할 수 있습니다(해당 값이 다중 필드에 저장된 경우에도 적용됨). 예를 들어, 5년 넘게 보유한 카드 수를 계산하려면 다음을 사용하십시오.

```
count_greater_than(5, ['cardtenure' 'card2tenure' 'card3tenure'])
```

같은 필드 세트에서 널 값을 계산하려면 다음을 사용하십시오.

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

이 예제에서는 카드를 보유한 사람 수는 계산하지 않고 보유 중인 카드 수를 계산합니다. 자세한 정보는 비교 함수의 내용을 참조하십시오.

지정된 값이 다중 필드에서 나타나는 횟수를 계산하려면 count_equal 함수를 사용할 수 있습니다. 다음 예제에서는 목록에서 Y 값이 포함된 필드 수를 계산합니다.

```
count_equal("Y",[Answer1, Answer2, Answer3])
```

목록의 필드에 다음 값이 주어지면 이 함수는 표시된 대로 Y 값에 대한 결과를 리턴합니다.

표 1. 함수 값

Answer1	Answer2	Answer3	수
Y	N	Y	2
Y	N	N	1

숫자 함수

sum_n, mean_n, sdev_n 함수를 사용하여 다중 필드 전반에서 통계를 확보할 수 있습니다.

```
sum_n(['card1bal' 'card2bal''card3bal'])
```

```
mean_n(['card1bal' 'card2bal''card3bal'])
```

자세한 정보는 숫자 함수의 내용을 참조하십시오.

필드 목록 생성

필드 목록을 입력으로 허용하는 함수를 사용하는 경우 특수 함수 @FIELDS_BETWEEN(start, end) 및 @FIELDS_MATCHING(pattern)을 입력으로서 사용할 수 있습니다. 예를 들어, 필드 순서가 앞선 sum_n 예제에 표시된 것과 같다고 가정하면 다음과 같습니다.

```
sum_n(@FIELDS_BETWEEN(card1bal, card3bal))
```

또는 "*card*"로 시작하는 모든 필드에서 널 값 수를 계산하려면 다음을 사용하십시오.

```
count_nulls(@FIELDS_MATCHING('card*'))
```

자세한 정보는 특수 필드의 내용을 참조하십시오.

(11) 다중반응 데이터에 대한 작업

다음과 같은 다수의 비교 함수를 사용하여 다중반응 데이터를 분석할 수 있습니다.

- value_at
- first_index / last_index
- first_non_null / last_non_null
- first_non_null_index / last_non_null_index
- min_index / max_index

예를 들어, 특정 구매 결정을 위해 첫 번째, 두 번째, 세 번째로 가중 중요한 이유(예: 가격, 개인적인 추천, 검토, 로컬 공급자 등)를 묻는 다중반응 질문을 가정하십시오. 이 경우, 처음 포함된 필드의 지수를 파생시켜 가격의 중요도를 결정합니다.

```
first_index("price", [Reason1 Reason2 Reason3])
```

마찬가지로, 고객에게 구매 우도가 있는 순서대로 자동차 3대를 순위화하도록 요청하고 다음과 같이 반응을 세 개의 개별 필드에 코드화했다고 가정하십시오.

표 1. 자동차 순위화 예제

고객 id	car1	car2	car3
101	1	3	2
102	3	2	1
103	2	3	1

이 경우, min_index 함수를 사용하여 가장 좋아하는 자동차(#1로 순위가 매겨짐, 그렇지 않으면 최저 순위)에 대한 필드의 지수를 판별할 수 있습니다.

```
min_index(['car1' 'car2' 'car3'])
```

자세한 정보는 비교 함수의 내용을 참조하십시오.

다중반응 세트 참조

특수 @MULTI_RESPONSE_SET 함수를 사용하여 다중반응 세트의 모든 필드를 참조할 수 있습니다. 예를 들어, 이전 예제에 있는 세 개의 *car* 필드가 *car_rankings*라는 다중반응 세트에 포함된 경우 다음은 동일한 결과를 리턴합니다.

```
max_index(@MULTI_RESPONSE_SET("car_rankings"))
```

(12) 표현식 작성기

CLEM 표현식을 수동으로 입력하거나 CLEM 함수와 연산자의 전체 목록뿐만 아니라 현재 스트림의 데이터 필드를 표시하는 표현식 작성기를 사용할 수 있습니다. 이를 통해 필드나 함수의 정확한 이름을 기억하지 않고도 표현식을 빠르게 작성할 수 있습니다. 또한 작성기는 필드와 값에 적합한 인용구를 자동으로 추가하여 통사적으로 정확한 표현식 작성을 보다 쉽게 해줍니다.

ⓘ **참고**: 표현식 작성기는 스크립트나 모수 설정에서는 지원되지 않습니다.

ⓘ **참고**: 데이터 소스를 변경하려는 경우에는 소스를 변경하기 전에 표현식 작성기가 사용자가 선택한 함수를 여전히 지원하는지를 확인해야 합니다. 모든 데이터베이스가 모든 함수를 지원하지는 않으므로 새 데이터 소스를 실행하는 경우 오류가 발생할 수도 있습니다.

① 표현식 작성기에 액세스

선택, 균형, 파생, 채움, 분석, 보고서 및 테이블 노드를 포함하여 CLEM 표현식이 사용하는 모든 노드에서 표현식 작성기를 사용할 수 있습니다. 수식 필드 오른쪽에 있는 계산기 단추를 클릭하여 열 수 있습니다.

② 표현식 작성

표현식 작성기는 필드, 함수 및 연산자 전체 목록은 물론 데이터가 인스턴스화된 경우 데이터 값에 대한 액세스도 제공합니다.

표현식 작성기를 사용하여 표현식을 작성하는 방법

1. 함수 및 필드 목록을 참조로 사용하여 표현식 필드를 입력하십시오.
   or
2. 스크롤 목록에서 필수 필드 및 함수를 선택하십시오.
3. 노란색 화살표 단추를 두 번 클릭하거나 클릭하여 표현식 필드에 필드나 함수를 추가하십시오.
4. 대화 상자 중앙의 피연산자 단추를 사용하여 표현식에 연산을 삽입하십시오.

③ **함수 선택**

함수 목록은 사용 가능한 모든 CLEM 함수 및 연산자를 표시합니다. 스크롤하여 목록에서 함수를 선택하거나 보다 쉽게 검색하려면 드롭 다운 목록을 사용하여 함수 또는 연산자의 서브세트를 표시하십시오. 사용 가능한 함수는 쉽게 검색할 수 있도록 범주로 그룹화되어 있습니다.

이러한 범주의 대부분은 CLEM 언어 설명의 참조 섹션에 설명되어 있습니다. 추가 정보는 함수 참조의 내용을 참조하십시오.

기타 범주는 다음과 같습니다.

- **일반 함수**에는 가장 자주 사용하는 함수 중 일부의 선택사항이 포함됩니다.
- **최근에 사용**에는 현재 세션 내에서 사용하는 CLEM 함수 목록이 포함됩니다.
- **@ 함수**에는 모든 특수 함수의 목록이 포함되며 각 이름 앞에 "@" 부호가 있습니다.

> ⓘ **참고**: @DIFF1(FIELD1,FIELD2) 및 @DIFF2(FIELD1,FIELD2) 함수의 경우 두 필드 유형이 동일해야 합니다(예: 둘 다 정수 또는 둘 다 Long 또는 Real).

- **데이터베이스 함수**. 스트림에 데이터베이스 연결이 포함되는 경우(데이터베이스 소스 노드 사용), 이 선택영역은 사용자 정의 함수(UDF)를 포함하여 해당 데이터베이스 내에서 사용 가능한 함수를 나열합니다. 추가 정보는 데이터베이스 함수의 내용을 참조하십시오.
- **데이터베이스 통합**. 스트림에 데이터베이스 연결이 포함되는 경우(데이터베이스 소스 노드 사용), 이 선택영역은 해당 데이터베이스 내에서 사용 가능한 통합 옵션을 나열합니다. 이러한 옵션은 통합 노드의 표현식 작성기에서 사용 가능합니다.
- **데이터베이스 창 통합**. 스트림에 데이터베이스 연결이 포함되는 경우(데이터베이스 소스 노드 사용), 이 선택은 해당 데이터베이스 내에서 사용할 수 있는 창 통합 옵션을 나열합니다. 이러한 옵션은 **필드 작업** 팔레트에서만 노드 내의 표현식 작성기에서 사용 가능합니다.

> ⓘ **참고:** SPSS® Modeler가 데이터베이스 시스템 보기에서 **창 통합 함수**를 획득하므로 사용 가능한 옵션은 데이터베이스 작동에 따라 다릅니다.

"통합"이라고 불리기는 하지만 이러한 옵션은 통합 노드 전용으로 지정되지 않았으므로 파생 또는 선택 등과 같은 노드에 보다 적합합니다. 이는 출력이 참 통합 대신에 스칼라이기 때문입니다. 즉, 통합 노드가 하는 것과 같은 방법으로 출력에 표시되는 데이터량을 줄이지 않습니다. 예를 들어, "현재 행과 모든 이전 행의 평균" 등과 같이 데이터 행을 아래쪽으로 이동 평균을 제공하기 위해 이 통합 정렬을 사용할 수 있습니다.

- **내장 통합**. 사용할 수 있는 가능한 통합 모드 목록을 포함합니다.
- **연산자**는 표현식을 작성할 때 사용할 수 있는 모든 연산자를 나열합니다. 연산자는 대화 상자의 중심에서 단추에서도 사용 가능합니다.
- **모든 함수**에는 사용 가능한 CLEM 함수의 전체 목록을 포함합니다.

함수 그룹을 선택한 후에는 두 번 클릭하여 함수를 커서 위치로 표시되는 지점의 표현식 필드에 삽입하십시오.

가. 데이터베이스 함수

데이터베이스 함수는 서로 다른 여러 위치에 나열할 수 있습니다. 다음 테이블은 SPSS® Modeler가 함수 세부사항을 찾을 때 검색하는 위치를 보여줍니다. 이 테이블은 데이터베이스 관리자가 필수 영역에 대한 액세스 권한이 있는 사용자가 서로 다른 함수를 사용할 수 있도록 확인하는 데 사용할 수 있습니다.

또한 테이블은 함수가 사용 가능할 때 데이터베이스와 함수 유형을 기반으로 필터링하는 데 사용하는 조건을 나열합니다.

> **참고**: Amazon Redshift에서 데이터베이스 함수를 사용하는 경우 데이터베이스 관리자가 다음 6가지 데이터베이스 오브젝트에 대한 권한을 부여해야 합니다. 첫 번째 4개는 시스템 카탈로그 테이블이며 나머지 2개는 스키마입니다.
> - pg_type
> - pg_proc
> - pg_namespace
> - pg_aggregate
> - information_schema
> - pg_catalog

표 1. 표현식 작성기에서 데이터베이스 함수

데이터베이스	함수유형	함수 찾는 위치	함수를 필터링하는 데 사용하는 조건
Db2 LUW	UDF	SYSCAT.ROUTINES SYSCAT.ROUTINEPARMS	ROUTINETYPE은 F이고 FUNCTIONTYPE은 S임
Db2 LUW	UDA	SYSCAT.ROUTINES SYSCAT.ROUTINEPARMS	ROUTINETYPE은 F이고 FUNCTIONTYPE은 C임
Db2 iSeries	UDF	QSYS2.SYSROUTINES QSYS2.SYSPARMS	ROUTINE_TYPE은 F이고 FUNCTION_TYPE은 S임
Db2 iSeries	UDA	QSYS2.SYSROUTINES QSYS2.SYSPARMS	ROUTINETYPE은 F이고 FUNCTIONTYPE은 C임
Db2 z/OS	UDF	SYSIBM.SYSROUTINES SYSIBM.SYSPARMS	ROUTINE_TYPE은 F이고 FUNCTION_TYPE은 S임
Db2 z/OS	UDA	SYSIBM.SYSROUTINES SYSIBM.SYSPARMS	ROUTINETYPE은 F이고 FUNCTIONTYPE은 C임
SQL Server	UDF	SYS.ALL_OBJECTS SYS.ALL_PARAMETERS SYS.TYPES	TYPE은 FN 또는 FS임
SQL Server	UDA	SYS.ALL_OBJECTS SYS.ALL_PARAMETERS SYS.TYPES	TYPE은 AF임

데이터베이스	함수유형	함수 찾는 위치	함수를 필터링하는 데 사용하는 조건
Oracle	UDF	ALL_ARGUMENTS ALL_PROCEDURES	다음 모든 조건을 충족합니다. - OBJECT_TYPE은 FUNCTION임 - AGGREGATE는 NO임 - PLS_TYPE은 NULL이 아님
Oracle	UDA	ALL_ARGUMENTS ALL_PROCEDURES	다음 모든 조건을 충족합니다. - ARGUMENT_NAME은 NULL임 - AGGREGATE는 YES임 - PLS_TYPE은 NULL이 아님
Teradata	UDF	DBC.FUNCTIONS DBC.ALLRIGHTS	다음 모든 조건을 충족합니다. - FUNCTIONTYPE은 F임 - COLUMNNAME은 RETURN0임 - SPPARAMETERTYPE은 O임 - ACCESSRIGHT는 EF임
Teradata	UDA	DBC.FUNCTIONS DBC.ALLRIGHTS	다음 모든 조건을 충족합니다. - FUNCTIONTYPE은 A임 - COLUMNNAME은 RETURN0임 - SPPARAMETERTYPE은 O임 - ACCESSRIGHT는 EF임
Netezza	UDF	####.._V_FUNCTION NZA.._V_FUNCTION INZA.._V_FUNCTION	####.._V_FUNCTION의 경우 다음 조건이 적용됩니다. - RESULT에는 다음과 같은 값이 있는 문자열을 포함하지 않습니다. TABLE% - FUNCTION에는 다음과 같은 값이 있는 문자열을 포함하지 않습니다. '/_%' escape '/' - VARARGS는 FALSE임 NZA.._V_FUNCTION 및 INZA.._V_FUNCTION 둘 모두의 경우 다음 조건이 적용됩니다. - RESULT에는 다음과 같은 값이 있는 문자열을 포함하지 않습니다. TABLE% - FUNCTION에는 다음과 같은 값이 있는 문자열을 포함하지 않습니다. '/_%' escape '/' - BUILTIN은 f임 - VARARGS는 FALSE임
Netezza	UDA	####.._V_FUNCTION NZA.._V_FUNCTION INZA.._V_FUNCTION	다음 조건을 둘 모두 충족합니다. - AGGTYPE은 ANY이거나 GROUPED임 - VARARGS는 FALSE임

데이터베이스	함수유형	함수 찾는 위치	함수를 필터링하는 데 사용하는 조건
Netezza	WUDA	####.._V_FUNCTION NZA.._V_FUNCTION INZA.._V_FUNCTION	####.._V_AGGREGATE의 경우 다음 조건이 적용됩니다. - AGGTYPE은 ANY이거나 ANALYTIC임 - AGGREGATE은 MAX_LABEL이 아님 - VARARGS는 FALSE임 NZA.._V_FUNCTION 및 INZA.._V_FUNCTION 둘 모두의 경우 다음 조건이 적용됩니다. - AGGTYPE은 ANY이거나 ANALYTIC임 - BUILTIN은 f임 - VARARGS는 FALSE임

테이블에 사용하는 주요 용어

- UDF 사용자 정의 함수(User Defined Function)
- UDA 사용자 정의 집계 함수(User Defined Aggregate Function)
- WUDA 사용자 정의 창 집계 함수(User Defined Window Aggregate Function)
- #### 현재 연결된 데이터베이스

④ 필드, 모수 및 글로벌 변수 선택

필드 목록은 데이터 스트림의 이 위치에서 사용 가능한 모든 필드를 표시합니다. 목록에서 필드를 선택하려면 스크롤하십시오. 표현식에 필드를 추가하려면 노란색 화살표 단추를 두 번 클릭하거나 클릭하십시오.

자세한 정보는 스트림, 세션, 수퍼 노드 모수의 내용을 참조하십시오.

필드에 추가로 다음 항목에서 선택할 수도 있습니다.

다중반응 세트. 자세한 정보는 *IBM SPSS Modeler 소스, 프로세스 및 출력 노드* 안내서를 참조하십시오.

최근에 사용에는 필드 목록, 다중반응 세트, 모수 및 현재 세션 내에서 사용된 글로벌 값이 포함됩니다.

모수. 자세한 정보는 스트림, 세션, 수퍼 노드 모수의 내용을 참조하십시오.

글로벌 값. 자세한 정보는 *IBM SPSS Modeler 소스, 프로세스 및 출력 노드* 안내서를 참조하십시오.

⑤ 값 보기 또는 선택

필드 값은 표현식 작성기, 데이터 검토 보고서 및 시간 간격 노드에서 미래 값을 편집할 때를 포함하여 시스템 내의 여러 장소에서 볼 수 있습니다. 데이터가 이 기능을 사용하기 위해서는 소스 또는 유형 노드에서 완전히 인스턴스화되어야 저장 공간, 유형 및 값이 알려질 수 있음을 유의하십시오.

표현식 작성기 또는 시간 간격 노드에서 필드의 값을 보려면 필수 필드를 선택하고 값 선택도구 단추를 클릭하여 선택된 필드의 값을 나열하는 대화 상자를 여십시오. 그런 다음 값을 선택하고 **삽입**을 클릭하여 값을 현재 표현식 또는 목록에 붙여넣을 수 있습니다.

그림 1. 값 선택도구 단추

데이터 검토 보고서에서 필드의 값을 보려면 필수 필드의 유형 셀 또는 고유 셀을 클릭하십시오.

플래그 및 명목 필드의 경우 정의된 모든 값이 나열됩니다. 연속(숫자 범위) 필드의 경우 최소 및 최대 값이 표시됩니다.

⑥ CLEM 표현식 검사

표현식 작성기(오른쪽 하단 모서리)에서 검사를 클릭하여 표현식을 검증하십시오. 검사하지 않은 표현식은 빨간색으로 표시됩니다. 오류가 발견되면 원인을 나타내는 메시지가 표시됩니다.

다음 항목을 검사합니다.

- 값 및 필드 이름의 올바른 인용
- 모수 및 글로벌 변수의 올바른 사용
- 유효한 연산자 사용
- 참조된 필드의 존재
- 참조된 전역값의 존재 및 정의

구문에서 오류가 발생하면 표현식을 수동으로 입력하지 말고 목록과 연산자 단추를 사용하여 표현식을 작성하십시오. 필드 및 값에 대한 적절한 인용을 자동으로 추가합니다.

> ⓘ **참고**: 구분 문자가 포함된 필드 이름은 작은따옴표로 묶어야 합니다. 따옴표를 자동으로 추가하려면 표현식을 수동으로 입력하는 대신 목록 및 연산자 단추를 사용하여 표현식을 작성하면 됩니다. 필드 이름에 다음과 같은 문자가 있을 경우 오류가 발생할 수 있습니다. • ! "# $% & '() = ~ |-^ ¥ @" "+ *" "<>? . ,/ :; → (화살표), □ △(도형 부호 등)

(13) 찾기 및 바꾸기

찾기/바꾸기 대화 상자는 스크립트 또는 표현식 텍스트를 편집하는 위치(스크립트 편집기, CLEM 표현식 작성기 포함)에서, 또는 보고서 노드에서 템플리트를 정의할 때 사용할 수 있습니다. 이러한 영역에서 편집하는 경우 Ctrl+F를 눌러 이 대화 상자에 액세스하고 커서가 텍스트 영역에서 초점을 갖고 있는지 확인하십시오. 예를 들어, 채움 노드에서 작업하는 경우 설정 탭의 텍스트 영역에서 또는 표현식 작성기의 텍스트 필드에서 대화 상자에 액세스할 수 있습니다.

1. 텍스트 영역에 커서를 두고 Ctrl+F를 눌러서 찾기/바꾸기 대화 상자에 액세스하십시오.
2. 검색할 텍스트를 입력하거나 최근에 검색한 항목 드롭 다운 목록에서 선택하십시오.
3. 대체 텍스트(있는 경우)를 입력하십시오.
4. **다음 찾기**를 클릭하여 검색을 시작하십시오.
5. **바꾸기**를 클릭하여 현재 선택을 바꾸거나 **모두 바꾸기**를 클릭하여 선택된 또는 모든 인스턴스를 업데이트하십시오.
6. 각 작업 후에 대화 상자가 닫힙니다. 텍스트 영역에서 F3을 눌러 마지막 찾기 작업을 반복하거나 Ctrl+F를 눌러 대화 상자에 다시 액세스하십시오.

검색 옵션

대소문자 구분. 찾기 작업이 대소문자를 구분하는지 여부를 지정합니다(예: *myvar*이 *myVar*과 매치하는지 여부). 대체 텍스트는 이 설정에 관계없이 항상 입력된 대로 정확하게 삽입됩니다.

전체 단어만. 찾기 작업이 단어 안에 임베드된 텍스트와 매치하는지 여부를 지정합니다. 선택된 경우, 예를 들어 *spider*에 대한 검색은 *spiderman* 또는 *spider-man*과 매치하지 않습니다.

정규식. 정규식 구문 사용 여부를 지정합니다(다음 절 참조). 선택된 경우에는 **전체 단어**만 옵션을 사용할 수 없으며 해당 값을 무시합니다.

선택된 텍스트만. **모두 바꾸기** 옵션 사용 시 검색 범위를 제어합니다.

정규식 구문

정규식을 사용하여 탭이나 줄 바꾸기 문자와 같은 특수 문자, a - d와 같은 문자 클래스나 범위, 숫자나 비숫자, 행 시작이나 끝과 같은 경계에 대한 검색을 수행할 수 있습니다. 지원되는 표현식 유형은 다음과 같습니다.

표 1. 문자 매치

문자	매치
x	x 문자

문자	매치
₩₩	백슬래시 문자
₩0n	8진 값 0n이 포함된 문자(0 <= n <= 7)
₩0nn	8진 값 0nn이 포함된 문자(0 <= n <= 7)
₩0mnn	8진 값 0mnn이 포함된 문자(0 <= m <= 3, 0 <= n <= 7)
₩xhh	16진값 0xhh가 포함된 문자
₩uhhhh	16진값 0xhhhh가 포함된 문자
₩t	탭 문자('₩u0009')
₩n	줄 바꾸기(줄 바꾸기) 문자('₩u000A')
₩r	캐리지 리턴 문자('₩u000D')
₩f	용지 넘김 문자('₩u000C')
₩a	경고(벨) 문자('₩u0007')
₩e	이스케이프 문자('₩u001B')
₩cx	x에 해당되는 제어 문자

표 2. 문자 클래스 매치

문자 클래스	매치
[abc]	a, b 또는 c(단순 클래스)
[^abc]	a, b 또는 c를 제외한 모든 문자(뺄셈)
[a-zA-Z]	a - z 또는 A - Z(경계값 포함)
[a-d[m-p]]	a - d 또는 m - p(합집합). 또는 이를 [a-dm-p]로 지정할 수 있음
[a-z&&[def]]	a - z, d, e 또는 f(교집합)
[a-z&&[^bc]]	b와 c를 제외하고 a - z(뺄셈). 또는 이를 [ad-z]로 지정할 수 있음
[a-z&&[^m-p]]	m - p를 제외하고 a - z(뺄셈). 또는 이를 [a-lq-z]로 지정할 수 있음

표 3. 사전 정의된 문자 클래스

사전 정의된 문자 클래스	매치
.	모든 문자(행 종결자와 매치하거나 매치하지 않을 수 있음)
₩d	모든 숫자: [0-9]

사전 정의된 문자 클래스	매치
₩D	비숫자: [^0-9]
₩s	공백 문자: [₩t₩n₩x0B₩f₩r]
₩S	비공백 문자: [^₩s]
₩w	단어 문자: [a-zA-Z_0-9]
₩W	비단어 문자: [^₩w]

표 4. 경계 매치

경계 매치기	매치
^	행의 시작
$	행의 끝
₩b	단어 경계
₩B	비단어 경계
₩A	입력의 시작
₩Z	입력의 끝이지만 최종 종결자로도 사용(해당하는 경우)
₩z	입력의 끝

9) CLEM 언어 참조

(1) CLEM 참조 개요

이 절에서는 IBM® SPSS® Modeler 스트림에서 사용하는 데이터를 분석하고 조작하는 데 사용하는 강력한 도구인 Control Language for Expression Manipulation(CLEM)에 대해 설명합니다. 노드 내에서 CLEM을 사용하여 조건 평가 또는 값 파생에서부터 보고서에 데이터 삽입에 이르기까지 다양한 작업을 수행할 수 있습니다.

CLEM 표현식은 값, 필드 이름, 연산자, 함수로 구성됩니다. 올바른 구문을 사용하면 강력한 데이터 연산을 다양하게 작성할 수 있습니다.

(2) CLEM 데이터 유형

CLEM 데이터 유형은 다음으로 구성될 수 있습니다.

- 정수
- 실수
- 문자
- 문자열
- 목록
- 필드
- 날짜/시간

인용 규칙

CLEM 표현식에서 사용하는 필드, 값, 모수, 문자열을 판별할 때 IBM® SPSS® Modeler가 유연하긴 하지만 다음 일반 규칙은 표현식 작성 시 사용할 "우수 사례" 목록을 제공합니다.

- 문자열 - 문자열 작성 시 항상 큰따옴표를 사용하십시오(예: "Type 2"). 작은따옴표를 대신 사용할 수 있지만 인용된 필드와 혼동될 위험이 있습니다.
- 필드 - 공백 또는 기타 특수 문자를 묶는 데 필요한 경우에만 작은따옴표를 사용하십시오(예: 'Order Number'). 인용되었지만 데이터 세트에 정의되지 않은 필드는 문자열로 잘못 읽힙니다.
- 모수 - 모수 사용 시 항상 작은따옴표를 사용하십시오(예: '$P-threshold').
- 문자 - 항상 단일 역따옴표(`)를 사용하십시오(예: stripchar(`d`, "drugA")).

이러한 규칙에 대해서는 다음 주제에서 자세히 설명합니다.

① 정수

정수는 십진 숫자의 시퀀스로 표시됩니다. 또는 정수 앞에 음의 부호(−)를 두면 음수를 표시할 수 있습니다(예: 1234, 999, −77).

CLEM 언어는 임의의 정밀도의 정수를 처리합니다. 최대 정수 크기는 플랫폼에 따라 다릅니다. 값이 정수 필드에 표시하기에 너무 크면 일반적으로 필드 유형을 Real로 변경하면 값이 복원됩니다.

② 실수

*실수*는 부동 소수점을 가리킵니다. 실수는 하나 이상의 숫자 뒤에 소수점과 하나 이상의 숫자로 표시됩니다. CLEM 실수는 이중 정밀도를 가집니다.
또는 실수 앞에 음의 부호(−)를 두고 음수를 표시할 수도 있습니다(예: 1.234, 0.999, −77.001). *<number>* e *<exponent>* 양식을 사용하여 실수를 지수 표기법으로 표시하십시오(예: 1234.0e5, 1.7e−2). IBM® SPSS® Modeler 애플리케이션이 파일에서 숫자 문자열을 읽고 이를 숫자로 자동으로 변환하면 소수점 앞에 선행하는 숫자가 없는 숫자나 소수점 뒤에 숫자가 없는 숫자도 허용됩니다(예: 999. 또는 .11). 그러나 이러한 양식은 CLEM 표현식에서는 올바르지 않습니다.

참고: CLEM 표현식에서 실수를 참조할 때는 현재 스트림이나 로케일의 설정과 관계없이 마침표를 소수점 구분자로 사용해야 합니다. 예를 들어, 다음을 지정하십시오.

```
Na > 0.6
```

다음이 아님

```
Na > 0,6
```

이는 스트림 특성 대화 상자에서 쉼표가 소수점 기호로 선택되었더라도 적용되며 코드 구문이 특정 로케일이나 규칙으로부터 독립적이어야 하는 일반적인 지침과도 매치합니다.

③ 문자

문자(대개 CHAR로 표시됨)는 일반적으로 CLEM 표현식 내에서 문자열에 대한 검정을 수행하는데 사용합니다. 예를 들어, isuppercode 함수를 사용하여 문자열의 첫 번째 문자가 대문자인지 여부를 판별할 수 있습니다. 다음 CLEM 표현식은 문자를 사용하여 문자열의 첫 번째 문자에 대해 검정을 수행해야 함을 표시합니다.

```
isuppercode(subscrs(1, "MyString"))
```

CLEM 표현식에서 특정 문자의 코드(위치와 대조)를 표시하려면, `*<character>*` 양식의 단일 역따옴표(예: `A`, `Z`)를 사용하십시오.

참고: 필드의 경우 CHAR 저장 유형이 없으므로 필드가 CHAR을 초래하는 표현식으로 채워지거나 파생된 경우 해당 결과는 문자열로 변환됩니다.

④ 문자열

일반적으로 문자열을 큰따옴표로 묶어야 합니다. 문자열의 예는 "c35product2" 및 "referrerID" 입니다. 문자열에서 특수 문자를 표시하려면 백슬래시를 사용하십시오(예: "₩$65443"). (백슬래시 문자를 표시하려면 이중 백슬래시, ₩₩를 사용하십시오.) 문자열을 작은따옴표로 묶을 수도 있지만 결과는 인용된 필드('referrerID')와 구별이 되지 않습니다. 자세한 정보는 문자열 함수 주제를 참조하십시오.

⑤ 목록

목록은 요소의 순서화된 시퀀스이며 이는 혼합 유형일 수도 있습니다. 목록은 꺽쇠 대괄호([])로 싸여 있습니다. 목록의 예제는 [1 2 4 16] 및 ["abc" "def"]입니다. 목록은 IBM® SPSS®

Modeler 필드의 값으로 사용하지 않습니다. 이들은 member 및 oneof 등과 같이 함수에 인수를 제공하는 데 사용합니다.

> ⓘ **참고**: 목록은 함수 호출이 아니라 정적 오브젝트(예: 문자열, 번호 또는 필드 이름)로만 구성될 수 있습니다.

⑥ 필드

함수의 이름이 아닌 CLEM 표현식에서의 이름은 필드 이름으로 가정됩니다. 이들을 단순하게 Power, val27, state_flag 등으로 작성할 수 있지만, 이름이 숫자로 시작하거나 공백 등과 같이 영문자가 아닌 문자를 포함하면(밑줄은 예외) 이름을 작은따옴표 안에 두십시오(예: 'Power Increase', '2nd answer', '#101', '$P-NextField').

참고: 인용되지만 데이터 세트에 정의되어 있지 않은 필드는 문자열로 잘못 읽힙니다.

⑦ 날짜

날짜 계산은 스트림 특성 대화 상자에 지정된 "기준" 날짜를 기반으로 합니다. 기본 기준 날짜는 1900년 1월 1일입니다.

CLEM 언어가 지원하는 날짜 형식은 다음과 같습니다.

표 1. CLEM 언어 날짜 형식

형식	예
DDMMYY	150163
MMDDYY	011563
YYMMDD	630115
YYYYMMDD	19630115
YYYYDDD	연도의 일을 나타내는 3자리 숫자가 뒤에 오는 4자리 연도(예를 들어, 2000032는 2000년의 32번째 날 또는 2000년 2월 1일을 나타냄)
DAY	현재 로케일로 된 요일(예: 영어로 Monday, Tuesday, ...)
MONTH	현재 로케일로 된 월(예: January, February, …)
DD/MM/YY	15/01/63
DD/MM/YYYY	15/01/1963

형식	예
MM/DD/YY	01/15/63
MM/DD/YYYY	01/15/1963
DD-MM-YY	15-01-63
DD-MM-YYYY	15-01-1963
MM-DD-YY	01-15-63
MM-DD-YYYY	01-15-1963
DD.MM.YY	15.01.63
DD.MM.YYYY	15.01.1963
MM.DD.YY	01.15.63
MM.DD.YYYY	01.15.1963
DD-MON-YY	15-JAN-63, 15-jan-63, 15-Jan-63
DD/MON/YY	15/JAN/63, 15/jan/63, 15/Jan/63
DD.MON.YY	15.JAN.63, 15.jan.63, 15.Jan.63
DD-MON-YYYY	15-JAN-1963, 15-jan-1963, 15-Jan-1963
DD/MON/YYYY	15/JAN/1963, 15/jan/1963, 15/Jan/1963
DD.MON.YYYY	15.JAN.1963, 15.jan.1963, 15.Jan.1963
MON YYYY	Jan 2004
q Q YYYY	문자 Q와 4자리 연도가 뒤에 오는 분기를 나타내는 숫자(1 - 4)로 표시된 날짜(예: 2004년 12월 25일은 4 Q 2004로 표시됨)
ww WK YYYY	문자 WK와 4자리 연도가 뒤에 오는 연간 통산주를 나타내는 2자리 숫자. 연간 통산주는 첫 번째 요일이 월요일이고 첫 번째 주에 최소 1일이 있다고 가정하고 계산됩니다.

⑧ 시간

CLEM 언어는 다음 시간 형식을 지원합니다.

표 1. CLEM 언어 시간 형식

형식	예
HHMMSS	120112, 010101, 221212
HHMM	1223, 0745, 2207

형식	예
MMSS	5558, 0100
HH:MM:SS	12:01:12, 01:01:01, 22:12:12
HH:MM	12:23, 07:45, 22:07
MM:SS	55:58, 01:00
(H)H:(M)M:(S)S	12:1:12, 1:1:1, 22:12:12
(H)H:(M)M	12:23, 7:45, 22:7
(M)M:(S)S	55:58, 1:0
HH.MM.SS	12.01.12, 01.01.01, 22.12.12
HH.MM	12.23, 07.45, 22.07
MM.SS	55.58, 01.00
(H)H.(M)M.(S)S	12.1.12, 1.1.1, 22.12.12
(H)H.(M)M	12.23, 7.45, 22.7
(M)M.(S)S	55.58, 1.0

(3) CLEM 연산자

다음 연산자가 사용 가능합니다.

표 1. CLEM 언어 연산자

연산	설명	우선순위 (다음 섹션 참조)
or	두 개의 CLEM 표현식 사이에 사용합니다. 둘 중 하나라도 참이거나 둘 모두가 참이면 참 값을 리턴합니다.	10
and	두 개의 CLEM 표현식 사이에 사용합니다. 둘 모두 참이면 참 값을 리턴합니다.	9
=	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 ITEM2와 동일하면 참을 리턴합니다.	7
=	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 ITEM2와 동일하면 참을 리턴합니다.	7
==	=과 동일합니다.	7
/=	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 ITEM2와 동일하지 않으면 참을 리턴합니다.	7

연산	설명	우선순위 (다음 섹션 참조)
/==	/=과 동일합니다.	6
>	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 엄밀하게 ITEM2를 초과하면 참을 리턴합니다.	6
>=	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 ITEM2 이상이면 참을 리턴합니다.	6
<	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 엄밀하게 ITEM2 미만이면 참을 리턴합니다.	6
<=	두 개의 비교 가능한 항목 사이에 사용합니다. ITEM1이 ITEM2 이하이면 참을 리턴합니다.	6
&&=_0	두 개의 정수 사이에 사용합니다. 부울 표현식 INT1 && INT2 = 0에 해당됩니다.	6
&&/=_0	두 개의 정수 사이에 사용합니다. 부울 표현식 INT1 && INT2 /= 0에 해당됩니다.	6
+	두 개의 수를 더합니다. NUM1 + NUM2	5
><	2개의 문자열을 연결합니다. 예를 들면, STRING1 >< STRING2.	5
-	하나의 수를 다른 하나에서 뺍니다. NUM1 - NUM2. 또한 숫자의 앞에 사용할 수도 있습니다. - NUM	6
*	2개의 수를 곱하는 데 사용합니다. NUM1 * NUM2	4
&&	두 개의 정수 사이에 사용합니다. 결과는 정수 INT1과 INT2의 bitwise 'and'입니다.	4
&&~~	두 개의 정수 사이에 사용합니다. 결과는 INT1의 bitwise 'and'와 INT2의 bitwise 보수입니다.	4
||	두 개의 정수 사이에 사용합니다. 결과는 INT1과 INT2의 bitwise 'inclusive or'입니다.	4
~~	정수 앞에 사용합니다. INT의 bitwise 보수를 생성합니다.	4
||/&	두 개의 정수 사이에 사용합니다. 결과는 INT1과 INT2의 bitwise 'exclusive or'입니다.	4
INT1 << N	두 개의 정수 사이에 사용합니다. N 위치만큼 왼쪽으로 이동된 INT의 비트 패턴을 생성합니다.	4
INT1 >> N	두 개의 정수 사이에 사용합니다. N 위치만큼 오른쪽으로 이동된 INT의 비트 패턴을 생성합니다.	4
/	한 수를 다른 수로 나누는 데 사용합니다. NUM1 / NUM2	4

연산	설명	우선순위 (다음 섹션 참조)
**	두 개의 수 사이에 사용합니다. BASE ** POWER. POWER로 거듭제곱된 BASE를 리턴합니다.	3
rem	두 개의 정수 사이에 사용합니다. INT1 rem INT2. 나머지를 리턴합니다. INT1 - (INT1 div INT2) * INT2	2
div	두 개의 정수 사이에 사용합니다. INT1 div INT2. 정수 나눗셈을 수행합니다.	2

연산자 우선순위

우선순위는 특히 둘 이상의 인픽스 연산자가 있는 비브래킷 표현식과 같은 복합 표현식의 구문 분석을 결정합니다. 예를 들어,

```
3 + 4 * 5
```

상대적인 우선순위에서 *가 +보다 먼저 구문 분석된다고 지시되므로 (3 + 4) * 5가 아니라 3 + (4 * 5)로 구문 분석됩니다. CLEM 언어의 모든 연산자에는 이와 연관된 우선순위 값이 있습니다. 이 값이 낮을 수록 구문 분석 목록에서 더 중요합니다. 즉, 우선순위 값이 높은 다른 연산자보다 먼저 처리된다는 의미입니다.

(4) 함수 참조

IBM® SPSS® Modeler에서 데이터에 대한 작업에 사용할 수 있는 CLEM 함수는 다음과 같습니다. 이러한 함수를 여러 대화 상자(예: 파생 및 플래그로 설정 노드)에서 코드로 입력하거나 표현식 작성기를 사용하여 함수 목록이나 필드 이름을 암기하지 않고 유효한 CLEM 표현식을 작성할 수 있습니다.

표 1. IBM SPSS Modeler 데이터에 사용할 CLEM 함수

함수 유형	설명
정보	필드 값을 간파하는 데 사용합니다. 예를 들어, is_string 함수는 유형이 문자열인 모든 레코드에 대해 참을 리턴합니다.
변환	새 필드를 구성하거나 저장 유형을 변환하는 데 사용합니다. 예를 들어, to_timestamp 함수는 선택된 필드를 시간소인으로 변환합니다.
비교	필드 값을 서로 비교하거나 지정된 문자열과 비교하는 데 사용합니다. 예를 들어, <=은 두 필드의 값이 작거나 같은지 여부를 비교하는 데 사용합니다.

함수 유형	설명
논리	논리 연산(예: if, then, else 연산)을 수행하는 데 사용합니다.
숫자	수치 연산(예: 필드 값의 자연 로그)을 수행하는 데 사용합니다.
삼각	삼각 연산(예: 지정된 각의 아크코사인)을 수행하는 데 사용합니다.
확률	다양한 분포를 기반으로 하는 확률(예: 스튜던트의 t 분포 값이 지정된 값보다 작은 확률)을 리턴합니다.
공간	지리 공간 데이터에 대한 공간 연산을 수행하는 데 사용합니다.
Bitwise	정수를 비트 패턴으로 조작하는 데 사용합니다.
Random	임의로 항목을 선택하거나 숫자를 생성하는 데 사용합니다.
문자열	문자열에 대한 다양한 연산(예: 지정된 문자를 제거할 수 있게 하는 stripchar)을 수행하는 데 사용합니다.
SoundEx	정확한 맞춤법을 알 수 없는 경우 문자열을 찾는 데 사용되며, 특정 문자가 발음되는 방법에 대한 음성 가정을 기반으로 합니다.
날짜 및 시간	날짜, 시간 및 시간소인 필드에 대한 다양한 연산을 수행하는 데 사용합니다.
시퀀스	데이터 세트의 레코드 시퀀스를 간파하거나 해당 시퀀스를 기반으로 하는 연산을 수행하는 데 사용합니다.
글로벌	전역값 설정 노드에 의해 작성된 글로벌 값에 액세스하는 데 사용합니다. 예를 들어, @MEAN은 전체 데이터 세트에서 필드의 모든 값의 평균을 참조하는 데 사용합니다.
공백 및 널	사용자 지정 공백 또는 시스템 결측값에 액세스하고 플래그를 지정하며 자주 채우는 데 사용합니다. 예를 들어, @BLANK(FIELD)는 공백이 있는 레코드에 true 플래그를 지정하는 데 사용합니다.
특수 필드	검사 중인 특정 필드를 표시하는 데 사용합니다. 예를 들어, @FIELD는 다중 필드 파생 시 사용합니다.

① 함수 설명 규칙

함수의 항목을 참조할 때 이 안내서 전반에서 사용하는 규칙은 다음과 같습니다.

표 1. 함수 설명 규칙

규칙	설명
BOOL	부울 또는 플래그(예: true 또는 false)입니다.

규칙	설명
NUM, NUM1, NUM2	숫자입니다.
REAL, REAL1, REAL2	실수입니다(예: 1.234 또는 -77.01).
INT, INT1, INT2	정수입니다(예: 1 또는 -77).
CHAR	문자 코드입니다(예: `A`).
STRING	문자열입니다(예: "referrerID").
LIST	항목 목록입니다(예: ["abc" "def"]).
ITEM	필드입니다(예: Customer 또는 extract_concept).
DATE	start_date와 같은 날짜 필드입니다(여기서 값은 DD-MON-YYYY와 같은 형식임).
TIME	power_flux와 같은 시간 필드입니다(여기서 값은 HHMMSS와 같은 형식임).

이 안내서에서 함수는 한 열에는 함수, 다른 열에는 결과 유형(정수, 문자열 등), 세 번째 열에는 설명(사용 가능한 경우)과 함께 나열되어 있습니다. 예를 들어, rem 함수에 대한 설명은 다음과 같습니다.

표 2. rem 함수 설명

기능	결과	설명
INT1 rem INT2	*숫자*	*INT1*을 *INT2*로 나눈 나머지를 리턴합니다 (예: INT1 - (INT1 div INT2) * INT2).

항목 나열 또는 함수의 문자 지정과 같은 사용 규칙에 대한 세부사항은 다른 곳에 설명되어 있습니다. 자세한 정보는 CLEM 데이터 유형의 내용을 참조하십시오.

② 정보 함수

정보 함수는 특정 필드 값을 간파하는 데 사용합니다. 일반적으로 플래그 필드를 파생시키는 데 사용합니다. 예를 들어, @BLANK 함수를 사용하여 선택된 필드에 대해 값이 공백인 레코드를 표시하는 플래그 필드를 작성할 수 있습니다. 마찬가지로 is_string과 같은 저장 유형 함수를 사용하여 필드의 저장 유형을 확인할 수 있습니다.

표 1. CLEM 정보 함수

기능	결과	설명
@BLANK(FIELD)	*Boolean*	업스트림 유형 노드 또는 소스 노드(유형 탭)에 설정된 공백 처리 규칙에 따라 값이 공백인 모든 레코드에 대해 참을 리턴합니다.
@NULL(ITEM)	*Boolean*	값이 정의되지 않은 모든 레코드에 대해 참을 리턴합니다. 정의되지 않은 값은 시스템 널값입니다(IBM® SPSS® Modeler에 $null$로 표시됨).
is_date(ITEM)	*Boolean*	유형이 날짜인 모든 레코드에 대해 참을 리턴합니다.
is_datetime(ITEM)	*Boolean*	유형이 날짜, 시간 또는 시간소인인 모든 레코드에 대해 참을 리턴합니다.
is_integer(ITEM)	*Boolean*	유형이 정수인 모든 레코드에 대해 참을 리턴합니다.
is_number(ITEM)	*Boolean*	유형이 숫자인 모든 레코드에 대해 참을 리턴합니다.
is_real(ITEM)	*Boolean*	유형이 실수인 모든 레코드에 대해 참을 리턴합니다.
is_string(ITEM)	*Boolean*	유형이 문자열인 모든 레코드에 대해 참을 리턴합니다.
is_time(ITEM)	*Boolean*	유형이 시간인 모든 레코드에 대해 참을 리턴합니다.
is_timestamp(ITEM)	*Boolean*	유형이 시간소인인 모든 레코드에 대해 참을 리턴합니다.

③ 변환 함수

변환 함수를 사용하여 새 필드를 구성하고 기존 파일의 저장 유형을 변환할 수 있습니다. 예를 들어, 문자열을 함께 결합하거나 문자열을 분해하여 새 문자열을 형성할 수 있습니다. 두 개의 문자열을 결합하려면 >< 연산자를 사용하십시오. 예를 들어, Site 필드 값이 "BRAMLEY"인 경우 "xx" >< Site는 "xxBRAMLEY"를 리턴합니다. 인수가 문자열이 아니더라도 ><의 결과는 항상 문자열입니다. 따라서 V1 필드가 3이고 V2 필드가 5인 경우 V1 >< V2는 "35"(숫자가 아닌 문자열)를 리턴합니다.

변환 함수(및 날짜 또는 시간 값과 같은 특정 유형의 입력을 필요로 하는 기타 함수)는 스트림 옵션 대화 상자에 지정된 현재 형식에 따라 다릅니다. 예를 들어, 값이 Jan 2003, Feb 2003 등인 문자열 필드를 변환하려면 매치하는 날짜 형식 **MON YYYY**를 스트림의 기본 날짜 형식으로 선택하십시오.

표 1. CLEM 변환 함수

기능	결과	설명
ITEM1 >< ITEM2	문자열	두 필드의 값을 연결하고 결과 문자열을 ITEM1ITEM2로 리턴합니다.
to_integer(ITEM)	Integer	지정된 필드의 저장 공간을 정수로 변환합니다.
to_real(ITEM)	Real	지정된 필드의 저장 공간을 실수로 변환합니다.
to_number(ITEM)	숫자	지정된 필드의 저장 공간을 문자열로 변환합니다. 이 함수를 사용하여 실수가 문자열로 변환되면 소수점 뒤 6자리의 값을 리턴합니다.
to_time(ITEM)	시간	지정된 필드의 저장 공간을 시간으로 변환합니다.
to_date(ITEM)	날짜	지정된 필드의 저장 공간을 날짜로 변환합니다.
to_timestamp(ITEM)	시간소인	지정된 필드의 저장 공간을 시간소인으로 변환합니다.
to_datetime(ITEM)	날짜/시간	지정된 필드의 저장 공간을 날짜, 시간 또는 시간소인 값으로 변환합니다.
datetime_date(ITEM)	날짜	숫자, 문자열 또는 시간소인의 날짜 값을 리턴합니다. 이는 숫자(초)를 날짜로 다시 변환할 수 있게 하는 유일한 함수입니다. ITEM이 문자열인 경우 현재 날짜 형식으로 문자열을 구문 분석하여 날짜를 생성합니다. 이 함수가 성공하려면 스트림 특성 대화 상자에 지정된 날짜 형식이 올발라야 합니다. ITEM이 숫자인 경우 기준 날짜(또는 에포크) 이후의 초 단위로 해석됩니다. 1일의 분수는 잘립니다. ITEM이 시간소인인 경우 시간소인의 날짜 부분이 리턴됩니다. ITEM이 날짜인 경우 변경되지 않고 리턴됩니다.
stb_centroid_latitude (ITEM)	Integer	Geohash 인수의 중심값에 해당하는 위도의 정수 값을 리턴합니다.
stb_centroid_longitude (ITEM)	Integer	Geohash 인수의 중심값에 해당하는 경도의 정수 값을 리턴합니다.
to_geohash(ITEM)	문자열	밀도에 대해 지정된 비트 수를 사용하여 위도 및 경도에 해당하는 Geohash 문자열을 리턴합니다. Geohash는 위도 및 경도 세부사항을 기반으로 지리 좌표 세트를 식별하는 데 사용하는 코드입니다. to_geohash에 대한 세 모수는 다음과 같습니다. - 위도: (-180, 180) 범위이고, 단위는 WGS84 좌표계에서 도입니다. - 경도: (-90, 90) 범위이고, 단위는 WGS84 좌표계에서 도입니다. - 비트: 해시 저장에 사용할 비트 수입니다. 범위는 [1,75]입니다. 리턴된 문자열의 길이(5비트당 1자가 사용됨)와 해시 정확도에 모두 영향을 미칩니다. 예를 들어 5비트(1자)는 약 2500킬로미터를 나타내고, 45비트(9자)는 약 2.3미터를 나타냅니다.

④ 비교 함수

비교 함수는 필드 값을 서로 비교하거나 지정된 문자열과 비교하는 데 사용합니다. 예를 들어, =을 사용하여 문자열이 같은지 확인할 수 있습니다. 문자열이 같은지 확인하는 예제는 Class = "class 1"입니다.

수치 비교를 위해 greater는 양의 무한대에 가까움을 의미하고 lesser는 음의 무한대에 가까움을 의미합니다. 즉, 모든 음수는 임의의 양수보다 작습니다.

표 1. CLEM 비교 함수

기능	결과	설명
count_equal (ITEM1, LIST)	*Integer*	*ITEM1*과 같은 필드 목록의 값 수를 리턴하거나 *ITEM1*이 널인 경우 널을 리턴합니다.
count_greater_than (ITEM1, LIST)	*Integer*	*ITEM1*보다 큰 필드 목록의 값 수를 리턴하거나 *ITEM1*이 널인 경우 널을 리턴합니다.
count_less_than (ITEM1, LIST)	*Integer*	*ITEM1*보다 작은 필드 목록의 값 수를 리턴하거나 *ITEM1*이 널인 경우 널을 리턴합니다.
count_not_equal (ITEM1, LIST)	*Integer*	*ITEM1*과 같지 않은 필드 목록의 값 수를 리턴하거나 *ITEM1*이 널인 경우 널을 리턴합니다.
count_nulls(LIST)	*Integer*	필드 목록의 널값 수를 리턴합니다.
count_non_nulls(LIST)	*Integer*	필드 목록의 널이 아닌 값 수를 리턴합니다.
date_before (DATE1, DATE2)	*Boolean*	날짜 값 순서를 확인하는 데 사용합니다. *DATE1*이 *DATE2* 이전이면 참 값을 리턴합니다.
first_index(ITEM, LIST)	*Integer*	필드 LIST에서 ITEM이 들어 있는 첫 번째 필드의 지수를 리턴하거나 값을 찾을 수 없으면 0을 리턴합니다. 문자열, 정수 및 실수 유형에만 지원됩니다.
first_non_null(LIST)	*모두*	제공된 필드 목록에서 널이 아닌 첫 번째 값을 리턴합니다. 모든 저장 유형이 지원됩니다.
first_non_null_index (LIST)	*Integer*	지정된 LIST에서 널이 아닌 값이 들어 있는 첫 번째 필드의 지수를 리턴하거나 모든 값이 널인 경우 0을 리턴합니다. 모든 저장 유형이 지원됩니다.
ITEM1 = ITEM2	*Boolean*	*ITEM1*이 *ITEM2*와 같은 레코드에 대해 참을 리턴합니다.
ITEM1 /= ITEM2	*Boolean*	두 문자열이 동일하지 않으면 참을 리턴하거나 동일하면 0을 리턴합니다.
ITEM1 < ITEM2	*Boolean*	*ITEM1*이 *ITEM2*보다 작은 레코드에 대해 참을 리턴합니다.
ITEM1 <= ITEM2	*Boolean*	*ITEM1*이 *ITEM2*보다 작거나 같은 레코드에 대해 참을 리턴합니다.

기능	결과	설명
ITEM1 > ITEM2	*Boolean*	*ITEM1*이 *ITEM2*보다 큰 레코드에 대해 참을 리턴합니다.
ITEM1 >= ITEM2	*Boolean*	*ITEM1*이 *ITEM2*보다 크거나 같은 레코드에 대해 참을 리턴합니다.
last_index(ITEM, LIST)	*Integer*	필드 LIST에서 ITEM이 들어 있는 마지막 필드의 지수를 리턴하거나 값을 찾을 수 없으면 0을 리턴합니다. 문자열, 정수 및 실수 유형에만 지원됩니다.
last_non_null(LIST)	*모두*	제공된 필드 목록에서 널이 아닌 마지막 값을 리턴합니다. 모든 저장 유형이 지원됩니다.
last_non_null_index (LIST)	*Integer*	지정된 LIST에서 널이 아닌 값이 들어 있는 마지막 필드의 지수를 리턴하거나 모든 값이 널인 경우 0을 리턴합니다. 모든 저장 유형이 지원됩니다.
max(ITEM1, ITEM2)	*모두*	두 개의 항목 중 더 큰 항목을 리턴합니다(*ITEM1* 또는 *ITEM2*).
max_index(LIST)	*Integer*	숫자 필드 목록에서 최대값이 들어 있는 필드의 지수를 리턴하거나 모든 값이 널인 경우 0을 리턴합니다. 예를 들어, 나열된 세 번째 필드에 최대값이 포함된 경우 지수 값 3이 리턴됩니다. 여러 필드에 최대값이 포함된 경우 첫 번째로 나열된 (맨 왼쪽) 필드가 리턴됩니다.
max_n(LIST)	*숫자*	숫자 필드 목록의 최대값을 리턴하거나 모든 필드 값이 널인 경우 널을 리턴합니다.
member(ITEM, LIST)	*Boolean*	*ITEM*이 지정된 *LIST*의 멤버이면 참을 리턴합니다. 그렇지 않으면, 거짓 값이 리턴됩니다. 필드 이름 목록도 지정할 수 있습니다.
min(ITEM1, ITEM2)	*모두*	두 개의 항목 중 더 작은 항목을 리턴합니다(*ITEM1* 또는 *ITEM2*).
min_index(LIST)	*Integer*	숫자 필드 목록에서 최소값이 들어 있는 필드의 지수를 리턴하거나 모든 값이 널인 경우 0을 리턴합니다. 예를 들어, 나열된 세 번째 필드에 최소값이 포함된 경우 지수 값 3이 리턴됩니다. 여러 필드에 최소값이 포함된 경우 첫 번째로 나열된 (맨 왼쪽) 필드가 리턴됩니다.
min_n(LIST)	*숫자*	숫자 필드 목록의 최소값을 리턴하거나 모든 필드 값이 널인 경우 널을 리턴합니다.
time_before (TIME1, TIME2)	*Boolean*	시간 값 순서를 확인하는 데 사용합니다. TIME1이 TIME2 이전이면 참 값을 리턴합니다.
value_at(INT, LIST)		오프셋 INT에서 나열된 각 필드의 값을 리턴하거나 오프셋이 유효한 값 범위를 벗어난 경우(즉, 1보다 작거나 나열된 필드 수보다 큼) 널(NULL)을 리턴합니다. 모든 저장 유형이 지원됩니다.

⑤ 논리 함수

CLEM 표현식을 사용하여 논리 연산을 수행할 수 있습니다.

표 1. CLEM 논리 함수

기능	결과	설명
COND1 and COND2	*Boolean*	이 연산은 논리곱이며 *COND1*과 *COND2*가 둘 다 참이면 참 값을 리턴합니다. *COND1*이 거짓이면 *COND2*는 평가되지 않습니다. 따라서 *COND1*이 *COND2*의 연산이 올바른지 처음으로 검정하는 경우 논리곱을 가질 수 있습니다 (예: length(Label) >=6 및 Label(6) = 'x').
COND1 or COND2	*Boolean*	이 연산은 논리합(포함적 분리)이며 *COND1* 또는 *COND2*가 참이거나 둘 다 참이면 참 값을 리턴합니다. *COND1*이 참이면 *COND2*가 평가되지 않습니다.
not(COND)	*Boolean*	이 연산은 논리 부정이며 *COND*가 거짓이면 참 값을 리턴합니다. 그렇지 않으면 이 연산은 0 값을 리턴합니다.
if COND then EXPR1 else EXPR2 endif	*모두*	이 연산은 조건부 평가입니다. *COND*가 참인 경우 이 연산은 *EXPR1*의 결과를 리턴합니다. 그렇지 않으면 *EXPR2* 평가 결과가 리턴됩니다.
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	*모두*	이 연산은 다중 분기 조건부 평가입니다. *COND1*이 참인 경우 이 연산은 *EXPR1*의 결과를 리턴합니다. 그렇지 않으면 *COND2*가 참인 경우 이 연산은 *EXPR2* 평가 결과를 리턴합니다. 그렇지 않으면 *EXPR_N* 평가 결과가 리턴됩니다.

⑥ 숫자 함수

CLEM에는 일반적으로 사용하는 다수의 숫자 함수가 포함되어 있습니다.

표 1. CLEM 숫자 함수

기능	결과	설명
-NUM	*숫자*	*NUM*을 부정하는 데 사용합니다. 반대 부호를 가진 해당 숫자를 리턴합니다.
NUM1 + NUM2	*숫자*	*NUM1*과 *NUM2*의 합계를 리턴합니다.
NUM1 -NUM2	*숫자*	*NUM1*에서 *NUM2*를 뺀 값을 리턴합니다.
NUM1 * NUM2	*숫자*	*NUM1*에 *NUM2*를 곱한 값을 리턴합니다.
NUM1 / NUM2	*숫자*	*NUM1*을 *NUM2*로 나눈 값을 리턴합니다.

기능	결과	설명
INT1 div INT2	*숫자*	정수 나눗셈을 수행하는 데 사용합니다. *INT1*을 *INT2*로 나눈 값을 리턴합니다.
INT1 rem INT2	*숫자*	*INT1*을 *INT2*로 나눈 나머지를 리턴합니다(예: INT1 - (INT1 div INT2) * INT2).
INT1 mod INT2	*숫자*	이 함수는 더 이상 사용되지 않습니다. rem 함수를 대신 사용하십시오.
BASE ** POWER	*숫자*	거듭제곱 *POWER*로 제곱된 *BASE*를 리턴합니다. 여기서 둘 중 하나는 임의의 숫자입니다(단, *POWER*가 정수 0이 아닌 다른 유형의 0이면 *BASE*가 0이 아니어야 함). *POWER*가 정수이면 연속적으로 *BASE*의 거듭제곱을 곱하여 계산이 수행됩니다. 따라서 *BASE*가 정수이면 결과는 정수입니다. *POWER*가 정수 0이면 결과는 항상 *BASE*와 동일한 유형의 1입니다. 그렇지 않으면 *POWER*가 정수가 아닌 경우 결과는 exp(POWER * log(BASE))로 계산됩니다.
abs(NUM)	*숫자*	항상 같은 유형의 숫자인 *NUM*의 절대값을 리턴합니다.
exp(NUM)	*Real*	거듭제곱 *NUM*으로 제곱된 e를 리턴합니다(여기서 e는 자연로그의 기수임).
fracof(NUM)	*Real*	NUM-intof(NUM)으로 정의된 *NUM*의 분수 부분을 리턴합니다.
intof(NUM)	*Integer*	인수를 정수로 자릅니다. *NUM*과 동일 부호의 정수 및 abs(INT) <= abs(*NUM*)과 같은 가장 큰 크기의 정수를 리턴합니다.
log(NUM)	*Real*	0이 아니어야 하는 *NUM*의 자연(기수 e) 로그를 리턴합니다.
log10(NUM)	*Real*	0이 아니어야 하는 *NUM*의 기수 10 로그를 리턴합니다. 이 함수는 log(NUM) / log(10)으로 정의됩니다.
negate(NUM)	*숫자*	*NUM*을 부정하는 데 사용합니다. 반대 부호를 가진 해당 숫자를 리턴합니다.
round(NUM)	*Integer*	*NUM*이 양수인 경우 intof(NUM+0.5)를 사용하고 *NUM*이 음수인 경우 intof(NUM-0.5)를 사용하여 *NUM*을 정수로 반올림하는 데 사용됩니다.
sign(NUM)	*숫자*	*NUM*의 부호를 판별하는 데 사용합니다. 이 연산은 *NUM*이 정수이면 -1, 0 또는 1을 리턴합니다. *NUM*이 실수이면 *NUM*이 음수인지, 0인지, 양수인지 여부에 따라 -1.0, 0.0 또는 1.0을 리턴합니다.
sqrt(NUM)	*Real*	*NUM*의 제곱근을 리턴합니다. *NUM*은 양수여야 합니다.

기능	결과	설명
sum_n(LIST)	*숫자*	숫자 필드 목록의 합계를 리턴하거나 모든 필드 값이 널인 경우 널을 리턴합니다.
mean_n(LIST)	*숫자*	숫자 필드 목록의 평균 값을 리턴하거나 모든 필드 값이 널인 경우 널을 리턴합니다.
sdev_n(LIST)	*숫자*	숫자 필드 목록의 표준 편차를 리턴하거나 모든 필드 값이 널인 경우 널을 리턴합니다.

⑦ 삼각 함수

이 절의 모든 함수는 각을 인수로 사용하거나 각을 결과로 리턴합니다. 두 경우 모두 관련 스트림 옵션 설정을 통해 각 단위(라디안 또는 도)를 제어합니다.

표 1. CLEM 삼각 함수

기능	결과	설명
arccos(NUM)	*Real*	지정된 각의 아크코사인을 계산합니다.
arccosh(NUM)	*Real*	지정된 각의 쌍곡선 아크코사인을 계산합니다.
arcsin(NUM)	*Real*	지정된 각의 아크사인을 계산합니다.
arcsinh(NUM)	*Real*	지정된 각의 쌍곡선 아크사인을 계산합니다.
arctan(NUM)	*Real*	지정된 각의 아크탄젠트를 계산합니다.
arctan2(NUM_Y, NUM_X)	*Real*	NUM_Y / NUM_X의 아크탄젠트를 계산하며 두 수의 부호를 사용하여 사분면 정보를 도출합니다. 결과는 - pi < ANGLE <= pi (radians) - 180 < ANGLE <= 180 (degrees) 범위의 실수입니다.
arctanh(NUM)	*Real*	지정된 각의 쌍곡아크탄젠트를 계산합니다.
cos(NUM)	*Real*	지정된 각의 코사인을 계산합니다.
cosh(NUM)	*Real*	지정된 각의 쌍곡선 코사인을 계산합니다.
pi	*Real*	이 상수는 파이(pi)에 가장 가까운 값입니다.
sin(NUM)	*Real*	지정된 각의 사인을 계산합니다.
sinh(NUM)	*Real*	지정된 각의 쌍곡선 사인을 계산합니다.
tan(NUM)	*Real*	지정된 각의 탄젠트를 계산합니다.
tanh(NUM)	*Real*	지정된 각의 쌍곡 탄젠트를 계산합니다.

⑧ 확률 함수

확률 함수는 다양한 분포를 기반으로 하는 확률(예: 스튜던트의 t 분포 값이 특정 값보다 작은 확률)을 리턴합니다.

표 1. CLEM 확률 함수

기능	결과	설명
cdf_chisq(NUM, DF)	*Real*	지정된 자유도를 가진 카이제곱 분포 값이 지정된 숫자보다 작을 확률을 리턴합니다.
cdf_f(NUM, DF1, DF2)	*Real*	자유도가 *DF1* 및 *DF2*인 *F* 분포의 값이 지정된 숫자보다 작을 확률을 리턴합니다.
cdf_normal(NUM, MEAN, STDDEV)	*Real*	지정된 평균과 표준 편차를 가진 정규 분포 값이 지정된 숫자보다 작을 확률을 리턴합니다.
cdf_t(NUM, DF)	*Real*	지정된 자유도를 가진 스튜던트의 *t* 분포 값이 지정된 숫자보다 작을 확률을 리턴합니다.

⑨ 공간 함수

지리 공간 데이터에 공간 함수를 사용할 수 있습니다. 예를 들어, 공간 함수를 사용하여 두 지점 간의 거리, 다각형의 영역 등을 계산할 수 있습니다. 공간 술어(within, close to 등)를 기반으로 하는 다중 지리 공간 데이터 세트 병합이 필요한 상황도 있을 수 있습니다(병합 조건을 통해 이를 수행할 수 있음).

이러한 공간 함수는 **도구** > **스트림 특성** > **옵션** > **지리 공간**에 지정된 좌표계와 함께 작동합니다.

> ⓘ **참고**: 이러한 공간 함수는 3차원 데이터에는 적용되지 않습니다. 3차원 데이터를 시스템으로 가져오는 경우 이러한 함수에서는 처음 두 차원만 사용합니다. Z축 값은 무시합니다.

표 1. CLEM 공간 함수

기능	결과	설명
close_to (SHAPE,SHAPE,NUM)	*Boolean*	두 개의 모양이 서로 일정한 거리만큼 떨어져 있는지 여부를 검정합니다. 투영된 좌표계를 사용하는 경우 거리는 미터 단위입니다. 좌표계를 사용하지 않으면 임의의 단위입니다.
crosses(SHAPE,SHAPE)	*Boolean*	두 개의 모양이 서로 교차하는지 여부를 검정합니다. 이 함수는 두 개의 행 문자열 모양 또는 하나의 행 문자열과 하나의 다각형에 적합합니다.

기능	결과	설명
overlap(SHAPE,SHAPE)	*Boolean*	SHAPE1 전체가 POLYGON에 포함되는지 여부를 검정합니다.
within(SHAPE,SHAPE)	*Boolean*	SHAPE1 전체가 POLYGON에 포함되는지 여부를 검정합니다.
area(SHAPE)	*Real*	지정된 POLYGON의 영역을 리턴합니다. 투영된 시스템을 사용하는 경우 이 함수는 제곱 미터를 리턴합니다. 좌표계를 사용하지 않으면 임의의 단위입니다. 모양은 POLYGON 또는 MULTIPOLYGON이어야 합니다.
num_points(SHAPE,LIST)	*Integer*	POLYGON 범위 내에 포함된 포인트 필드의 포인트 수(MULTIPOINT)를 리턴합니다. SHAPE1은 POLYGON 또는 MULTIPOLYGON이어야 합니다.
distance(SHAPE,SHAPE)	*Real*	SHAPE1과 SHAPE2 사이의 거리를 리턴합니다. 투영된 좌표계를 사용하는 경우 이 함수는 미터를 리턴합니다. 좌표계를 사용하지 않으면 임의의 단위입니다. SHAPE1과 SHAPE2는 임의의 지리 측정 유형입니다.

⑩ Bitwise 정수 연산

이러한 함수를 사용하여 정수를 2의 보수 값을 나타내는 비트 패턴으로 조작할 수 있습니다(여기서 비트 위치 N은 가중치 2**N을 가짐). 비트는 0부터 위로 번호가 매겨집니다. 이러한 연산은 정수의 부호 비트가 왼쪽으로 무한정 확장되는 것처럼 작동합니다. 따라서 최상위 비트 위 어디서나 양의 정수는 0 비트를, 음의 정수는 1 비트를 갖습니다.

표 1. CLEM bitwise 정수 연산

기능	결과	설명
~~ INT1	*Integer*	정수 INT1의 bitwise 보수를 생성합니다. 즉, INT1에 0이 있는 각 비트 위치에 대한 결과에 1이 있습니다. ~~ INT = -(INT + 1)는 항상 참입니다.
INT1 || INT2	*Integer*	이 연산의 결과는 INT1 및 INT2의 bitwise "inclusive or" 입니다. 즉, INT1이나 INT2 또는 둘 다에 1이 있는 각 비트 위치에 대한 결과에 1이 있습니다.
INT1 ||/& INT2	*Integer*	이 연산의 결과는 INT1 및 INT2의 bitwise "exclusive or"입니다. 즉, INT1이나 INT2에는 1이 있지만 둘 다에는 있지 않은 각 비트 위치에 대한 결과에 1이 있습니다.
INT1 && INT2	*Integer*	정수 INT1 및 INT2의 bitwise "and"를 생성합니다. 즉, INT1과 INT2 둘 다에 1이 있는 각 비트 위치에 대한 결과에 1이 있습니다.

기능	결과	설명
INT1 &&~~ INT2	*Integer*	INT1의 bitwise "and" 및 INT2의 bitwise 보수를 생성합니다. 즉, INT1에는 1이 있고 INT2에는 0이 있는 각 비트 위치에 대한 결과에 1이 있습니다. 이는 INT1 && (~~INT2)와 동일하며 INT2에 설정된 INT1의 비트를 지우는 데 유용합니다.
INT << N	*Integer*	N 위치만큼 왼쪽으로 이동한 INT1의 비트 패턴을 생성합니다. N에 대한 음수값은 오른쪽 이동을 생성합니다.
INT >> N	*Integer*	N 위치만큼 오른쪽으로 이동한 INT1의 비트 패턴을 생성합니다. N에 대한 음수값은 왼쪽 이동을 생성합니다.
INT1 &&=_0 INT2	*Boolean*	부울 표현식 INT1 && INT2 /== 0과 같지만 보다 효율적입니다.
INT1 &&/=_0 INT2	*Boolean*	부울 표현식 INT1 && INT2 == 0과 같지만 보다 효율적입니다.
integer_bitcount(INT)	*Integer*	INT에 대한 2의 보수 표시에서 1 또는 0 비트 수를 계산합니다. INT가 음수가 아닌 경우 N은 1 비트 수입니다. INT가 음수이면 0 비트 수입니다. 부호 확장으로 인해 음수가 아닌 정수에 무제한 수의 0 비트가 있거나 음의 정수에 1 비트가 있습니다. 항상 integer_bitcount(INT) = integer_bitcount(-(INT+1))입니다.
integer_leastbit(INT)	*Integer*	정수 INT에 설정된 최하위 비트의 비트 위치 N을 리턴합니다. N은 INT가 정확하게 나누는 2의 최대 거듭제곱입니다.
integer_length(INT)	*Integer*	2의 보수 정수로 INT의 비트 길이를 리턴합니다. 즉, N은 INT < (1 << N) if INT >= 0 INT >= (-1 << N) if INT < 0인 가장 작은 정수입니다. INT가 음수가 아닌 경우, 부호 없는 정수로 INT를 표시하려면 최소 N 비트의 필드가 필요합니다. 또는 부호에 관계없이 부호 있는 정수로 INT를 표시하려면 최소 N+1 비트가 필요합니다.
testbit(INT, N)	*Boolean*	정수 INT의 N 위치에서 비트를 검정하고 1의 경우 참이고 0의 경우 거짓인 부울 값으로 비트 N의 상태를 리턴합니다.

⑪ 랜덤 함수

임의로 항목을 선택하거나 임의로 숫자를 생성하는 데 사용하는 함수는 다음과 같습니다.

표 1. CLEM 랜덤 함수

기능	결과	설명
oneof(LIST)	*모두*	임의로 선택된 *LIST*의 요소를 리턴합니다. 목록 항목은 [ITEM1,ITEM2,...,ITEM_N]으로 입력해야 합니다. 필드 이름 목록도 지정할 수 있습니다.
random(NUM)	*숫자*	1에서 시작하여 *NUM*까지 같은 유형(*INT* 또는 *REAL*)의 균등 분포된 난수를 리턴합니다. 정수를 사용하는 경우 정수만 리턴됩니다. 실수(10진수)를 사용하는 경우 실수가 리턴됩니다(스트림 옵션이 판별한 10진수 정밀도). 이 함수가 리턴하는 최대 난수는 *NUM*과 같을 수 있습니다.
random0(NUM)	*숫자*	random(NUM)과 같은 특성을 갖지만 0에서 시작합니다. 이 함수가 리턴하는 최대 난수는 *NUM*과 같지 않습니다.

⑫ 문자열 함수

CLEM에서 문자열에 대해 다음 작업을 수행할 수 있습니다.

- 문자열 비교
- 문자열 작성
- 문자 액세스

CLEM에서 문자열은 매치하는 큰따옴표("string quotes") 사이의 문자열 시퀀스입니다. 문자(CHAR)는 단일 영숫자입니다. `<character>` 양식으로(예: `z`, `A` 또는 `2`) 작은 역따옴표를 사용하여 CLEM 표현식에 선언됩니다. 범위를 벗어나거나 문자열에 음수 지수인 문자를 사용하면 정의되지 않은 동작이 발생합니다.

ⓘ **참고**: SQL 푸시백을 사용하는 문자열과 사용하지 않는 문자열 간의 비교는 후미 공백이 존재하는 서로 다른 결과를 생성할 수 있습니다.

표 1. CLEM 문자열 함수

기능	결과	설명
allbutfirst(N, STRING)	*문자열*	처음 *N*개의 문자가 제거된 *STRING*인 문자열을 리턴합니다.

기능	결과	설명
allbutlast(N, STRING)	*문자열*	마지막 문자가 제거된 *STRING인* 문자열을 리턴합니다.
alphabefore (STRING1, STRING2)	*Boolean*	문자열의 알파벳 순서를 확인하는 데 사용합니다. *STRING1*이 *STRING2*에 선행하면 참을 리턴합니다.
endstring (LENGTH, STRING)	*문자열*	지정된 문자열에서 마지막 *N*개의 문자를 추출합니다. 문자열 길이가 지정된 길이보다 작거나 같으면 변경되지 않습니다.
hasendstring (STRING, SUBSTRING)	*Integer*	이 함수는 isendstring(SUBSTRING, STRING)과 같습니다.
hasmidstring (STRING, SUBSTRING)	*Integer*	이 함수는 ismidstring(SUBSTRING, STRING)(내장된 하위 문자열)과 같습니다.
hasstartstring (STRING, SUBSTRING)	*Integer*	이 함수는 isstartstring(SUBSTRING, STRING)과 같습니다.
hassubstring (STRING, N, SUBSTRING)	*Integer*	이 함수는 issubstring(SUBSTRING, N, STRING)과 같습니다(여기서 *N* 기본값은 1임).
count_substring (STRING, SUBSTRING)	*Integer*	문자열 내에서 지정된 하위 문자열이 발생하는 횟수를 리턴합니다. 예를 들어, count_substring ("foooo.txt", "oo")은 3을 리턴합니다.
hassubstring (STRING, SUBSTRING)	*Integer*	이 함수는 issubstring(SUBSTRING, 1, STRING)과 같습니다(여기서 *N* 기본값은 1임).
isalphacode(CHAR)	*Boolean*	*CHAR*이 문자 코드가 문자인 지정된 문자열(대개 필드 이름)의 문자인 경우 참 값을 리턴합니다. 그렇지 않으면 이 함수는 0값을 리턴합니다(예: isalphacode(produce_num(1))).
isendstring (SUBSTRING, STRING)	*Integer*	*STRING* 문자열이 *SUBSTRING* 하위 문자열로 끝나는 경우 이 함수는 *STRING*의 *SUBSTRING*의 정수 아래첨자를 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다.
islowercode(CHAR)	*Boolean*	*CHAR*이 지정된 문자열(대개 필드 이름)의 소문자인 경우 참 값을 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다. 예를 들어, islowercode(``)와 islowercode (country_name(2))는 둘 다 유효한 표현식입니다.
isstartstring (SUBSTRING, STRING)	*Integer*	*STRING* 문자열이 *SUBSTRING* 하위 문자열로 시작하는 경우 이 함수는 아래첨자 1을 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다.
issubstring(SUBSTRING, N, STRING)	*Integer*	*N*번째 문자에서부터 시작하여 *SUBSTRING* 문자열에서 *STRING* 문자열과 같은 하위 문자열을 검색합니다. 찾으면 이 함수는 매치하는 하위 문자열이 시작되는 정수 아래첨자를 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다. *N*이 주어지지 않으면 이 함수의 기본값은 1로 설정됩니다.

기능	결과	설명
issubstring (SUBSTRING, STRING)	*Integer*	*N*번째 문자에서부터 시작하여 *SUBSTRING* 문자열에서 *STRING* 문자열과 같은 하위 문자열을 검색합니다. 찾으면 이 함수는 매치하는 하위 문자열이 시작되는 정수 아래첨자를 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다. *N*이 주어지지 않으면 이 함수의 기본값은 1로 설정됩니다.
issubstring_count (SUBSTRING, N, STRING):	*Integer*	지정된 *STRING* 내에서 *SUBSTRING*의 *N*번째 발생 지수를 리턴합니다. *SUBSTRING*의 *N*번 발생보다 적으면 0이 리턴됩니다.
issubstring_lim (SUBSTRING, N, STARTLIM, ENDLIM, STRING)	*Integer*	이 함수는 issubstring과 같지만 매치는 아래첨자 *STARTLIM*이나 그 앞에서 시작하거나 아래첨자 *ENDLIM*이나 그 앞에서 끝나도록 제한됩니다. 둘 중 어느 한 인수에 거짓 값을 제공하여 *STARTLIM* 또는 *ENDLIM* 제한조건을 사용하지 않도록 설정할 수 있습니다. 예를 들어, is*substring*_lim(SUBSTRING, N, false, false, STRING)은 issubstring과 같습니다.
isuppercode(CHAR)	*Boolean*	*CHAR*이 대문자인 경우 참 값을 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다. 예를 들어, isuppercode(``)와 isuppercode(country_name(2))는 둘 다 유효한 표현식입니다.
last(CHAR)	*문자열*	*STRING*(한 문자 이상이어야 함)의 마지막 문자 *CHAR*을 리턴합니다.
length(STRING)	*Integer*	*STRING* 문자열의 길이(즉, 문자열 안의 문자 수)를 리턴합니다.
locchar (CHAR, N, STRING)	*Integer*	기호 필드에서 문자 위치를 식별하는 데 사용합니다. 이 함수는 *STRING* 문자열에서 *CHAR* 문자를 검색합니다(*STRING*의 *N*번째 문자에서 검색을 시작함). 이 함수는 문자를 찾은 위치(*N*에서 시작)를 표시하는 값을 리턴합니다. 문자를 찾을 수 없는 경우 이 함수는 0 값을 리턴합니다. 이 함수에 유효하지 않은 오프셋 (*N*)이 있는 경우(예: 문자열 길이를 벗어난 오프셋) 이 함수는 $null$을 리턴합니다. 예를 들어, locchar(`n`, 2, web_page)는 *web_page* 필드에서 필드 값의 두 번째 문자에서 시작하여 `n` 문자를 검색합니다. 참고: 지정된 문자를 캡슐화하려면 작은 역따옴표를 사용하십시오.
locchar_back (CHAR, N, STRING)	*Integer*	locchar과 비슷하나, *N*번째 문자에서 시작하여 역방향으로 검색이 수행됩니다. 예를 들어, loc*char*_back(`n`, 9, web_page)는 9번째 문자에서 시작하여 문자열의 시작 방향으로 거꾸로 이동하여 *web_page* 필드를 검색합니다. 이 함수에 유효하지 않은 오프셋(예: 문자열 길이를 벗어난 오프셋)이 있는 경우 이 함수는 $null$을 리턴합니다. 이상적으로는 필드 현재 값의 길이를 동적으로 사용하려면 locchar_back을 length(<field>) 함수와 함께 사용해야 합니다(예: loc*char*_back(`n`, (length(web_page)), web_page)).

기능	결과	설명
lowertoupper(CHAR) lowertoupper (STRING)	*CHAR* 또는 *문자열*	입력은 이 함수에서 같은 유형의 새 항목을 리턴하는 데 사용하는 문자열 또는 문자입니다(소문자는 해당 대문자로 변환됨). 예를 들어, lowertoupper(`a`), lowertoupper("My string"), lowertoupper(field_name(2))는 모두 유효한 표현식입니다.
matches	*Boolean*	문자열이 지정된 패턴과 매치하면 참을 리턴합니다. 패턴은 문자열 리터럴이어야 합니다. 패턴을 포함하는 필드 이름이 아니어야 합니다. 물음표(?)를 패턴에 포함시켜 하나의 문자와 정확하게 매치할 수 있습니다. 별표(*)는 0개 이상의 문자와 매치됩니다. 리터럴 물음표나 별표와 매치하려면(이를 와일드카드로 사용하지 않음) 백슬래시(₩)를 이스케이프 문자로 사용할 수 있습니다.
replace(SUBSTRING, NEWSUBSTRING, STRING)	*문자열*	지정된 *STRING* 내에서 *SUBSTRING*의 인스턴스를 *NEWSUBSTRING*으로 모두 바꿉니다.
replicate(COUNT, STRING)	*문자열*	지정된 횟수만큼 복사된 원래 문자열로 구성된 문자열을 리턴합니다.
stripchar(CHAR,STRING)	*문자열*	지정된 문자를 문자열 또는 필드에서 제거할 수 있게 합니다. 이 함수를 사용하여 통화 표기법과 같은 추가 기호를 데이터에서 제거하여 단순한 숫자나 이름을 만들 수 있습니다. 예를 들어, stripchar(`$`, 'Cost') 구문을 사용하면 모든 값에서 달러 부호가 제거된 새 필드가 리턴됩니다. *참고*: 지정된 문자를 캡슐화하려면 작은 역따옴표를 사용하십시오.
skipchar (CHAR, N, STRING)	*Integer*	*N*번째 문자에서 시작하여 *STRING* 문자열에서 *CHAR* 이외의 다른 문자를 검색합니다. 이 함수는 문자를 찾은 위치를 표시하는 정수 하위 문자열을 리턴하거나 *N*번째 문자 이후의 모든 문자가 *CHAR*이면 0을 리턴합니다. 이 함수에 유효하지 않은 오프셋(예: 문자열 길이를 벗어난 오프셋)이 있는 경우 이 함수는 $null$을 리턴합니다. locchar은 종종 skip*char* 함수와 함께 사용되어 *N*(문자열 검색을 시작할 위치) 값을 판별합니다(예: skipchar(`s`, (locchar(`s`, 1, "MyString")), "MyString")).
skipchar_back (CHAR, N, STRING)	*Integer*	skipchar과 비슷하나, *N*번째 문자에서 시작하여 역방향으로 검색이 수행됩니다.
startstring (LENGTH, STRING)	*문자열*	지정된 문자열에서 처음 *N*개의 문자를 추출합니다. 문자열 길이가 지정된 길이보다 작거나 같으면 변경되지 않습니다.
strmember(CHAR, STRING)	*Integer*	locchar(CHAR, 1, STRING)과 같습니다. *CHAR*이 처음 나타나는 위치를 표시하는 정수 하위 문자열이나 0을 리턴합니다. 이 함수에 유효하지 않은 오프셋(예: 문자열 길이를 벗어난 오프셋)이 있는 경우 이 함수는 $null$을 리턴합니다.

기능	결과	설명
subscrs(N, STRING)	*CHAR*	입력 문자열 *STRING*의 *N*번째 문자 *CHAR*을 리턴합니다. 이 함수를 간단한 양식으로 STRING(N)으로 작성할 수도 있습니다. 예를 들어, lowertoupper("name"(1))은 유효한 표현식입니다.
substring (N, LEN, STRING)	*문자열*	아래첨자 *N*의 문자에서 시작하여 *STRING* 문자열의 *LEN* 문자로 구성되는 *SUBSTRING* 문자열을 리턴합니다.
substring_between (N1, N2, STRING)	*문자열*	아래첨자 *N1*에서 시작하고 아래첨자 *N2*에서 끝나는 *STRING*의 하위 문자열을 리턴합니다.
trim(STRING)	*문자열*	지정된 문자열에서 선행 및 후행 공백 문자를 제거합니다.
trim_start(STRING)	*문자열*	지정된 문자열에서 선행 공백 문자를 제거합니다.
trimend(STRING)	*문자열*	지정된 문자열에서 후행 공백 문자를 제거합니다.
unicode_char(NUM)	*CHAR*	입력은 16진 값이 아닌 10진수여야 합니다. 유니코드 값 *NUM*을 포함한 문자를 리턴합니다.
unicode_value(CHAR)	*NUM*	*CHAR*의 유니코드 값을 리턴합니다.
uppertolower(CHAR) uppertolower (STRING)	*CHAR 또는 문자열*	입력은 이 함수에서 같은 유형의 새 항목을 리턴하는 데 사용하는 문자열 또는 문자입니다(대문자는 해당 소문자로 변환됨). *참고*: 큰따옴표로 문자열을 지정하고, 작은 역따옴표로 문자를 지정하십시오. 단순 필드 이름은 따옴표 없이 지정해야 합니다.

⑬ SoundEx 함수

SoundEx는 사운드는 알려져 있지만 정확한 맞춤법을 모를 경우 문자열을 찾는 데 사용하는 메소드입니다. 1918에 개발되었으며 특정 문자가 발음되는 방법에 대한 음성학적 가정을 기반으로 사운드가 유사한 단어를 검색합니다. 데이터베이스에서 이름을 검색하는 데 사용할 수 있습니다(예를 들어, 유사한 이름에 대해 맞춤법과 발음이 다른 경우). 기본 SoundEx 알고리즘은 알려진 제한사항(예를 들어, ph 및 f와 같은 선행 문자 조합은 동일한 것처럼 들리더라도 매치하지 않음)에도 불구하고 수많은 소스에 설명되어 있으며 대부분의 데이터베이스에서 일부 양식으로 지원됩니다.

표 1. CLEM soundex 함수

기능	결과	설명
soundex(STRING)	Integer	지정된 STRING의 4자 SoundEx 코드를 리턴합니다.

기능	결과	설명
soundex_difference (STRING1, STRING2)	Integer	두 개의 문자열에 대한 SoundEx 인코딩이 동일한 문자 수를 표시하는 0과 4 사이의 정수를 리턴합니다(여기서, 0은 유사성이 없음을 표시하고 4는 강력한 유사성 또는 동일한 문자열을 표시함).

⑭ 날짜 및 시간 함수

CLEM은 날짜 및 시간을 나타내는 문자열 변수의 날짜/시간 저장 공간을 가진 필드를 처리하기 위한 함수 패밀리를 포함합니다. 사용하는 날짜 및 시간 형식은 각 스트림에 고유하며 스트림 특성 대화 상자에 지정됩니다. 날짜 및 시간 함수는 현재 선택된 형식에 따라 날짜 및 시간 문자열을 구문 분석합니다.

날짜에 두 자리만 사용하는 연도를 지정하는 경우(즉, 세기는 지정되지 않음) IBM® SPSS® Modeler은 스트림 특성 대화 상자에 지정되는 기본 세기를 사용합니다.

ⓘ **참고**: 데이터 함수가 SQL 또는 IBM SPSS Analytic Server에 푸시백되는 경우, Analytic Server 데이터 소스를 뒤따르는 분기에서 해당 데이터 내 날짜 형식 문자열(to_date)은 SPSS Modeler 스트림에 지정된 날짜 형식과 매치해야 합니다.

표 1. CLEM 날짜 및 시간 함수

기능	결과	설명
@TODAY	*문자열*	스트림 특성 대화 상자에서 **롤오버 일/분(O)**을 선택하는 경우 이 함수는 현재 날짜를 현재 날짜 형식의 문자열로 리턴합니다. 두 자리 숫자 날짜 형식을 사용하고 **롤오버 일/분(O)**을 선택하지 않는 경우 이 함수는 현재 서버에서 $null$을 리턴합니다.
to_time(ITEM)	*시간*	지정된 필드의 저장 공간을 시간으로 변환합니다.
to_date(ITEM)	*날짜*	지정된 필드의 저장 공간을 날짜로 변환합니다.
to_timestamp(ITEM)	*시간소인*	지정된 필드의 저장 공간을 시간소인으로 변환합니다.
to_datetime(ITEM)	*날짜/시간*	지정된 필드의 저장 공간을 날짜, 시간 또는 시간소인 값으로 변환합니다.

기능	결과	설명
datetime_date(ITEM)	*날짜*	*숫자*, *문자열* 또는 *시간소인*의 날짜 값을 리턴합니다. 이는 숫자(초)를 날짜로 다시 변환할 수 있게 하는 유일한 함수입니다. ITEM이 문자열인 경우 현재 날짜 형식으로 문자열을 구문 분석하여 날짜를 생성합니다. 이 함수가 성공하려면 스트림 특성 대화 상자에 지정된 날짜 형식이 올바라야 합니다. ITEM이 숫자인 경우 기준 날짜(또는 에포크) 이후의 초 단위로 해석됩니다. 1일의 분수는 잘립니다. ITEM이 시간소인인 경우 시간소인의 날짜 부분이 리턴됩니다. ITEM이 날짜인 경우 변경되지 않고 리턴됩니다.
date_before(DATE1, DATE2)	*Boolean*	*DATE1*이 *DATE2*가 나타내는 날짜 또는 시간소인 이전의 날짜 또는 시간소인을 나타내는 경우 참 값을 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다.
date_days_difference (DATE1, DATE2)	*정수*	*DATE1*이 나타내는 날짜 또는 시간소인부터 *DATE2*가 나타내는 날짜 또는 시간소인까지의 시간(일)을 정수로 리턴합니다. *DATE2*가 *DATE1* 이전인 경우 이 함수는 음수를 리턴합니다.
date_in_days(DATE)	*정수*	기준 날짜부터 *DATE*가 나타내는 날짜 또는 시간소인까지의 시간(일)을 정수로 리턴합니다. *DATE*가 기준 날짜 이전인 경우 이 함수는 음수를 리턴합니다. 계산이 적절히 작동하려면 유효한 날짜를 포함시켜야 합니다. 예를 들어, 2001년 2월 29일을 날짜로 지정해서는 안 됩니다. 2001년은 윤년이 아니기 때문에 이 날짜가 존재하지 않습니다.
date_in_months(DATE)	*실수*	기준 날짜부터 *DATE*가 나타내는 날짜 또는 시간소인까지의 시간(월)을 실수로 리턴합니다. 이는 30.4375일을 1개월로 한 대략적인 숫자입니다. *DATE*가 기준 날짜 이전인 경우 이 함수는 음수를 리턴합니다. 계산이 적절히 작동하려면 유효한 날짜를 포함시켜야 합니다. 예를 들어, 2001년 2월 29일을 날짜로 지정해서는 안 됩니다. 2001년은 윤년이 아니기 때문에 이 날짜가 존재하지 않습니다.
date_in_weeks(DATE)	*실수*	기준 날짜부터 *DATE*가 나타내는 날짜 또는 시간소인까지의 시간(주)을 실수로 리턴합니다. 이는 7.0일을 1주로 합니다. *DATE*가 기준 날짜 이전인 경우 이 함수는 음수를 리턴합니다. 계산이 적절히 작동하려면 유효한 날짜를 포함시켜야 합니다. 예를 들어, 2001년 2월 29일을 날짜로 지정해서는 안 됩니다. 2001년은 윤년이 아니기 때문에 이 날짜가 존재하지 않습니다.

기능	결과	설명
date_in_years(DATE)	*실수*	기준 날짜부터 *DATE*가 나타내는 날짜 또는 시간소인까지의 시간(년)을 실수로 리턴합니다. 이는 365.25일을 1년으로 한 대략적인 숫자입니다. *DATE*가 기준 날짜 이전인 경우 이 함수는 음수를 리턴합니다. 계산이 적절히 작동하려면 유효한 날짜를 포함시켜야 합니다. 예를 들어, 2001년 2월 29일을 날짜로 지정해서는 안 됩니다. 2001년은 윤년이 아니기 때문에 이 날짜가 존재하지 않습니다.
date_months_difference (DATE1, DATE2)	*실수*	*DATE1*이 나타내는 날짜 또는 시간소인부터 *DATE2*가 나타내는 날짜 또는 시간소인까지의 시간(월)을 실수로 리턴합니다. 이는 30.4375일을 1개월로 한 대략적인 숫자입니다. *DATE2*가 *DATE1* 이전인 경우 이 함수는 음수를 리턴합니다.
datetime_date(YEAR, MONTH, DAY)	*날짜*	주어진 *YEAR*, *MONTH*, *DAY*의 날짜 값을 생성합니다. 인수는 정수여야 합니다.
datetime_day(DATE)	*정수*	주어진 *DATE* 또는 시간소인의 날짜(일)를 리턴합니다. 결과는 1 - 31 범위의 정수입니다.
datetime_day_name(DAY)	*문자열*	주어진 *DAY*의 전체 이름을 리턴합니다. 인수는 1(일요일) - 7(토요일) 범위의 정수여야 합니다.
datetime_hour(TIME)	*정수*	*TIME* 또는 시간소인의 시간을 리턴합니다. 결과는 0 - 23 범위의 정수입니다.
datetime_in_seconds(TIME)	*실수*	*TIME*에 저장된 초 부분을 리턴합니다.
datetime_in_seconds(DATE), datetime_in_seconds(DATETIME)	*실수*	현재 *DATE* 또는 *DATETIME*과 기준 날짜(1900-01-01) 간의 차이로부터 누적된 숫자(초로 변환됨)입니다.
datetime_minute(TIME)	*정수*	*TIME* 또는 시간소인의 분을 리턴합니다. 결과는 0 - 59 범위의 정수입니다.
datetime_month(DATE)	*정수*	*DATE* 또는 시간소인의 월을 리턴합니다. 결과는 1 - 12 범위의 정수입니다.
datetime_month_name (MONTH)	*문자열*	주어진 *MONTH*의 전체 이름을 리턴합니다. 인수는 1 - 12 범위의 정수여야 합니다.
datetime_now	*시간소인*	현재 시간을 시간소인으로 리턴합니다.
datetime_second(TIME)	*정수*	*TIME* 또는 시간소인의 초를 리턴합니다. 결과는 0 - 59 범위의 정수입니다.
datetime_day_short_name (DAY)	*문자열*	주어진 *DAY*의 축약 이름을 리턴합니다. 인수는 1(일요일) - 7(토요일) 범위의 정수여야 합니다.
datetime_month_short_name (MONTH)	*문자열*	주어진 *MONTH*의 축약 이름을 리턴합니다. 인수는 1 - 12 범위의 정수여야 합니다.

기능	결과	설명
datetime_time(HOUR, MINUTE, SECOND)	*시간*	지정된 *HOUR*, *MINUTE*, *SECOND*의 시간 값을 리턴합니다. 인수는 정수여야 합니다.
datetime_time(ITEM)	*시간*	주어진 *ITEM*의 시간 값을 리턴합니다.
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	*시간소인*	주어진 *YEAR*, *MONTH*, *DAY*, *HOUR*, *MINUTE*, *SECOND*의 시간소인 값을 리턴합니다.
datetime_timestamp(DATE, TIME)	*시간소인*	주어진 *DATE* 및 *TIME*의 시간소인 값을 리턴합니다.
datetime_timestamp (NUMBER)	*시간소인*	주어진 초 단위의 시간소인 값을 리턴합니다.
datetime_weekday(DATE)	*정수*	주어진 *DATE* 또는 시간소인의 요일을 리턴합니다.
datetime_year(DATE)	*정수*	*DATE* 또는 시간소인의 연도를 리턴합니다. 결과는 2002와 같은 정수입니다.
date_weeks_difference (DATE1, DATE2)	*실수*	*DATE1*이 나타내는 날짜 또는 시간소인부터 *DATE2*가 나타내는 날짜 또는 시간소인까지의 시간(주)을 실수로 리턴합니다. 이는 7.0일을 1주로 합니다. *DATE2*가 *DATE1* 이전인 경우 이 함수는 음수를 리턴합니다.
date_years_difference (DATE1, DATE2)	*실수*	*DATE1*이 나타내는 날짜 또는 시간소인부터 *DATE2*가 나타내는 날짜 또는 시간소인까지의 시간(년)을 실수로 리턴합니다. 이는 365.25일을 1년으로 한 대략적인 숫자입니다. *DATE2*가 *DATE1* 이전인 경우 이 함수는 음수를 리턴합니다.
date_from_ywd(YEAR, WEEK, DAY)	*Integer*	ISO 8601 표준을 사용하여 연도, 주(한 해 기준), 요일을 날짜로 변환합니다.
date_iso_day(DATE)	*Integer*	ISO 8601 표준을 사용하여 날짜의 요일을 리턴합니다.
date_iso_week(DATE)	*Integer*	ISO 8601 표준을 사용하여 날짜의 주(한 해 기준)를 리턴합니다.
date_iso_year(DATE)	*Integer*	ISO 8601 표준을 사용하여 날짜의 연도를 리턴합니다.
time_before(TIME1, TIME2)	*Boolean*	*TIME1*이 *TIME2*가 나타내는 시간 또는 시간소인 이전의 시간 또는 시간소인을 나타내는 경우 참 값을 리턴합니다. 그렇지 않으면 이 함수는 0 값을 리턴합니다.
time_hours_difference (TIME1, TIME2)	*실수*	*TIME1*과 *TIME2*가 나타내는 시간 또는 시간소인 간의 시차(시간)를 실수로 리턴합니다. 스트림 특성 대화 상자에서 **롤오버 일/분(O)**을 선택하는 경우 이전 일을 참조하기 위해 상위 값 *TIME1*을 사용합니다. 롤오버 옵션을 선택하지 않으면 상위 값 *TIME1*으로 인해 리턴된 값은 음수가 됩니다.

기능	결과	설명
time_in_hours(TIME)	*실수*	*TIME*이 나타내는 시간(시간)을 실수로 리턴합니다. 예를 들어, HHMM 시간 형식에서 time_in_hours('0130') 표현식은 1.5로 평가됩니다. *TIME*은 시간 또는 시간소인을 나타낼 수 있습니다.
time_in_mins(TIME)	*실수*	*TIME*이 나타내는 시간(분)을 실수로 리턴합니다. *TIME*은 시간 또는 시간소인을 나타낼 수 있습니다.
time_in_secs(TIME)	*정수*	*TIME*이 나타내는 시간(초)을 정수로 리턴합니다. *TIME*은 시간 또는 시간소인을 나타낼 수 있습니다.
time_mins_difference(TIME1, TIME2)	*실수*	*TIME1*과 *TIME2*가 나타내는 시간 또는 시간소인 간의 시차(분)를 실수로 리턴합니다. 스트림 특성 대화 상자에서 **롤오버 일/분(O)**을 선택하는 경우 이전 일(또는 현재 형식에 분과 초만 지정된 경우 이전 시간)을 참조하기 위해 상위 값 *TIME1*을 사용합니다. 롤오버 옵션을 선택하지 않으면 상위 값 *TIME1*으로 인해 리턴된 값은 음수가 됩니다.
time_secs_difference(TIME1, TIME2)	*정수*	*TIME1*과 *TIME2*가 나타내는 시간 또는 시간소인 간의 시차(초)를 정수로 리턴합니다. 스트림 특성 대화 상자에서 **롤오버 일/분(O)**을 선택하는 경우 이전 일(또는 현재 형식에 분과 초만 지정된 경우 이전 시간)을 참조하기 위해 상위 값 *TIME1*을 사용합니다. 롤오버 옵션을 선택하지 않으면 상위 값 *TIME1*으로 인해 리턴된 값은 음수가 됩니다.

가. 날짜 및 시간 값 변환

변환 함수(및 날짜 또는 시간 값과 같은 특정 유형의 입력을 필요로 하는 기타 함수)는 스트림 옵션 대화 상자에 지정된 현재 형식에 따라 다릅니다. 예를 들어, 값이 Jan 2003, Feb 2003 등인 문자열로 저장된 DATE 필드가 있으면 다음과 같이 날짜 저장 공간으로 변환할 수 있습니다.

```
to_date(DATE)
```

이 변환이 작동하려면 매치하는 날짜 형식 **MON YYYY**를 스트림의 기본 날짜 형식으로 선택하십시오.

채움 노드를 사용하여 문자열 값을 날짜로 변환하는 예제는 ₩Demos 폴더의 streams 서브폴더에 설치된 스트림 broadband_create_models.str을 참조하십시오.

숫자로 저장된 날짜. 이전 예제에서 DATE는 필드 이름인 반면, to_date는 CLEM 함수입니다.

숫자로 저장된 날짜가 있으면 datetime_date 함수를 사용하여 변환할 수 있습니다(여기서 숫자는 기준 날짜(또는 에포크) 이후의 초 단위로 해석됨).

```
datetime_date(DATE)
```

날짜를 다수의 초로, 초를 날짜로 변환하여 현재 날짜 더하기 또는 빼기 고정 일 수와 같은 계산을 수행할 수 있습니다. 예를 들어, 다음과 같습니다.

```
datetime_date((date_in_days(DATE)-7)*60*60*24)
```

⑮ 시퀀스 함수

일부 연산의 경우 이벤트 시퀀스가 중요합니다. 애플리케이션을 통해 다음 레코드 시퀀스에 대한 작업을 할 수 있습니다.

- 시퀀스 및 시계열
- 시퀀스 함수
- 레코드 색인화
- 값 평균, 합계 및 비교
- 변경 모니터링 - 차별
- @SINCE
- 오프셋 값
- 추가 시퀀스 기능

많은 애플리케이션의 경우, 스트림을 통과하는 각 레코드를 다른 모든 케이스와 독립된 개별 케이스로 간주할 수 있습니다. 이러한 상황에서는 일반적으로 레코드 순서가 중요하지 않습니다.

그러나 일부 문제점 클래스의 경우 레코드 시퀀스가 매우 중요합니다. 이는 일반적으로 레코드 시퀀스가 정렬된 이벤트 또는 발생 시퀀스를 나타내는 시계열 상황입니다. 각 레코드는 특정 순간의 스냅샷을 나타냅니다. 그러나 많은 양의 풍부한 정보가 순시값에 포함되지 않고 시간이 경과하면서 이러한 값이 변화하고 동작합니다.

물론 관련 모수가 시간이 아닌 다른 어떤 것일 수 있습니다. 예를 들어, 레코드가 선을 따라 거리를 두고 수행된 분석을 나타낼 수 있지만 동일한 원칙이 적용됩니다.

다음 특성을 통해 시퀀스 및 특수 함수를 즉시 인식할 수 있습니다.

- 모두 앞에 @가 붙습니다.
- 이름이 대문자입니다.

시퀀스 함수는 노드가 현재 처리 중인 레코드, 노드를 이미 통과한 레코드, 한 케이스에선 아직 노드를 통과해야 하는 레코드를 참조할 수 있습니다. 인수로 사용할 수 있는 것에 약간의 제한이 있긴 하지만 시퀀스 함수를 CLEM 표현식의 기타 구성요소와 자유롭게 혼합할 수 있습니다.

예

특정 이벤트가 발생했거나 조건이 참인 이후로 얼마나 경과했는지를 아는 것이 유용합니다. 이를 수행하려면 @SINCE 함수를 사용하십시오. 예를 들어, 다음과 같습니다.

```
@SINCE(Income > Outgoings)
```

이 함수는 이 조건이 참인 마지막 레코드의 오프셋을 리턴합니다(즉, 조건이 참이었던 이 레코드 이전의 레코드 수). 조건이 참이었던 적이 없으면 @SINCE가 @INDEX + 1을 리턴합니다.

@SINCE에서 사용하는 표현식에서 현재 레코드 값을 참조하려는 경우가 있습니다. 필드 이름이 항상 현재 레코드에 적용됨을 지정하는 @THIS 함수를 사용하여 이를 수행할 수 있습니다. 현재 레코드의 값보다 두 배가 넘는 Concentration 필드 값을 가진 마지막 레코드의 오프셋을 찾으려면 다음을 사용할 수 있습니다.

```
@SINCE(Concentration > 2 * @THIS(Concentration))
```

일부 경우 @SINCE에 지정된 조건이 정의상 현재 레코드에도 적용됩니다. 예를 들어, 다음과 같습니다.

```
@SINCE(ID == @THIS(ID))
```

이러한 이유로 @SINCE는 현재 레코드에 대해 조건을 평가하지 않습니다. 이전 레코드는 물론 현재 레코드에 대해서도 조건을 평가하려면 유사 함수인 @SINCE0을 사용하십시오. 조건이 현재 레코드에서 참이면 @SINCE0이 0을 리턴합니다.

표 1. CLEM 시퀀스 함수

기능	결과	설명
MEAN(FIELD)	*Real*	지정된 *FIELD* 또는 *FIELDS* 값의 평균을 리턴합니다.
@MEAN(FIELD, EXPR)	*Real*	현재 레코드를 포함하여 현재 노드가 수신한 마지막 *EXPR* 레코드에 대한 *FIELD* 값의 평균을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 평균이 리턴됩니다.
@MEAN (FIELD, EXPR, INT)	*Real*	현재 레코드를 포함하여 현재 노드가 수신한 마지막 *EXPR* 레코드에 대한 *FIELD* 값의 평균을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 평균이 리턴됩니다. *INT*는 되돌아볼 값의 최대 수를 지정합니다. 이는 단지 두 개의 인수를 사용하는 것보다 훨씬 효율적입니다.

기능	결과	설명
@DIFF1(FIELD)	*Real*	*FIELD*의 첫 번째 미분을 리턴합니다. 따라서 단일 인수 양식은 필드의 현재 값과 이전 값 간의 차이만 리턴합니다. 관련된 이전 레코드가 없으면 **$null$**을 리턴합니다.
@DIFF1 (FIELD1, FIELD2)	*Real*	이 2인수 양식은 *FIELD2*와 관련하여 *FIELD1*의 첫 번째 미분을 제공합니다. 관련된 이전 레코드가 없으면 **$null$**을 리턴합니다. @DIFF1(FIELD1)/@DIFF1(FIELD2)로 계산됩니다.
@DIFF2(FIELD)	*Real*	*FIELD*의 두 번째 미분을 리턴합니다. 따라서 단일 인수 양식은 필드의 현재 값과 이전 값 간의 차이만 리턴합니다. 관련된 이전 레코드가 없으면 **$null$**을 리턴합니다. @DIFF2는 @DIFF(@DIFF(FIELD))로 계산됩니다.
@DIFF2 (FIELD1, FIELD2)	*Real*	이 2인수 양식은 *FIELD2*와 관련하여 *FIELD1*의 두 번째 미분을 제공합니다. 관련된 이전 레코드가 없으면 **$null$**을 리턴합니다. 이는 복잡한 계산입니다 (@DIFF1(FIELD1)/@DIFF1(FIELD2) - @OFFSET(@DIFF1(FIELD1),1)/@OFFSET(@DIFF1(FIELD2))) / @DIFF1(FIELD2)).
@INDEX	*Integer*	현재 레코드의 지수를 리턴합니다. 지수는 현재 노드에 도착하면 레코드에 할당됩니다. 첫 번째 레코드에는 지수 1이 부여되며 지수는 각각의 후속 레코드에 대해 1씩 증분됩니다.
@LAST_NON_BLANK (FIELD)	*모두*	업스트림 소스 또는 유형 노드에 정의된 대로 공백이 아닌 *FIELD*의 마지막 값을 리턴합니다. 지금까지 읽은 레코드에 공백이 아닌 *FIELD* 값이 없으면 **$null$**이 리턴됩니다. 사용자 결측값이라고도 하는 공백값을 각 필드에 대해 개별적으로 정의할 수 있습니다.
@MAX(FIELD)	*숫자*	지정된 *FIELD*의 최대값을 리턴합니다.
@MAX(FIELD, EXPR)	*숫자*	현재 레코드를 포함하여 지금까지 수신된 마지막 *EXPR* 레코드에 대한 *FIELD* 최대값을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다.
@MAX (FIELD, EXPR, INT)	*숫자*	현재 레코드를 포함하여 지금까지 수신된 마지막 *EXPR* 레코드에 대한 *FIELD* 최대값을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 최대값이 리턴됩니다. *INT*는 되돌아볼 값의 최대 수를 지정합니다. 이는 단지 두 개의 인수를 사용하는 것보다 훨씬 효율적입니다.
@MIN(FIELD)	*숫자*	지정된 *FIELD*의 최소값을 리턴합니다.

기능	결과	설명
@MIN(FIELD, EXPR)	*숫자*	현재 레코드를 포함하여 지금까지 수신된 마지막 *EXPR* 레코드에 대한 *FIELD* 최소값을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다.
@MIN (FIELD, EXPR, INT)	*숫자*	현재 레코드를 포함하여 지금까지 수신된 마지막 *EXPR* 레코드에 대한 *FIELD* 최소값을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 최소값이 리턴됩니다. *INT*는 되돌아볼 값의 최대 수를 지정합니다. 이는 단지 두 개의 인수를 사용하는 것보다 훨씬 효율적입니다.
@OFFSET (FIELD, EXPR)	*모두*	*EXPR* 값으로 현재 레코드의 레코드 오프셋에 있는 *FIELD* 값을 리턴합니다. 양수 오프셋은 이미 전달한("되돌아보기") 레코드를 참조하는 반면, 음수 오프셋은 도착해야 하는 레코드에 "미리보기"를 지정합니다. 예를 들어, @OFFSET(Status, 1)은 이전 레코드의 Status 필드 값을 리턴하는 반면, @OFFSET(Status, -4)는 값을 획득하기 위해 시퀀스에서 4개의 레코드(즉, 아직 이 노드를 통과하지 못한 레코드)를 "미리 봅니다". *음수(미리보기) 오프셋을 상수로 지정해야 합니다.* 양수 오프셋의 경우만 *EXPR*은 오프셋을 제공하기 위해 현재 레코드에 대해 평가되는 임의의 CLEM 표현식이기도 합니다. 이 경우, 이 함수의 3인수 버전은 성능을 향상시켜야 합니다(다음 함수 참조). 표현식이 음수가 아닌 정수 이외의 것을 리턴하는 경우 이로 인해 오류가 발생합니다. 즉, 계산된 미리보기 오프셋을 갖는 것이 올바르지 않습니다. *참조*: 자기 참조 @OFFSET 함수는 리터럴 미리보기를 사용할 수 없습니다. 예를 들어, 채움 노드에서는 @OFFSET(field1,-2)와 같은 표현식을 사용하여 field1의 값을 대체할 수 없습니다. *참고*: 채움 노드에서 필드를 채울 때 실질적으로 해당 필드의 두 개의 다른 값(즉, 사전에 채워진 값과 사후 채워진 값)이 있습니다. @OFFSET이 자체를 참조할 때는 사후 채워진 값을 참조합니다. 사후 채워진 이 값은 과거 행에 대해서만 존재하므로 자기 참조 @OFFSET은 과거 행만 참조할 수 있습니다. 자기 참조 @OFFSET은 미래를 참조할 수 없으므로 오프셋에 대한 다음 검사를 수행합니다. - 오프셋이 리터럴이고 미래로 평가되는 경우 실행이 시작되기 전에 오류가 보고됩니다. - 오프셋이 표현식이고 런타임 시 미래로 평가되는 경우 @OFFSET은 $null$을 리턴합니다. *참고*: 한 노드에서 "미리보기"와 "되돌아보기"를 둘 다 사용하는 것은 지원되지 않습니다.

기능	결과	설명
@OFFSET (FIELD, EXPR, INT)	*모두*	되돌아볼 값의 최대 수를 지정하는 세 번째 인수인 *INT*를 추가하여 @OFFSET 함수와 동일한 작업을 수행합니다. 오프셋이 표현식에서 계산되는 경우 이 세 번째 인수는 성능을 향상시켜야 합니다. 예를 들어, @OFFSET(Foo, Month, 12)와 같은 표현식에서 시스템은 Foo의 마지막 12개 값만 유지하도록 알고 있습니다. 그렇지 않으면 만약을 위해 모든 값을 저장해야 합니다. 오프셋 값이 음수의 "미리보기" 오프셋(상수여야 함)을 포함하여 상수인 경우 세 번째 인수는 무의미하며 이 함수의 2인수 버전을 사용해야 합니다. 앞서 설명한 2인수 버전의 자기 참조 함수에 대한 참고도 참조하십시오. *참고*: 한 노드에서 "미리보기"와 "되돌아보기"를 둘 다 사용하는 것은 지원되지 않습니다.
@SDEV(FIELD)	*Real*	지정된 *FIELD* 또는 *FIELDS* 값의 표준 편차를 리턴합니다.
@SDEV(FIELD, EXPR)	*Real*	현재 레코드를 포함하여 현재 노드가 수신한 마지막 *EXPR* 레코드에 대한 *FIELD* 값의 표준 편차를 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 표준 편차가 리턴됩니다.
@SDEV (FIELD, EXPR, INT)	*Real*	현재 레코드를 포함하여 현재 노드가 수신한 마지막 *EXPR* 레코드에 대한 *FIELD* 값의 표준 편차를 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 표준 편차가 리턴됩니다. *INT*는 되돌아볼 값의 최대 수를 지정합니다. 이는 단지 두 개의 인수를 사용하는 것보다 훨씬 효율적입니다.
@SINCE(EXPR)	*모두*	임의의 CLEM 표현식인 *EXPR*이 참인 이후로 전달한 레코드 수를 리턴합니다.
@SINCE(EXPR, INT)	*모두*	두 번째 인수인 *INT*를 추가하면 되돌아볼 레코드의 최대 수를 지정합니다. *EXPR*이 참이었던 적이 없으면 *INT*는 @INDEX+1입니다.
@SINCE0(EXPR)	*모두*	@SINCE는 고려하지 않는 반면 현재 레코드를 고려합니다. 따라서 *EXPR*이 현재 레코드에 대해 참이면 @SINCE0이 0을 리턴합니다.
@SINCE0(EXPR, INT)	*모두*	두 번째 인수인 *INT*를 추가하면 되돌아볼 레코드의 최대 수를 지정합니다.
@SUM(FIELD)	*숫자*	지정된 *FIELD* 또는 *FIELDS* 값의 합계를 리턴합니다.

기능	결과	설명
@SUM(FIELD, EXPR)	*숫자*	현재 레코드를 포함하여 현재 노드가 수신한 마지막 *EXPR* 레코드에 대한 *FIELD* 값의 합계를 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 합계가 리턴됩니다.
@SUM (FIELD, EXPR, INT)	*숫자*	현재 레코드를 포함하여 현재 노드가 수신한 마지막 *EXPR* 레코드에 대한 *FIELD* 값의 합계를 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. *EXPR*은 0보다 큰 정수로 평가되는 표현식입니다. *EXPR*이 생략되었거나 지금까지 수신된 레코드 수를 초과하는 경우 지금까지 수신된 모든 레코드에 대한 합계가 리턴됩니다. *INT*는 되돌아볼 값의 최대 수를 지정합니다. 이는 단지 두 개의 인수를 사용하는 것보다 훨씬 효율적입니다.
@THIS(FIELD)	*모두*	현재 레코드에서 *FIELD* 필드의 값을 리턴합니다. @SINCE 표현식에서만 사용합니다.

⑯ 글로벌 함수

@MEAN, @SUM, @MIN, @MAX, @SDEV 함수는 최대한으로 보면, 현재 레코드를 포함하여 지금까지 읽은 모든 레코드에서만 작동합니다. 그러나 일부 경우 현재 레코드의 값을 전체 데이터 세트에 표시된 값과 비교하는 방법을 이해하는 것이 유용합니다. 전역값 설정 노드를 사용하여 전체 데이터 세트에서 값을 생성하는 경우 글로벌 함수를 사용하여 CLEM 표현식에서 이러한 값에 액세스할 수 있습니다.

예를 들어,

```
@GLOBAL_MAX(Age)
```

데이터 세트의 Age 최고값을 리턴하는 반면, 다음 표현식은

```
(Value - @GLOBAL_MEAN(Value)) / @GLOBAL_SDEV(Value)
```

이 레코드의 Value와 글로벌 평균 간의 차이를 다수의 표준 편차로 표현합니다. 전역값 설정 노드가 계산한 이후에만 글로벌 값을 사용할 수 있습니다. 스트림 특성 대화 상자의 전역값 탭에서 **글로벌 값 지우기** 단추를 클릭하여 현재 글로벌 값을 모두 취소할 수 있습니다.

표 1. CLEM 글로벌 함수

기능	결과	설명
@GLOBAL_MAX(FIELD)	*숫자*	전역값 설정 노드에 의해 이전에 생성된 전체 데이터 세트에 대한 *FIELD*의 최대값을 리턴합니다. *FIELD*는 숫자, 날짜/시간/날짜시간 또는 문자열 필드의 이름이어야 합니다. 해당 글로벌 값이 설정되지 않은 경우에는 오류가 발생합니다.
@GLOBAL_MIN(FIELD)	*숫자*	전역값 설정 노드에 의해 이전에 생성된 전체 데이터 세트에 대한 *FIELD*의 최소값을 리턴합니다. *FIELD*는 숫자, 날짜/시간/날짜시간 또는 문자열 필드의 이름이어야 합니다. 해당 글로벌 값이 설정되지 않은 경우에는 오류가 발생합니다.
@GLOBAL_SDEV(FIELD)	*숫자*	전역값 설정 노드에 의해 이전에 생성된 전체 데이터 세트에 대한 *FIELD*의 표준 편차 값을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. 해당 글로벌 값이 설정되지 않은 경우에는 오류가 발생합니다.
@GLOBAL_MEAN(FIELD)	*숫자*	전역값 설정 노드에 의해 이전에 생성된 전체 데이터 세트에 대한 *FIELD* 값의 평균을 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. 해당 글로벌 값이 설정되지 않은 경우에는 오류가 발생합니다.
@GLOBAL_SUM(FIELD)	*숫자*	전역값 설정 노드에 의해 이전에 생성된 전체 데이터 세트에 대한 *FIELD*의 합계를 리턴합니다. *FIELD*는 숫자 필드의 이름이어야 합니다. 해당 글로벌 값이 설정되지 않은 경우에는 오류가 발생합니다.

⑰ 공백 및 널값 처리 함수

CLEM을 사용하여 필드의 특정 값이 "공백" 또는 결측값으로 간주되도록 지정할 수 있습니다. 공백에 대해 작동하는 함수는 다음과 같습니다.

표 1. CLEM 공백 및 널값 함수

기능	결과	설명
@BLANK(FIELD)	*Boolean*	업스트림 유형 노드 또는 소스 노드(유형 탭)에 설정된 공백 처리 규칙에 따라 값이 공백인 모든 레코드에 대해 참을 리턴합니다.

기능	결과	설명
@LAST_NON_BLANK(FIELD)	*모두*	업스트림 소스 또는 유형 노드에 정의된 대로 공백이 아닌 *FIELD*의 마지막 값을 리턴합니다. 지금까지 읽은 레코드에 공백이 아닌 *FIELD* 값이 없으면 $null$이 리턴됩니다. 사용자 결측값이라고도 하는 공백값을 각 필드에 대해 개별적으로 정의할 수 있습니다.
@NULL(FIELD)	*Boolean*	*FIELD* 값이 시스템 결측값 $null$이면 참을 리턴합니다. 사용자 정의 공백을 포함하여 기타 모든 값에 대해서는 거짓을 리턴합니다. 둘 다를 확인하려면 @BLANK(FIELD) 및 @NULL(FIELD)를 사용하십시오.
undef	*모두*	$null$ 값을 입력하기 위해(예: 공백값을 채움 노드의 널값으로 채우기 위해) CLEM에서 일반적으로 사용합니다.

공백 필드는 채움 노드로 "채워"집니다. 채움 노드와 파생 노드(다중 모드만) 둘 다에서 특수 CLEM 함수 @FIELD는 검사 중인 현재 필드를 참조합니다.

⑱ 특수 필드

특수 함수는 검사 중인 특정 필드를 표시하거나 필드 목록을 입력으로 생성하는 데 사용합니다. 예를 들어, 동시에 여러 필드를 파생시키는 경우 @FIELD를 사용하여 "선택된 필드에 대해 이 파생 동작을 수행함"을 표시해야 합니다. log(@FIELD) 표현식을 사용하면 선택된 각 필드에 대해 새 로그 필드가 파생됩니다.

표 1. CLEM 특수 필드

기능	결과	설명
@FIELD	*모두*	표현식 컨텍스트에 지정된 모든 필드에 대해 동작을 수행합니다.
@TARGET	*모두*	CLEM 표현식을 사용자 정의 분석 함수에서 사용하는 경우 @TARGET은 목표 필드 또는 분석 중인 목표/예측 쌍의 "올바른 값"을 나타냅니다. 이 함수는 일반적으로 분석 노드에서 사용합니다.
@PREDICTED	*모두*	CLEM 표현식을 사용자 정의 분석 함수에서 사용하는 경우 @PREDICTED는 분석 중인 목표/예측 쌍의 예측값을 나타냅니다. 이 함수는 일반적으로 분석 노드에서 사용합니다.

기능	결과	설명
@PARTITION_FIELD	*모두*	현재 파티션 필드의 이름을 대체합니다.
@TRAINING_PARTITION	*모두*	현재 훈련 파티션 값을 리턴합니다. 예를 들어, 선택 노드를 사용하여 훈련 레코드를 선택하려면 CLEM 표현식 (@PARTITION_FIELD = @TRAINING_PARTITION)을 사용하십시오. 이렇게 하면 데이터의 각 파티션을 나타내는 데 사용하는 값에 관계없이 선택 노드가 항상 작동합니다.
@TESTING_PARTITION	*모두*	현재 검정 파티션 값을 리턴합니다.
@VALIDATION_PARTITION	*모두*	현재 검증 파티션 값을 리턴합니다.
@FIELDS_BETWEEN (start, end)	*모두*	데이터 필드의 자연(즉, 삽입) 순서를 기반으로 지정된 시작 필드와 종료 필드(포함) 사이의 필드 이름 목록을 리턴합니다.
@FIELDS_MATCHING (pattern)	*모두*	지정된 패턴과 매치하는 필드 이름 목록을 리턴합니다. 물음표(?)를 패턴에 포함시켜 하나의 문자와 정확하게 매치할 수 있습니다. 별표(*)는 0개 이상의 문자와 매치됩니다. 리터럴 물음표나 별표와 매치하려면(이를 와일드카드로 사용하지 않음) 백슬래시(₩)를 이스케이프 문자로 사용할 수 있습니다. ⓘ **참고**: 이 경우 문자열 리터럴이 인수로 필요합니다(중첩 표현식을 사용하여 인수를 생성할 수 없음).
@MULTI_RESPONSE_SET	*모두*	이름 지정된 다중 응답 세트의 필드 목록을 리턴합니다.

10) IBM Cloud Pak for Data에 스트림 저장

스트림을 IBM Cloud Pak for Data Server에 저장하고 여기서 스트림을 플로우로 열어 Cloud Pak for Data에서 사용 가능한 여러 기능을 활용할 수 있습니다.

Cloud Pak for Data는 데이터 수집, 구성 및 분석 방법을 단순화하고 비즈니스 전반에 걸쳐 AI를 주입합니다. 광범위한 데이터 및 AI 소프트웨어, 선도적인 거버넌스 및 보안, 협업을 위한 통합 환경을 원활하게 통합합니다. 추가 정보는 https://www.ibm.com/products/cloud-pak-for-data의 내용을 참조하십시오. Cloud Pak for Data에 관심이 있는 경우 IBM 담당자에게 문의하십시오.

서버 연결을 구성하는 방법

이 지시사항은 Firefox를 기반으로 합니다. 사용하는 웹 브라우저에 따라 단계가 다를 수 있습니다.

1. Firefox를 열고 주소 표시줄에 Cloud Pak for Data URL을 붙여넣은 후 Enter 키를 누르십시오. URL을 확인하려면 Cloud Pak for Data 웹 클라이언트를 열고 브라우저 주소 표시줄의 URL(/zen 앞까지의 모든 항목)을 검토하십시오. 기본적으로 URL 형식은 https://<namespace>-cpd-<namespace>.apps.<cluster-subdomain>입니다.

   각 항목에 대한 설명은 다음과 같습니다.

   - <namespace>는 사용자의 네임스페이스입니다(예: ab123).
   - <cluster-subdomain>은 클러스터의 하위 도메인 이름입니다 (예: ocp454x86-aqt-mycompany.com).

   이 예의 경우 URL은 https://ab123-cpd-ab123.apps.ocp454x86-aqt-mycompany.com입니다.

   Cloud Pak for Data URL에 대한 자세한 정보는 Cloud Pak for Data 문서를 참조하십시오.

2. 주소 표시줄 왼쪽에 있는 노란색 자물쇠 아이콘을 클릭한 후 **안전하지 않은 연결** 옆의 화살표를 클릭하십시오.
3. **자세한 정보**를 클릭하십시오. 페이지 정보 대화 상자가 열립니다.
4. 페이지 정보 대화 상자에서 **보안**으로 이동한 후 **인증서 보기**를 클릭하십시오.
5. **기타**까지 아래로 스크롤하여 PEM 인증 파일을 다운로드하십시오.
6. 명령행에서 다음과 같이 인증 파일을 클라이언트 디렉토리로 가져오십시오.
   Windows:

```
"%JAVA_HOME%"₩bin₩keytool.exe -importcert -trustcacerts -file <Certification file absolute path> -alias <alias> -keystore "<Modeler client installation path>"₩jre₩lib₩security₩cacerts
```

   macOS:

```
"%JAVA_HOME%"/bin/keytool -importcert -trustcacerts -file <Certification file absolute path> -alias <alias> -keystore "<Modeler client installation path>"/jre/lib/security/cacerts
```

프롬프트가 표시되면 키 저장소 비밀번호로 changeit을 입력한 후 yes를 입력하여 인증을 신뢰하십시오.

7. 의 기본 메뉴에서 다음을 클릭하십시오.
   도구 > IBM Cloud Pak for Data Server > IBM Cloud Pak for Data Server 연결...

8. 서버에 대한 연결 설정을 지정하십시오. 1단계에서 사용한 것과 동일한 URL을 사용하십시오. 사용할 사용자 이름 또는 비밀번호를 모를 경우 로컬 시스템 관리자에게 문의하십시오. API 키 로그인 방법을 선택한 경우 다음 방법을 사용하여 URL을 가져오십시오.
   - Cloud Pak for Data 클러스터에 로그인하고 oc get routes -n ibm-common-services 명령을 실행한 후 cp-console의 HOST/PORT 값을 복사하십시오. 다음 예의 경우 URL은 https://cp-console.apps.wml46x04.cp.myserver.com입니다.

```
[root@wml46x04-inf ocp-scripts]# oc get routes -n ibm-common-services
NAME        HOST/PORT                          PATH                 SERVICES  PORT        TERMINATION        WILDCARD
cp-console  cp-console.apps.wml46x04.cp.myserver.com  icp-management-ingress  https     reencrypt/Redirect  None
cp-proxy    cp-proxy.apps.wml46x04.cp.myserver.com  nginx-ingress-controller  https     passthrough/Redirect  None
```

9. **연결**을 클릭하여 서버에 연결하십시오.

Cloud Pak for Data에 스트림을 저장하는 방법

1. 기본 메뉴에서 다음을 클릭하십시오.
   파일 > 플로우로 보관... Cloud Pak for Data로 마이그레이션된 후에는 *스트림*을 *플로우*라고 합니다.

   또는 스트림 관리자 분할창을 마우스 오른쪽 단추로 클릭한 후 **플로우로 보관...**을 선택할 수도 있습니다.

2. 앞에서 구성한 Cloud Pak for Data Server를 선택하십시오. 그런 다음 플로우를 저장할 프로젝트를 선택하십시오.

> **팁**: 저장할 스트림을 한 번에 여러 개 선택할 수도 있습니다. 기본 메뉴에서 **도구 > IBM Cloud Pak for Data Server... > IBM Cloud Pak for Data Server에 스트림 대량 업로드**를 클릭하십시오. 그러면 로컬 파일 시스템에서 스트림을 선택하고 서버를 선택한 후 프로젝트를 선택하라는 프롬프트가 표시됩니다.

11) 외부 애플리케이션으로 내보내기

(1) 외부 애플리케이션으로 내보내기 정보

IBM® SPSS® Modeler에서는 전체 데이터 마이닝 프로세스를 외부 애플리케이션으로 내보내기 위한 많은 메커니즘을 제공하므로 데이터를 준비하고 모델을 작성하기 위해 수행하는 작업을 IBM SPSS Modeler 외부에서도 사용자의 편의를 위해 사용할 수 있습니다.

이전 섹션에서는 다중 사용자 액세스, 작업 스케줄링 및 기타 기능을 활용하여 IBM SPSS Collaboration and Deployment Services 리포지토리에 스트림을 배포하는 방법을 보여주었습니다. 이와 비슷한 방법으로 IBM SPSS Modeler 스트림은 다음과 결합하여 사용할 수도 있습니다.

- IBM SPSS Modeler Advantage
- PMML 형식으로 파일을 가져오고 내보낼 수 있는 애플리케이션

IBM SPSS Modeler Advantage와 함께 스트림 사용에 대한 자세한 정보는 스트림을 IBM SPSS Modeler Advantage에서 열기의 내용을 참조하십시오.

모델을 PMML 파일로 내보내고 가져오고, 모델을 이 형식을 지원하는 다른 애플리케이션과 공유 가능하게 만드는 방법에 대한 자세한 정보는 모델을 PMML로 가져오기 및 내보내기의 내용을 참조하십시오.

(2) 스트림을 IBM SPSS Modeler Advantage에서 열기

IBM® SPSS® Modeler 스트림은 씬 클라이언트 애플리케이션 IBM SPSS Modeler Advantage와 함께 사용할 수 있습니다. 사용자 정의된 애플리케이션을 IBM SPSS Modeler Advantage 내에서 전적으로 작성하는 것이 가능하긴 하지만 IBM SPSS Modeler에서 이미 작성된 스트림을 애플리케이션 작업흐름의 기초로 사용할 수도 있습니다.

IBM SPSS Modeler Advantage에서 스트림 열기

1. 스트림을 IBM SPSS Collaboration and Deployment Services 리포지토리에 배포하고, **스트림으로 배포** 옵션을 반드시 클릭하십시오. 자세한 정보는 스트림 배포의 내용을 참조하십시오.
2. IBM SPSS Modeler Advantage 도구 모음 단추에서 열기를 클릭하거나 주 메뉴에서 다음을 클릭하십시오.

파일 > IBM SPSS Modeler Advantage에서 열기

1. 필요한 경우 리포지토리에 연결 설정을 지정하십시오. 자세한 정보는 리포지토리에 연결의 내용을 참조하십시오. 특정 포트, 비밀번호 및 기타 연결 세부사항에 대해서는 로컬 시스템 관리자에게 문의하십시오.

   참고: 리포지토리 서버에는 또한 IBM SPSS Modeler Advantage 소프트웨어가 설치되어 있어야 합니다.

   1. 리포지토리: 저장 대화 상자에서 오브젝트를 저장하려는 폴더를 선택하고 기록할 다른 정보를 지정한 다음 **저장** 단추를 클릭하십시오. 자세한 정보는 오브젝트 특성 설정의 내용을 참조하십시오.

   그러면 스트림이 이미 열린 상태로 IBM SPSS Modeler Advantage가 실행됩니다. 스트림은 IBM SPSS Modeler에서 닫힙니다.

12) 프로젝트 및 보고서

(1) 프로젝트에 대한 소개

프로젝트는 데이터 마이닝 태스크에 관련된 파일 그룹입니다. 프로젝트에는 데이터 스트림, 그래프, 모형 생성, 보고서 및 그 밖에 IBM® SPSS® Modeler에서 작성된 것이 포함됩니다. 처음에는 IBM SPSS Modeler 프로젝트가 단순히 결과를 구성하는 방법이지만, 실제로는 더 많은 것을 할 수 있습니다. 프로젝트를 사용하여 할 수 있는 것은 다음과 같습니다.

- 프로젝트 파일의 각 오브젝트에 주석을 달 수 있습니다.
- 데이터 마이닝에 대한 노력을 이끌어 주는 CRISP-DM 방법론을 사용할 수 있습니다. 또한 프로젝트에는 CRISP-DM을 사용한 데이터 마이닝에 대한 세부사항 및 실제 예를 제공하는 CRISP-DM 도움말 시스템이 포함되어 있습니다.
- 비IBM SPSS Modeler 오브젝트를 프로젝트(예: 데이터 마이닝 목표를 나타내는 데 사용되는 PowerPoint 슬라이드쇼 또는 사용하려는 알고리즘에 대한 백서)에 추가할 수 있습니다.
- 주석을 기반으로 한 포괄적이고 단순한 업데이트 보고서를 만들 수 있습니다. 이 보고서는 조직의 인트라넷에 쉽게 게재할 수 있도록 HTML로 생성될 수 있습니다.

참고: 프로젝트 분할창이 IBM SPSS Modeler 창에 표시되지 않는 경우, 보기 메뉴에서 프로젝트를 클릭하십시오.

프로젝트에 추가하는 오브젝트는 두 가지 방식, **클래스 보기**와 **CRISP-DM 보기**로 표시될 수 있습니다. 사용자가 프로젝트에 추가하는 것은 두 개의 보기에 추가되며, 보기 간에 전환이 가능하므로 가장 잘 작동하는 구성을 작성할 수 있습니다.

① CRISP-DM 보기

CRISP-DM(Cross-Industry Standard Process for Data Mining)을 지원하면 IBM® SPSS® Modeler 프로젝트는 데이터 마이닝 노력의 일부를 구성하는 업계에서 증명된 비독점 방법을 제공합니다. CRISP-DM은 6개의 단계를 사용하여 시작(비즈니스 요구사항 수집)부터 마침(결과 배포)까지의 프로세스를 설명합니다. 몇몇 단계는 일반적으로 IBM SPSS Modeler에서의 작업과 관련되지 않지만 프로젝트 분할창에는 프로젝트와 연관된 모든 자료를 저장하고 추적하기 위한 중앙 위치를 가질 수 있도록 6개의 단계가 모두 포함됩니다. 예를 들어, 비즈니스 이해 단계는 일반적으로 요구사항 수집과 IBM SPSS Modeler에서 데이터에 대해 작업하는 대신에 목표를 결정하기 위해 동료와 미팅하는 것이 관련됩니다. 프로젝트 분할창을 사용하면 이러한 미팅에서 나온 사용자의 노트를 나중에 참조하거나 보고서에 포함하기 위해 *비즈니스 이해* 폴더에 저장할 수 있습니다.

프로젝트 분할창에서 CRISP-DM 보기는 또한 데이터 마이닝 라이프 사이클을 안내하기 위해 자체 도움말 시스템을 갖추고 있습니다. IBM SPSS Modeler에서 이 도움말은 도움말 메뉴에서 **CRISP-DM 도움말**을 클릭하여 액세스할 수 있습니다.

참고: 프로젝트 분할창이 창에 표시되지 않으면 보기 메뉴에서 **프로젝트**를 클릭하십시오.

가. 기본 프로젝트 단계 설정

프로젝트에 추가된 오브젝트는 CRISP-DM의 기본 단계에 추가됩니다. 즉 오브젝트를 사용하는 데이터 마이닝 단계에 따라 수동으로 오브젝트를 구성해야 함을 의미합니다. 현재 작업 중인 단계에 기본 폴더를 설정하는 것이 바람직합니다.

기본값으로 사용할 단계 선택

1. CRISP-DM 보기에서 기본값으로 설정할 단계의 폴더를 마우스 오른쪽 단추로 클릭하십시오.
2. 메뉴에서 **기본값으로 설정**을 클릭하십시오.

기본 폴더는 굵은 글자로 표시됩니다.

② 클래스 보기

프로젝트 분할창에서 클래스 보기는 사용자의 작업을 IBM® SPSS® Modeler에서 작성된 오브젝트의 유형별로 범주별로 구성합니다. 저장된 오브젝트는 다음 범주에 추가할 수 있습니다.

- 스트림
- 노드
- 모델
- 테이블, 그래프, 보고서
- 기타(슬라이드 쇼 또는 데이터 마이닝 작업과 관련된 백서 등과 같이 IBM SPSS Modeler이 아닌 파일)

클래스 보기에 오브젝트를 추가하면 CRISP-DM 보기에서 기본 단계 폴더에도 추가됩니다.

참고: 프로젝트 분할창이 창에 표시되지 않으면 보기 메뉴에서 **프로젝트**를 클릭하십시오.

(2) 프로젝트 작성

프로젝트는 본질적으로 프로젝트와 연관시키는 모든 파일에 대한 참조를 포함하는 파일입니다. 즉 프로젝트 항목이 개별적으로 저장되고 프로젝트 파일(*.cpj*)에 참조로서도 저장된다는 의미입니다. 이 참조 구조 때문에 다음을 참조하십시오.

- 프로젝트 항목은 먼저 프로젝트에 추가되기 전에 개별적으로 저장되어야 합니다. 항목이 저장되지 않은 경우에는 현재 프로젝트에 추가하기 전에 먼저 저장하라는 메시지가 표시됩니다.
- 스트림 등과 같이 개별적으로 업데이트된 오브젝트는 또한 프로젝트 파일에서도 업데이트됩니다.
- 파일 시스템에서 스트림, 노드 및 출력 오브젝트 등과 같은 오브젝트를 수동으로 이동하거나 삭제하면 프로젝트 파일에 있는 링크가 무효한 것으로 렌더링됩니다.

① 새 프로젝트 작성

새 프로젝트는 IBM® SPSS® Modeler 창에서 쉽게 작성할 수 있습니다. 열려 있는 프로젝트가 없으면 새로 작성하거나, 기존 프로젝트를 닫고 처음부터 다시 시작할 수 있습니다.

기본 메뉴에서 다음을 클릭하십시오.

파일 > 프로젝트 > 새 프로젝트...

② 프로젝트에 추가

프로젝트를 작성하거나 연 후에는 여러 방법을 사용하여 데이터 스트림, 노드 및 보고서 등과 같은 오브젝트를 추가할 수 있습니다.

관리자에서 오브젝트 추가

IBM® SPSS® Modeler 창의 오른쪽 상단 구석에서 관리자를 사용하여 스트림이나 출력을 추가할 수 있습니다.

1. 관리자 탭 중 하나에서 테이블 또는 스트림 등과 같은 오브젝트를 선택할 수 있습니다.
2. 마우스 오른쪽 단추로 클릭하고 **프로젝트에 추가**를 클릭하십시오.
   오브젝트가 이미 저장된 경우에는 적합한 오브젝트 폴더(클래스 보기에서)나 기본 단계 폴더(CRISP-DM 보기)로 자동으로 추가됩니다.

3. 또는 관리자에서 프로젝트 분할창까지 오브젝트를 끌어다 놓을 수 있습니다.

참고: 먼저 오브젝트를 저장하도록 요청받을 수 있습니다. 저장할 때 저장 대화 상자에서 **프로젝트에 파일 추가**가 선택되었는지 확인하십시오. 그러면 저장 후에 프로젝트에 자동으로 오브젝트가 추가됩니다.

캔버스에서 노드 추가

저장 대화 상자를 사용하여 스트림 캔버스에서 개별 노드를 추가할 수 있습니다.

1. 캔버스에서 노드를 선택하십시오.
2. 마우스 오른쪽 단추를 클릭하고 **노드 저장**을 클릭하십시오. 또는 기본 메뉴에서 다음을 클릭하십시오.
   편집 > 노드 > 노드 저장...

3. 저장 대화 상자에서 **프로젝트에 파일 추가**를 선택하십시오.
4. 노드의 이름을 작성하고 **저장**을 클릭하십시오.

그러면 파일이 저장되고 프로젝트에 추가됩니다. 노드는 클래스 보기에서는 *노드* 폴더에 추가되고 CRISP-DM 보기에서는 기본 단계 폴더에 추가됩니다.

외부 파일 추가

다양한 비IBM SPSS Modeler 오브젝트를 프로젝트에 추가할 수 있습니다. 이는 IBM SPSS Modeler 내에서 전체 데이터 마이닝 프로세스를 관리할 때 유용합니다. 예를 들어, 데이터, 노트, 프리젠테이션 및 그래픽에 대한 링크를 프로젝트에 저장할 수 있습니다. CRISP-DM 보기에서 외부 파일은 사용자가 선택하는 폴더에 추가할 수 있습니다. 클래스 보기에서 외부 파일은 *기타* 폴더에만 저장할 수 있습니다.

외부 파일을 프로젝트에 추가

1. 파일을 데스크탑에서 프로젝트까지 끄십시오.
   or

2. CRISP-DM 또는 클래스 보기에서 대상 폴더를 마우스 오른쪽 단추로 클릭하십시오.
3. 메뉴에서 **폴더에 추가**를 클릭하십시오.
4. 대화 상자에서 파일을 선택하고 **열기**를 클릭하십시오.

그러면 IBM SPSS Modeler 프로젝트 내에 선택한 오브젝트에 대한 참조가 추가됩니다.

③ 프로젝트를 IBM SPSS Collaboration and Deployment Services Repository에 전송

모든 구성요소 파일을 포함하여 전체 프로젝트를 한 단계에서 IBM® SPSS® Collaboration and Deployment Services Repository로 전송할 수 있습니다. 대상 위치에 이미 있는 모든 오브젝트는 이동되지 않습니다. 이 기능은 반대로도 작동합니다. 전체 프로젝트를 IBM SPSS Collaboration and Deployment Services Repository에서 로컬 파일 시스템으로 전송할 수도 있습니다.

프로젝트 전송

전송하려는 프로젝트가 프로젝트 분할창에 열려 있는지 확인하십시오.

프로젝트 전송

1. 루트 프로젝트 폴더를 마우스 오른쪽 단추로 클릭하고 **프로젝트 전송**을 클릭하십시오.
2. 프롬프트가 표시되면 IBM SPSS Collaboration and Deployment Services Repository에 로그인하십시오.
3. 프로젝트의 새 위치를 지정하고 **확인**을 클릭하십시오.

④ 프로젝트 특성 설정

프로젝트 특성 대화 상자를 사용하여 프로젝트의 컨텐츠와 문서를 사용자 정의할 수 있습니다.
프로젝트 특성에 액세스

1. 프로젝트 분할창에서 오브젝트나 폴더를 마우스 오른쪽 단추로 클릭하고 **프로젝트 특성**을 클릭하십시오.
2. **프로젝트** 탭을 클릭하여 기본 프로젝트 정보를 지정하십시오.

작성 날짜. 프로젝트의 작성 날짜를 보여줍니다(편집 불가능).

요약. 프로젝트 보고서에 표시될 데이터 마이닝 프로젝트의 요약을 입력할 수 있습니다.
컨텐츠. 프로젝트 파일이 참조하는 구성요소 유형과 수를 나열합니다(편집 불가능).

저장되지 않은 오브젝트를 다른 이름으로 저장. 저장되지 않은 오브젝트를 로컬 파일 시스템에 저장하는지 리포지토리에 저장하는지 여부를 지정합니다. 자세한 정보는 정보 주제를 참조하십시오.

프로젝트를 로드할 때 오브젝트 참조를 업데이트. 프로젝트의 참조를 해당 구성요소로 업데이트하려면 이 옵션을 선택하십시오. *참고*: 프로젝트에 추가된 파일은 프로젝트 파일 자체에 저장되지 않습니다. 대신 파일에 대한 참조가 프로젝트에 저장됩니다. 즉 파일을 이동하거나 삭제하면 프로젝트에서 오브젝트가 제거됩니다.

⑤ 프로젝트에 주석달기

프로젝트 분할창은 데이터 마이닝 노력에 주석(Annotation)을 달기 위한 수많은 방법을 제공합니다. 프로젝트 수준 주석(Annotation)은 종종 "큰 그림" 목표와 의사결정을 추적하는 데 사용하는 반면에, 폴더나 노드 주석(Annotation)은 추가 세부사항을 제공합니다. 주석(Annotation) 탭은 검색할 수 없는 결측 데이터가 있는 데이터 제외 또는 데이터 탐색 동안 형성된 유망한 가설 등과 같은 프로젝트 수준 세부사항을 문서화하기에 충분한 공간을 제공합니다.

프로젝트에 주석(Annotation) 달기:

1. CRISP-DM 또는 클래스 보기에서 프로젝트 폴더를 선택하십시오.
2. 폴더를 마우스 오른쪽 단추로 클릭하고 **프로젝트 특성**을 클릭하십시오.
3. **주석(Annotation)** 탭을 클릭하십시오.
4. 프로젝트를 설명하는 키워드와 텍스트를 입력하십시오.

가. 폴더 특성 및 주석(Annotation)

개별 프로젝트 폴더(CRISP-DM 및 클래스 보기 둘 모두에서)에 주석(Annotation)이 작성됩니다. CRISP-DM 보기에서 이는 데이터 마이닝의 각 단계별 조직의 목표를 문서화하는 매우 효과적인 방법일 수 있습니다. 예를 들어, *비즈니스 이해* 폴더에 주석(Annotation) 도구를 사용하면 "이 연구의 비즈니스 목표는 VIP 고객 중의 이탈을 줄이기 위한 것임" 등과 같은 문서를 포함할 수 있습니다. 이 텍스트는 그런 다음 **보고서에 포함** 옵션을 선택하여 프로젝트 보고서에 자동으로 포함될 수 있습니다.

폴더에 주석(Annotation) 달기:

1. 프로젝트 분할창에서 폴더를 선택하십시오.
2. 폴더를 마우스 오른쪽 단추로 클릭하고 **폴더 특성**을 클릭하십시오.

CRISP-DM 보기에서 폴더에는 각 단계의 목적의 요약뿐만 아니라 관련 데이터 마이닝 작업 완료에 대한 안내가 주석(Annotation)으로 작성됩니다. 이러한 주석(Annotation)을 제거하거나 편집할 수 있습니다.

이름. 이 영역은 선택된 필드의 이름을 표시합니다.

도구 팁 텍스트. 마우스 포인터를 프로젝트 폴더 위에서 움직일 때 표시되는 사용자 정의 도구 팁을 작성하십시오. 이는 CRISP-DM 보기에서 유용합니다. 예를 들어, 각 단계의 목표의 빠른 개요를 제공하거나 단계의 상태를 표시하려고 할 때입니다(예: "진행 중" 또는 "완료").

주석(Annotation) 필드. 프로젝트 보고서에서 대조할 수 있는 보다 길이가 긴 주석(Annotation)에는 이 필드를 사용하십시오. CRISP-DM 보기에는 주석(Annotation)에 각 데이터 마이닝 단계의 설명이 포함되지만 사용자 고유의 프로젝트에 맞게 이를 자유롭게 사용자 정의할 수 있습니다.

보고서에 포함. 보고서에 주석(Annotation)을 포함하려면 **보고서에 포함**을 선택하십시오.

⑥ 오브젝트 특성

오브젝트 특성을 보고 개별 오브젝트를 프로젝트 보고서에 포함할지 여부를 선택할 수 있습니다. 오브젝트 특성에 액세스

1. 프로젝트 분할창에서 오브젝트를 마우스 오른쪽 단추로 클릭합니다.
2. 메뉴에서 **오브젝트 특성**을 클릭합니다.

이름. 이 영역은 저장된 오브젝트의 이름을 나열합니다.

경로. 이 영역은 저장된 오브젝트의 위치를 나열합니다.

보고서에 포함. 생성된 보고서에 오브젝트 세부사항을 포함하려면 이 옵션을 선택하십시오.

⑦ 프로젝트 닫기

IBM® SPSS® Modeler를 종료하거나 새 프로젝트를 열 때 기존 프로젝트 파일(*.cpj*)이 닫힙니다.

프로젝트와 연관된 몇몇 파일(예: 스트림, 노드 또는 그래프)은 여전히 열려 있을 수 있습니다. 이러한 파일을 열어 두려면 **... 이러한 파일을 저장하고 닫으시겠습니까?** 메시지에 아니오로 응답하십시오.
프로젝트를 닫은 후에 연관된 파일을 수정하고 저장하는 경우에는 이러한 업데이트된 버전은 다음 번에 프로젝트를 열 때 프로젝트에 포함됩니다. 이러한 작동을 막기 위해서는 프로젝트에서 파일을 제거하거나 이를 다른 파일 이름으로 저장하십시오.

(3) 보고서 생성

프로젝트의 가장 유용한 기능 중 하나는 프로젝트 항목과 주석(Annotation)을 기반으로 보고서를 생성하는 기능입니다. CRISP-DM 방법론에서 논의된 것처럼 이는 효과적인 데이터 마이닝을 위한 중요한 구성요소입니다. 보고서를 몇몇 파일 유형 중 하나로 직접 생성하거나 즉각적으로 볼 수 있도록 화면의 출력 창으로 생성할 수 있습니다. 여기에서 보고서를 웹 브라우저에서 인쇄, 저장 또는 볼 수 있습니다. 사용자의 조직에서 다른 사람에게 저장된 보고서를 배포할 수 있습니다.

보고서는 종종 프로젝트에 관련된 사람에게 배포하기 위해 데이터 마이닝 프로세스 중에 프로젝트 파일에서부터 몇 차례 생성되기도 합니다. 보고서는 프로젝트 파일뿐만 아니라 작성된 주석(Annotation)에서 참조된 오브젝트에 대한 정보를 모읍니다. 클래스 보기 또는 CRISP-DM 보기를 기반으로 보고서를 작성할 수 있습니다.

보고서 생성:

1. CRISP-DM 또는 클래스 보기에서 프로젝트 폴더를 선택하십시오.
2. 폴더를 마우스 오른쪽 단추로 클릭하고 **프로젝트 보고서**를 클릭하십시오.
3. 보고서 옵션을 지정하고 **보고서 생성**을 클릭하십시오.

보고서 대화 상자에 있는 옵션은 필요한 보고서 유형을 생성하는 몇 가지 방법을 제공합니다.

출력 이름. 보고서의 출력을 화면에 전송하기로 선택한 경우에는 출력 창의 이름을 지정하십시오. 사용자 정의 이름을 지정하거나 IBM® SPSS® Modeler가 자동으로 창에 이름을 지정하도록 두십시오.

화면으로 출력. 보고서를 생성하고 출력 창에 표시하려면 이 옵션을 선택하십시오. 출력 창으로부터 보고서를 다양한 파일 유형으로 내보내는 옵션이 있음을 유의하십시오.

파일로 출력. 보고서를 파일 유형 목록에 지정된 파일 유형으로 생성하고 저장하려면 이 옵션을 선택하십시오.

파일 이름. 생성된 보고서의 파일 이름을 지정하십시오. 파일은 기본적으로 IBM SPSS Modeler ₩*bin* 디렉토리에 저장됩니다. 생략 기호**(...)**를 사용하여 다른 위치를 지정하십시오.

파일 유형. 사용 가능한 파일 유형은 다음과 같습니다.

- **HTML 문서**. 보고서는 단일 HTML 파일로 저장됩니다. 보고서에 그래프가 포함된 경우에는 이들은 PNG 파일로 저장되고 HTML 파일에 의해 참조됩니다. 인터넷에서 보고서를 공개할 때는 HTML 파일과 함께 이 파일이 참조하는 이미지도 업로드하십시오.
- **텍스트 문서**. 보고서는 단일 텍스트 파일로 저장됩니다. 보고서에 그래프가 포함된 경우에는 파일 이름과 경로 참조만이 보고서에 포함됩니다.
- **Microsoft Word 문서**. 보고서는 단일 문서로 저장되고 그래프가 문서에 직접 임베드됩니다.
- **Microsoft Excel 문서**. 보고서가 단일 스프레드시트로 저장되고 그래프가 스프레드시트에 직접 임베드됩니다.
- **Microsoft PowerPoint 문서**. 각 단계가 새 슬라이드에 표시됩니다. 모든 그래프는 PowerPoint 슬라이드에 직접 임베드됩니다.
- **출력 오브젝트**. IBM SPSS Modeler에서 열리면 이 파일(*.cou*)은 **보고서 형식** 그룹의 **화면으로 출력** 옵션과 같습니다.

참고: Microsoft Office 파일로 내보내려면 해당하는 애플리케이션이 설치되어 있어야 합니다.

제목. 보고서의 제목을 지정하십시오.

보고서 구조. **CRISP-DM** 또는 **클래스**를 선택하십시오. CRISP-DM 보기는 "큰 그림" 시놉시스뿐만 아니라 데이터 마이닝의 각 단계에 대한 세부사항이 있는 상태 보고서를 제공합니다. 클래스 보기는 데이터와 스트림의 내부 추적에 보다 적합한 오브젝트 기반 보기입니다.

작성자. 기본 사용자 이름이 표시되지만 이를 변경할 수 있습니다.

보고서 포함. 보고서에 오브젝트를 포함하기 위한 방법을 선택하십시오. 프로젝트 파일에 추가된 모든 항목을 포함하려면 **모든 폴더 및 오브젝트**를 선택하십시오. **보고서에 포함**이 오브젝트 특성에 선택되어 있는지 여부를 기반으로 항목을 포함할 수도 있습니다. 또는 보고되지 않은 항목을 확인하려면 제외 표시된 항목만을 포함하도록 선택할 수도 있습니다(여기서 **보고서에 포함**은 선택되어 있지 않음).

선택. 이 옵션을 사용하면 보고서에서 **최근 항목**만을 선택하여 프로젝트 업데이트를 제공할 수 있습니다. 또는 **이전 항목**의 모수를 설정하여 오래되었거나 아마도 해결되지 않은 문제점을 추적할 수도 있습니다. 시간을 보고서의 모수로서 제거하려면 **모든 항목**을 선택하십시오.

출력순서. 다음 오브젝트 특성의 조합을 선택하여 이들을 폴더 내에 순서화할 수 있습니다.

- **유형**. 오브젝트를 유형별로 그룹화합니다.
- **이름**. 오브젝트를 문자순으로 구성합니다.
- **추가된 날짜**. 프로젝트에 추가된 날짜를 사용하여 오브젝트를 정렬합니다.

① 생성된 보고서 저장 및 내보내기

화면에 생성된 보고서가 새 출력 창에 표시됩니다. 보고서에 포함된 모든 그래프가 인라인 이미지로 표시됩니다.

보고서 용어

각 스트림에서 총 노드 수는 보고서 내에 나열됩니다. 숫자는 CRISP-DM 용어가 아니라 IBM® SPSS® Modeler 용어를 사용하는 다음 머리말 아래에 표시됩니다.

- 데이터 리더기. 소스 노드.
- **데이터 기록기**. 내보내기 노드.
- **모델 작성기**. 작성 또는 모델링, 노드
- **모델 적용자**. 너깃으로도 알려진 생성된 모델.
- **출력 작성기**. 그래프 또는 출력 노드.
- **기타**. 프로젝트와 관련된 다른 노드. 예를 들어, 노드 팔레트의 필드 작업 탭 또는 레코드 작업 탭에서 사용 가능한 노드입니다.

보고서 저장

1. 파일 메뉴에서 **저장**을 클릭하십시오.
2. 파일 이름을 지정하십시오.
   보고서가 출력 오브젝트로 저장됩니다.

보고서 내보내기

3. 파일 메뉴에서 **내보내기**와 내보낼 파일 유형을 클릭하십시오.
4. 파일 이름을 지정하십시오.

보고서는 선택한 형식으로 저장됩니다.

다음 파일 유형으로 내보낼 수 있습니다.

- HTML
- 텍스트
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

참고: Microsoft Office 파일로 내보내려면 해당하는 애플리케이션이 설치되어 있어야 합니다.

창 위쪽에 있는 단추를 사용하여 다음을 수행하십시오.

- 보고서를 인쇄합니다.
- 보고서를 외부 웹 브라우저에서 HTML로 봅니다.

13) IBM® SPSS® Modeler 사용자 정의

(1) IBM SPSS Modeler 옵션 사용자 정의

필요에 맞게 IBM® SPSS® Modeler를 사용자 정의하기 위해 수행할 수 있는 많은 작업이 있습니다. 주로 이 사용자 정의는 메모리 할당, 기본 디렉토리, 사운드 및 색상 사용과 같은 특정 사용자 옵션 설정으로 이루어집니다. IBM SPSS Modeler 창 맨 아래에 있는 노드 팔레트도 사용자 정의할 수 있습니다.

(2) IBM SPSS Modeler 옵션 설정

IBM® SPSS® Modeler 옵션을 사용자 정의하고 설정하는 몇 가지 방법이 있습니다.

- **도구 > 옵션** 메뉴에서 **시스템 옵션**을 클릭하여 메모리 사용, 로케일과 같은 시스템 옵션을 설정하십시오.
- **도구 > 옵션** 메뉴에서 **사용자 옵션**을 클릭하여 디스플레이 글꼴, 색상과 같은 사용자 옵션을 설정하십시오.

- **도구 > 옵션** 메뉴에서 **헬퍼 애플리케이션**을 클릭하여 IBM SPSS Modeler에 대해 작업하는 애플리케이션의 위치를 지정하십시오.
- 파일 메뉴에서 **디렉토리 설정** 또는 **서버 디렉토리 설정**을 클릭하여 IBM SPSS Modeler에서 사용하는 기본 디렉토리를 지정하십시오.

일부 또는 모든 스트림에 적용되는 옵션도 설정할 수 있습니다. 자세한 정보는 스트림의 옵션 설정의 내용을 참조하십시오.

① 시스템 옵션

도구 > 옵션 메뉴에서 **시스템 옵션**을 클릭하여 IBM® SPSS® Modeler의 선호하는 언어 또는 로케일을 지정할 수 있습니다. 여기서 SPSS Modeler의 최대 메모리 사용량을 설정하고 자동으로 스트림을 저장하는 빈도를 지정할 수도 있습니다. 이 대화 상자에서 작성된 변경사항은 SPSS Modeler를 다시 시작하기 전까지는 적용되지 않음을 유의하십시오.

최대 메모리. IBM SPSS Modeler의 메모리 사용량의 한계를 부과하려면 선택하십시오(MB 단위). 일부 플랫폼에서는 SPSS Modeler는 해당 프로세스 크기를 자원이 제한되어 있거나 로드가 많은 컴퓨터에서 통행량을 줄이기 위해 제한합니다. 대량의 데이터를 처리하고 있으면 이로 인해 "메모리 부족" 오류가 발생할 수도 있습니다. 새 임계값을 지정하여 메모리 로드를 완화할 수 있습니다.

예를 들어, 매우 큰 의사결정 트리를 표시하려고 시도하면 메모리 오류가 발생할 수 있습니다. 이 경우에는 메모리를 최대 4096Mb로 늘리는 것이 권장됩니다. 대량의 데이터를 처리하게 되는 경우에는 메모리 허용량을 늘린 후에 SPSS Modeler를 종료한 후 명령행에서 시작하여 데이터 처리 시 최대 메모리 양을 사용하는지 확인하십시오.

명령행에서 시작하려면(SPSS Modeler가 기본 위치에 설치된 것으로 가정), 명령 프롬프트 창에서 다음을 입력하십시오.

```
C:₩Program Files₩IBM₩SPSS₩Modeler₩18.3.0₩bin₩modelerclient.exe" -J-Xss4096M
```

시스템 로케일 사용. 이 옵션은 기본적으로 선택되어 있고 영어(미국)로 설정됩니다. 사용 가능한 언어 및 로케일의 목록에서 또 다른 언어를 지정하려면 선택 취소하십시오.

스트림 자동 저장 간격(분). SPSS Modeler에서 자동으로 스트림을 저장할 빈도를 지정하십시오. 최대값은 60분이고 최소값은 1분이며 기본값은 5분입니다.

가. 메모리 관리

시스템 옵션 대화 상자에 지정된 **최대 메모리** 설정에 추가로 메모리 사용을 최적화할 수 있는 몇 가지 방법이 있습니다.

- 스트림 특성 대화 상자에서 **명목 필드의 최대 멤버** 옵션을 조정하십시오. 이 옵션은 필드의 측정 수준이 *유형 없음*이 된 후에 명목 필드의 최대 멤버 수를 지정합니다. 자세한 정보는 스트림의 일반 옵션 설정의 내용을 참조하십시오.
- IBM® SPSS® Modeler이 사용 중인 메모리와 할당된 양이 표시되는 창(*xx*MB / *xx*MB) 오른쪽 하단 구석을 클릭하여 IBM SPSS Modeler이 메모리를 강제로 해제하도록 하십시오. 이 영역을 클릭하면 메모리 할당 수치가 떨어진 후에 어두운 그림자로 변합니다. 영역이 일반 색상으로 돌아가면 IBM SPSS Modeler이 가능한 모든 메모리를 해제한 것입니다.

② 기본 디렉토리 설정

파일 메뉴에서 **디렉토리 설정** 또는 **서버 디렉토리 설정**을 선택하여 파일 브라우저 및 출력에 사용하는 기본 디렉토리를 지정할 수 있습니다.

- **디렉토리 설정**. 이 옵션을 사용하여 작업 디렉토리를 설정할 수 있습니다. 기본 작업 디렉토리는 IBM® SPSS® Modeler 버전 설치 경로 또는 명령행에서 IBM SPSS Modeler를 실행하는 데 사용하는 경로를 기반으로 합니다. 로컬 모드에서 작업 디렉토리는 모든 클라이언트 측 작업 및 출력 파일(상대 경로로 참조된 경우)에 사용하는 경로입니다.
- **서버 디렉토리 설정**. 원격 서버 연결이 있을 때마다 파일 메뉴에서 서버 디렉토리 설정 옵션을 사용할 수 있습니다. 이 옵션을 사용하여 입력 또는 출력에 대해 지정된 모든 서버 파일 및 데이터 파일의 기본 디렉토리를 지정하십시오. 기본 서버 디렉토리는 *$CLEO/data*입니다(여기서 *$CLEO*는 서버 버전의 IBM SPSS Modeler가 설치된 디렉토리임). 명령행을 사용하는 경우 -server_directory 플래그를 modelerclient 명령행 인수와 함께 사용하여 이 기본값을 대체할 수도 있습니다.

③ 사용자 옵션 설정

도구 > 옵션 메뉴에서 **사용자 옵션**을 선택하여 IBM® SPSS® Modeler의 일반 옵션을 설정할 수 있습니다. 이러한 옵션은 IBM SPSS Modeler에서 사용된 모든 스트림에 적용됩니다.

다음 옵션 유형은 해당하는 탭을 클릭하여 설정할 수 있습니다.

- 모델 덮어쓰기 및 오류 메시지 등과 같은 알림 옵션.
- 그래프 및 배경 색상 등과 같은 표시 옵션.
- 구문 색상 표시 옵션.
- 모델을 예측 모델 마크업 언어(PMML)로 내보낼 때 사용하는 PMML 내보내기 옵션.
- 사용자의 이름, 이니셜 및 이메일 주소 등과 같은 사용자 또는 작성자 정보. 이 정보는 노드 및 사용자가 작성하는 기타 오브젝트의 주석(Annotation) 탭에 표시될 수 있습니다.
- 기존 모드와 분석 서버 모드 간에 전환

소수점 구분자, 시간 및 데이터 형식, 최적화, 스트림 레이아웃 및 스트림 스크립트 등과 같은 스트림 특정 옵션을 설정하려면 파일 및 도구 메뉴에서 사용 가능한 스트림 특성 대화 상자를 사용하십시오.

가. 알림 옵션 설정

사용자 옵션 대화 상자의 알림 탭을 사용하면 IBM® SPSS® Modeler에서 발생과 경고 유형 및 확인 창과 관련하여 다양한 옵션을 설정할 수 있습니다. 새 출력과 모델이 생성될 때 관리자 분할창에서 출력과 모델 탭의 작동을 지정할 수도 있습니다.

스트림 실행 피드백 대화 상자 표시 스트림이 3초 동안 실행된 후 진행 표시기를 포함하는 대화 상자를 표시하려면 선택하십시오. 대화 상자에는 또한 스트림이 작성한 출력 오브젝트의 세부사항도 포함됩니다.

- **완료 시 대화 상자 닫기** 기본적으로 스트림이 실행을 완료하면 대화 상자가 닫힙니다. 스트림이 마칠 때 대화 상자가 표시 상태로 남아 있게 하려면 이 상자를 지우십시오.

노드가 파일을 덮어쓸 때 경고 노드 작업이 기존 파일을 덮어쓸 때 오류 메시지와 함께 경고하려면 선택하십시오.

노드가 데이터베이스 테이블을 덮어쓸 때 경고 노드 작업이 기존 데이터베이스 테이블을 덮어쓸 때 오류 메시지와 함께 경고하려면 선택하십시오.

사운드 알림

이벤트나 오류가 발생할 때 사운드로 알릴지 여부를 지정하려면 목록을 사용하십시오. 여러 사운드를 사용할 수 있습니다. 선택한 사운드를 재생하려면 재생(확성기) 단추를 사용하십시오. 사운드를 찾아서 선택하려면 생략 기호**(...)**를 사용하십시오.

> **참고**: IBM SPSS Modeler에서 사운드를 작성하는 데 사용하는 .wav 파일은 설치의 /media/sounds 디렉토리에 저장됩니다.

- **모든 사운드 음소거** 모든 이벤트의 사운드 알림을 끄려면 선택하십시오.

시각적 알림

이 그룹의 옵션은 새 항목이 생성될 때 표시의 오른쪽 상단에 있는 관리자 분할창의 출력 및 모델 탭의 작동을 지정하는 데 사용합니다. 해당 탭의 작동을 지정하려면 목록에서 **새 모델** 또는 **새 출력**을 선택하십시오.

다음 옵션을 **새 모델**에 사용 가능합니다.

이전 모델 바꾸기 선택된 경우(기본값) 모델 탭의 이 스트림과 스트림 캔버스에서 기존 모델을 덮어씁니다. 이 상자가 선택 해제되면 모델은 탭과 캔버스에서 기존 모델에 추가됩니다. 이 설정은 모델 링크의 모델 대체 설정으로 대체됨을 유의하십시오.

다음 옵션을 **새 출력**에 사용 가능합니다.

출력이 [n] 초과 시 경고 출력 탭의 항목 수가 사전에 지정된 양을 초과할 때 경고를 표시할지 여부를 선택하십시오. 기본 양은 20입니다. 그러나 필요한 경우 이를 변경할 수 있습니다.

다음 옵션은 모든 경우에 사용 가능합니다.

선택 탭 스트림이 실행되는 동안 해당하는 오브젝트가 생성될 때 출력 또는 모델 탭으로 전환할지 여부를 선택합니다.

- 관리자 분할창에서 해당하는 탭으로 전환하려면 **항상**을 선택하십시오.
- 캔버스에서 현재 표시되는 스트림이 생성한 오브젝트에 대해서만 해당 탭으로 전환하려면 **현재 스트림에 의해 생성된 경우**를 선택하십시오.
- 소프트웨어가 생성된 출력이나 모델에 대해 사용자에게 알리기 위해 해당 탭으로 전환되는 것을 막으려면 **사용 안 함**을 선택하십시오.

플래시 탭 새 출력이나 모델이 생성될 때 관리자 분할창에서 출력이나 모델 탭을 플래시할지 여부를 선택하십시오.

- 관리자 분할창에서 새 오브젝트가 생성될 때마다 해당 탭을 플래시하려면(아직 선택되지 않은 경우) **선택되지 않은 경우**를 선택하십시오.
- 소프트웨어가 생성된 오브젝트에 대해 사용자에게 알리기 위해 해당 탭을 플래시하는 것을 막으려면 **사용 안 함**을 선택하십시오.

팔레트를 스크롤하여 표시(**새 모델**에만 해당). 관리자 분할창에서 가장 최신 모델을 표시하기 위해 모델 탭을 자동으로 스크롤하는지 여부를 선택하십시오.

- 스크롤링을 사용 설정하려면 **항상**을 선택하십시오.
- 캔버스에서 현재 표시되는 스트림이 생성한 오브젝트에 대해서만 스크롤하려면 **현재 스트림이 생성한 경우**를 선택하십시오.
- 소프트웨어가 모델 탭을 자동으로 스크롤링하는 것을 막으려면 **사용 안 함**을 선택하십시오.

창 열기(**새 출력**에만 해당). 생성 시 출력 창을 자동으로 열지 여부를 선택하십시오.

- 새 출력 창을 항상 열려면 **항상**을 선택하십시오.

- 캔버스에서 현재 표시되는 스트림이 생성한 출력에 대해 새 창을 열려면 **현재 스트림이 생성한 경우**를 선택하십시오.
- 소프트웨어가 생성된 출력에 대해 자동으로 새 창을 여는 것을 막으려면 **사용 안 함**을 선택하십시오.

이 탭의 시스템 기본 설정으로 돌아가려면 **기본값**을 클릭하십시오.

나. 표시 옵션 설정

사용자 옵션 대화 상자의 표시 탭을 사용하면 IBM® SPSS® Modeler에서 글꼴과 색상의 표시에 옵션을 설정할 수 있습니다.

시작할 때 시작 대화 상자 표시: 시작 시 시작 대화 상자를 표시하려면 선택하십시오. 시작 대화 상자에는 애플리케이션 예제 자습서를 실행하고, 시연 스트림이나 기존 스트림 또는 프로젝트를 열거나 새 스트림을 작성하기 위한 옵션이 있습니다.

스트림 및 수퍼노드 마크업 보이기: 선택할 경우 스트림 및 수퍼노드에 대한 마크업(있는 경우)이 기본적으로 표시됩니다. 마크업에는 스트림 주석, 모델 링크 및 스코어링 분기의 강조 표시가 포함됩니다.

표준 글꼴 및 색상(재시작 시 적용): 이 제어 상자의 옵션은 IBM SPSS Modeler 화면 디자인, 색 구성표 및 표시되는 글꼴의 크기를 지정하는 데 사용됩니다. 여기에서 선택한 옵션은 IBM SPSS Modeler를 닫고 다시 시작할 때까지 적용되지 않습니다.

- **모양과 느낌**: 표준 색 구성표와 화면 디자인을 선택하십시오. 다음에서 선택할 수 있습니다.
  - **SPSS Standard** - 기본 디자인입니다.
  - **SPSS Classic** - SPSS Modeler의 이전 버전 사용자에게 친숙한 디자인.
  - **Windows** - 스트림 캔버스와 팔레트에서 대비 증가에 사용할 수 있는 Windows 디자인.
  - **Analytics Carbon** - 매끈한 아이콘과 색상을 사용한 모던 디자인입니다.

- **노드의 기본 글꼴 크기**: 노드 팔레트 및 스트림 캔버스에 표시되는 노드에 사용할 글꼴 크기를 지정하십시오.
- **고정 너비 글꼴 지정**: 고정 너비 글꼴 및 스크립팅과 CLEM 표현식 제어에 사용할 연관된 글꼴 **크기**를 선택하려면 이 확인 상자를 선택하십시오. 기본 글꼴은 Monospace plain입니다. 선택할 수 있는 다른 글꼴 목록을 표시하려면 **변경...**을 클릭하십시오.

> **참고**: 스트림 특성 대화 상자의 옵션 탭의 레이아웃 분할창에서 스트림의 노드 아이콘 크기를 설정할 수 있습니다. 기본 메뉴에서 **도구** > **스트림 특성** > **옵션** > **레이아웃**을 선택하십시오.

사용자 정의 색상: 이 테이블에는 다양한 표시 항목에 사용되는 현재 선택된 색상이 나열됩니다. 테이블에 나열된 각 항목마다 **색상** 열에서 해당하는 행을 두 번 클릭하고 목록에서 색상을 선택하여 현재 색상을 변경할 수 있습니다. 사용자 정의 색상을 지정하려면 목록에서 맨 아래로 스크롤하고 **색상...** 항목을 클릭하십시오.

차트 범주 색상 순서: 이 테이블에는 새로 작성된 그래프에 표시하는 데 사용되는 현재 선택된 색상이 나열됩니다. 색상의 순서는 차트에 사용하는 순서를 반영합니다. 예를 들어, 색상 오버레이로 사용하는 명목 필드에 4개의 고유한 값이 포함되는 경우에는 여기에 나열되는 처음 4개의 색상만을 사용합니다. 테이블에 나열된 각 항목마다 **색상** 열에서 해당하는 행을 두 번 클릭하고 목록에서 색상을 선택하여 현재 색상을 변경할 수 있습니다. 사용자 정의 색상을 지정하려면 목록에서 맨 아래로 스크롤하고 **색상...** 항목을 클릭하십시오. 여기에서 작성한 변경사항은 이전에 작성한 그래프에는 영향을 미치지 않습니다.
이 탭의 시스템 기본 설정으로 돌아가려면 **기본값**을 클릭하십시오.

다. 구문 표시 옵션 설정

사용자 옵션 대화 상자의 구문 탭을 사용하면 IBM® SPSS® Modeler에서 작성하는 스크립트에서 글꼴 속성과 표시 색상의 옵션을 설정할 수 있습니다.

구문 강조 표시. 이 테이블은 글꼴과 표시 창 둘 모두를 포함하여 다양한 구문 항목에 사용하는 현재 선택된 색상을 나열합니다. 테이블에 나열되는 각 항목에 대해 행에서 해당하는 드롭 다운 목록을 클릭하고 목록에서 색상을 선택하여 색상을 변경할 수 있습니다. 또한 글꼴 항목의 경우 굵은체와 기울임꼴 강조를 추가할 수 있습니다.

미리보기. 이 테이블은 **구문 강조 표시** 테이블에서 선택한 색상 및 글꼴 속성을 사용하는 예제 구문 표시를 보여줍니다. 이 미리보기는 선택을 변경하는 즉시 업데이트합니다.

이 탭의 시스템 기본 설정으로 돌아가려면 **기본값**을 클릭하십시오.

라. PMML 내보내기 옵션 설정

PMML 탭에서 IBM® SPSS® Modeler가 모델을 예측 모델 마크업 언어(PMML)로 내보내는 방법을 제어할 수 있습니다. 자세한 정보는 모델을 PMML로 가져오기 및 내보내기의 내용을 참조하십시오.

PMML 내보내기. 여기에서 목표 애플리케이션과 가장 잘 작동하는 PMML의 변형을 구성할 수 있습니다.

- 표준 PMML에 상당하는 표준이 없는 특수한 경우 PMML 확장을 허용하려면 **확장 사용**을 선택하십시오. 대부분의 경우 이는 표준 PMML과 동일한 결과를 생성함을 유의하십시오.

- PMML 표준을 가능한 가장 밀접하게 고수하는 PMML을 내보내려면 **표준 PMML로...**를 선택하십시오.

표준 PMML 옵션. **표준 PMML로...** 옵션이 선택된 경우에는 선형 및 로지스틱 회귀 모델을 내보내기 위한 두 개의 유효한 방법 중 하나를 선택할 수 있습니다.

- **PMML 〈GeneralRegression〉 모델**로
- **PMML 〈Regression〉 모델**로

PMML에 대한 자세한 정보는 데이터 마이닝 그룹 웹 사이트(http://www.dmg.org)를 참조하십시오.

마. 사용자 정보 설정

사용자/작성자 정보. 여기에 입력하는 정보는 사용자가 작성하는 노드와 기타 오브젝트의 주석 탭에 표시할 수 있습니다.

(3) 노드 팔레트 사용자 정의

스트림은 노드를 사용하여 작성됩니다. IBM® SPSS® Modeler 창 맨 아래의 노드 팔레트는 스트림 작성에서 사용 가능한 모든 노드를 포함합니다. 자세한 정보는 노드 팔레트의 내용을 참조하십시오.

다음 두 가지 방법으로 노드 팔레트를 재구성할 수 있습니다.

- 팔레트 관리자 사용자 정의. 자세한 정보는 팔레트 관리자 사용자 정의의 내용을 참조하십시오.
- 하위 팔레트가 포함된 팔레트 탭이 노드 팔레트에 표시되는 방법 변경. 자세한 정보는 하위 팔레트 작성 주제를 참조하십시오.

① 팔레트 관리자 사용자 정의

팔레트 관리자는 IBM® SPSS® Modeler의 사용법에 맞춰 사용자 정의할 수 있습니다. 예를 들어, 데이터베이스에서 시계열 데이터를 자주 분석하는 경우에는 데이터베이스 소스 노드, 시간 구간 노드, 시계열 노드 및 시간 도표 그래프 노드가 함께 고유한 팔레트 탭에서 사용 가능한지 확인하려고 할 수 있습니다. 팔레트 관리자를 사용하면 노드 팔레트에서 사용자 정의 팔레트 탭을 작성하여 이러한 조정을 쉽게 할 수 있습니다.

팔레트 관리자를 사용하면 다양한 작업을 수행할 수 있습니다.

- 어떤 팔레트 탭이 스트림 캔버스 아래 노드 팔레트에 표시되는지를 제어합니다.
- 노드 팔레트에 팔레트 탭이 표시되는 순서를 변경합니다.
- 사용자 고유의 팔레트 탭 및 연관된 하위 팔레트를 작성하고 편집합니다.
- 탭에서 기본 노드 선택을 편집합니다.

팔레트 관리자에 액세스하려면 도구 메뉴에서 **팔레트 관리**를 클릭하십시오.

팔레트 이름. 노드 팔레트에 표시되는지 여부와 관계없이 사용 가능한 각 팔레트 탭이 나열됩니다. 여기에는 사용자가 작성한 모든 팔레트 탭이 포함됩니다. 자세한 정보는 팔레트 탭 작성 주제를 참조하십시오.

노드 수. 각 팔레트 탭에 표시되는 노드 수입니다. 숫자가 높을수록 탭에서 노드를 나누는 하위 팔레트를 작성하는 것이 보다 편리할 수 있습니다. 자세한 정보는 하위 팔레트 작성 주제를 참조하십시오.

표시 여부 노드 팔레트에서 팔레트 탭을 표시하려면 이 필드를 선택하십시오. 자세한 정보는 노드 팔레트에 팔레트 탭 표시 주제를 참조하십시오.

하위 팔레트. 팔레트 탭에 표시할 하위 팔레트를 선택하려면 필요한 팔레트 이름을 강조 표시하고 이 단추를 클릭하여 하위 팔레트 대화 상자를 표시하십시오. 자세한 정보는 하위 팔레트 작성 주제를 참조하십시오.

기본값 복원. 팔레트와 하위 팔레트에 작성한 모든 변경사항과 추가사항을 완전히 제거하고 기본 팔레트 설정으로 돌아가려면 이 단추를 클릭하십시오.

가. 팔레트 탭 작성

사용자 정의 팔레트 탭 작성하기

1. 도구 메뉴에서 팔레트 관리자를 여십시오.
2. *표시?* 열의 오른쪽에서 팔레트 추가 단추를 클릭하십시오. 팔레트 작성/편집 대화 상자가 표시됩니다.
3. 고유한 **팔레트 이름**을 입력하십시오.
4. **사용 가능한 노드** 영역에서 팔레트 탭에 추가할 노드를 선택하십시오.
5. 노드 추가 오른쪽 화살표 단추를 클릭하여 강조 표시된 노드를 **선택 노드** 영역으로 이동하십시오. 원하는 모든 노드를 추가할 때까지 반복하십시오.
   필요한 모든 노드를 추가한 후에는 팔레트 탭에 노드가 표시되는 순서를 변경할 수 있습니다.

6. 단순 화살표 단추를 사용하여 노드를 한 행 위나 아래로 이동하십시오.
7. 선 화살표 단추를 사용하여 노드를 목록의 맨 아래나 맨 위로 이동하십시오.
8. 팔레트에서 노드를 제거하려면 노드를 강조 표시하고 **선택 노드** 영역 오른쪽의 삭제 단추를 클릭하십시오.

나. 노드 팔레트에 팔레트 탭 표시

IBM® SPSS® Modeler 내에 사용자가 결코 사용하지 않는 옵션이 사용 가능할 수 있습니다. 이 경우 팔레트 관리자를 사용하여 이러한 노드를 포함하는 탭을 숨기면 됩니다.

노드 팔레트에 어떤 탭이 표시되는지 선택

1. 도구 메뉴에서 팔레트 관리자를 여십시오.
2. *표시?* 열의 선택란을 사용하여 각 팔레트 탭을 포함하거나 숨길지 여부를 선택합니다.

노드 팔레트에서 팔레트 탭을 영구적으로 제거하려면 노드를 강조 표시하고 *표시?* 열 오른쪽의 삭제 단추를 클릭하십시오. 팔레트 탭은 한 번 삭제하면 복구할 수 없습니다.

참고: 즐겨찾기 탭을 제외하고는 IBM SPSS Modeler과 함께 제공된 기본 팔레트 탭은 삭제할 수 없습니다.

노드 팔레트에서 표시 순서 변경

표시할 팔레트 탭을 선택한 후에는 노드 팔레트에 표시되는 순서를 변경할 수 있습니다.

1. 단순 화살표 단추를 사용하여 팔레트 탭을 한 행씩 위나 아래로 이동하십시오. 위로 이동하면 노드 팔레트의 왼쪽으로 이동되고, 역 또한 같습니다.
2. 선 화살표 단추를 사용하여 팔레트 탭을 목록의 맨 아래나 맨 위로 이동하십시오. 목록의 맨 위에 있는 팔레트는 노드 팔레트의 왼쪽에 표시됩니다.

다. 팔레트 탭에 하위 팔레트 표시

같은 방법으로 노드 팔레트에 어떤 팔레트 탭이 표시되는지를 제어하고, 해당 상위 팔레트 탭에서 어떤 하위 팔레트를 사용 가능한지를 제어할 수 있습니다.

팔레트 탭에 표시할 하위 팔레트 선택

1. 도구 메뉴에서 팔레트 관리자를 여십시오.
2. 필요한 팔레트를 선택하십시오.

3. 하위 팔레트 단추를 클릭하면 하위 팔레트 대화 상자가 표시됩니다.
4. *표시?* 열의 선택란을 사용하여 각 하위 팔레트를 팔레트 탭에 포함할지 여부를 선택합니다. **모두** 하위 팔레트는 항상 표시되고 삭제할 수 없습니다.
5. 팔레트 탭에서 하위 팔레트를 영구적으로 제거하려면 하위 팔레트를 강조 표시하고 *표시?* 열 오른쪽의 삭제 단추를 클릭하십시오.

참고: 모델링 팔레트 탭과 함께 제공된 기본 하위 팔레트는 삭제할 수 없습니다.

팔레트 탭에서 표시 순서 변경

표시할 하위 팔레트를 선택한 후에는 상위 팔레트 탭에 표시되는 순서를 변경할 수 있습니다.

1. 단순 화살표 단추를 사용하여 하위 팔레트를 한 행 위나 아래로 이동하십시오.
2. 선 화살표 단추를 사용하여 하위 팔레트를 목록의 맨 아래나 맨 위로 이동하십시오.

사용자가 작성하는 하위 팔레트는 상위 팔레트 탭을 선택할 때 노드 팔레트에 표시됩니다. 자세한 정보는 팔레트 탭 보기 변경 주제를 참조하십시오.

라. 하위 팔레트 작성

기존 노드를 사용자가 작성하는 사용자 정의 팔레트 탭에 추가할 수 있으므로 스크롤링하지 않고도 화면에 쉽게 표시할 수 있는 노드보다 많은 노드를 선택하는 것이 가능합니다. 스크롤링할 필요를 막기 위해서는 팔레트 탭에서 선택한 노드를 배치할 하위 팔레트를 작성하면 됩니다. 예를 들어, 스트림을 작성하기 위해 자주 사용하는 노드를 포함하는 팔레트 탭을 작성한 경우에는 선택사항을 소스 노드, 필드 작업, 모델링 및 출력으로 구분하는 4개의 하위 팔레트를 작성할 수 있습니다.

참고: 상위 팔레트 탭에 추가한 하위 팔레트 노드에서만 선택할 수 있습니다.

하위 팔레트 작성

1. 도구 메뉴에서 팔레트 관리자를 여십시오.
2. 하위 팔레트를 추가하려는 팔레트를 선택하십시오.
3. 하위 팔레트 단추를 클릭하면 하위 팔레트 대화 상자가 표시됩니다.
4. *표시?* 열의 오른쪽에서 하위 팔레트 추가 단추를 클릭하십시오. 하위 팔레트 작성/편집 대화 상자가 표시됩니다.
5. 고유한 **하위 팔레트 이름**을 입력하십시오.
6. **사용 가능한 노드** 영역에서 하위 팔레트 탭에 추가할 노드를 선택하십시오.
7. 노드 추가 오른쪽 화살표 단추를 클릭하여 선택한 노드를 **선택 노드** 영역으로 이동하십시오.
8. 필수 노드를 추가한 후 **확인**을 클릭하여 하위 팔레트 대화 상자로 돌아가십시오.

사용자가 작성하는 하위 팔레트는 상위 팔레트 탭을 선택할 때 노드 팔레트에 표시됩니다. 자세한 정보는 팔레트 탭 보기 변경 주제를 참조하십시오.

② 팔레트 탭 보기 변경

IBM® SPSS® Modeler에서 사용 가능한 노드가 아주 많기 때문에, 노드 팔레트의 왼쪽 또는 오른쪽으로 스크롤하지 않으면 작은 화면에 노드가 모두 표시되지는 않을 수 있습니다. 모델링 팔레트 탭에서는 특히 그렇습니다. 스크롤 필요성을 줄이기 위해 하위 팔레트(사용 가능한 경우)에 포함된 노드만 표시하도록 선택할 수 있습니다. 자세한 정보는 하위 팔레트 작성 주제를 참조하십시오.

팔레트 탭에 표시된 노드를 변경하려면 팔레트 탭을 선택한 후 왼쪽 메뉴에서 모든 노드 또는 특정 하위 팔레트의 해당 노드만 표시하도록 선택하십시오.

14) 스트림과 노드의 성능 고려사항

노드를 가장 효율적인 구성으로 배열하고 해당하는 경우 노드 캐시를 사용 가능으로 설정하고, 이 섹션에 설명된 다른 고려사항에 유의하여 성능을 최대화하도록 스트림을 계획할 수 있습니다.

여기에서 설명된 고려사항은 제외하고, 추가적 및 보다 실질적인 성능 개선사항은 일반적으로 특히 SQL 최적화를 통해 데이터베이스를 효율적으로 사용함으로써 얻을 수 있습니다.

(1) 노드의 순서

SQL 최적화를 사용 중이 아니더라도 스트림에서 노드의 순서는 성능에 영향을 미칩니다. 일반 목표는 다운스트림 처리를 최소화하는 것입니다. 그러므로 데이터 양을 줄이는 노드가 있으면 이를 스트림의 시작 부분 근처에 두십시오. IBM® SPSS® Modeler Server는 컴파일링 중에 몇몇 순서 조정 규칙을 자동으로 적용하여 특정 노드를 앞으로 가져와도 안전한 것으로 증명할 수 있을 때는 그렇게 합니다. (이 기능은 기본적으로 사용 안함으로 설정됩니다. 시스템 관리자에게 문의하여 설치에서 사용 가능한지 확인하십시오.)

SQL 최적화를 사용할 때 가용성과 효율을 극대화하려고 합니다. 최적화는 스트림에 데이터베이스에서 수행할 수 없는 작업이 포함된 경우 정지되므로 SQL 최적화된 작업은 스트림의 시작 부분에 함께 그룹화하는 것이 좋습니다. 이 전략은 데이터베이스에서 더 많은 처리를 보존하므로 IBM SPSS Modeler에 전달되는 데이터 양은 줄어듭니다.

다음 작업은 대부분의 데이터베이스에서 수행할 수 있습니다. 스트림의 *시작 부분*에 이들을 그룹화하십시오.

- 키를 기준으로 병합(결합)
- 선택
- 통합
- 정렬
- 샘플
- 붙여쓰기
- 모든 필드가 선택된 *포함* 모드에서 고유 작업
- 채움 작업
- (데이터베이스가 어떤 작업을 지원하는지에 따라) 표준 산술 또는 문자열 조작을 사용하는 기본 파생 작업
- 설정 플래그

다음 작업은 대부분의 데이터베이스에서 수행할 수 없습니다. 이들은 앞의 목록에서 작업 후 스트림에 배치해야 합니다.

- 플랫 파일 등과 같은 비데이터베이스 데이터에 대한 작업
- 순서별 병합
- 균형
- *삭제* 모드 또는 필드의 서브세트만이 고유로 선택된 곳에서 고유 작업
- 처리 중인 레코드가 아닌 레코드에서 데이터에 액세스해야 하는 모든 작업
- 상태 및 개수 필드 파생
- 히스토리 노드 작업
- "@"(시계열) 함수와 관련된 작업
- 유형 확인 모드 *경고* 및 *중단*
- 모델 구성, 애플리케이션 및 분석
  참고: 의사결정 트리, 규칙세트, 선형 회귀 및 요소 생성 모델은 SQL을 생성할 수 있으므로 데이터베이스로 다시 넣을 수 있습니다.

- 데이터를 처리 중인 동일 데이터베이스가 아닌 다른 데이터베이스로 데이터 출력

(2) 노드 캐시

실행 중인 스트림을 최적화하기 위해 비터미널 노드에서 캐시를 설정할 수 있습니다. 노드에서 캐시를 설정하면 다음 번에 데이터 스트림을 실행할 때 노드를 통과하는 데이터로 캐시가 채워집니다. 그때부터는 데이터 소스가 아닌 캐시(디스크의 임시 디렉토리에 저장됨)에서 데이터를 읽습니다.

캐싱은 정렬, 병합 또는 통합과 같이 시간이 걸리는 작업 다음으로 가장 유용합니다. 예를 들어, 데이터베이스에서 판매 데이터를 읽는 소스 노드 세트와 지역별로 판매액을 요약하는 통합 노드가 있다고 가정하십시오. 전체 데이터 세트가 아니라 통합된 데이터를 저장할 캐시를 원하기 때문에 소스 노드가 아닌 통합 노드에서 캐시를 설정할 수 있습니다.

> **참고**: 소스 노드에서 캐싱하면(IBM® SPSS® Modeler로 데이터를 읽을 때 원 데이터 세트의 사본만 저장함) 대부분의 상황에서 성능이 향상되지는 않습니다.

캐싱을 사용하는 노드에는 오른쪽 상단 모서리에 작은 문서 아이콘이 표시됩니다. 노드에서 데이터가 캐싱될 때 문서 아이콘은 녹색입니다.

캐시 사용 방법

1. 스트림 캔버스에서 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 **캐시**를 클릭하십시오.
2. 캐싱 하위 메뉴에서 **사용**을 클릭하십시오.
3. 노드를 마우스 오른쪽 단추로 클릭하고 캐싱 하위 메뉴에서 **사용 안함**을 클릭하여 캐시를 설정 해제할 수 있습니다.

데이터베이스의 노드 캐싱

데이터베이스에서 실행되는 스트림의 경우, 스트림 중간에 파일 시스템이 아닌 데이터베이스의 임시 테이블에 데이터를 캐싱할 수 있습니다. SQL 최적화와 결합되면 이로 인해 성능이 상당히 향상됩니다. 예를 들어, 다중 테이블을 병합하여 데이터 마이닝 보기를 작성하는 스트림의 출력을 캐싱하고 필요에 따라 재사용할 수 있습니다. 모든 다운스트림 노드에 대해 SQL을 자동으로 생성하면 성능이 추가로 향상될 수 있습니다.

데이터베이스 캐싱을 이용하려면 SQL 최적화와 데이터베이스 캐싱 둘 다를 사용으로 설정해야 합니다. 서버 최적화 설정은 클라이언트에서 해당 설정을 대체합니다. 자세한 정보는 스트림의 최적화 옵션 설정의 내용을 참조하십시오.

데이터베이스 캐싱을 사용으로 설정한 상태에서는 비터미널 노드를 마우스 오른쪽 단추로 클릭하여 해당 지점에서 데이터를 캐싱하면 되며, 다음 번에 스트림이 실행될 때 데이터베이스에 자동으로 캐시가 직접 작성됩니다. 데이터베이스 캐싱 또는 SQL 최적화를 사용할 수 없으면 캐시는 파일 시스템에 대신 기록됩니다.

> **참고**: 캐싱 목적으로 임시 테이블을 지원하는 데이터베이스는 Db2, Oracle, SQL Server, Teradata입니다. Netezza 등의 기타 데이터베이스는 데이터베이스 캐싱에 정규 테이블을 사용합니다. 특정 데이터베이스에 대해 SQL 코드를 사용자 정의할 수 있습니다. 지원을 받으려면 서비스팀에 문의하십시오.

(3) 성능: 프로세스 노드

정렬. 정렬 노드는 정렬하려면 먼저 전체 입력 데이터 세트를 읽어야 합니다. 데이터는 일부 한계까지 메모리에 저장되고 초과분은 디스크로 유출됩니다. 정렬 알고리즘은 조합 알고리즘입니다. 데이터는 한계까지 메모리로 읽혀지고 빠른 하이브리드 빠른 정렬 알고리즘을 사용하여 정렬

됩니다. 모든 데이터가 메모리에 맞는 경우에는 정렬이 완료됩니다. 그렇지 않은 경우에는 병합-정렬 알고리즘이 적용됩니다. 정렬된 데이터가 파일에 작성되고 데이터의 다음 청크가 메모리에 읽혀지고, 정렬된 후 디스크에 작성됩니다. 이는 모든 데이터가 읽혀질 때까지 반복된 다음 정렬된 청크가 병합됩니다. 병합하려면 디스크에 저장된 데이터 위에서 반복적으로 전달해야 합니다. 최고 사용량 시에, 정렬 노드에는 디스크에 두 개의 전체 데이터 세트 사본이 생깁니다(정렬 및 정렬되지 않음).

알고리즘의 전체 실행 시간은 대략 N*log(N)입니다. 여기서 *N*은 레코드 수입니다. 메모리에서 정렬하는 것은 디스크에서 병합하는 것보다 빠르므로 실제 실행 시간은 정렬에 더 많은 메모리를 할당하여 줄일 수 있습니다. 알고리즘은 IBM® SPSS® Modeler Server 구성 옵션 *메모리 사용 승수*에 의해 제어되는 물리적 RAM의 일부를 자신에게 할당합니다. 정렬에 사용된 메모리를 늘리기 위해서는 더 많은 물리적 RAM을 제공하거나 이 값을 늘리십시오. 사용한 메모리의 비율이 프로세스의 작업 세트를 초과하면 메모리의 일부가 디스크에 페이징될 수 있도록 성능이 저하됨을 유의하십시오. 인메모리 정렬 알고리즘의 메모리 액세스 패턴은 임의적이고 과도한 페이징을 야기할 수 있기 때문입니다. 정렬 알고리즘은 정렬 노드가 아닌 몇몇 노드에서 사용하지만 동일한 성능 규칙이 적용됩니다.

구간화. 구간화 노드는 구간에 레코드를 할당하기 전에 구간 경계를 계산하기 위해 전체 입력 데이터 세트를 읽습니다. 경계가 계산되는 동안 데이터 세트가 캐싱됩니다. 그런 다음 할당을 위해 다시 스캔됩니다. 구간화 방법이 *고정 너비*이거나 *평균+표준 편차*이면 데이터 세트는 디스크에 직접 캐싱됩니다. 이러한 방법에는 선형 실행 시간이 있고 전체 데이터 세트를 저장하기 위한 충분한 디스크 공간이 필요합니다. 구간화 방법이 *순위*이거나 *분위수*이면 데이터 세트는 앞서 설명된 정렬 알고리즘을 사용하여 정렬되고 정렬된 데이터 세트는 캐시로 사용합니다. 정렬은 이러한 방법에 M*N*log(N) 실행 시간을 제공합니다. 여기서 *M*은 구간화 필드의 수이고 *N*은 레코드 수입니다. 데이터 세트 크기의 두 배에 해당하는 디스크 공간이 필요합니다.

생성된 구간을 기반으로 파생 노드를 생성하면 후속 전달 시에 성능이 향상됩니다. 파생 작업은 구간화보다 훨씬 빠릅니다.

키별로 병합(결합). 병합 방법이 *키*이면(데이터베이스 결합에 해당) 병합 노드는 해당 입력 데이터 세트를 키 필드별로 정렬합니다. 프로시저의 이 부분은 실행 시간이 M*N*log(N)입니다. 여기서 *M*은 입력 수이고 *N*은 최신 입력에 있는 레코드 수입니다. 해당 입력 데이터 세트와 가장 큰 데이터 세트의 두 번째 사본을 모두 저장하려면 충분한 디스크 공간이 필요합니다. 병합 자체의 실행 시간은 출력 데이터 세트의 크기에 비례하며 이는 매치하는 키의 빈도에 따라 다릅니다. 출력이 입력의 데카르트 곱인 최악의 경우 실행 시간은 NM에 접근할 수 있습니다. 이는 드문 경우이며 대부분의 결합은 매치하는 키 수가 훨씬 적습니다. 하나의 데이터 세트가 다른 데이터 세트보다 상대적으로 크거나 수신 데이터가 이미 키 필드별로 정렬되어 있으면 최적화 탭을 사용하여 이 노드의 성능을 개선할 수 있습니다.

통합. *키가 인접함* 옵션이 설정되지 않은 경우 이 노드는 통합 출력을 생성하기 전에 전체 입력 데이터 세트를 읽지만 저장하지는 않습니다. 통합 데이터의 크기가 한계(IBM SPSS Modeler

Server 구성 옵션 *메모리 사용 승수*에 의해 결정됨)에 도달하는 보다 극단적인 경우 나머지 데이터 세트는 *키가 인접함* 옵션이 설정된 것처럼 정렬되고 처리됩니다. 이 옵션이 설정되면 통합 출력 레코드는 입력 데이터를 읽을 때 생성되므로 데이터가 저장되지 않습니다.

고유. 고유 노드는 모든 고유한 키 필드를 입력 데이터 세트에 저장합니다. 모든 필드가 키 필드이고 모든 레코드가 고유한 경우 이는 전체 데이터 세트를 저장합니다. 기본적으로 고유 노드는 키 필드에서 데이터를 정렬한 다음 각 그룹에서 맨 처음 고유 레코드를 선택 또는 삭제합니다. 고유 키 개수가 적은 더 작은 데이터 세트 또는 미리 정렬된 데이터 세트의 경우 처리의 속도와 효율성을 높이는 옵션을 선택할 수 있습니다.

유형. 몇몇 경우에는 유형 노드는 값을 읽을 때 입력 데이터를 캐싱합니다. 캐시는 다운스트림 처리에 사용합니다. 캐시에는 전체 데이터 세트를 저장하려면 충분한 디스크 공간이 필요하지만 처리 속도를 높입니다.

평가. 평가 노드는 분위수를 계산하기 위해 입력 데이터를 정렬해야 합니다. 스코어와 결과적 레코드 순서가 서로 다르기 때문에 평가된 각 모델마다 정렬이 반복됩니다. 실행 시간은 M*N*log(N)입니다. 여기서 *M*은 모델 수이고 *N*은 레코드 수입니다.

(4) 성능: 모델링 노드

신경망 및 코호넨. 코호넨 알고리즘을 포함하여 신경망 학습 알고리즘은 학습 데이터 위에서 전달하는 것이 많습니다. 데이터는 일부 한계까지 메모리에 저장되고 초과분은 디스크로 유출됩니다. 디스크에서 학습 데이터에 액세스하는 방법은 액세스 방법이 임의적이어서 과도한 디스크 활동으로 이어질 수 있으므로 비용이 많이 듭니다. 이러한 알고리즘의 경우 노드의 대화 상자의 모델 탭에서 **속도 최적화** 옵션을 선택하여 디스크 저장 공간 사용을 사용 안함으로 설정하여 모든 데이터가 메모리에 저장되도록 강제할 수 있습니다. 데이터를 저장하는 데 필요한 메모리 양이 서버가 처리하는 작업 세트보다 큰 경우 데이터의 일부는 디스크로 페이징되고 성능도 이에 따라 저하됩니다.

메모리 최적화가 사용 가능하면 물리적 RAM의 퍼센트는 IBM® SPSS® Modeler Server 구성 옵션 *모델링 메모리 한계 퍼센트*의 값에 따라 알고리즘에 할당됩니다. 신경망 훈련을 위해 더 많은 메모리를 사용하려면 더 많은 RAM을 제공하거나 이 옵션의 값을 늘리십시오. 그러나 값을 너무 높게 설정하면 페이징이 발생함을 유의하십시오.

신경망 알고리즘의 실행 시간은 필요한 정확도 수준에 따라 다릅니다. 노드의 대화 상자에서 중지 조건을 설정하여 실행 시간을 제어할 수 있습니다.

K-평균. K-평균 군집 알고리즘에는 메모리 사용을 제어하기 위한 신경망 알고리즘과 동일한 옵션이 있습니다. 그러나 데이터에 대한 액세스가 순차적이므로 디스크에 저장된 데이터에 대한 성능이 더 낫습니다.

(5) 성능: CLEM 표현식

데이터 스트림을 되돌아보는 CLEM 순차규칙 함수("@ 함수")는 가장 긴 되돌아보기를 충족하기 위해서는 충분한 데이터를 저장해야 합니다. 되돌아보기의 정도가 제한이 없는 작업의 경우에는 필드의 모든 값을 저장해야 합니다. 제한이 없는 작업은 오프셋 값이 리터럴 정수가 아닌 작업입니다(예: @OFFSET(Sales, Month)). 오프셋 값은 값이 실행될 때까지 알려지지 않은 필드 이름 *월*입니다. 서버는 정확한 결과를 위해서는 *판매* 필드의 모든 값을 저장해야 합니다. 상한선이 알려진 곳에서는 값을 추가 인수로서 제공해야 합니다(예: @OFFSET(Sales, Month, 12)). 이 작업은 서버에게 *판매*의 최근 12개 값을 초과하여 저장하지 않도록 지시합니다. 순차규칙 함수는 제한이 있거나, 그렇지 않은 경우에는 거의 항상 SQL 생성을 금지합니다.

15) IBM® SPSS® Modeler의 내게 필요한 옵션

(1) IBM SPSS Modeler의 내게 필요한 옵션 개요

IBM® SPSS® Modeler는 시각적 및 기타 기능적 장애가 있는 사용자에 대한 특정 지원은 물론 모든 사용자에게 내게 필요한 옵션 지원을 제공합니다. 이 절에서는 스크린 리더, 키보드 단축키와 같은 내게 필요한 옵션 향상을 사용하여 작업하는 방법 및 기능에 대해 설명합니다.

(2) 내게 필요한 옵션 지원 유형

시각 장애가 있거나 키보드에 조작을 의존하는 경우 이 데이터 마이닝 툴킷을 사용하기 위한 매우 다양한 대체 방법이 있습니다. 예를 들어, 스트림 빌드, 옵션 지정, 출력 읽기를 마우스를 사용하지 않고 모두 수행할 수 있습니다. 사용 가능한 키보드 단축키는 뒤따르는 주제에 나열되어 있습니다. 또한 IBM® SPSS® Modeler는 Windows용 JAWS와 같은 스크린 리더에 대한 광범위한 지원을 제공합니다. 또한 색 구성표를 최적화하여 추가 대비를 제공할 수 있습니다. 이러한 유형의 지원은 다음 주제에서 설명합니다.

① 시각 장애인을 위한 내게 필요한 옵션

소프트웨어 사용 능력을 향상시킬 다수의 특성을 IBM® SPSS® Modeler에서 지정할 수 있습니다.

표시 옵션

그래프 표시 색상을 선택할 수 있습니다. 또한 소프트웨어 자체에 특정 Windows 설정을 사용하도록 선택할 수 있습니다. 이렇게 하면 시각적 대비를 증대하는 데 도움이 됩니다.

1. 표시 옵션을 설정하려면 도구 메뉴에서 **사용자 옵션**을 클릭하십시오.
2. **디스플레이** 탭을 클릭하십시오. 이 탭의 옵션으로는 소프트웨어 색 구성표, 차트 색상, 노드 글꼴 크기가 있습니다.

> **참고**: 스크린 리더에서 그래프를 읽을 수 없으므로 시각 장애 사용자가 액세스할 수 없습니다.

알림을 위한 소리 사용

소리를 켜거나 꺼서 소프트웨어의 특정 조작에 대한 경고를 받는 방법을 제어할 수 있습니다. 예를 들어, 노드 작성 및 삭제 또는 새 출력이나 모델 생성과 같은 이벤트에 대한 소리를 활성화할 수 있습니다.

1. 알림 옵션을 설정하려면 도구 메뉴에서 **사용자 옵션**을 클릭하십시오.
2. **알림** 탭을 클릭하십시오.

새 창 자동 시작 제어

사용자 옵션 대화 상자의 알림 탭은 테이블 및 차트와 같이 새로 생성된 출력이 별도의 창에서 시작되는지 여부를 제어하는 데도 사용합니다. 이 옵션을 사용 안함으로 설정하고 필요할 때만 출력 창을 여는 것이 용이합니다.

1. 이러한 옵션을 설정하려면 도구 메뉴에서 **사용자 옵션**을 클릭하십시오.
2. **알림** 탭을 클릭하십시오.
3. 대화 상자의 **시각적 알림** 그룹에 있는 목록에서 **새 출력**을 선택하십시오.
4. **창 열기** 아래에서 **사용 안 함**을 선택하십시오.

노드 크기

표준 또는 작은 크기를 사용하여 노드를 표시할 수 있습니다. 필요에 맞게 이러한 크기를 조정할 수 있습니다.

1. 노드 크기 옵션을 설정하려면 파일 메뉴에서 **스트림 특성**을 클릭하십시오.
2. **레이아웃** 탭을 클릭하십시오.
3. **아이콘 크기** 목록에서 **표준**을 선택하십시오.

② 맹인 사용자를 위한 내게 필요한 옵션

맹인 사용자를 위한 지원은 대부분 Windows용 JAWS와 같은 스크린 리더 사용에 의존합니다. IBM® SPSS® Modeler에서 스크린 리더 사용을 최적화하기 위해 다수의 설정을 지정할 수 있습니다.

표시 옵션

스크린 리더는 화면에서 시각적 대비가 클때 최상의 성능을 보이는 경향이 있습니다. 이미 고대비 Windows 설정이 있는 경우, 소프트웨어 자체에 이러한 Windows 설정을 사용하도록 선택할 수 있습니다.

1. 표시 옵션을 설정하려면 도구 메뉴에서 **사용자 옵션**을 클릭하십시오.
2. **디스플레이** 탭을 클릭하십시오.

ⓘ **참고**: 스크린 리더에서 그래프를 읽을 수 없으므로 맹인 사용자가 액세스할 수 없습니다.

알림을 위한 소리 사용

소리를 켜거나 꺼서 소프트웨어의 특정 조작에 대한 경고를 받는 방법을 제어할 수 있습니다. 예를 들어, 노드 작성 및 삭제 또는 새 출력이나 모델 생성과 같은 이벤트에 대한 소리를 활성화할 수 있습니다.

1. 알림 옵션을 설정하려면 도구 메뉴에서 **사용자 옵션**을 클릭하십시오.
2. **알림** 탭을 클릭하십시오.

새 창 자동 시작 제어

사용자 옵션 대화 상자의 알림 탭은 새로 생성된 출력이 별도의 창에서 시작되는지 여부를 제어하는 데도 사용합니다. 이 옵션을 사용 안함으로 설정하고 필요에 따라 출력 창을 여는 것이 용이합니다.

1. 이러한 옵션을 설정하려면 도구 메뉴에서 **사용자 옵션**을 클릭하십시오.
2. **알림** 탭을 클릭하십시오.
3. 대화 상자의 **시각적 알림** 그룹에 있는 목록에서 **새 출력**을 선택하십시오.
4. **창 열기** 아래에서 **사용 안 함**을 선택하십시오.

③ 내게 필요한 옵션의 키보드 기능

키보드에서 제품 기능에 액세스할 수 있습니다. 가장 기본적인 레벨에서 Alt + 해당 키를 눌러 창 메뉴를 활성화(예: Alt+F를 눌러 파일 메뉴에 액세스)하거나 Tab 키를 눌러 대화 상자 제어를 스크롤할 수 있습니다. 그러나 제품의 각 기본 창과 관련된 특정 문제와 대화 상자 탐색에 유용한 힌트가 있습니다.

이 절에서는 스트림 열기에서 노드 대화 상자를 사용하여 출력에 대해 작업하는 방법에 이르기까지 내게 필요한 옵션의 키보드 기능의 강조 표시에 대해 설명합니다. 또한 보다 효율적인 탐색을 위해 키보드 단축키 목록도 제공됩니다.

가. 기본 창 탐색을 위한 단축키

IBM® SPSS® Modeler의 기본 창에서 데이터 마이닝 작업을 대부분 수행합니다. 기본 영역을 **스트림 캔버스**라고 하며 데이터 스트림을 빌드하고 실행하는 데 사용합니다. 창의 하단부는 사용 가능한 노드를 포함하는 **노드 팔레트**를 포함합니다. 팔레트는 노드의 각 그룹에 대한 데이터 마이닝 작업 유형에 해당하는 탭에 구성됩니다. 예를 들어, 데이터를 IBM SPSS Modeler로 가져오는 데 사용하는 노드는 소스 탭에 그룹화되고 필드 파생, 필터링 또는 입력에 사용하는 노드는 필드 작업 탭에 그룹화됩니다.
창의 오른쪽에는 스트림, 출력, 프로젝트 관리를 위한 몇 개의 도구가 있습니다. 오른쪽의 상단 절반은 **관리자**를 포함하며 스트림, 출력, 생성된 모델을 관리하는 데 사용하는 세 개의 탭이 있습니다. 탭과 오브젝트를 목록에서 선택하여 이러한 오브젝트에 액세스할 수 있습니다. 오른쪽의 하단 절반은 **프로젝트 분할창**을 포함하는데, 이를 사용하여 프로젝트에 대한 작업을 구성할 수 있습니다. 이 영역에는 프로젝트에 대한 서로 다른 두 개의 보기를 반영하는 두 개의 탭이 있습니다. **클래스 보기**는 프로젝트 오브젝트를 유형별로 정렬하는 반면, **CRISP-DM 보기**는 데이터 준비 또는 모델링과 같은 관련 데이터 마이닝 단계별로 오브젝트를 정렬합니다. IBM SPSS Modeler 창의 이러한 다양한 측면은 도움말 시스템과 사용자 안내서 전반에 설명되어 있습니다.

다음은 기본 IBM SPSS Modeler 창 내에서의 이동 및 스트림 빌드에 사용하는 단축키 표입니다. 대화 상자 및 출력 단축키는 뒤따르는 주제에 나열되어 있습니다. 이러한 단축키는 기본 창에서만 사용할 수 있습니다.

표 1. 기본 창 단축키

단축키	기능
Ctrl+F5	노드 팔레트로 초점을 이동합니다.
Ctrl+F6	스트림 캔버스로 초점을 이동합니다.
Ctrl+F7	관리자 분할창으로 초점을 이동합니다.
Ctrl+F8	프로젝트 분할창으로 초점을 이동합니다.

표 2. 노드 및 스트림 단축키

단축키	기능
Ctrl+N	빈 스트림 캔버스를 새로 작성합니다.
Ctrl+O	열기 대화 상자를 표시합니다(여기서 기존 스트림을 선택하고 열 수 있음).
Ctrl+숫자 키	창 또는 분할창의 해당 탭으로 초점을 이동합니다. 예를 들어, 탭 지정된 분할창 또는 창 내에서 Ctrl+1은 왼쪽에서 시작하여 첫 번째 탭으로 이동하고 Ctrl+2는 두 번째 탭으로 이동하는 식입니다.
Ctrl+아래로 화살표	팔레트 탭에서 해당 탭 아래의 첫 번째 노드로 초점을 이동하기 위해 노드 팔레트에서 사용합니다.
Ctrl+위쪽 화살표	노드에서 팔레트 탭으로 초점을 이동하기 위해 노드 팔레트에서 사용합니다.
Enter	노드 팔레트에서 노드가 선택된 경우(생성된 모델 팔레트의 세분화된 모델 포함) 이 키 입력은 스트림 캔버스에 노드를 추가합니다. 캔버스에서 노드가 이미 선택된 경우 **Enter**를 누르면 해당 노드에 대한 대화 상자가 열립니다.
Ctrl+Enter	팔레트에서 노드가 선택된 경우 노드를 선택하지 않고 스트림 캔버스에 해당 노드를 추가하고 팔레트의 첫 번째 노드로 초점을 이동합니다.
Alt+Enter	팔레트에서 노드가 선택된 경우 스트림 캔버스에 해당 노드를 추가하고 팔레트의 첫 번째 노드로 초점을 이동하는 동안 노드를 선택합니다.
Shift+스페이스바	노드나 주석이 팔레트에서 초점을 가진 경우 해당 노드나 주석 선택과 선택 취소 간에 토글합니다. 다른 노드나 주석도 선택된 경우 선택 취소됩니다.
Ctrl+Shift+스페이스바	노드나 주석이 스트림에서 초점을 가진 경우 또는 노드나 주석이 팔레트에서 초점을 가진 경우 노드나 주석 선택과 선택 취소 간에 토글합니다. 선택된 다른 노드나 주석에는 영향을 주지 않습니다.
왼쪽/오른쪽 화살표	스트림 캔버스가 초점을 가진 경우 전체 스트림을 화면에서 수평으로 이동합니다. 팔레트 탭이 초점을 가진 경우 탭 간에 순환합니다. 팔레트 노드가 초점을 가진 경우 팔레트의 노드 간에 이동합니다.
위로/아래로 화살표	스트림 캔버스가 초점을 가진 경우 전체 스트림을 화면에서 수직으로 이동합니다. 팔레트 노드가 초점을 가진 경우 팔레트의 노드 간에 이동합니다. 하위 팔레트가 초점을 가진 경우 이 팔레트 탭의 다른 하위 팔레트 간에 이동합니다.

단축키	기능
Alt+왼쪽/오른쪽 화살표	스트림 캔버스에서 선택된 노드와 주석을 화살표 키 방향으로 수평으로 이동합니다.
Alt+위로/아래로 화살표	스트림 캔버스에서 선택된 노드와 주석을 화살표 키 방향으로 수직으로 이동합니다.
Ctrl+A	스트림의 모든 노드를 선택합니다.
Ctrl+Q	노드가 초점을 가진 경우 해당 노드와 모든 노드 다운스트림을 선택하고 모든 노드 업스트림을 선택 취소합니다.
Ctrl+W	선택된 노드가 초점을 가진 경우 해당 노드와 선택된 모든 노드 다운스트림을 선택 취소합니다.
Ctrl+Alt+D	선택된 노드를 복제합니다.
Ctrl+Alt+L	스트림에서 모델 너깃이 선택된 경우 삽입 대화 상자를 열어 저장된 모델을 .nod 파일에서 스트림으로 로드할 수 있게 합니다.
Ctrl+Alt+R	선택된 노드에 대한 주석(Annotation) 탭을 표시하여 노드 이름을 변경할 수 있게 합니다.
Ctrl+Alt+U	사용자 입력 소스 노드를 작성합니다.
Ctrl+Alt+C	노드 캐시 설정 또는 해제를 토글합니다.
Ctrl+Alt+F	노드 캐시를 비웁니다.
Tab	스트림 캔버스에서 현재 스트림의 모든 소스 노드와 주석을 순환합니다. 노드 팔레트에서 팔레트의 노드 간에 이동합니다. 선택된 하위 팔레트에서 하위 팔레트의 첫 번째 노드로 이동합니다.
Shift+Tab	Tab과 동일한 작업을 역순으로 수행합니다.
Ctrl+Tab	관리자 분할창 또는 프로젝트 분할창에 초점을 맞춘 상태에서 스트림 캔버스로 초점을 이동합니다. 노드 팔레트에 초점을 맞춘 상태에서 노드와 해당 팔레트 탭 간에 초점을 이동합니다.
임의의 알파벳 키	현재 스트림의 노드에 초점을 맞춘 상태에서 이름이 눌려진 키로 시작하는 다음 노드에 초점을 맞추고 이동합니다.
F1	초점 관련 주제에서 도움말 시스템을 엽니다.
F2	캔버스에서 선택된 노드에 대한 연결 프로세스를 시작합니다. Tab 키를 사용하여 캔버스의 필수 노드로 이동하고 Shift+스페이스바를 눌러 연결을 마치십시오.
F3	캔버스에서 선택된 노드에 대한 연결을 모두 삭제합니다.
F6	관리자 분할창, 프로젝트 분할창, 노드 팔레트 간에 초점을 이동합니다.

단축키	기능
F10	파일 메뉴를 엽니다.
Shift+F10	노드 또는 스트림에 대한 팝업 메뉴를 엽니다.
Delete	캔버스에서 선택된 노드를 삭제합니다.
Esc	팝업 메뉴 또는 대화 상자를 닫습니다.
Ctrl+Alt+X	수퍼노드를 펼칩니다.
Ctrl+Alt+Z	수퍼노드를 확대합니다.
Ctrl+Alt+Shift+Z	수퍼노드를 축소합니다.
Ctrl+E	스트림 캔버스에 초점을 맞춘 상태에서 현재 스트림을 실행합니다.

IBM SPSS Modeler에서는 다수의 표준 단축키도 사용합니다(예: 복사의 경우 Ctrl+C). 자세한 정보는 단축키 사용의 내용을 참조하십시오.

나. 대화 상자 및 테이블 단축키

대화 상자, 테이블, 대화 상자의 테이블에 대한 작업 시 몇 개의 단축키와 스크린 리더 키가 유용합니다. 특수 키보드 및 스크린 리더 단축키 전체 목록은 다음과 같습니다.

표 1. 대화 상자 및 표현식 작성기 단축키

단축키	기능
Alt+4	열린 대화 상자 또는 출력 창을 모두 종료하는 데 사용합니다. 관리자 분할창의 출력 탭에서 출력을 검색할 수 있습니다.
Ctrl+End	표현식 작성기의 임의 제어에 초점을 맞춘 상태에서 표현식의 끝으로 삽입점을 이동합니다.
Ctrl+1	표현식 작성기에서 표현식 편집 제어로 초점을 이동합니다.
Ctrl+2	표현식 작성기에서 함수 목록으로 초점을 이동합니다.
Ctrl+3	표현식 작성기에서 필드 목록으로 초점을 이동합니다.

테이블 단축키

테이블 단축키는 유형, 필터, 병합과 같은 노드에 대해 대화 상자의 테이블 제어에는 물론 출력 테이블에도 사용합니다. 일반적으로 테이블 셀 간의 이동에는 Tab 키를 사용하고 테이블 제어를 종료하는 데는 Ctrl+Tab을 사용합니다. *참고*: 때때로 스크린 리더가 셀 컨텐츠 읽기를 즉시 시작하지 않을 수도 있습니다. 화살표 키를 한 번 또는 두 번 누르면 소프트웨어가 재설정되고 음성을 시작합니다.

표 2. 테이블 단축키

단축키	기능
Ctrl+W	테이블에 대해 선택된 **행**에 대한 간단한 설명을 읽습니다. 예를 들어, "선택된 행 2 값은 성, 플래그, 남/여 등입니다."
Ctrl+Alt+W	테이블에 대해 선택된 **행**에 대한 자세한 설명을 읽습니다. 예를 들어, "선택된 행 2 값은 필드 = 성, 유형 = 플래그, 성 = 남/여 등입니다."
Ctrl+D	테이블에 대해 선택된 영역에 대한 간단한 **설명**을 읽습니다(예: "선택은 6열당 1행임").
Ctrl+Alt+D	테이블에 대해 선택된 영역에 대한 자세한 **설명**을 제공합니다(예: "선택은 6열당 1행임. 선택된 열은 필드, 유형, 결측값임. 선택 행은 1임").
Ctrl+T	테이블에 대해 선택된 열에 대한 간단한 설명을 제공합니다(예: "필드, 유형, 결측값").
Ctrl+Alt+T	테이블에 대해 선택된 열에 대한 자세한 설명을 제공합니다(예: "선택된 열은 필드, 유형, 결측값임").
Ctrl+R	테이블에 대해 테이블의 **레코드** 수를 제공합니다.
Ctrl+Alt+R	테이블에 대해 테이블의 **레코드** 수는 물론 열 이름도 제공합니다.
Ctrl+I	테이블에 대해 초점이 있는 셀의 셀 **정보** 또는 컨텐츠를 읽습니다.
Ctrl+Alt+I	테이블에 대해 초점이 있는 셀의 셀 **정보**(열 이름 및 셀 컨텐츠)에 대한 자세한 설명을 읽습니다.
Ctrl+G	테이블에 대해 간단한 **일반** 선택 정보를 제공합니다.
Ctrl+Alt+G	테이블에 대해 자세한 **일반** 선택 정보를 제공합니다.
Ctrl+Q	테이블에 대해 테이블 셀 **빠른** 토글을 제공합니다. Ctrl+Q는 화살표 키를 사용하여 테이블을 이동할 때 "성=여성"과 같은 자세한 설명을 읽습니다. Ctrl+Q를 다시 선택하면 간단한 설명(셀 컨텐츠)으로 토글됩니다.
F8	테이블에 대해 초점이 테이블이면 열 헤더로 초점을 설정합니다.
스페이스바	테이블에 대해 초점이 열 헤더이면 열 정렬을 사용 가능하게 합니다.

다. 주석 단축키

온스크린 주석에 대한 작업 시 다음 단축키를 사용할 수 있습니다.

표 1. 주석 단축키

단축키	기능
Alt+C	주석 표시/숨기기 기능을 토글합니다.
Alt+M	주석이 현재 표시된 경우 새 주석을 삽입합니다. 주석이 현재 숨겨진 경우 주석을 표시합니다.
Tab	스트림 캔버스에서 현재 스트림의 모든 소스 노드와 주석을 순환합니다.
Enter	주석에 초점이 있으면 편집 시작을 표시합니다.
Alt+Enter 또는 Ctrl+Tab	편집을 종료하고 편집 변경사항을 저장합니다.
Esc	편집을 취소합니다. 편집 중에 작성된 변경사항은 유실됩니다.
Alt+Shift+위쪽 화살표	눈금에 정렬이 설정(또는 설정 해제)되어 있으면 1눈금 셀(또는 1픽셀)씩 텍스트 영역의 높이를 줄입니다.
Alt+Shift+아래로 화살표	눈금에 정렬이 설정(또는 설정 해제)되어 있으면 1눈금 셀(또는 1픽셀)씩 텍스트 영역의 높이를 늘립니다.
Alt+Shift+왼쪽 화살표	눈금에 정렬이 설정(또는 설정 해제)되어 있으면 1눈금 셀(또는 1픽셀)씩 텍스트 영역의 너비를 줄입니다.
Alt+Shift+오른쪽 화살표	눈금에 정렬이 설정(또는 설정 해제)되어 있으면 1눈금 셀(또는 1픽셀)씩 텍스트 영역의 너비를 늘립니다.

라. 군집 뷰어 및 모델 뷰어 단축키

군집 뷰어 및 모델 뷰어 창 주위를 이동하는 데 단축키를 사용할 수 있습니다.

표 1. 일반 단축키 - 군집 뷰어 및 모델 뷰어

단축키	기능
Tab	다음 화면 제어로 초점을 이동합니다.
Shift+Tab	이전 화면 제어로 초점을 이동합니다.

단축키	기능
아래로 화살표	드롭 다운 목록에 초점이 있으면 목록을 열거나 목록의 다음 항목으로 이동합니다. 메뉴에 초점이 있으면 메뉴의 다음 항목으로 이동합니다. 썸네일 그래프에 초점이 있으면 세트의 다음 항목(또는 마지막 썸네일에 초점이 있으면 첫 번째 항목)으로 이동합니다.
위쪽 화살표	드롭 다운 목록이 열려 있으면 목록의 이전 항목으로 이동합니다. 메뉴에 초점이 있으면 메뉴의 이전 항목으로 이동합니다. 썸네일 그래프에 초점이 있으면 세트의 이전 항목(또는 첫 번째 썸네일에 초점이 있으면 마지막 항목)으로 이동합니다.
Enter	열려 있는 드롭 다운 목록을 닫거나 열린 메뉴를 선택합니다.
F6	창의 왼쪽 및 오른쪽 분할창 사이에서 초점을 토글합니다.
왼쪽 및 오른쪽 화살표	탭에 초점이 있으면 이전 또는 다음 탭으로 이동합니다. 메뉴에 초점이 있으면 이전 또는 다음 메뉴로 이동합니다.
Alt+*문자*	해당 이름의 이 문자에 밑줄이 쳐져 있는 단추 또는 메뉴를 선택합니다.
Esc	열린 메뉴 또는 드롭 다운 목록을 닫습니다.

군집 뷰어만

군집 뷰어에는 기능별 군집 눈금이 포함된 군집 뷰어가 있습니다.

모델 요약 보기 대신 군집 보기를 선택하려면 다음을 수행하십시오.

1. 보기 단추가 선택될 때까지 반복해서 탭을 누르십시오.
2. 아래로 화살표를 두 번 눌러 군집을 선택하십시오.
   여기서부터 눈금 내 개별 셀을 선택할 수 있습니다.
3, 시각화 도구 모음의 마지막 아이콘에 도착할 때까지 반복해서 탭을 누르십시오.

   그림 1. 시각화 트리 표시 아이콘

4. Tab을 한 번 더 누른 후 스페이스바를 누르고 화살표 키를 누르십시오.
   이제 다음 키보드 단축키를 사용할 수 있습니다.

표 2. 군집 뷰어 단축키

단축키	기능
화살표 키	눈금의 개별 셀 간에 초점을 이동합니다. 초점이 이동함에 따라 오른쪽 분할창의 셀 분포 표시가 변경됩니다.

단축키	기능
Ctrl+,(쉼표)	셀에 초점이 있는 눈금의 전체 열을 선택하거나 선택 취소합니다. 선택에 열을 추가하려면 화살표 키를 사용하여 해당 열의 셀로 이동한 후 Ctrl+,를 다시 누르십시오.
Tab	초점이 눈금에서 벗어나 다음 화면 제어로 이동합니다.
Shift+Tab	초점이 눈금에서 벗어나 이전 화면 제어로 다시 이동합니다.
F2	편집 모드에 진입합니다(레이블 및 설명 셀만).
Enter	편집 변경사항을 저장하고 편집 모드를 종료합니다(레이블 및 설명 셀만).
Esc	편집 변경사항을 저장하지 않고 편집 모드를 종료합니다(레이블 및 설명 셀만).

마. 단축키 예제: 스트림 빌드

키보드나 스크린 리더에 의존하는 사용자를 위해 스트림 빌드 프로세스를 보다 명확하게 하기 위해 다음은 마우스를 사용하지 않고 스트림을 빌드하는 예제입니다. 이 예제에서는 다음 단계를 사용하여 가변파일 노드, 파생 노드, 히스토그램 노드가 포함된 스트림을 빌드합니다.

1. **IBM® SPSS® Modeler 시작**. IBM SPSS Modeler가 처음 시작될 때 초점은 노드 팔레트의 즐겨찾기 탭에 맞춰져 있습니다.
2. **Ctrl+아래로 화살표**. 탭 자체에서 탭 본문으로 초점을 이동합니다.
3. **오른쪽 화살표**. 가변파일 노드로 초점을 이동합니다.
4. **스페이스바**. 가변파일 노드를 선택합니다.
5. **Ctrl+Enter**. 스트림 캔버스에 가변파일 노드를 추가합니다. 이 키 조합은 가변파일 노드에 대한 선택도 유지하므로 추가된 다음 노드는 가변파일 노드에 연결됩니다.
6. **Tab**. 노드 팔레트로 초점을 다시 이동합니다.
7. **오른쪽 화살표 네 번**. 파생 노드로 이동합니다.
8. **스페이스바**. 파생 노드를 선택합니다.
9. **Alt+Enter**. 캔버스에 파생 노드를 추가하고 파생 노드로 선택을 이동합니다. 이제 이 노드는 추가된 다음 노드에 연결할 준비가 되었습니다.
10. **Tab**. 노드 팔레트로 초점을 다시 이동합니다.
11. **오른쪽 화살표 다섯 번**. 팔레트의 히스토그램 노드로 초점을 이동합니다.
12. **스페이스바**. 히스토그램 노드를 선택합니다.
13. **Enter**. 스트림에 노드를 추가하고 스트림 캔버스로 초점을 이동합니다.

다음 예제를 계속하거나 나중에 다음 예제를 시도하려면 스트림을 저장하십시오.

바. 단축키 예제: 노드 편집

이 예제에서는 이전 예제에 빌드된 스트림을 사용합니다. 스트림은 가변파일 노드, 파생 노드, 히스토그램 노드로 구성되어 있습니다. 지시사항은 스트림의 세 번째 노드인 히스토그램 노드에 초점을 맞추고 시작합니다.

1. **Ctrl+왼쪽 화살표 두 번**. 초점을 가변파일 노드로 다시 이동합니다.
2. **Enter**. 가변파일 대화 상자를 엽니다. Tab 키를 사용하여 파일 필드로 이동하고 텍스트 파일 경로와 이름을 입력하여 해당 파일을 선택하십시오. Ctrl+Tab을 눌러 대화 상자의 하단 부분으로 이동하고 확인 단추까지 Tab 키로 이동한 후 Enter를 눌러 대화 상자를 닫으십시오.
3. **Ctrl+오른쪽 화살표**. 두 번째 노드인 파생 노드에 초점을 맞춥니다.
4. **Enter**. 파생 노드 대화 상자를 엽니다. Tab 키로 이동하여 필드를 선택하고 파생 조건을 지정하십시오. Ctrl+Tab을 눌러 확인 단추까지 이동하고 Enter를 눌러 대화 상자를 닫으십시오.
5. **Ctrl+오른쪽 화살표**. 세 번째 노드인 히스토그램 노드에 초점을 맞춥니다.
6. **Enter**. 히스토그램 노드 대화 상자를 엽니다. Tab 키로 이동하여 필드를 선택하고 그래프 옵션을 지정하십시오. 드롭 다운 목록의 경우 아래로 화살표를 눌러 목록을 열고 목록 항목을 강조 표시한 후 Enter를 눌러 목록 항목을 선택하십시오. 확인 단추까지 Tab 키로 이동한 후 Enter를 눌러 대화 상자를 닫으십시오.

이 시점에서 추가 노드를 추가하거나 현재 스트림을 실행할 수 있습니다. 스트림을 빌드하고 있을 때 다음 팁을 명심하십시오.

- 수동으로 노드를 연결하는 경우 F2를 사용하여 연결 시작점을 작성하고 Tab 키를 사용하여 끝점으로 이동한 후 Shift+스페이스바를 사용하여 연결을 마무리하십시오.
- 캔버스의 선택된 노드에 대한 연결을 모두 영구 삭제하려면 F3을 사용하십시오.
- 스트림을 작성했으면 Ctrl+E를 사용하여 현재 스트림을 실행하십시오.

단축키 전체 목록을 사용할 수 있습니다. 자세한 정보는 기본 창 탐색을 위한 단축키의 내용을 참조하십시오.

④ 스크린 리더 사용

시장에서 다수의 스크린 리더를 사용할 수 있습니다. IBM® SPSS® Modeler는 IBM SPSS Modeler와 함께 설치되는 Java Access Bridge를 사용하여 Windows용 JAWS를 지원하도록 구성되어 있습니다. JAWS가 설치된 경우에는 IBM SPSS Modeler를 시작하기 전에 JAWS를 시작하여 이 제품을 사용하면 됩니다.

ⓘ **참고**: SPSS Modeler에서 JAWS를 실행하려면 최소 6GB의 공간을 권장합니다.

IBM SPSS Modeler의 데이터 마이닝 프로세스에 대한 고유한 그래픽 표시 특성으로 인해 차트와 그래프는 시각적으로 최적으로 사용합니다. 그러나 스크린 리더를 사용하여 텍스트로 보여지는 출력과 모델을 기반으로 이해하고 결정할 수 있습니다.

> **참고**: 64비트 클라이언트 시스템을 사용할 경우 일부 지원 기술 기능이 작동하지 않습니다. 그 이유는 Java Access Bridge가 64비트 작업에 맞게 설계되지 않았기 때문입니다.

IBM SPSS Modeler 사전 파일 사용

IBM SPSS Modeler 사전 파일(*Awt.JDF*)을 JAWS에 사용할 수 있습니다. 이 파일을 사용하려면 다음을 수행하십시오.

1. IBM SPSS Modeler 설치의 */accessibility* 서브디렉토리로 이동하여 사전 파일(*Awt.JDF*)을 복사하십시오.
2. JAWS 스크립트가 있는 디렉토리로 복사하십시오.

다른 JAVA 애플리케이션이 실행 중인 경우에는 시스템에 *Awt.JDF* 파일이 이미 있을 수 있습니다. 이 경우, 사전 파일을 수동으로 편집해야 이 사전 파일을 사용할 수 있습니다.

가. HTML 출력에 스크린 리더 사용

스크린 리더를 사용하여 IBM® SPSS® Modeler에 HTML로 표시된 출력을 볼 때 일부 난관에 봉착할 수 있습니다. 다음을 포함하여 다수의 출력이 영향을 받습니다.

- 회귀분석, 로지스틱 회귀분석, 요인/PCA 노드의 고급 탭에서 보여지는 출력
- 보고서 노드 출력

이러한 창이나 대화 상자 각각에는 기본 브라우저에서 출력을 실행하는 데 사용할 수 있는 도구가 도구 모음에 있습니다. 그러면 스크린 리더를 사용하여 출력 정보를 전달할 수 있습니다.

나. 대화식 트리 창에서 내게 필요한 옵션

대화식 트리 창에서 의사결정 트리 모형의 표준 표시는 스크린 리더에 문제를 일으킬 수 있습니다. 액세스 가능한 버전에 액세스하려면 대화식 트리 메뉴에서 다음을 클릭하십시오.

보기 > 액세스 가능 창

이는 표준 트리 맵과 유사한 보기를 표시하지만 JAWS가 올바르게 읽을 수 있는 보기입니다. 표준 화살표 창을 사용하여 위, 아래, 오른쪽 또는 왼쪽으로 이동할 수 있습니다. 액세스 가능 창을 탐색할 때 대화식 트리 창의 초점도 이에 따라서 이동합니다. 스페이스바를 사용하여 선택을 변경하거나 Ctrl+스페이스바를 사용하여 현재 선택을 확장하십시오.

(3) 사용 팁

IBM® SPSS® Modeler 환경에 보다 쉽게 액세스할 수 있는 몇 가지 팁이 있습니다. IBM SPSS Modeler에서 작업 시 일반적인 힌트는 다음과 같습니다.

- **확장 텍스트 상자 종료**. Ctrl+Tab을 사용하여 확장 텍스트 상자를 종료하십시오. 참고: Ctrl+Tab은 테이블 제어를 종료하는 데도 사용합니다.
- **화살표 키보다 Tab 키 사용**. 대화 상자 옵션 선택 시 Tab 키를 사용하여 옵션 단추 간에 이동하십시오. 화살표 키는 이 컨텍스트에서 작동하지 않습니다.
- **드롭 다운 목록**. 대화 상자에 대한 드롭 다운 목록에서 Escape 키 또는 스페이스바를 사용하여 항목을 선택한 후 목록을 닫을 수 있습니다. 또한 Escape 키를 사용하여 다른 제어에 탭을 지정한 경우 닫지 않는 드롭 다운 목록을 닫을 수 있습니다.
- **실행 상태**. 대형 데이터베이스에서 스트림을 실행 중인 경우 스트림 상태를 읽을 때 JAWS가 뒤쳐질 수 있습니다. 주기적으로 Ctrl 키를 눌러 상태 보고를 업데이트하십시오.
- **노드 팔레트 사용**. 노드 팔레트의 탭에 처음 진입할 때 JAWS가 노드 이름 대신 "그룹 상자"를 읽는 경우가 있습니다. 이 경우, Ctrl+오른쪽 화살표와 Ctrl+왼쪽 화살표를 차례로 사용하여 스크린 리더를 재설정하고 노드 이름을 들을 수 있습니다.
- **메뉴 읽기**. 때때로 메뉴를 처음 열 때 JAWS가 첫 번째 메뉴 항목을 읽지 않을 수 있습니다. 이런 상황이 발생했다고 의심되면 아래로 화살표와 위쪽 화살표를 차례로 사용하여 메뉴의 첫 번째 항목을 들으십시오.
- **계단식 메뉴**. JAWS는 계단식 메뉴의 첫 번째 레벨을 읽지 않습니다. 메뉴를 이동하는 동안 소리가 중단되면 오른쪽 화살표 키를 눌러 하위 메뉴 항목을 들으십시오.

또한 IBM SPSS Modeler Text Analytics가 설치된 경우 다음 팁을 사용하면 대화식 워크벤치 인터페이스에 보다 쉽게 액세스할 수 있습니다.

- **대화 상자 진입**. 대화 상자에 진입할 때 첫 번째 제어에 초점을 맞추려면 Tab 키를 눌러야 합니다.
- **확장 텍스트 상자 종료**. Ctrl+Tab을 사용하여 확장 텍스트 상자를 종료하고 다음 제어로 이동하십시오. 참고: Ctrl+Tab은 테이블 제어를 종료하는 데도 사용합니다.
- **트리 목록의 요소를 찾기 위해 첫 번째 문자 입력**. 범주 분할창, 추출된 결과 분할창 또는 라이브러리 트리에서 요소를 찾을 때 분할창이 초점을 가진 경우 요소의 첫 번째 문자를 입력할 수 있습니다. 이렇게 하면 입력한 문자로 시작하는 요소의 다음 발생이 선택됩니다.
- **드롭 다운 목록**. 대화 상자에 대한 드롭 다운 목록에서 스페이스바를 사용하여 항목을 선택한 후 목록을 닫을 수 있습니다.

추가 사용 팁은 다음 주제에서 자세히 설명합니다.

① 다른 소프트웨어와의 충돌

IBM® SPSS® Modeler에서 JAWS와 같은 스크린 리더 사용을 검정할 때 개발 팀은 조직 내에서 SMS(Systems Management Server)를 사용하면 IBM SPSS Modeler와 같은 Java 기반 애플리케이션을 읽을 수 있는 JAWS의 기능을 방해할 수 있음을 발견했습니다. SMS를 사용하지 않으면 이런 상황이 발생하지 않습니다. SMS에 대한 자세한 정보를 알려면 Microsoft 웹 사이트를 방문하십시오.

② JAWS 및 Java

여러 버전의 JAWS는 Java 기반 소프트웨어 애플리케이션에 대한 다양한 지원 레벨을 제공합니다. IBM® SPSS® Modeler에서 최신 버전의 모든 JAWS가 작동하긴 하지만 일부 버전에서는 Java 기반 시스템에서 사용할 경우 사소한 문제점이 발생할 수 있습니다. Windows용 JAWS 웹 사이트(http://www.FreedomScientific.com)를 방문하십시오.

③ IBM® SPSS® Modeler에서 그래프 사용

히스토그램, 평가 차트, 다중 도표, 산점도와 같은 정보에 대한 시각적 표시는 스크린 리더로 해석하기 어렵습니다. 그러나 출력 창에서 사용 가능한 텍스트 요약을 사용하여 해당 웹 그래프와 분포를 볼 수는 있습니다.

16) 유니코드 지원

(1) IBM SPSS Modeler에서 유니코드 지원

IBM® SPSS® Modeler는 IBM SPSS Modeler 및 IBM SPSS Modeler Server 둘 모두에 대해 완전히 유니코드 사용이 가능합니다. 이로 인해 다중 언어 데이터베이스를 포함하여 유니코드를 지원하는 다른 애플리케이션과 데이터를 로케일 특정 인코딩 스키마로/에서의 변환으로 초래될 수 있는 데이터의 유실 없이 교환하는 것이 가능합니다.

- IBM SPSS Modeler는 유니코드 데이터를 내부적으로 저장하고 유니코드로 저장된 다중 언어 데이터를 유실 없이 데이터베이스에 씁니다.
- IBM SPSS Modeler는 UTF-8 인코딩 텍스트 파일을 읽고 쓸 수 있습니다. 텍스트 파일 가져오기 및 내보내기는 기본적으로 로케일 인코딩으로 설정되지만 대안으로 UTF-8을 지원합니다. 이 설정은 파일 가져오기와 내보내기 노드에 지정되거나 기본 인코딩을 스트림 특성 대화 상자에서 변경할 수 있습니다. 자세한 정보는 스트림의 일반 옵션 설정의 내용을 참조하십시오.

- 로케일 인코딩에 저장된 Statistics, SAS 및 텍스트 데이터 파일은 가져올 때 UTF-8으로 변환되고 내보낼 때 다시 돌아갑니다. 파일에 쓸 때, 로케일 문자 세트에 없는 유니코드 문자가 있는 경우에는 이들은 대체되고 경고가 표시됩니다. 이는 데이터를 유니코드를 지원하고(데이터베이스 또는 UTF-8 텍스트 파일) 다른 로케일이나 다중 로케일 또는 문자 세트의 문자를 포함하는 데이터 소스로부터 가져오는 경우에만 발생합니다.
- IBM SPSS Modeler Solution Publisher 이미지는 UTF-8 인코딩이 되어 있고 플랫폼과 로케일 사이에 휴대가 가능합니다.

유니코드 정보

유니코드 표준의 목표는 경계선, 로케일 및 애플리케이션 간에 쉽게 공유할 수 있도록 다국어 텍스트를 인코딩하는 일관된 방법을 제공합니다. 현재 버전 4.0.1인 유니코드 표준은 오늘날 세계에서 공통으로 사용하고 있는 모든 문자 세트의 수퍼세트인 문자 세트를 정의하고 각 문자에 고유한 이름과 코드 포인트를 지정합니다. 문자 및 해당 코드 포인트는 ISO-10646에 의해 정의된 범용 문자 세트(UCS)의 그것과 동일합니다. 자세한 정보는 유니코드 홈 페이지를 참조하십시오.

2. 소스, 프로세스 및 출력 노드

1) 소스 노드

(1) 개요

소스 노드를 사용하면 플랫 파일, IBM® SPSS® Statistics(.sav), SAS, Microsoft Excel 및 ODBC 준수 관계형 데이터베이스 등의 여러 가지 형식으로 저장된 데이터를 가져올 수 있습니다. 또한 사용자 입력 노드를 사용하여 합성 데이터를 생성할 수 있습니다.

소스 팔레트에는 다음 노드가 포함됩니다.

데이터베이스 노드를 사용하면 ODBC(Open Database Connectivity)를 사용하여 Microsoft SQL Server, Db2, Oracle 및 기타를 포함한 다양한 다른 패키지로부터 데이터를 가져올 수 있습니다. 자세한 정보는 데이터베이스 소스 노드의 내용을 참조하십시오.

가변파일 노드는 자유 필드 텍스트 파일, 즉 레코드가 일정한 수의 필드를 포함하지만 변하는 문자를 포함하는 파일로부터 데이터를 읽습니다. 이 노드는 또한 고정 길이 헤더 텍스트와 특정 유형의 주석을 갖는 파일에도 유용합니다. 자세한 정보는 가변파일 노드의 내용을 참조하십시오.

고정 파일 노드는 고정 필드 텍스트 파일, 즉 그의 필드가 구분되지 않고 동일한 위치에서 시작하며 고정된 길이의 파일로부터 데이터를 가져옵니다. 머신 생성 또는 레거시 데이터가 자주 고정 필드 형식으로 저장됩니다. 자세한 정보는 고정 파일 노드의 내용을 참조하십시오.

통계량 파일 노드는 IBM SPSS Statistics에서 사용하는 .sav 또는 .zsav 파일 형식뿐 아니라 동일한 형식을 사용하는 IBM SPSS Modeler에 저장된 캐시 파일로부터 데이터를 읽습니다.

Data Collection 노드는 Data Collection 데이터 모델을 준수하는 시장 조사 소프트웨어에서 사용하는 다양한 형식에서 설문조사 데이터를 가져옵니다. 이 노드를 사용하려면 Data Collection Developer Library가 설치되어 있어야 합니다. 자세한 정보는 Data Collection 노드의 내용을 참조하십시오.

IBM Cognos 소스 노드는 Cognos Analytics 데이터베이스에서 데이터를 가져옵니다.

IBM Cognos TM1 소스 노드는 Cognos TM1 데이터베이스에서 데이터를 가져옵니다.

SAS 파일 노드는 SAS 데이터를 IBM SPSS Modeler로 가져옵니다. 자세한 정보는 SAS 소스 노드의 내용을 참조하십시오.

Excel 노드는 Microsoft Excel로부터 .xlsx 파일 형식으로 데이터를 가져옵니다. ODBC 데이터 소스는 필요하지 않습니다. 자세한 정보는 Excel 소스 노드의 내용을 참조하십시오.

XML 소스 노드는 XML 형식의 데이터를 스트림으로 가져옵니다. 단일 파일 또는 디렉토리의 모든 파일을 가져올 수 있습니다. 선택적으로 XML 구조를 읽을 스키마 파일을 지정할 수 있습니다.

사용자 입력 노드는 스크래치로부터 또는 기존 데이터를 변경하여 합성 데이터를 작성하는 쉬운 방법을 제공합니다. 이것은 예를 들어 모델링을 위한 검정 데이터 세트를 작성할 때 유용합니다. 자세한 정보는 사용자 입력 노드의 내용을 참조하십시오.

시뮬레이션 생성 노드는 사용자가 지정한 통계 분포를 사용하는 스크래치로부터 또는 기존 히스토리 데이터에 대해 시뮬레이션 적합 노드를 실행하여 얻은 분포를 자동으로 사용하여 시뮬레이션된 데이터를 생성하는 쉬운 방법을 제공합니다. 이것은 모델 입력에 불확실성이 존재하는 상황에서 예측 모델의 결과를 평가하기 원할 때 유용합니다.

맵 또는 공간 데이터를 데이터 마이닝 세션으로 가져오려면 지리 공간적 소스 노드를 사용하십시오. 자세한 정보는 지리 공간적 소스 노드의 내용을 참조하십시오.

JSON 소스 노드가 JSON 파일에서 데이터를 가져옵니다. 자세한 정보는 JSON 소스 노드의 내용을 참조하십시오.

스트림을 시작하려면 스트림 캔버스에 소스 노드를 추가하십시오. 그런 다음, 노드를 두 번 클릭하여 대화 상자를 여십시오. 대화 상자의 다양한 탭을 사용하여 데이터를 읽고 필드 및 값을 보고 다양한 옵션(필터, 데이터 유형, 필드 역할 및 결측값 확인 등)을 설정할 수 있습니다.

(2) 필드 저장 공간 및 형식화 설정

고정 파일, 가변파일, XML 소스 및 사용자 입력 노드의 데이터 탭에 있는 옵션을 사용하여 IBM® SPSS® Modeler에서 필드를 가져오거나 작성할 때 필드의 저장 유형을 지정할 수 있습니다. 고정 파일, 가변파일 및 사용자 입력 노드의 경우 필드 형식화 및 다른 메타데이터도 지정할 수 있습니다.

다른 소스에서 읽히는 데이터의 경우, 저장 공간은 자동으로 결정되지만 채움 노드 또는 파생 노드에서 변환 함수(예: to_integer)를 사용하여 변경할 수 있습니다.

필드 현재 데이터 세트의 필드를 보고 선택하려면 **필드** 열을 사용하십시오.

대체 저장 공간 및 **입력 형식** 열의 옵션을 활성화하려면 **대체** 열의 선택란을 선택하십시오.

데이터 저장 공간

저장 공간은 데이터를 필드에 저장하는 방식을 설명합니다. 예를 들어, 값이 1 및 0인 필드는 정수 데이터를 저장합니다. 이는 데이터 사용에 대해 설명하고 저장 공간에 영향을 미치지 않는 측정 수준과 구별됩니다. 예를 들어, 값이 1 및 0인 정수 필드에 대한 측정 수준을 플래그로 설정할 수 있습니다. 이는 일반적으로 1 = *True* 및 0 = *False*임을 나타냅니다. 저장 공간은 소스에서 결정해야 하지만 측정 수준은 스트림의 어느 지점에서나 유형 노드를 사용하여 변경할 수 있습니다. 자세한 정보는 측정 수준의 내용을 참조하십시오.

사용 가능한 저장 유형은 다음과 같습니다.

- **문자열** 영숫자 데이터라고도 하는 숫자가 아닌 데이터가 포함된 필드에 사용됩니다. 문자열은 *fred*, *Class 2* 또는 *1234* 등의 문자열 시퀀스를 포함할 수 있습니다. 문자열의 숫자는 계산에서 사용할 수 없습니다.
- **정수** 값이 정수인 필드입니다.

- **실수** 값은 10진수를 포함할 수 있는 숫자입니다(정수로 제한되지 않음). 표시 형식은 스트림 특성 대화 상자에서 지정되며 유형 노드(형식 탭)에서 개별 필드에 대해 대체될 수 있습니다.
- **날짜** 연도, 월 및 일 등의 표준 형식으로 지정된 날짜 값입니다(예: 2007-09-26). 구체적인 형식은 스트림 특성 대화 상자에서 지정됩니다.
- **시간** 기간으로 측정된 시간입니다. 예를 들어, 1시간 26분 38초 동안 지속되는 서비스 호출은 스트림 특성 대화 상자에서 지정된 대로 현재 시간 형식에 따라 01:26:38로 표시될 수 있습니다.
- **시간소인** 스트림 특성 대화 상자의 현재 날짜 및 시간 형식에 따라 날짜와 시간 구성요소를 모두 포함하는 값입니다(예: 2007-09-26 09:04:00). 시간소인 값을 별도의 날짜 및 시간 값 대신 단일 값으로 해석하기 위해 시간소인 값을 큰따옴표로 묶어야 할 수 있습니다. (이는 예를 들어, 사용자 입력 노드에서 값을 입력하는 경우에 적용됩니다.)
- **목록** 지리 공간 및 콜렉션이라는 새로운 측정 수준과 함께 SPSS Modeler 버전 17에서 도입된 목록 저장 공간 필드에는 단일 레코드에 대한 다중 값이 포함되어 있습니다. 모든 기타 저장 유형의 목록 버전이 있습니다.

표 1. 목록 저장 유형 아이콘

아이콘	저장유형
[A]	문자열목록
[◇]	정수목록
[#]	실수목록
[🕒]	시간목록
[📅]	날짜목록
[📅]	시간소인목록
[[]]	0(영)보다 큰 깊이를 가진 목록

또한 콜렉션 측정 수준과 함께 사용하기 위해 다음과 같은 측정 수준의 목록 버전이 있습니다.

표 2. 목록 측정 수준 아이콘

아이콘	측정수준
[📏]	연속형 목록
[🚦]	범주형 목록
[🚩]	플래그 목록
[●●●]	명목 목록
[📊]	순서 목록

목록은 세 가지 소스 노드(Analytic Server, 지리 공간 또는 가변파일) 중 하나에서 SPSS Modeler로 가져오거나 파생 또는 채움 필드 작업 노드를 사용하여 스트림 내에서 작성할 수 있습니다.

목록 및 해당 콜렉션 및 지리 공간 측정 수준과의 상호작용에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

저장 공간 변환. 채움 노드에서 to_string 및 to_integer 등의 다양한 변환 함수를 사용하여 필드에 대한 저장 공간을 변환할 수 있습니다. 자세한 정보는 채움 노드를 사용한 저장 공간 변환의 내용을 참조하십시오. 변환 함수(및 날짜 또는 시간 값 등의 특정 입력 유형이 필요한 기타 함수)는 스트림 특성 대화 상자에서 지정된 현재 형식에 따라 다릅니다. 예를 들어, 값이 *Jan 2018, Feb 2018* 등인 문자열 필드를 날짜 저장 공간으로 변환하려면 **MON YYYY**를 스트림의 기본 날짜 형식으로 선택하십시오. 파생 계산 중 임시 변환의 경우 파생 노드에서도 변환 함수를 사용할 수 있습니다. 파생 노드를 사용하여 범주형 값을 가진 문자열 필드 코딩 변경 등의 기타 조작을 수행할 수도 있습니다. 자세한 정보는 파생 노드를 사용하여 값 코딩변경의 내용을 참조하십시오.

혼합 데이터 읽기. 숫자 저장 공간(정수, 실수, 시간, 시간소인 또는 날짜)을 가진 필드에서 읽을 때 숫자가 아닌 값은 널값 또는 시스템 결측값으로 설정됩니다. 이는 일부 애플리케이션과는 달리 IBM SPSS Modeler가 필드 내에서 혼합 저장 유형을 허용하지 않기 때문입니다. 이를 방지하기 위해 혼합 데이터가 포함된 필드는 필요에 따라 소스 노드 또는 외부 애플리케이션에서 저장 유형을 변경하여 문자열로 읽어야 합니다.

필드 입력 형식(고정 파일, 가변파일 및 사용자 입력 노드에만 해당)

문자열 및 정수를 제외한 모든 저장 유형의 경우, 드롭 다운 목록을 사용하여 선택된 필드에 대해 형식화 옵션을 지정할 수 있습니다. 예를 들어, 다양한 로케일의 데이터를 병합할 때, 하나의 필드에 대해 소수점 구분 문자로 마침표(.)를 지정해야 하지만 다른 필드는 쉼표 구분 문자를 필요로 할 수 있습니다.

소스 노드에 지정된 입력 옵션은 스트림 특성 대화 상자에 지정된 형식화 옵션을 대체합니다(단, 나중에 스트림에서 지속되지 않음). 이러한 옵션은 데이터에 대한 지식을 기반으로 입력을 올바르게 구문 분석하는 데 사용됩니다. 지정된 형식은 IBM SPSS Modeler로 데이터를 읽은 후 데이터를 형식화하는 방법을 결정하는 데 사용되지 않고 IBM SPSS Modeler로 데이터를 읽을 때 데이터를 구문 분석하기 위한 지침으로 사용됩니다. 스트림의 다른 위치에서 필드별로 형식화를 지정하려면 유형 노드의 형식 탭을 사용하십시오. 자세한 정보는 필드 형식 설정 탭 주제를 참조하십시오.

옵션은 저장 유형에 따라 다릅니다. 예를 들어, 실수 저장 유형의 경우, 소수점 구분 문자로 마침표(.) 또는 쉼표(,)를 선택할 수 있습니다. 시간소인 필드의 경우, 드롭 다운 목록에서 지정을 선택하면 별도의 대화 상자가 열립니다. 자세한 정보는 필드 형식 옵션 설정 주제를 참조하십시오.

모든 저장 유형의 경우, **스트림 기본값**을 선택하여 가져오기에 스트림 기본값 설정을 사용할 수도 있습니다. 스트림 설정은 스트림 특성 대화 상자에 지정됩니다.

추가 옵션

데이터 탭을 사용하여 일부 다른 옵션을 지정할 수 있습니다.

- 현재 노드를 통해 더 이상 연결되지 않는 데이터에 대한 저장 공간 설정을 보려면(예를 들어, 훈련 데이터) **사용하지 않는 필드 설정 보기**를 선택하십시오. **지우기**를 클릭하여 레거시 필드를 지울 수 있습니다.
- 이 대화 상자에서 작업하는 중 언제든 **새로 고치기**를 클릭하여 데이터 소스에서 필드를 다시 로드할 수 있습니다. 이는 소스 노드와의 데이터 연결을 변경하거나 대화 상자의 탭들 사이에서 작업 중일 때 유용합니다.

① 목록 저장 공간 및 연관된 측정 수준

새 측정 수준인 지리 공간적 및 콜렉션에 대해 작업하기 위해 SPSS® Modeler 버전 17에 도입된 목록 저장 공간 필드에는 단일 레코드에 대한 다중 값이 포함되어 있습니다. 목록은 꺽쇠 대괄호([])로 싸여 있습니다. 목록의 예로는 [1,2,4,16]과 ["abc", "def"]가 있습니다.

목록은 세 가지 소스 노드(Analytic Server, 지리 공간적 또는 가변파일) 중 하나에서 SPSS Modeler로 가져오거나 파생 또는 채움 필드 조작 노드를 사용하여 스트림 내에서 작성되거나 순위화된 조건 병합 방법 사용 시 병합 노드에 의해 생성될 수 있습니다.

목록은 깊이를 가지는 것으로 간주됩니다. 예를 들어, [1,3] 형식의 단일 대괄호로 묶인 항목을 가진 단순 목록은 깊이 0(영)을 사용하여 IBM® SPSS Modeler에서 기록됩니다. 깊이 0(영)을 가진 단순 목록 외에도 목록 내 각각의 값이 목록 자체인 중첩된 목록을 사용할 수 있습니다.

중첩된 목록의 깊이는 연관된 측정 수준에 따라 다릅니다. 유형 없음의 경우에는 설정된 깊이 제한이 없고 콜렉션의 경우 깊이는 0(영)이며 지리 공간적 측정 수준의 경우 깊이는 중첩된 항목의 수에 따라 0(영)과 2 사이(경계값 포함)여야 합니다.

깊이가 0(영)인 목록의 경우 측정 수준을 지리 공간적 또는 콜렉션으로 설정할 수 있습니다. 이 수준은 모두 상위 측정 수준이므로 사용자는 값 대화 상자에서 측정 하위 수준 정보를 설정합니다. 콜렉션의 측정 하위 수준은 해당 목록에 있는 요소의 측정 수준을 결정합니다. 모든 측정 수준(유형 없음 및 지리 공간적 측정 수준 제외)을 콜렉션의 하위 수준으로 사용할 수 있습니다. 지리 공간적 측정 수준은 점, 선 스트링, 다각형, 다중 점, 다중 선 스트림 및 다중 다각형이라는 6개의 하위 수준을 가지고 있습니다. 자세한 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.

> ⓘ **참고:** 콜렉션 측정 수준은 깊이가 0(영)인 목록에만 사용할 수 있고 지리 공간적 측정 수준은 최대 깊이가 2인 목록에만 사용할 수 있으며 유형 없음 측정 수준은 모든 목록 깊이와 함께 사용할 수 있습니다.

다음 예제에서는 지리 공간적 측정 하위 수준 점 및 선 스트링의 구조를 사용하여 깊이가 0(영)인 목록과 중첩된 목록 사이의 차이를 보여줍니다.

- 지리 공간적 측정 하위 수준 점의 필드 깊이는 0(영)입니다.

  [1,3] 2개의 좌표

  [1,3,-1] 3개의 좌표

- 지리 공간적 측정 하위 수준 선 스트림의 필드 깊이는 1입니다.

  [[1,3], [5,0]] 2개의 좌표

  [[1,3,-1], [5,0,8]] 3개의 좌표

깊이가 0(영)인 점 필드는 각각의 값이 2개 또는 3개의 좌표로 구성되는 일반적인 목록입니다. 깊이가 1인 선 스트림 필드는 각각의 점이 추가적인 일련의 목록 값으로 구성되는 점의 목록입니다.

목록 작성에 대한 자세한 정보는 목록 또는 지리 공간적 필드 파생의 내용을 참조하십시오.

(3) 지원되지 않는 제어 문자

SPSS® Modeler의 일부 프로세스에서는 다양한 제어 문자가 포함된 데이터를 처리할 수 없습니다. 데이터가 이 문자를 사용하는 경우에는 다음 예제와 같은 오류 메시지가 표시될 수 있습니다.

```
Unsupported control characters found in values of field {0}
```

지원되지 않는 문자는 0x0부터 0x3F까지(0x3F 포함) 및 0x7F이지만 탭(0x9(₩t)), 줄 바꾸기(0xA(₩n)) 및 캐리지 리턴(0xD(₩r)) 문자는 문제를 유발하지 않습니다.

지원되지 않는 문자와 관련된 오류 메시지가 표시되는 경우에는 스트림의 소스 노드 뒤에서 채움 노드 및 CLEM 표현식 **stripctrlchars**를 사용하여 해당 문자를 바꾸십시오.

(4) 데이터베이스 소스 노드

데이터베이스 소스 노드를 사용하면 ODBC(Open Database Connectivity)를 사용하여 Microsoft SQL Server, Db2, Oracle 및 기타를 포함한 다양한 다른 패키지로부터 데이터를 가져올 수 있습니다.

데이터베이스를 읽거나 데이터베이스에 쓰려면 필요에 따라 읽기 또는 쓰기 권한을 가지고 관련 데이터베이스에 대해 ODBC 데이터 소스가 설치 및 구성되어 있어야 합니다. IBM SPSS Data Access Pack에는 이 용도로 사용할 수 있는 ODBC 드라이버 세트가 포함되어 있으며 이 드라이버는 다운로드 사이트에서 사용 가능합니다. ODBC 데이터 소스에 대한 작성 및 설정에 관한 문의사항이 있으면 데이터베이스 관리자에게 문의하십시오.

지원되는 ODBC 드라이버

IBM® SPSS® Modeler와 함께 사용하기 위해 지원되고 테스트되는 데이터베이스 및 ODBC 드라이버에 대한 최신 정보는 회사 지원 사이트(http://www.ibm.com/support)에서 제품 호환성 교차표를 참조하십시오.
드라이버 설치 위치

> **참고**: 처리가 발생할 수 있는 각각의 컴퓨터에 ODBC 드라이버가 설치 및 구성되어 있어야 합니다.

- 로컬(독립형) 모드에서 IBM SPSS Modeler를 실행 중인 경우에는 로컬 컴퓨터에 드라이버가 설치되어야 합니다.
- 원격 IBM SPSS Modeler Server에 대해 분산 모드에서 IBM SPSS Modeler를 실행 중인 경우에는 IBM SPSS Modeler Server가 설치되는 컴퓨터에 ODBC 드라이버가 설치되어야 합니다. UNIX 시스템의 IBM SPSS Modeler Server에 대해서는 이 절의 뒷부분에 있는 "UNIX 시스템에서 ODBC 드라이버 구성"도 참조하십시오.
- IBM SPSS Modeler와 IBM SPSS Modeler Server 모두에서 동일한 소스에 액세스해야 하는 경우에는 두 컴퓨터 모두에 ODBC 드라이버가 설치되어야 합니다.
- 터미널 서비스를 통해 IBM SPSS Modeler를 실행 중인 경우에는 IBM SPSS Modeler를 설치한 터미널 서비스 서버에 ODBC 드라이버가 설치되어야 합니다.

데이터베이스의 데이터에 액세스

데이터베이스의 데이터에 액세스하려면 다음의 단계를 완료하십시오.

- ODBC 드라이버를 설치하고 원하는 데이터베이스에 데이터 소스를 구성하십시오.
- 데이터베이스 노드 대화 상자에서 테이블 모드 또는 SQL 쿼리 모드를 사용하여 데이터베이스에 연결하십시오.
- 데이터베이스에서 테이블을 선택하십시오.
- 데이터베이스 노드 대화 상자의 탭을 사용하여 사용 유형을 변경하고 데이터 필드를 필터링할 수 있습니다.

선행 단계에 대한 자세한 내용은 관련 문서 주제에서 제공됩니다.

> **참고**: SPSS Modeler에서 데이터베이스 스토어드 프로시저(SP)를 호출하는 경우에는 SP의 예상 출력 대신 RowsAffected라는 단일 출력 필드가 리턴되는 것을 볼 수 있습니다. 이는 ODBC가 SP의 출력 데이터 모델을 판별하기에 충분한 정보를 리턴하지 않는 경우 발생합니다. SPSS Modeler는 출력을 리턴하는 SP에 대해 제한된 지원만 제공하므로 SP를 사용하는 대신 SP에서 SELECT를 추출한 후 다음 조치 중 하나를 사용하는 것이 좋습니다.
> - SELECT를 기반으로 보기를 작성하고 데이터베이스 소스 노드에서 보기 선택
> - 데이터베이스 소스 노드에서 직접 SELECT 사용

① 데이터베이스 노드 옵션 설정

데이터베이스 소스 노드 대화 상자의 데이터 탭에 있는 옵션을 사용하여 데이터베이스에 대한 액세스를 얻고 선택된 테이블에서 데이터를 읽어올 수 있습니다.

모드. 대화 상자 제어를 사용하여 테이블에 연결하려면 **테이블**을 선택하십시오.

SQL을 사용하여 아래에서 선택된 데이터베이스를 쿼리하려면 **SQL 쿼리**를 선택하십시오. 자세한 정보는 데이터베이스 조회의 내용을 참조하십시오.

데이터 소스. 테이블 모드와 SQL 쿼리 모드 모두에 대해 데이터 소스 필드에서 이름을 입력하거나 드롭 다운 목록에서 **새 데이터베이스 연결 추가**를 선택할 수 있습니다.

데이터베이스에 연결하고 대화 상자를 사용하여 테이블을 선택하는 데 사용되는 옵션은 다음과 같습니다.

테이블 이름. 액세스할 테이블의 이름을 알고 있는 경우에는 테이블 이름 필드에 해당 이름을 입력하십시오. 그렇지 않으면 **선택** 단추를 클릭하여 사용 가능한 테이블이 나열되는 대화 상자를 여십시오.

테이블 및 열 이름 따옴표로 묶기. 쿼리가 데이터베이스에 전송될 때 예를 들어, 쿼리에 공백 또는 구두점이 포함된 경우 테이블 및 열 이름을 따옴표로 묶을지 여부를 지정하십시오.

- **필요 시** 옵션은 테이블 및 필드 이름이 비표준 문자를 포함하는 경우에만 테이블 및 필드 이름을 따옴표로 묶습니다. 비표준 문자에는 비ASCII 문자, 공백 문자, 마침표(.) 이외의 영숫자가 아닌 문자가 포함됩니다.
- *모든* 테이블 및 필드 이름을 따옴표로 묶으려면 **항상**을 선택하십시오.
- 테이블 및 필드 이름을 따옴표로 묶지 *않으려면* **사용 안 함**을 선택하십시오.

선행 및 후미 공백 제거. 문자열에서 선행 및 후미 공백을 삭제하는 옵션을 선택하십시오.

> ⓘ **참고**: SQL 푸시백을 사용하는 문자열과 사용하지 않는 문자열 간의 비교는 후미 공백이 존재하는 서로 다른 결과를 생성할 수 있습니다.

Oracle에서 빈 문자열 읽기. Oracle 데이터베이스에서 읽거나 Oracle 데이터베이스에 쓸 때 IBM® SPSS® Modeler 및 대부분의 다른 데이터베이스와 달리 Oracle은 빈 문자열 값을 널값과 같다고 간주하고 저장합니다. 이는 Oracle 데이터베이스에서 추출된 동일한 데이터가 파일 또는 다른 데이터베이스에서 추출된 경우와 다르게 동작하고 데이터가 다른 결과를 리턴할 수 있음을 의미합니다.

② 데이터베이스 연결 추가

데이터베이스를 열려면 먼저 연결할 데이터 소스를 선택하십시오. 데이터 탭의 데이터 소스 드롭다운 목록에서 **새 데이터베이스 연결 추가**를 선택하십시오.

데이터베이스 연결 대화 상자가 열립니다.

> ⓘ **참고**: 이 대화 상자를 여는 다른 방법은 기본 메뉴에서 **도구** > **데이터베이스...**를 선택하는 것입니다.

데이터 소스. 사용 가능한 데이터 소스를 나열합니다. 원하는 데이터베이스가 표시되지 않으면 아래로 스크롤하십시오. 데이터 소스를 선택하고 비밀번호를 입력한 후 **연결**을 클릭하십시오. **새로 고치기**를 클릭하여 목록을 업데이트하십시오.

모드 다음 모드 중 하나를 선택하십시오.
- **사용자 이름 및 비밀번호**. 데이터 소스가 비밀번호로 보호되어 있으면 사용자 이름 및 연관된 비밀번호를 입력하십시오.
- **저장된 신임 정보**. IBM® SPSS® Collaboration and Deployment Services에서 신임 정보가 구성되어 있는 경우 이 옵션을 선택하여 리포지토리에서 해당 신임 정보를 찾아볼 수 있습니다. 신임 정보의 사용자 이름 및 비밀번호는 데이터베이스에 액세스하기 위해 필요한 사용자 이름 및 비밀번호와 일치해야 합니다.

연결 현재 연결된 데이터베이스를 표시합니다.

- **기본값**. 선택적으로 하나의 연결을 기본값으로 선택할 수 있습니다. 그러면 데이터베이스 소스 또는 내보내기 노드에 이 연결이 데이터 소스로 사전 정의되지만 이는 원하는 경우 편집할 수 있습니다.
- **저장**. 선택적으로 후속 세션에서 다시 표시할 하나 이상의 연결을 선택하십시오.
- **데이터 소스**. 현재 연결된 데이터베이스에 대한 연결 문자열입니다.
- **사전 설정**. 사전 설정 값이 데이터베이스 연결에 대해 지정되었는지를 표시합니다(* 문자 사용). 사전 설정 값을 지정하려면 데이터베이스 연결에 해당하는 행에서 이 열을 클릭한 후 목록에서 지정을 선택하십시오. 자세한 정보는 데이터베이스 연결에 대한 사전 설정된 값 지정의 내용을 참조하십시오.

연결을 제거하려면 목록에서 하나의 연결을 선택한 후 **제거**를 클릭하십시오.

드라이버. 모드에 대해 **데이터 소스** 대신 **드라이버**를 선택한 경우 목록에서 원하는 드라이버를 선택하십시오.

- **속성** 필드에 데이터베이스 연결 문자열을 입력하십시오. 사용할 적절한 문자열은 해당 데이터베이스 문서를 참조하십시오.
- 표시 이름을 입력하십시오.
- 데이터베이스 사용자 이름 및 비밀번호를 입력하고 **연결**을 클릭하십시오.

데이터베이스를 읽거나 데이터베이스에 쓰려면 필요에 따라 읽기 또는 쓰기 권한을 가지고 관련 데이터베이스에 대해 ODBC 데이터 소스가 설치 및 구성되어 있어야 합니다. IBM SPSS Data Access Pack에는 이 용도로 사용할 수 있는 ODBC 드라이버 세트가 포함되어 있으며 이 드라이버는 다운로드 사이트에서 사용 가능합니다. ODBC 데이터 소스에 대한 작성 및 설정에 관한 문의사항이 있으면 데이터베이스 관리자에게 문의하십시오.

지원되는 ODBC 드라이버

IBM SPSS Modeler와 함께 사용하기 위해 지원되고 테스트되는 데이터베이스 및 ODBC 드라이버에 대한 최신 정보는 회사 지원 사이트(http://www.ibm.com/support)에서 제품 호환성 교차표를 참조하십시오.

드라이버 설치 위치

> **참고**: 처리가 발생할 수 있는 각각의 컴퓨터에 ODBC 드라이버가 설치 및 구성되어 있어야 합니다.

- 로컬(독립형) 모드에서 IBM SPSS Modeler를 실행 중인 경우에는 로컬 컴퓨터에 드라이버가 설치되어야 합니다.

- 원격 IBM SPSS Modeler Server에 대해 분산 모드에서 IBM SPSS Modeler를 실행 중인 경우에는 IBM SPSS Modeler Server가 설치되는 컴퓨터에 ODBC 드라이버가 설치되어야 합니다. UNIX 시스템의 IBM SPSS Modeler Server에 대해서는 이 절의 뒷부분에 있는 "UNIX 시스템에서 ODBC 드라이버 구성"도 참조하십시오.
- IBM SPSS Modeler와 IBM SPSS Modeler Server 모두에서 동일한 소스에 액세스해야 하는 경우에는 두 컴퓨터 모두에 ODBC 드라이버가 설치되어야 합니다.
- 터미널 서비스를 통해 IBM SPSS Modeler를 실행 중인 경우에는 IBM SPSS Modeler를 설치한 터미널 서비스 서버에 ODBC 드라이버가 설치되어야 합니다.

UNIX 시스템에서 ODBC 드라이버 구성

기본적으로 DataDirect 드라이버 관리자는 UNIX 시스템의 IBM SPSS Modeler Server에 대해 구성되어 있지 않습니다. DataDirect 드라이버 관리자를 로드하도록 UNIX를 구성하려면 다음의 명령을 입력하십시오.

```
cd <modeler_server_install_directory>/bin
rm -f libspssodbc.so
```

그런 다음 UTF8 드라이버 랩퍼를 사용하려면 다음 명령을 실행하십시오.

```
ln -s libspssodbc_datadirect.so libspssodbc.so
```

또는 UTF16 드라이버 랩퍼를 사용하려면 다음 명령을 대신 실행하십시오.

```
ln -s libspssodbc_datadirect_utf16.so libspssodbc.so
```

그러면 기본 링크가 제거되고 DataDirect 드라이버 관리자에 대한 링크가 작성됩니다.

> **참고**: 일부 데이터베이스의 경우 SAP HANA 또는 IBM Db2 CLI 드라이버를 사용하려면 UTF16 드라이버 랩퍼가 필요합니다. DashDB에는 IBM Db2 CLI 드라이버가 필요합니다.

SPSS Modeler Server를 구성하려면 다음을 수행하십시오.

1. modelersrv.sh에 다음 행을 추가하여 IBM SPSS Data Access Pack odbc.sh 환경 파일을 제공하도록 SPSS Modeler Server 시작 스크립트 modelersrv.sh를 구성하십시오.

```
. /<pathtoSDAPinstall>/odbc.sh
```

여기서 <pathtoSDAPinstall>은 IBM SPSS Data Access Pack 설치의 전체 경로입니다.

2. SPSS Modeler Server를 다시 시작하십시오.

또한 odbc.ini 파일의 DSN에 다음 모수 정의를 추가하여 연결 중 버퍼 오버플로우를 방지하십시오(SAP HANA 및 IBM Db2의 경우에만).

```
DriverUnicodeType=1
```

ⓘ **참고**: libspssodbc_datadirect_utf16.so 랩퍼는 다른 SPSS Modeler Server 지원 ODBC 드라이버와도 호환 가능합니다.

③ 잠재적 데이터베이스 문제

사용하는 데이터베이스에 따라 알아두어야 할 잠재적 문제가 있습니다.

IBM Db2

Db2 데이터베이스에서 데이터를 읽는 스트림에서 노드를 캐시하려고 하면 다음 오류 메시지가 표시될 수 있습니다.

```
A default table space could not be found with a pagesize of at least 4096 that authorization
  ID TEST is authorized to use
```

SPSS® Modeler에서 제대로 작동하도록 데이터베이스 내 캐싱을 사용하게 Db2를 구성하려면 데이터베이스 관리자가 "사용자 임시" 테이블스페이스를 작성하고 관련 Db2 계정에 이 테이블스페이스에 대한 액세스 권한을 부여해야 합니다.
성공적으로 캐싱할 수 있는 필드 수의 제한이 증가되므로 새 테이블스페이스에서 페이지 크기 32768을 사용하는 것이 좋습니다.

z/OS용 IBM Db2

- 생성된 SQL을 사용하여 신뢰도를 사용한 상태로 알고리즘의 서브세트를 스코어링하면 실행 시 오류가 리턴될 수 있습니다. 이 문제는 Db2 for z/OS에 특정합니다. 이 문제를 해결하려면 z/OS용 SPSS Modeler Server Scoring Adapter for Db2를 사용하십시오.
- Db2 for z/OS에 대해 스트림을 실행할 때 유휴 데이터베이스 연결의 제한시간이 사용되고 너무 낮게 설정된 경우 데이터베이스 오류가 발생할 수 있습니다. Db2 for z/OS 버전 8에서 기본값은 제한시간 없음에서 2분으로 변경되었습니다. 솔루션은 Db2 시스템 매개변수 IDLE THREAD TIMEOUT (IDTHTOIN)의 값을 늘리거나 값을 0으로 설정하는 것입니다.

Oracle

통합 노드가 있는 스트림을 실행하는 경우 SQL을 Oracle 데이터베이스에 푸시백할 때 첫 번째 및 세 번째 사분위수에 대해 리턴되는 값이 원시 모드에서 리턴되는 값과 다를 수 있습니다.

④ 데이터베이스 연결에 대한 사전 설정된 값 지정

일부 데이터베이스의 경우 데이터베이스 연결에 대해 다수의 기본 설정을 지정할 수 있습니다. 이 설정은 모두 데이터베이스 내보내기에 적용됩니다.

이 기능을 지원하는 데이터베이스 유형은 다음과 같습니다.

- SQL Server Enterprise 및 Developer Edition. 자세한 정보는 SQL Server에 대한 설정의 내용을 참조하십시오.
- Oracle Enterprise 또는 Personal Edition. 자세한 정보는 Oracle에 대한 설정의 내용을 참조하십시오.
- z/OS용 IBM Db2 및 Teradata는 모두 비슷한 방식으로 데이터베이스 또는 스키마에 연결됩니다. 자세한 정보는 z/OS용 IBM Db2, IBM Db2 LUW 및 Teradata에 대한 설정의 내용을 참조하십시오.

이 기능을 지원하지 않는 데이터베이스 또는 스키마에 연결되는 경우에는 이 **데이터베이스 연결에는 사전 설정을 구성할 수 없음** 메시지가 표시됩니다.

가. SQL Server에 대한 설정

이 설정은 SQL Server Enterprise 및 Developer Edition에 대해 표시됩니다.

압축 사용. 선택된 경우 압축을 사용하여 내보낼 테이블을 작성합니다.

압축. 압축의 수준을 선택하십시오.

- **행**. 행 수준 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…) WITH (DATA_COMPRESSION = ROW);와 동등).
- **페이지**. 페이지 수준 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…) WITH (DATA_COMPRESSION = PAGE);).

나. Oracle에 대한 설정

Oracle 설정 - 기본 옵션

이 설정은 기본 옵션을 사용하는 Oracle Enterprise 또는 Personal Edition에 대해 표시됩니다.

압축 사용. 선택된 경우 압축을 사용하여 내보낼 테이블을 작성합니다.

압축. 압축의 수준을 선택하십시오.

- **기본값**. 기본 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…) COMPRESS;). 이 케이스에서 이는 **기본** 옵션과 동일한 효과를 가집니다.
- **기본**. 기본 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…) COMPRESS BASIC;).

Oracle 설정 - 고급 옵션

이 설정은 고급 옵션을 사용하는 Oracle Enterprise 또는 Personal Edition에 대해 표시됩니다.

압축 사용. 선택된 경우 압축을 사용하여 내보낼 테이블을 작성합니다.

압축. 압축의 수준을 선택하십시오.

- **기본값**. 기본 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…) COMPRESS;). 이 케이스에서 이는 **기본** 옵션과 동일한 효과를 가집니다.
- **기본**. 기본 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…) COMPRESS BASIC;).

- **OLTP**. OLTP 압축을 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE (…)COMPRESS FOR OLTP;).

- **쿼리 낮음/높음**. (Exadata 서버 전용) 쿼리에 대해 HCC(Hybrid Columnar Compression)를 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…)COMPRESS FOR QUERY LOW; 또는 CREATE TABLE MYTABLE(…)COMPRESS FOR QUERY HIGH;). 쿼리에 대한 압축은 데이터 웨어하우징 환경에서 유용합니다. HIGH는 LOW보다 높은 압축 비율을 제공합니다.
- **아카이브 낮음/높음**. (Exadata 서버 전용) 아카이브에 대해 HCC(Hybrid Columnar Compression)를 사용으로 설정합니다(예: SQL의 CREATE TABLE MYTABLE(…)COMPRESS FOR ARCHIVE LOW; 또는 CREATE TABLE MYTABLE(…)COMPRESS FOR ARCHIVE HIGH;). 아카이브에 대한 압축은 장기간 저장될 데이터를 압축하는 경우에 유용합니다. HIGH는 LOW보다 높은 압축 비율을 제공합니다.

다. z/OS용 IBM Db2, IBM Db2 LUW 및 Teradata에 대한 설정

z/OS용 IBM Db2, IBM Db2 LUW 또는 Teradata에 대한 사전 설정을 지정하면 다음 항목 중에서 선택하라는 프롬프트가 표시됩니다.

서버 스코어링 어댑터 데이터베이스 사용 또는 **서버 스코어링 어댑터 스키마 사용**. 선택된 경우 **서버 스코어링 어댑터 데이터베이스** 또는 **서버 스코어링 어댑터 스키마** 옵션을 사용할 수 있게 합니다.

서버 스코어링 어댑터 데이터베이스 또는 **서버 스코어링 어댑터 스키마** 드롭 다운 목록에서 필요한 연결을 선택하십시오.

또한 Teradata의 경우 쿼리 밴딩 세부사항을 설정하여 워크로드 관리, 쿼리 조합, 식별 및 분석, 데이터베이스 사용 추적 등의 항목을 지원하기 위한 추가적인 메타데이터를 제공할 수도 있습니다.

쿼리 밴딩 스펠링. Teradata 데이터베이스 연결에 대해 작업하는 전체 시간에 대해 한 번 쿼리 밴딩을 설정할 경우(**세션의 경우**) 또는 스트림을 실행할 때마다 쿼리 밴딩을 설정할 경우(**트랜잭션의 경우**) 선택하십시오.

> **참고**: 스트림에서 쿼리 밴딩을 설정하는 경우에는 해당 스트림을 다른 시스템으로 복사하면 밴딩이 유실됩니다. 이를 방지하기 위해 스크립팅을 사용하여 스트림을 실행하고 스크립트에 키워드 *querybanding*을 사용하여 필요한 설정을 적용할 수 있습니다.

필요한 데이터베이스 권한

SPSS® Modeler 데이터베이스 기능이 올바르게 작동하게 하려면 다음과 같은 항목에 대한 액세스를 사용된 모든 사용자 ID에 부여하십시오.

Db2 LUW

SYSIBM.SYSDUMMY1
SYSIBM.SYSFOREIGNKEYS
SYSIBM.SYSINDEXES
SYSIBM.SYSKEYCOLUSE
SYSIBM.SYSKEYS
SYSIBM.SYSPARMS
SYSIBM.SYSRELS
SYSIBM.SYSROUTINES
SYSIBM.SYSROUTINES_SRC
SYSIBM.SYSSYNONYMS
SYSIBM.SYSTABCONST
SYSIBM.SYSTABCONSTPKC
SYSIBM.SYSTABLES
SYSIBM.SYSTRIGGERS
SYSIBM.SYSVIEWDEP
SYSIBM.SYSVIEWS
SYSCAT.TABLESPACES
SYSCAT.SCHEMATA

Db2/z

SYSIBM.SYSDUMMY1
SYSIBM.SYSFOREIGNKEYS
SYSIBM.SYSINDEXES
SYSIBM.SYSKEYCOLUSE
SYSIBM.SYSKEYS
SYSIBM.SYSPARMS
SYSIBM.SYSRELS
SYSIBM.SYSROUTINES
SYSIBM.SYSROUTINES_SRC
SYSIBM.SYSSYNONYMS
SYSIBM.SYSTABCONST
SYSIBM.SYSTABLES
SYSIBM.SYSTRIGGERS
SYSIBM.SYSVIEWDEP
SYSIBM.SYSVIEWS
SYSIBM.SYSDUMMYU
SYSIBM.SYSPACKSTMT

Teradata

DBC.Functions
DBC.USERS

⑤ 데이터베이스 테이블 선택

데이터 소스에 연결한 후 특정 테이블 또는 보기에서 필드를 가져오도록 선택할 수 있습니다. 데이터베이스 대화 상자의 데이터 탭에서 테이블 이름 필드에 테이블의 이름을 입력하거나 선택을 클릭하여 사용 가능한 테이블 및 보기를 나열하는 테이블/보기 **선택** 대화 상자를 열 수 있습니다.

테이블 소유자 표시. 사용자가 테이블에 액세스하려면 먼저 테이블의 소유자를 지정하도록 데이터 소스에서 요구하는 경우 선택하십시오. 이 요구사항을 가지고 있지 않은 데이터 소스의 경우에는 이 옵션을 선택 취소하십시오.

> ⓘ **참고**: SAS 및 Oracle 데이터베이스는 일반적으로 테이블 소유자를 표시하도록 요구합니다.

테이블/보기. 가져올 테이블 또는 보기를 선택하십시오.

표시. 현재 연결된 데이터 소스의 열을 나열합니다. 다음 옵션 중 하나를 클릭하여 사용 가능한 테이블의 보기를 사용자 정의하십시오.

- **사용자 테이블**을 클릭하여 데이터베이스 사용자가 작성한 일반적인 데이터베이스 테이블을 보십시오.
- **시스템 테이블**을 클릭하여 시스템이 소유한 데이터베이스 테이블을 보십시오(예: 인덱스 세부사항 등의 데이터베이스에 대한 정보를 제공하는 테이블). 이 옵션은 Excel 데이터베이스에서 사용되는 탭을 보는 데 사용할 수 있습니다. (별도의 Excel 소스 노드도 사용할 수 있습니다. 자세한 정보는 Excel 소스 노드의 내용을 참조하십시오.)

- **보기**를 클릭하여 하나 이상의 일반적인 테이블과 관련된 쿼리를 기반으로 가상 테이블을 보십시오.
- **동의어**를 클릭하여 기존 테이블에 대해 데이터베이스에서 작성된 동의어를 보십시오.

이름/소유자 필터. 이 필드를 사용하면 이름 또는 소유자별로 표시된 테이블의 목록을 필터링할 수 있습니다. 예를 들어, SYS를 입력하여 해당 소유자를 가진 테이블만 나열하십시오. 와일드카드 검색의 경우 밑줄(_)은 단일 문자를 나타내는 데 사용할 수 있고 퍼센트 부호(%)는 0개 이상 문자의 시퀀스를 나타낼 수 있습니다.

기본값으로 설정. 현재 설정을 현재 사용자의 기본값으로 저장합니다. 이 설정은 사용자가 나중에 *동일한 데이터 소스 이름 및 사용자 로그인* 전용 새 테이블 선택기 대화 상자를 열 때 복원됩니다.

⑥ 데이터베이스 조회

일단 데이터 소스에 연결되면 SQL 쿼리를 사용하여 필드를 가져올 수 있습니다. 주 대화 상자에서 연결 모드로 **SQL 쿼리**를 선택하십시오. 그러면 쿼리 편집기 창이 대화 상자에 추가됩니다. 쿼리 편집기를 사용하면 결과 세트를 데이터 스트림으로 읽어 올 하나 이상의 SQL 쿼리를 작성하거나 로드할 수 있습니다.

다중 SQL 쿼리를 지정하는 경우 세미콜론(;)으로 구분하고 다중 SELECT문이 없는지 확인하십시오.

쿼리 편집기 창을 취소하거나 닫으려면 연결 모드로 **테이블**을 선택하십시오.

SQL 쿼리에 SPSS® Modeler 스트림 모수(사용자 정의 변수의 유형)를 포함할 수 있습니다. 자세한 정보는 SQL 쿼리에서 스트림 모수 사용의 내용을 참조하십시오.

쿼리 로드. 파일 브라우저를 열고 이전에 저장한 쿼리를 로드하려면 클릭하십시오.

쿼리 저장. 쿼리 저장 대화 상자를 열어 현재 쿼리를 저장하려면 클릭하십시오.

기본값 가져오기. 대화 상자에서 선택된 테이블 및 열을 사용하여 자동으로 구성된 SQL SELECT 명령문 예를 가져오려면 클릭하십시오.

지우기. 작업 영역의 컨텐츠를 지웁니다. 다시 시작하려면 이 옵션을 사용하십시오.

텍스트 분할. 기본 옵션인 **사용 안 함**은 쿼리가 데이터베이스에 하나로 전송됨을 의미합니다. 또는 SPSS Modeler에서 쿼리를 구문 분석하려고 시도하고 데이터베이스에 하나씩 차례로 전송되어야 하는 SQL문이 있는지 식별하는 **필요시 사용**을 선택할 수 있습니다.

중요사항: 사용 중인 데이터베이스에 따라 SPSS Modeler에서 SQL 결과가 있는 데이터 모델을 얻기 위해 사용자가 입력한 사용자 정의 SQL을 실행할 수 있습니다. 예를 들어 MySQL 또는 Google BigQuery를 사용하는 경우 테이블 체계를 가져올 때 기본적으로 SQLExecute()가 호출됩니다. 따라서 SPSS Modeler에서는 SQL을 실행하고 데이터 모델을 가져옵니다. 대부분의 데이터베이스 드라이버에는 해당되지 않습니다.
이 동작을 방지하려면 이와 같은 경우 SPSS Modeler에서 SQL을 처리하는 방법을 사용자 정의하는 데 관해 사용자 정의 데이터베이스 구성 파일 사용에서 자세히 알아보십시오.

가. SQL 쿼리에서 스트림 모수 사용

필드를 가져오기 위해 SQL 쿼리를 작성하는 경우 이전에 정의된 SPSS® Modeler 스트림 매개변수를 포함할 수 있습니다. 모든 유형의 스트림 매개변수가 지원됩니다.

다음 표에는 SQL 쿼리에서 스트림 매개변수의 일부 예제가 해석되는 방식을 보여줍니다.

표 1. 스트림 매개변수의 예

스트림모수이름(예)	저장공간	스트림모수값	해석
PString	문자열	ss	'ss'
PInt	정수	5	5
PReal	실수	5.5	5.5
PTime	시간	23:05:01	t{'23:05:01'}
PDate	날짜	2011-03-02	d{'2011-03-02'}
PTimeStamp	시간소인	2011-03-02 23:05	ts{'2011-03-02 23:05:01'}
PColumn	알 수 없음	IntValue	IntValue

SQL 쿼리에서는 '$P-<parameter_name>'이라는 CLEM 표현식과 동일한 방식으로 스트림 매개변수를 지정합니다. 여기서 <parameter_name>은 스트림 매개변수에 대해 정의된 이름입니다.

필드를 참조하는 경우에는 저장 유형이 알 수 없음으로 정의되어야 하며 필요한 경우 매개변수 값을 따옴표로 묶어야 합니다. 따라서 표에 있는 예제를 사용하여 다음의 SQL 쿼리를 입력한 경우:

```
select "IntValue" from Table1 where "IntValue" < '$P-PInt';
```

다음과 같이 평가됩니다.

```
select "IntValue" from Table1 where "IntValue" < 5;
```

PColumn 매개변수를 사용하여 IntValue 필드를 참조해야 하는 경우에는 동일한 결과를 얻기 위해 다음과 같이 쿼리를 지정해야 합니다.

```
select "IntValue" from Table1 where "'$P-PColumn'" < '$P-PInt';
```

⑦ 사용자 정의 데이터베이스 구성 파일 사용

SPSS® Modeler에서 SQL을 처리하는 방법을 사용자 정의해야 하는 경우 사용자 정의 데이터베이스 구성 파일을 사용할 수 있습니다. 그러면 특정 환경에 따라 필요한 경우 다양하게 사용자 정의할 수 있습니다. 이 기능의 전체 문서는 지원되지 않습니다. 이 기능과 관련된 고유 요구사항에 관한 지원은 데이터베이스 관리자 또는 IBM 지원 센터에 문의하십시오. 하지만 IBM에서는 IBM이 제공하지 않는 구성 파일을 통해 수행하는 사용자 정의는 지원하지 않습니다.

사용자 정의 데이터베이스 구성 파일 구현

1. 고유 데이터베이스에 맞는 .cfg 파일을 작성하십시오.
2. 필요한 대로 옵션을 추가하십시오. 광범위하게 사용자 정의할 수 있습니다. 고유 데이터베이스에 맞게 파일의 이름과 형식을 적절하게 지정해야 합니다.
3. SPSS Modeler 서버 및/또는 SPSS Modeler 클라이언트 설치 디렉토리의 config 폴더에 파일을 두십시오.

.cfg 파일 이름은 odbc-<db>-custom-properties.cfg 형식을 따라야 합니다. 여기서 <db>는 다음 데이터베이스 이름 중 하나입니다.

- bigquery
- db2
- greenplum
- hana
- hive
- impala
- informix
- mssql
- mysql
- neoview
- netezza
- oracle
- postgresql
- redshift
- soliddb
- sybase
- teradata
- vertica

예: execute_while_getting_schema

데이터베이스 소스 노드에서 사용자 정의 SQL과 함께 MySQL 데이터베이스를 사용하는 경우 테이블 체계를 가져올 때 기본적으로 MySQL 데이터베이스 드라이버에서 SQLExecute()를 호출합니다. 결과적으로 SPSS Modeler에서 SQL을 실행하고 데이터 모델을 가져와야 합니다. SPSS Modeler에서 데이터 모델을 얻지 않게 하려면 다음 단계를 따르십시오.

1. odbc-mysql-custom-properties.cfg라는 파일을 작성합니다.

2. 다음 행을 추가하고 N으로 설정하여 MySQL의 기본 동작을 대체합니다.

```
execute_while_getting_schema, N
```

3. SPSS Modeler 서버 및 SPSS Modeler 클라이언트 설치의 config 디렉토리에 파일을 복사합니다.

예: sqlmx_sort_by

기본적으로 order_by는 SQL 푸시백 중에 중첩 SQL에서 허용되지 않습니다. order_by를 사용으로 설정하면 정렬 노드가 샘플 노드 앞에 있을 때 SQL 푸시백이 샘플 노드에 대해 작동하게 설정할 수 있습니다. 예를 들어 Google BigQuery 데이터베이스에서 order_by를 사용하려면 다음 단계를 따르십시오.

1. odbc-bigquery-custom-properties.cfg라는 파일을 작성합니다.
2. 다음 행을 추가하고 Y로 설정하여 Google BigQuery의 기본 동작을 대체합니다.

```
sqlmx_sort_by, Y
```

3. SPSS Modeler 서버 및 SPSS Modeler 클라이언트 설치의 config 디렉토리에 파일을 복사합니다.

예: uda_list_sql_basic and uda_list_sql_parameter

사용자 정의 SQL을 사용하여 데이터베이스 함수 및 집계 함수를 검색할 수 있습니다. 예를 들어 Oracle 데이터베이스에서 다음 단계를 따르십시오.

1. odbc-oracle-custom-properties.cfg라는 파일을 작성합니다.
2. 파일에 다음 행을 추가합니다.

```
#Define the UDA (database window aggregates) sqls

uda_list_sql_basic, "SELECT '<src_database_name>',OBJECT_NAME,OWNER,'', '', CASE WHEN OWNER='SYS' THEN 1 ELSE 0 END BUILTIN FROM ALL_ARGUMENTS WHERE OBJECT_ID IN (SELECT OBJECT_ID FROM ALL_PROCEDURES WHERE AGGREGATE = 'YES') AND OBJECT_ID NOT IN (SELECT DISTINCT OBJECT_ID FROM ALL_ARGUMENTS WHERE PLS_TYPE IS NULL) AND ARGUMENT_NAME IS NULL ORDER BY OBJECT_ID"

uda_list_sql_parameter, "SELECT POSITION, DATA_PRECISION,DATA_SCALE,DATA_TYPE,'',DATA_TYPE,'',0 FROM ALL_ARGUMENTS WHERE OBJECT_ID IN (SELECT OBJECT_ID FROM ALL_PROCEDURES WHERE AGGREGATE = 'YES') AND OBJECT_ID NOT IN (SELECT DISTINCT OBJECT_ID FROM ALL_ARGUMENTS WHERE PLS_TYPE IS NULL) ORDER BY OBJECT_ID,POSITION"
```

3. SPSS Modeler 서버 및 SPSS Modeler 클라이언트 설치의 config 디렉토리에 파일을 복사합니다.

uda_list_sql_basic 및 uda_list_sql_parameter에서는 다른 사용자 정의 SQL을 사용할 수 있습니다. 단, 해당 SQL이 다음 테이블 체계를 준수하는 경우에 한합니다(2단계는 예임).

```
#table schema for uda_list_sql_basic
databaseName,function,schema,catalog,description,isBuiltIn

#table schema for uda_list_sql_parameter
position,precision,scale,returnType,returnTypeName,parameterTypes,parameterTypeNames,isVarArg
```

(5) 가변파일 노드

가변파일 노드를 사용하여 자유 필드 텍스트 파일(해당 레코드에 일정한 수의 숫자와 일정하지 않은 수의 문자가 포함되는 파일이며 분리 텍스트 파일이라고도 함)에서 데이터를 읽을 수 있습니다. 이 유형의 노드는 또한 고정 길이 헤더 텍스트와 특정 유형의 주석(Annotation)이 있는 파일에도 유용합니다. 레코드는 한 번에 하나씩 읽히며 전체 파일이 읽혀질 때가지 스트림을 통해 전달됩니다.

지리 공간적 데이터 읽기 관련 주의사항

노드에 지리 공간적 데이터가 포함되고 노드가 플랫 파일에서 내보내기로 작성된 경우, 일부 추가 단계를 수행하여 지리 공간적 메타데이터를 설정해야 합니다. 추가 정보는 가변파일 노드에 지리 공간적 데이터 가져오기의 내용을 참조하십시오.

분리 텍스트 데이터 읽기 주의사항

- 레코드는 각 행의 끝에서 줄 바꾸기 문자를 사용하여 구분해야 합니다. 다른 목적으로(예를 들어, 필드 이름 또는 값 내에서) 줄 바꾸기 문자를 사용해서는 안 됩니다. 꼭 필요한 것은 아니지만 공간을 절약하기 위해 선행 및 후미 공백을 제거하는 것이 좋습니다. 선택적으로 노드에서 이러한 공백을 제거할 수 있습니다.
- 필드는 쉼표 또는 다른 문자(필드 이름 또는 값에서 사용되지 않고 구분자로만 사용되는 문자가 좋음)로 구분해야 합니다. 이것이 불가능한 경우, 큰따옴표가 포함된 필드 이름 또는 텍스트 값이 없는 한, 모든 텍스트 필드를 큰따옴표로 묶을 수 있습니다. 필드 이름 또는 값에 큰따옴표가 있으면, 값에서 작은따옴표가 사용되지 않는 한, 대안으로서 텍스트 필드를 작은따옴표로 묶을 수 있습니다. 큰따옴표와 작은따옴표 둘 다 사용할 수 없는 경우에는 텍스트 값을 수정하여 구분 문자나 작은따옴표 또는 큰따옴표를 제거 또는 대체해야 합니다.
- 헤더 행을 포함하여 각 행은 같은 수의 필드를 포함해야 합니다.

- 첫 번째 행은 필드 이름을 포함해야 합니다. 그렇지 않은 경우, **파일에서 필드 이름 읽기**를 선택 취소하여 각 필드에 Field1, Field2 등의 일반 이름을 부여하십시오.
- 두 번째 행은 데이터의 첫 번째 레코드를 포함해야 합니다. 공백 행 또는 주석이 없어야 합니다.
- 숫자 값에는 천단위 구분 문자 또는 그룹화 기호가 없어야 합니다(예를 들어, 3,000.00에서 쉼표가 없어야 함). 소수점 표시기(US 또는 UK에서는 마침표)는 적합한 곳에서만 사용해야 합니다.
- 날짜 및 시간 값은 스트림 옵션 대화 상자에서 인식되는 형식 중 하나여야 합니다(예: DD/MM/YYYY 또는 HH:MM:SS). 파일의 모든 날짜 및 시간 필드는 동일한 형식을 따르는 것이 좋으며, 날짜를 포함하는 필드는 해당 필드 내의 모든 값에 동일한 형식을 사용해야 합니다.

① 가변파일 노드의 옵션 설정

가변파일 대화 상자의 파일 탭에 있는 옵션을 설정합니다.

파일 파일의 이름을 지정하십시오. 파일 이름을 입력하거나 생략 기호 단추(**...**) 를 클릭하여 파일을 선택할 수 있습니다. 파일을 선택하면 파일 경로가 표시되고 아래의 분할창에 해당 컨텐츠가 구분자와 함께 표시됩니다.

데이터 소스에서 표시되는 표본 텍스트는 복사하여 EOL 주석 문자 및 사용자 지정 구분자에 붙여넣을 수 있습니다. Ctrl-C 및 Ctrl-V를 사용하여 복사하고 붙여넣으십시오.

파일에서 필드 이름 읽기 기본적으로 선택되는 이 옵션은 데이터 파일의 첫 번째 행을 열의 레이블로 처리합니다. 첫 번째 행이 헤더가 아닌 경우에는 이 옵션을 선택 취소하십시오. 그러면 데이터 세트의 필드 수만큼 각 필드에 일반 이름(예: *Field1*, *Field2*)이 자동으로 부여됩니다.

필드 수 지정. 각 레코드의 필드 수를 지정하십시오. 레코드가 줄 바꾸기로 종료되는 경우 필드 수를 자동으로 발견할 수 있습니다. 수동으로 수를 설정할 수도 있습니다.

헤더 문자 건너뛰기. 첫 번째 레코드의 시작 부분에서 무시할 문자 수를 지정하십시오.

EOL 주석 문자. 문자(예: # 또는 !)를 지정하여 데이터에서 주석을 표시하십시오. 데이터 파일에서 이 문자 중 하나가 표시되는 경우마다 다음 줄 바꾸기 문자를 제외하고 이 문자까지의 모든 문자가 무시됩니다.

선행 및 후미 공백 제거. 가져올 때 문자열의 선행 및 후미 공백을 삭제하는 옵션을 선택하십시오.

> ⓘ **참고**: SQL 푸시백을 사용하는 문자열과 사용하지 않는 문자열 간의 비교는 후미 공백이 존재하는 서로 다른 결과를 생성할 수 있습니다.

유효하지 않은 문자. 데이터 소스에서 유효하지 않은 문자를 제거하려면 **삭제**를 선택하십시오. 유효하지 않은 문자를 지정된 기호(한 문자만)로 바꾸려면 **바꿀 문자열**을 선택하십시오. 유효하지 않은 문자는 널 문자이거나 지정된 인코딩 방법에 존재하지 않는 문자입니다.

인코딩. 사용되는 텍스트 인코딩 방법을 지정합니다. 시스템 기본값, 스트림 기본값 또는 UTF-8 중에서 선택할 수 있습니다.

- 시스템 기본값은 Windows 제어판에 지정되어 있거나 분산 모드에서 실행 중인 경우 서버 컴퓨터에 지정되어 있습니다.
- 스트림 기본값은 스트림 특성 대화 상자에서 지정됩니다.

소수점 기호 데이터 소스에서 사용되는 소수점 구분 문자의 유형을 선택하십시오. **스트림 기본값**은 스트림 특성 대화 상자의 옵션 탭에서 선택된 문자입니다. 또는 **마침표(.)** 또는 **쉼표(,)**를 선택하고 선택된 문자를 소수점 구분 문자로 사용하여 이 대화 상자의 모든 데이터를 읽을 수 있습니다.

행 구분자가 줄 바꾸기 문자임 필드 구분자 대신, 행 구분자로 줄 바꾸기 문자를 사용하려면 이 옵션을 선택하십시오. 이 옵션은 하나의 행에 줄 바꿈을 초래하는 홀수 개의 구분자가 있는 경우에 유용할 수 있습니다. 이 옵션을 선택하는 경우 구분자 목록에서 **줄 바꾸기**를 선택할 수 없습니다.

> ⓘ **참고**: 이 옵션을 선택하면 데이터 행 끝에 있는 공백 값이 제거됩니다.

구분자. 이 제어에 대해 나열된 선택란을 사용하면 파일에서 필드 경계를 정의하는 문자(예: 쉼표(,))를 지정할 수 있습니다. 여러 구분자를 사용하는 레코드의 경우 둘 이상의 구분자(예: ", |")를 지정할 수도 있습니다. 기본 구분자는 쉼표입니다.

> ⓘ **참고**: 쉼표가 **소수점 기호**로도 정의되어 있는 경우에는 여기서의 기본 설정이 작동하지 않습니다. 쉼표가 **필드 구분자**이자 **소수점 기호**이면 **필드 구분자** 목록에서 **기타**를 선택하십시오. 그런 다음 입력 필드에서 수동으로 쉼표를 지정하십시오.

인접한 여러 공백 구분 문자를 하나의 구문자로 처리하려면 **여러 공백 구분자 허용**을 선택하십시오. 예를 들어, 한 데이터 값 뒤에 4개의 공백이 있고 그 뒤에 다른 데이터 값이 있는 경우 이 그룹은 5개가 아니라 2개의 필드로 처리됩니다.

열 및 유형에 대해 스캔할 행 수 지정된 데이터 유형을 찾기 위해 스캔할 행 및 열 수를 지정하십시오.
날짜 및 시간 자동 인식 IBM® SPSS® Modeler에서 데이터 항목을 날짜 또는 시간으로 자동으로 인식할 수 있게 하려면 이 선택란을 선택하십시오. 예를 들어, 이는 07-11-1965와 같은 항목이 날짜로 식별되고 02:35:58은 시간으로 식별됨을 의미합니다. 그러나 07111965 또는 023558과 같은 모호한 항목은 숫자 사이에 구분자가 없으므로 정수로 표시됩니다.

ⓘ **참고**: 이전 버전의 IBM SPSS Modeler에 있는 데이터 파일을 사용할 때 발생할 수 있는 잠재적 데이터 문제점을 피하기 위해, 13 이전의 버전에 저장된 정보에 대해서는 기본적으로 이 선택란이 꺼져 있습니다.

대괄호를 목록으로 처리 이 선택란을 선택하면, 여는 대괄호와 닫는 대괄호 사이에 있는 데이터는 해당 컨텐츠에 쉼표 및 큰따옴표와 같은 구분 문자가 포함된 경우에도 단일 값으로 처리됩니다. 예를 들어, 2차원 또는 3차원 지리 공간적 데이터가 포함될 수 있으며, 이 경우 대괄호 안에 있는 좌표는 단일 목록 항목으로 처리됩니다. 추가 정보는 가변파일 노드에 지리 공간적 데이터 가져오기의 내용을 참조하십시오.

따옴표. 드롭 다운 목록을 사용하면 가져오기를 수행할 때 작은따옴표와 큰따옴표가 처리되는 방식을 지정할 수 있습니다. 모든 따옴표를 **삭제**하거나 필드 값에 포함시켜 **텍스트로 포함**하거나 **대응 및 삭제**하여 따옴표 쌍을 일치시킨 후 제거하도록 선택할 수 있습니다. 따옴표가 일치되지 않는 경우 오류 메시지가 수신됩니다. **삭제**와 **대응 및 삭제**는 모두 필드 값(따옴표 제외)을 문자열로 저장합니다.

ⓘ **참고**: **대응 및 삭제**를 사용하는 경우에는 공백이 유지됩니다. **삭제**를 사용하는 경우에는 따옴표 내부 및 외부의 후미 공백이 제거됩니다(예: **' " ab c" , "d ef " , " gh i " '**는 **'ab c, d ef, gh i'**라는 결과를 생성함). **텍스트로 포함**을 사용하는 경우에는 따옴표가 정상적인 문자로 처리되므로 선행 및 후미 공백은 자연스럽게 제거됩니다.

이 대화 상자에서 작업하는 중 언제든 **새로 고치기**를 클릭하여 데이터 소스에서 필드를 다시 로드할 수 있습니다. 이는 소스 노드와의 데이터 연결을 변경하거나 대화 상자의 탭들 사이에서 작업 중일 때 유용합니다.

② 가변파일 노드에 지리 공간적 데이터 가져오기

노드가 지리 공간적 데이터를 포함하고 플랫 파일에서 내보내기로 작성되었으며 작성된 스트림에서 사용되는 경우, 노드는 지리 공간적 메타데이터를 유지하며 추가 구성 단계가 필요하지 않습니다.

그러나, 노드를 내보내어 다른 스트림에서 사용하는 경우 지리 공간적 목록 데이터는 문자열 형식으로 자동으로 변환됩니다. 일부 추가 단계를 수행하여 목록 저장 유형 및 연관된 지리 공간적 메타데이터를 복원해야 합니다.

목록에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

지리 공간적 메타데이터로 설정할 수 있는 세부사항에 대한 자세한 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.

지리 공간적 메타데이터를 설정하려면 다음 단계를 사용하십시오.

1. 가변파일 노드의 파일 탭에서 **대괄호를 목록으로 처리** 선택란을 선택하십시오. 이 선택란을 선택하면, 여는 대괄호와 닫는 대괄호 사이에 있는 데이터는 해당 컨텐츠에 쉼표 및 큰따옴표와 같은 구분 문자가 포함된 경우에도 단일 값으로 처리됩니다. 이 선택란을 선택하지 않는 경우, 데이터는 문자열 저장 유형으로 읽혀지고 필드의 모든 쉼표는 분리자로 처리되어 데이터 구조가 올바르지 않게 해석됩니다.
2. 데이터에 작은따옴표 또는 큰따옴표가 포함된 경우, 필요에 따라 **작은따옴표** 및 **큰따옴표** 필드에서 **짝 짓기 및 삭제** 옵션을 선택하십시오.
3. 가변파일 노드의 데이터 탭에서, 지리 공간적 데이터 필드에 대해, **대체** 선택란을 선택하고 **저장** 유형을 문자열에서 목록으로 변경하십시오.
4. 기본적으로, 목록 **저장** 유형은 *실수 목록*으로서 설정되고 목록 필드의 기본 값 저장 유형은 *실수*로 설정됩니다. 기본 값 저장 유형 또는 깊이를 변경하려면 **지정…**을 클릭하여 저장 공간 부속 대화 상자를 표시하십시오.
5. 저장 공간 부속 대화 상자에서 다음 설정을 수정할 수 있습니다.
   - **저장 공간** 데이터 필드의 전체 저장 유형을 지정하십시오. 기본적으로 저장 유형이 목록으로 설정됩니다. 그러나 드롭 다운 목록에 기타 모든 저장 유형(문자열, 정수, 실수, 날짜, 시간 및 시간소인)이 포함됩니다. 목록 이외의 저장 유형을 선택하는 경우에는 **값 저장 공간** 및 **깊이** 옵션을 사용할 수 없습니다.
   - **값 저장 공간** 필드 전체가 아니라 목록에 있는 요소의 저장 유형을 지정하십시오. 지리 공간적 필드를 가져올 때 관련된 유일한 저장 유형은 실수 및 정수입니다. 기본 설정은 실수입니다.
   - **깊이** 목록 필드의 깊이를 지정하십시오. 필요한 깊이는 지리 공간적 필드의 유형에 따라 다르며 다음 기준을 따릅니다.
     - 점 - 0
     - LineString - 1
     - 다각형 - 1
     - 다중 점 - 1
     - 다중 LineString - 2
     - 다중 다각형 - 2

> **경고**: 다 시 목록으로 변환할 지리 공간적 필드의 유형과 그러한 종류의 필드에 필요한 깊이를 알아야 합니다. 이 정보가 올바르게 설정되어 있지 않으면 해당 필드를 사용할 수 없습니다.

6. 가변파일 노드의 유형 탭에서 지리 공간적 데이터 필드에 대한 **측정** 셀에 올바른 측정 수준이 포함되는지 확인하십시오. 측정 수준을 변경하려면 **측정** 셀에서 **지정…**을 클릭하여 값 대화 상자를 표시하십시오.
7. 값 대화 상자에서 목록에 대한 **측정**, **저장 공간** 및 **깊이**가 표시됩니다. **값 및 레이블 지정** 옵션을 선택하고 **유형** 드롭 다운 목록에서 **측정**에 올바른 유형을 선택하시시오. **유형**에 따라, 추가 세부사항(예: 데이터가 2차원을 나타내는지 또는 3차원을 나타내는지 여부와 사용되는 좌표계)을 요구하는 프롬프트가 표시될 수 있습니다.

좌표계에 대한 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오.

(6) 고정 파일 노드

고정 파일 노드를 사용하여 고정 필드 텍스트 파일(필드가 구분되지 않지만 동일한 위치에서 시작하며 길이가 고정된 파일)에서 데이터를 가져올 수 있습니다. 머신 생성 또는 레거시 데이터가 자주 고정 필드 형식으로 저장됩니다. 고정 파일 노드의 파일 탭을 사용하면 데이터에서 열의 위치 및 길이를 쉽게 지정할 수 있습니다.

① 고정 파일 노드의 옵션 설정

고정 파일 노드의 파일 탭을 사용하면 데이터를 IBM® SPSS® Modeler로 가져오고 레코드의 길이 및 열의 위치를 지정할 수 있습니다. 대화 상자의 가운데에 있는 데이터 미리보기 분할창을 사용하면 클릭하여 필드 간 중단점을 지정하는 화살표를 추가할 수 있습니다.

파일. 파일의 이름을 지정합니다. 파일 이름을 입력하거나 생략 기호 단추(...) 를 클릭하여 파일을 선택할 수 있습니다. 파일을 선택하고 나면 파일 경로가 표시되고 해당 컨텐츠가 구분자와 함께 아래의 패널에 표시됩니다.

데이터 미리보기 분할창은 열 위치 및 길이를 지정하는 데 사용할 수 있습니다. 미리보기 창의 맨 위에 있는 눈금자를 사용하면 변수의 길이를 측정하고 변수 간 중단점을 지정할 수 있습니다. 필드 위의 눈금자 영역을 클릭하여 중단점 행을 지정할 수 있습니다. 중단점은 끌어서 이동할 수 있으며 데이터 미리보기 영역 밖으로 끌어서 삭제할 수 있습니다. 눈금자는 ASCII 문자를 처리하도록 설계되었습니다.

- 각각의 중단점 행은 아래의 필드 테이블에 새 필드를 자동으로 추가합니다.
- 화살표에 의해 표시된 시작 위치는 아래 테이블의 시작 열에 자동으로 추가됩니다.

행 지향. 각 레코드의 끝에서 줄 바꾸기 문자를 건너뛰려면 선택하십시오.

헤더 행 건너뛰기. 첫 번째 레코드의 시작 부분에서 무시할 행 수를 지정하십시오. 이는 열 헤더를 무시하는 경우에 유용합니다.

레코드 길이. 각 레코드의 문자 수를 지정하십시오.

필드이 데이터 파일에 대해 정의한 모든 필드가 여기에 나열됩니다. 두 가지 방법으로 필드를 정의합니다.

- 위의 데이터 미리보기 분할창을 사용하여 대화식으로 필드를 지정하십시오.
- 아래의 테이블에 비어 있는 필드 행을 추가하여 수동으로 필드를 지정하십시오. 필드 분할창 오른쪽의 단추를 클릭하여 새 필드를 추가하십시오. 그런 다음 비어 있는 필드에서 필드 이름, 시작 위치 및 길이를 입력하십시오. 이 옵션은 데이터 미리보기 분할창에 화살표를 자동으로 추가하며 이는 쉽게 조정될 수 있습니다.

이전에 정의된 필드를 제거하려면 목록에서 해당 필드를 선택한 후 빨간색 삭제 단추를 클릭하십시오.

시작. 필드에서 첫 번째 문자의 위치를 지정하십시오. 예를 들어, 레코드의 두 번째 필드가 16번째 문자에서 시작하면 16을 시작점으로 입력합니다.

길이. 각 필드에 대해 가장 긴 값에 있는 문자의 수를 지정하십시오. 이는 다음 필드의 절사 지점을 결정합니다.

선행 및 후미 공백 제거. 가져올 때 문자열에서 선행 및 후미 공백을 삭제하려면 선택하십시오.

> ⓘ **참고**: SQL 푸시백을 사용하는 문자열과 사용하지 않는 문자열 간의 비교는 후미 공백이 존재하는 서로 다른 결과를 생성할 수 있습니다.

유효하지 않은 문자. 데이터 입력에서 유효하지 않은 문자를 제거하려면 **삭제**를 선택하십시오. 유효하지 않은 문자를 지정된 기호(한 문자만)로 바꾸려면 **바꿀 문자열**을 선택하십시오. 유효하지 않은 문자는 널(0) 문자 또는 현재 인코딩에 존재하지 않는 모든 문자입니다.

인코딩. 사용되는 텍스트 인코딩 방법을 지정합니다. 시스템 기본값, 스트림 기본값 또는 UTF-8 중에서 선택할 수 있습니다.

- 시스템 기본값은 Windows 제어판에 지정되어 있거나 분산 모드에서 실행 중인 경우 서버 컴퓨터에 지정되어 있습니다.
- 스트림 기본값은 스트림 특성 대화 상자에서 지정됩니다.

소수점 기호. 데이터 소스에서 사용되는 소수점 구분 문자의 유형을 선택하십시오. **스트림 기본값**은 스트림 특성 대화 상자의 옵션 탭에서 선택된 문자입니다. 또는 **마침표(.)** 또는 **쉼표(,)**를 선택하고 선택된 문자를 소수점 구분 문자로 사용하여 이 대화 상자의 모든 데이터를 읽을 수 있습니다.

날짜 및 시간 자동 인식. IBM SPSS Modeler가 자동으로 데이터 항목을 날짜 또는 시간으로 인식할 수 있게 하려면 이 선택란을 선택하십시오. 예를 들어, 이는 07-11-1965와 같은 항목이 날짜로 식별되고 02:35:58은 시간으로 식별됨을 의미합니다. 그러나 07111965 또는 023558과 같은 모호한 항목은 숫자 사이에 구분자가 없으므로 정수로 표시됩니다.

> ⓘ **참고**: 이전 IBM SPSS Modeler 버전의 데이터 파일을 사용할 때 잠재적인 데이터 문제점을 방지하기 위해 13 이전의 버전에서 저장된 정보에 대해서는 이 선택란이 기본적으로 비활성화되어 있습니다.

유형에 대해 스캔할 행. 지정된 데이터 유형에 대해 스캔할 행 수를 지정하십시오.

이 대화 상자에서 작업하는 중 언제든 **새로 고치기**를 클릭하여 데이터 소스에서 필드를 다시 로드할 수 있습니다. 이는 소스 노드에 대한 데이터 연결을 변경하거나 대화 상자의 탭 사이에서 작업할 때 유용합니다.

(7) 통계량 파일 노드

통계량 파일 노드를 사용하여 저장된 IBM® SPSS® Statistics 파일(*.sav* 또는 *.zsav*)에서 직접 데이터를 읽을 수 있습니다. 이제 이 형식이 IBM SPSS Modeler의 이전 버전의 캐시 파일을 바꾸는 데 사용됩니다. 저장된 캐시 파일을 가져오려면 IBM SPSS Statistics 파일 노드를 사용해야 합니다.

파일 가져오기. 파일의 이름을 지정합니다. 파일 이름을 입력하거나 생략 기호 단추(**...**) 를 클릭하여 파일을 선택할 수 있습니다. 파일을 선택하고 나면 경로가 표시됩니다.

파일이 비밀번호로 암호화됨. 파일이 비밀번호로 암호화되고 있음을 알고 있으면 이 선택란을 선택하십시오. **비밀번호**를 입력하도록 프롬프트됩니다. 파일이 비밀번호로 보호되고 있으나 비밀번호를 입력하지 않으면 다른 탭으로 변경하거나 데이터를 새로 고치거나 노드 컨텐츠를 미리 보거나 노드를 포함한 스트림을 실행하는 것과 같은 시도를 할 때 경고 메시지가 표시됩니다.

> ⓘ **참고**: 비밀번호로 보호된 파일은 IBM SPSS Modeler 버전 16 이상에서만 열 수 있습니다.

변수이름. IBM SPSS Statistics *.sav* 또는 *.zsav* 파일에서 가져올 때 변수 이름 및 레이블을 처리하는 방법을 선택합니다. 여기에 포함시키려고 선택하는 메타데이터는 IBM SPSS Modeler 내의 작업 전체에서 지속되며 IBM SPSS Statistics에서 사용하도록 다시 내보낼 수 있습니다.

- **이름 및 레이블 읽기**. 변수 이름 및 레이블을 둘 다 IBM SPSS Modeler에 읽어들일 수 있습니다. 기본적으로 이 옵션이 선택되고 변수 이름이 유형 노드에 표시됩니다. 레이블은 스트림 특성 대화 상자에서 지정된 옵션에 따라 도표, 모델 브라우저 및 기타 유형의 출력에 표시될 수 있습니다. 기본적으로 출력에 레이블 표시는 사용되지 않습니다.
- **이름으로 레이블 읽기**. IBM SPSS Statistics *.sav* 또는 *.zsav* 파일에서 짧은 필드 이름이 아니라 설명 변수 레이블을 읽고 해당 레이블을 IBM SPSS Modeler에서 변수 이름으로 사용하려면 선택하십시오.

값. IBM SPSS Statistics *.sav* 또는 *.zsav* 파일에서 가져올 때 값 및 레이블을 처리하는 방법을 선택합니다. 여기에 포함시키려고 선택하는 메타데이터는 IBM SPSS Modeler 내의 작업 전체에서 지속되며 IBM SPSS Statistics에서 사용하도록 다시 내보낼 수 있습니다.

- **데이터 및 레이블 읽기**. 실제 값 및 값 레이블을 둘 다 IBM SPSS Modeler에 읽어들일 수 있습니다. 기본적으로 이 옵션이 선택되고 값 자체가 유형 노드에 표시됩니다. 값 레이블은 스트림 특성 대화 상자에서 지정된 옵션에 따라 표현식 작성기, 도표, 모델 브라우저 및 기타 유형의 출력에 표시될 수 있습니다.
- **데이터로 레이블 읽기**. 값을 표시하는 데 수치 또는 기호 코드가 아니라 *.sav* 또는 *.zsav* 파일의 값 레이블을 사용하려면 선택하십시오. 예를 들어, 실제로는 남성 및 여성을 나타내는 1 및 2 값을 가진 성별 필드가 있는 데이터에 이 옵션을 선택하면 각 필드를 문자열로 변환하고 실제값으로 *남성* 및 *여성*을 가져옵니다.

이 옵션을 선택하기 전에 사용자의 IBM SPSS Statistics 데이터에서 결측값을 고려하는 것이 중요합니다. 예를 들어, 수치 필드가 결측값에 대해서만 레이블을 사용하는 경우(0 = *응답 없음*, -99 = *알 수 없음*), 위 옵션을 선택하면 *응답 없음* 및 *알 수 없음*이라는 값 레이블만 가져오고 필드를 문자열로 변환합니다. 이 경우, 값 자체를 가져오고 유형 노드에서 결측값을 설정해야 합니다.

필드 형식 정보를 사용하여 저장 공간 판별. 이 선택란을 선택 취소하면 *.sav* 파일에서 정수로 형식화된 필드 값(즉, F*n*으로 지정된 필드. IBM SPSS Statistics의 변수 보기의 0)을 정수 저장 공간을 사용하여 가져옵니다. 문자열을 제외한 모든 기타 필드 값은 실제 수로 가져옵니다.

이 상자를 선택하면(기본값), *.sav* 파일에서 정수로 형식화되었는지 여부에 상관없이 문자열을 제외한 모든 필드 값을 실제 수로 가져옵니다.

다중 응답 세트. 파일을 가져올 때 IBM SPSS Statistics 파일에서 정의한 모든 다중 응답 세트는 자동으로 유지됩니다. 필터 탭의 모든 노드에서 다중 응답 세트를 보고 편집할 수 있습니다. 자세한 정보는 다중 응답 세트 편집 주제를 참조하십시오.

(8) Data Collection 노드

Data Collection 소스 노드는 Data Collection 제품과 함께 제공되는 Survey Reporter Developer Kit을 기반으로 설문조사 데이터를 가져옵니다. 이 형식은 *케이스 데이터*(설문조사 중에 수집된 질문에 대한 실제 응답)를 케이스 데이터의 수집 및 구성 방식에 대해 설명하는 *메타데이터*와 구별합니다. 메타데이터는 케이스 데이터의 구조 정의, 질문 텍스트, 변수 이름 및 설명, 다중 응답 변수 정의, 텍스트 문자열의 변환 등의 정보로 구성됩니다.

> **참고**: 이 노드를 사용하려면 Data Collection 제품과 함께 배포되는 Survey Reporter Developer Kit이 필요합니다. 이 Developer Kit을 설치하는 것 외에는 추가적인 구성이 필요하지 않습니다.

설명

- 설문조사 데이터는 표 형식 플랫 VDATA 형식에서 읽어오며 메타데이터 소스를 포함하는 경우에는 계층적 HDATA 형식의 소스에서 읽어옵니다.
- 유형은 메타데이터의 정보를 사용하여 자동으로 인스턴스화됩니다.
- 설문조사 데이터를 SPSS® Modeler로 가져오면 질문이 각 응답자에 대한 레코드가 포함된 필드로 렌더링됩니다.

① Data Collection 파일 가져오기 옵션

Data Collection 노드의 파일 탭에서는 가져올 메타데이터 및 케이스 데이터에 대한 옵션을 지정할 수 있습니다.

메타데이터 설정

> **참고**: 사용 가능한 제공자 파일 유형의 전체 목록을 보려면 Data Collection 소프트웨어와 함께 사용 가능한 Survey Reporter Developer Kit을 설치해야 합니다.

메타데이터 제공자. Data Collection Survey Reporter Developer Kit에서 지원하는 대로 다수의 형식에서 설문조사 데이터를 가져올 수 있습니다. 사용 가능한 제공자 유형은 다음과 같습니다.

- **DataCollectionMDD**. 질문지 정의 파일(*.mdd*)에서 메타데이터를 읽어옵니다. 표준 Data Collection 데이터 모델 형식입니다.
- **ADO 데이터베이스**. ADO 파일에서 케이스 데이터 및 메타데이터를 읽어옵니다. 메타데이터가 포함된 *.adoinfo* 파일의 이름 및 위치를 지정하십시오. 이 DSC의 내부 이름은 *mrADODsc*입니다.
- **In2data 데이터베이스**. In2data 케이스 데이터 및 메타데이터를 읽습니다. 이 DSC의 내부 이름은 *mrI2dDsc*입니다.
- **Data Collection 로그 파일**. 표준 Data Collection 로그 파일에서 메타데이터를 읽어옵니다. 일반적으로 로그 파일의 파일 이름 확장자는 *.tmp*입니다. 하지만 일부 로그 파일의 파일 이름 확장자는 다를 수 있습니다. 필요할 경우 *.tmp* 파일 이름 확장자를 가지도록 파일의 이름을 바꿀 수 있습니다. 이 DSC의 내부 이름은 *mrLogDsc*입니다.
- **Quancept 정의 파일**. 메타데이터를 Quancept 스크립트로 변환합니다. Quancept *.qdi* 파일의 이름을 지정하십시오. 이 DSC의 내부 이름은 *mrQdiDrsDsc*입니다.
- **Quanvert 데이터베이스**. Quanvert 케이스 데이터 및 메타데이터를 읽습니다. *.qvinfo* 또는 *.pkd* 파일의 이름 및 위치를 지정하십시오. 이 DSC의 내부 이름은 *mrQvDsc*입니다.
- **Data Collection 참여 데이터베이스**. 프로젝트의 표본 및 히스토리 테이블 테이블을 읽고 해당 테이블의 열에 해당하는 파생된 범주형 변수를 작성합니다. 이 DSC의 내부 이름은 *mrSampleReportingMDSC*입니다.
- **Statistics 파일**. IBM® SPSS® Statistics *.sav* 파일에서 케이스 데이터 및 메타데이터를 읽어옵니다. IBM SPSS Statistics에서 분석을 위해 케이스 데이터를 IBM SPSS Statistics *.sav* 파일에 씁니다. IBM SPSS Statistics .sav 파일의 메타데이터를 .mdd 파일에 씁니다. 이 DSC의 내부 이름은 *mrSavDsc*입니다.
- **Surveycraft 파일**. SurveyCraft 케이스 데이터 및 메타데이터를 읽습니다. SurveyCraft *.vq* 파일의 이름을 지정하십시오. 이 DSC의 내부 이름은 *mrSCDsc*입니다.
- **Data Collection 스크립팅 파일**. mrScriptMetadata 파일의 메타데이터에서 읽어옵니다. 일반적으로 이 파일의 파일 이름 확장자는 *.mdd* 또는 *.dms*입니다. 이 DSC의 내부 이름은 *mrScriptMDSC*입니다.

- **Triple-S XML 파일**. XML 형식의 Triple-S 파일에서 메타데이터를 읽어옵니다. 이 DSC의 내부 이름은 *mrTripleSDsc*입니다.

메타데이터 특성. 선택적으로 특성을 선택하여 가져올 설문조사 버전과 사용할 언어, 컨텍스트 및 레이블 유형을 지정하십시오. 자세한 정보는 Data Collection 가져오기 메타데이터 특성의 내용을 참조하십시오.

케이스 데이터 설정

> ⓘ **참고**: 사용 가능한 제공자 파일 유형의 전체 목록을 보려면 Data Collection 소프트웨어와 함께 사용 가능한 Survey Reporter Developer Kit을 설치해야 합니다.

케이스 데이터 설정 가져오기. *.mdd* 파일에서만 메타데이터를 읽어오는 경우에는 **케이스 데이터 설정 가져오기**를 클릭하여 지정된 소스에 액세스하기 위해 필요한 특정 설정과 함께 선택된 메타데이터와 연관되는 케이스 데이터 소스를 판별하십시오. 이 옵션은 *.mdd* 파일에만 사용할 수 있습니다.

케이스 데이터 제공자. 다음과 같은 제공자 유형이 지원됩니다.

- **ADO 데이터베이스**. Microsoft ADO 인터페이스를 사용하여 케이스 데이터를 읽습니다. 케이스 데이터 유형에 대해 OLE-DB UDL을 선택하고 케이스 데이터 UDL 필드에서 연결 문자열을 지정하십시오. 자세한 정보는 데이터베이스 연결 문자열의 내용을 참조하십시오. 이 구성요소의 내부 이름은 *mrADODsc*입니다.
- **구분된 텍스트 파일(Excel)**. 쉼표로 구분된(.CSV) 파일에서 케이스 데이터를 읽어와서 Excel로 출력할 수 있습니다. 내부 이름은 *mrCsvDsc*입니다.
- **Data Collection 데이터 파일**. 원시 Data Collection 데이터 형식 파일에서 케이스 데이터를 읽어옵니다. 내부 이름은 *mrDataFileDsc*입니다.
- **In2data 데이터베이스**. In2data 데이터베이스(*.i2d*) 파일에서 케이스 데이터 및 메타데이터를 읽어옵니다. 내부 이름은 *mrI2dDsc*입니다.
- **Data Collection 로그 파일**. 표준 Data Collection 로그 파일에서 케이스 데이터를 읽어옵니다. 일반적으로 로그 파일의 파일 이름 확장자는 *.tmp*입니다. 하지만 일부 로그 파일의 파일 이름 확장자는 다를 수 있습니다. 필요할 경우 *.tmp* 파일 이름 확장자를 가지도록 파일의 이름을 바꿀 수 있습니다. 내부 이름은 *mrLogDsc*입니다.
- **Quantum 데이터 파일**. Quantum 형식 ASCII 파일(*.dat*)에서 케이스 데이터를 읽어옵니다. 내부 이름은 *mrPunchDsc*입니다.
- **Quancept 데이터 파일**. Quancept *.drs*, *.drz* 또는 *.dru* 파일에서 케이스 데이터를 읽어옵니다. 내부 이름은 *mrQdiDrsDsc*입니다.
- **Quanvert 데이터베이스**. Quanvert *qvinfo* 또는 *.pkd* 파일에서 케이스 데이터를 읽어옵니다. 내부 이름은 *mrQvDsc*입니다.

- **Data Collection 데이터베이스(MS SQL Server)**. 케이스 데이터를 관계형 Microsoft SQL Server 데이터베이스로 읽어옵니다. 자세한 정보는 데이터베이스 연결 문자열의 내용을 참조하십시오. 내부 이름은 *mrRdbDsc2*입니다.
- **Statistics 파일**. IBM SPSS Statistics .sav 파일에서 케이스 데이터를 읽어옵니다. 내부 이름은 *mrSavDsc*입니다.
- **Surveycraft 파일**. SurveyCraft *.qdt* 파일에서 케이스 데이터를 읽어옵니다. *.vq* 파일과 *.qdt* 파일은 모두 두 파일 모두에 대한 읽기 및 쓰기 액세스를 가진 동일한 디렉토리에 있어야 합니다. SurveyCraft를 사용할 때 기본적으로 이 방식으로 작성되지 않으므로 SurveyCraft 데이터를 가져오기 위해 파일 중 하나를 이동해야 합니다. 내부 이름은 *mrScDsc*입니다.
- **Triple-S 데이터 파일**. 고정 길이 또는 쉼표로 구분된 형식으로 Triple-S 데이터 파일에서 케이스 데이터를 읽어옵니다. 내부 이름은 *mr TripleDsc*입니다.
- **Data Collection XML**. Data Collection XML 데이터 파일에서 케이스 데이터를 읽어옵니다. 일반적으로 이 형식은 한 위치에서 다른 위치로 케이스 데이터를 전송하는 데 사용할 수 있습니다. 내부 이름은 *mrXmlDsc*입니다.

케이스 데이터 유형. 케이스 데이터를 파일, 폴더, OLE-DB UDL, ODBC DSN 중 어디에서 읽어오는지를 지정하고 대화 상자 옵션을 적절하게 업데이트합니다. 유효한 옵션은 제공자의 유형에 따라 다릅니다. 데이터베이스 제공자의 경우 OLE-DB 또는 ODBC 연결에 대한 옵션을 지정할 수 있습니다. 자세한 정보는 데이터베이스 연결 문자열의 내용을 참조하십시오.

케이스 데이터 프로젝트. Data Collection 데이터베이스에서 케이스 데이터를 읽어오는 경우 프로젝트의 이름을 입력할 수 있습니다. 다른 모든 케이스 데이터 유형의 경우 이 설정은 공백이어야 합니다.

변수 가져오기

시스템 변수 가져오기. 인터뷰 상태(진행 중, 완료됨, 완료 날짜 등)를 표시하는 변수를 포함한 시스템 변수를 가져오는지 여부를 지정합니다. **없음**, **모두** 또는 **공통**을 선택할 수 있습니다.

"Codes" 변수 가져오기. 범주형 변수에 대한 오픈 엔드 "기타" 반응에 사용되는 코드를 나타내는 변수의 가져오기를 제어합니다.

"SourceFile" 변수 가져오기. 스캔된 반응의 이미지 파일 이름이 포함된 변수의 가져오기를 제어합니다.

다중 응답 변수 가져오기. 다중 응답 변수를 새 스트림에 대한 기본 방법인 다중 플래그 필드(다중 이분형 세트)로 가져올 수 있습니다. 12.0 이전의 IBM SPSS Modeler 릴리스에서 작성된 스트림은 값을 쉼표로 구분하여 다중 응답을 단일 필드로 가져왔습니다. 기존 스트림이 이전과 마찬가지로 실행될 수 있게 이전의 방법이 계속 지원되지만 새로운 방법을 사용하도록 이전 스트림을 업데이트하는 것이 좋습니다. 자세한 정보는 다중 응답 세트 가져오기의 내용을 참조하십시오.

② Data Collection 가져오기 메타데이터 특성

Data Collection 설문조사 데이터를 가져오는 경우 메타데이터 특성 대화 상자에서 가져올 설문조사 버전과 사용할 언어, 컨텍스트 및 레이블 유형을 지정할 수 있습니다. 한 번에 하나의 언어, 컨텍스트 및 레이블 유형만 가져올 수 있습니다.

버전. 각각의 설문조사 버전을 특정 케이스 데이터 세트를 수집하는 데 사용되는 메타데이터의 스냅샷으로 간주할 수 있습니다. 질문지가 변경됨에 따라 여러 버전이 작성될 수 있습니다. 최신 버전, 모든 버전 또는 특정 버전을 가져올 수 있습니다.

- **모든 버전**. 사용 가능한 모든 버전의 조합(수퍼 세트)을 사용하려면 이 옵션을 선택하십시오. (이를 수퍼 버전이라고도 함). 버전 간 충돌이 있는 경우에는 일반적으로 최신 버전이 이전 버전보다 우선합니다. 예를 들어, 버전에서 범주 레이블이 다른 경우에는 최신 버전의 텍스트가 사용됩니다.
- **최신 버전**. 최신 버전을 사용하려면 이 옵션을 사용하십시오.
- **버전 지정**. 특정 설문조사 버전을 사용하려면 이 옵션을 선택하십시오.

예를 들어, 둘 이상의 버전에 대한 케이스 데이터를 내보내려고 하는데 한 버전으로 수집된 케이스 데이터가 다른 버전에서 유효하지 않음을 의미하는 변수 및 범주 정의에 대한 변경사항이 작성된 경우에는 모든 버전을 선택하는 것이 유용합니다. 케이스 데이터를 내보낼 모든 버전을 선택하는 것은 일반적으로 버전 간 차이로 인한 유효성 오류 없이 동시에 다양한 버전으로 수집된 케이스 데이터를 내보낼 수 있음을 의미합니다. 하지만 버전 변경사항에 따라 일부 유효성 오류가 여전히 발생할 수 있습니다.

언어. 질문 및 연관된 텍스트는 여러 언어로 메타데이터에 저장할 수 있습니다. 설문조사의 기본 언어를 사용하거나 특정 언어를 지정할 수 있습니다. 지정된 언어로 항목을 사용할 수 없는 경우에는 기본값이 사용됩니다.

컨텍스트. 사용할 사용자 컨텍스트를 선택하십시오. 사용자 컨텍스트는 표시되는 텍스트를 제어합니다. 예를 들어, 질문 텍스트를 표시하려면 **질문**을 선택하고 데이터 분석 시 표시하기에 적합한 더 짧은 텍스트를 표시하려면 **분석**을 선택하십시오.

레이블 유형. 정의된 레이블의 유형을 나열합니다. 기본값은 **레이블**이며 이는 분석 사용자 컨텍스트의 변수 설명과 질문 사용자 컨텍스트의 질문 텍스트에 사용됩니다. 지시사항, 설명 등에 대해 다른 레이블 유형을 정의할 수 있습니다.

③ 데이터베이스 연결 문자열

Data Collection 노드를 사용하여 OLE-DB 또는 ODBC를 통해 데이터베이스에서 케이스 데이터를 가져오는 경우 파일 탭에서 **편집**을 선택하여 연결 문자열 대화 상자에 액세스하십시오. 이 대화 상자에서는 연결을 미세 조정하기 위해 제공자에게 전달된 연결 문자열을 사용자 정의할 수 있습니다.

④ 고급 특성

Data Collection 노드를 사용하여 명시적 로그인이 필요한 데이터베이스로부터 케이스 데이터를 가져오는 경우 **고급**을 선택하여 데이터 소스에 액세스하는 데 필요한 사용자 ID 및 비밀번호를 제공하십시오.

⑤ 다중 응답 세트 가져오기

변수의 가능한 각각의 값에 대해 별도의 플래그 필드를 가진 다중 이분형 세트로 Data Collection에서 다중 응답 변수를 가져올 수 있습니다. 예를 들어, 응답자가 목록에서 방문한 박물관을 선택하도록 요청되는 경우 세트에는 나열된 각 박물관에 대한 별도의 플래그 필드가 포함됩니다.

데이터를 가져온 후 필터 탭을 포함하는 노드에서 다중 응답 세트를 추가하거나 편집할 수 있습니다. 자세한 정보는 다중 응답 세트 편집 주제를 참조하십시오.

단일 필드로 다중 응답 가져오기(이전 릴리스에서 작성된 스트림을 위해)

SPSS® Modeler의 이전 릴리스에서는 위에 설명된 대로 다중 응답을 가져오는 대신 쉼표로 구분된 값을 사용하여 단일 필드로 다중 응답을 가져왔습니다. 이 방법은 기존 스트림을 지원하기 위해 여전히 지원되지만 새로운 방법을 사용하도록 해당 스트림을 업데이트하는 것이 좋습니다.

⑥ Data Collection 열 가져오기 참고

Data Collection 데이터의 열은 다음 표에 요약된 대로 SPSS® Modeler로 읽어옵니다.

표 1. Data Collection 열 가져오기 요약

Data Collection 열 유형	SPSS ModelerStorage	측정수준
부울 플래그(예/아니오)	문자열	플래그(값 0 및 1)
범주형	문자열	명목
날짜 또는 시간소인	시간소인	연속
Double(지정된 범위 내의 부동 소수점 값)	실수	연속
Long(지정된 범위 내의 정수 값)	정수	연속
텍스트(자유 텍스트 설명)	문자열	유형 없음
수준(질문 내 눈금 또는 루프를 표시함)	VDATA에서는 발생하지 않으며 SPSS Modeler로 가져오지 않음	

Data Collection 열 유형	SPSS ModelerStorage	측정수준
오브젝트(낙서 텍스트 또는 음성 녹음을 표시하는 팩스 등의 이분형 데이터)	SPSS Modeler로 가져오지 않음	
없음(알 수 없는 유형)	SPSS Modeler로 가져오지 않음	
Respondent.Serial 열(고유 ID를 각 반응자와 연관시킴)	정수	유형 없음

실제 값과 메타데이터에서 읽어온 값 레이블 사이의 가능한 불일치를 방지하기 위해 모든 메타데이터 값이 소문자로 변환됩니다. 예를 들어, 값 레이블 E1720_years는 e1720_years로 변환됩니다.

(9) IBM Cognos 소스 노드

IBM Cognos 소스 노드에서는 Cognos 데이터베이스 데이터 또는 단일 목록 보고서를 데이터 마이닝 세션으로 가져올 수 있습니다. 이 방식으로 Cognos의 비즈니스 인텔리전스 기능을 IBM® SPSS® Modeler의 예측 분석 기능과 결합할 수 있습니다. 관계형, DMR(Dimensionally-Modeled Relational) 및 OLAP 데이터를 가져올 수 있습니다.

Cognos 서버 연결에서 먼저 데이터 또는 보고서를 가져올 위치를 선택하십시오. 한 위치에는 하나의 Cognos 모델과 해당 모델과 연관된 모든 폴더, 쿼리, 보고서, 보기, 단축키, URL 및 작업 정의가 포함되어 있습니다. Cognos 모델은 비즈니스 규칙, 데이터 설명, 데이터 관계, 비즈니스 차원 및 계층, 기타 관리 작업을 정의합니다.

데이터를 가져오는 경우에는 선택된 패키지에서 가져올 오브젝트를 선택하십시오. 가져올 수 있는 오브젝트에는 쿼리 제목(데이터베이스 테이블을 나타냄) 또는 개별 쿼리 항목(테이블 열을 나타냄)이 포함됩니다. 자세한 정보는 Cognos 오브젝트 아이콘의 내용을 참조하십시오.
패키지에 정의된 필터가 있으면 이 필터 중 하나 이상을 가져올 수 있습니다. 가져오는 필터가 가져온 데이터와 연관되는 경우에는 데이터를 가져오기 전에 필터가 적용됩니다. 가져올 데이터는 UTF-8 형식이어야 합니다.

보고서를 가져오는 경우에는 하나 이상의 보고서가 포함된 패키지(또는 패키지 내의 폴더)를 선택하십시오. 그런 다음 가져올 개별 보고서를 선택하십시오. 단일 목록 보고서만 가져올 수 있으며 다중 목록은 지원되지 않습니다.

데이터 오브젝트 또는 보고서에 대해 매개변수가 정의된 경우에는 오브젝트 또는 보고서를 가져오기 전에 이 매개변수에 대한 값을 지정할 수 있습니다.

> **참고**: Cognos 소스 노드는 Cognos CQM 패키지만 지원합니다. DQM 패키지는 지원되지 않습니다.

① Cognos 오브젝트 아이콘

Cognos Analytics 데이터베이스에서 가져올 수 있는 다양한 유형의 오브젝트는 다음 표에서와 같이 다양한 아이콘으로 표시됩니다.

표 1. Cognos 오브젝트 아이콘

아이콘	오브젝트
	패키지
	네임스페이스
	쿼리제목
	쿼리항목
	측도차원
	측도
	차원
	수준계층
	수준
	필터
	보고서
	독립형계산

② Cognos 데이터 가져오기

IBM Cognos Analytics 데이터베이스에서 데이터를 가져오려면 IBM Cognos 대화 상자의 데이터 탭에서 **모드**가 **데이터**로 설정되어 있는지 확인하십시오.

연결. **편집**을 클릭하여 데이터 또는 보고서를 가져올 새 Cognos 연결의 세부사항을 정의할 수 있는 대화 상자를 표시하십시오. IBM® SPSS® Modeler를 통해 이미 Cognos 서버에 로그인한 경우에는 현재 연결의 세부사항도 편집할 수 있습니다. 자세한 정보는 Cognos 연결의 내용을 참조하십시오.

위치. Cognos 서버 연결을 설정한 경우 이 필드 옆에 있는 **편집**을 클릭하여 컨텐츠를 가져올 사용 가능한 패키지의 목록을 표시하십시오. 자세한 정보는 Cognos 위치 선택를₩ 참조하십시오.

컨텐츠. 선택된 패키지와 연관된 네임스페이스와 함께 선택된 패키지의 이름을 표시합니다. 네임스페이스를 두 번 클릭하여 가져올 수 있는 오브젝트를 표시하십시오. 다양한 오브젝트 유형이 다양한 아이콘으로 표시됩니다. 자세한 정보는 Cognos 오브젝트 아이콘의 내용을 참조하십시오.

가져올 오브젝트를 선택하려면 오브젝트를 선택한 후 두 개의 오른쪽 화살표의 상단을 클릭하여 오브젝트를 **가져올 필드** 분할창으로 이동하십시오. 쿼리 제목을 선택하면 해당 쿼리 항목을 모두 가져옵니다. 쿼리 제목을 두 번 클릭하면 해당 개별 쿼리 항목을 하나 이상 선택할 수 있도록 쿼리 제목이 펼쳐집니다. Ctrl+클릭(개별 항목 선택), Shift+클릭(항목 블록 선택) 및 Ctrl+A(모든 항목 선택)를 사용하여 다중 선택을 수행할 수 있습니다.

적용할 필터를 선택하려면(패키지에 필터가 정의된 경우) 컨텐츠 분할창의 필터로 이동하여 필터를 선택한 후 두 개의 오른쪽 화살표의 하단을 클릭하여 필터를 **적용할 필터** 분할창으로 이동하십시오. Ctrl+클릭(개별 필터 선택) 및 Shift+클릭(필터 블록 선택)을 사용하여 다중 선택을 수행할 수 있습니다.

가져올 필드. 처리를 위해 IBM SPSS Modeler로 가져오기 위해 선택한 데이터베이스 오브젝트를 나열합니다. 특정 오브젝트가 더 이상 필요하지 않은 경우에는 해당 오브젝트를 선택한 후 왼쪽 화살표를 클릭하여 해당 오브젝트를 **컨텐츠** 분할창으로 반환하십시오. **컨텐츠**의 경우와 동일한 방식으로 다중 선택을 수행할 수 있습니다.

적용할 필터. 데이터를 가져오기 전에 데이터에 적용하기 위해 선택한 필터를 나열합니다. 특정 필터가 더 이상 필요하지 않은 경우에는 해당 필터를 선택한 후 왼쪽 화살표를 클릭하여 해당 필터를 **컨텐츠** 분할창으로 반환하십시오. **컨텐츠**의 경우와 동일한 방식으로 다중 선택을 수행할 수 있습니다.

모수 이 단추를 사용할 수 있으면 선택된 오브젝트에 매개변수가 정의되어 있습니다. 데이터를 가져오기 전에 매개변수를 사용하여 조정을 수행할 수 있습니다(예: 매개변수화된 계산 수행). 매개변수가 정의되어 있지만 기본값이 제공되지 않은 경우에는 이 단추에 경고 삼각형이 표시됩니다. 이 단추를 클릭하여 매개변수를 표시하고 선택적으로 편집하십시오. 이 단추가 사용 안함으로 설정되면 보고서에 매개변수가 정의되어 있지 않습니다.

가져오기 전에 데이터 통합. 원시 데이터 대신 통합 데이터를 가져오려면 이 선택란을 선택하십시오.

③ Cognos 보고서 가져오기

IBM Cognos 데이터베이스에서 사전 정의된 보고서를 가져오려면 IBM Cognos 대화 상자의 데이터 탭에서 **모드**가 **보고서**로 설정되어 있는지 확인하십시오. 단일 목록 보고서만 가져올 수 있으며 다중 목록은 지원되지 않습니다.

연결. **편집**을 클릭하여 데이터 또는 보고서를 가져올 새 Cognos 연결의 세부사항을 정의할 수 있는 대화 상자를 표시하십시오. IBM® SPSS® Modeler를 통해 이미 Cognos 서버에 로그인한 경우에는 현재 연결의 세부사항도 편집할 수 있습니다. 자세한 정보는 Cognos 연결의 내용을 참조하십시오.

위치. Cognos 서버 연결을 설정한 경우 이 필드 옆에 있는 **편집**을 클릭하여 컨텐츠를 가져올 사용 가능한 패키지의 목록을 표시하십시오. 자세한 정보는 Cognos 위치 선택을 참조하십시오.

내용. 보고서가 포함된 선택된 패키지 또는 폴더의 이름을 표시합니다. 특정 보고서로 이동하여 선택한 후 오른쪽 화살표를 클릭하여 보고서를 **가져올 보고서** 필드로 가져오십시오.

가져올 보고서. IBM SPSS Modeler로 가져오도록 선택한 보고서를 표시합니다. 보고서가 더 이상 필요하지 않은 경우에는 해당 보고서를 선택한 후 왼쪽 화살표를 클릭하여 **컨텐츠** 분할창에 리턴하거나 다른 보고서를 이 필드로 가져오십시오.

모수 이 단추가 사용으로 설정되면 선택된 보고서에 매개변수가 정의되어 있습니다. 매개변수를 사용하여 보고서를 가져오기 전에 조정(예: 보고서 데이터의 시작 및 종료 날짜 지정)을 수행할 수 있습니다. 매개변수가 정의되어 있지만 기본값이 제공되지 않은 경우에는 이 단추에 경고 삼각형이 표시됩니다. 이 단추를 클릭하여 매개변수를 표시하고 선택적으로 편집하십시오. 이 단추가 사용 안함으로 설정되면 보고서에 매개변수가 정의되어 있지 않습니다.

④ Cognos 연결

Cognos 연결 대화 상자에서는 데이터베이스 오브젝트를 가져오거나 내보낼 Cognos Analytics 서버를 선택할 수 있습니다.

Cognos 서버 URL 가져오거나 내보낼 Cognos Analytics 서버의 URL을 입력하십시오. 이는 Cognos 서버에서 IBM Cognos 구성의 "외부 디스패처 URI" 환경 특성의 값입니다. 사용할 URL이 확실하지 않으면 Cognos 시스템 관리자에게 문의하십시오.

모드 특정 Cognos 네임스페이스, 사용자 이름 및 비밀번호로 로그인하려는 경우(예: 관리자로서) **신임 정보 설정**을 선택하십시오. 사용자 신임 정보 없이 로그인하려면 **익명 연결 사용**을 선택하십시오. 이 경우에는 기타 필드를 채우지 않습니다.

또는 IBM® SPSS® Collaboration and Deployment Services 리포지토리에 IBM Cognos 신임 정보가 저장되어 있는 경우, 사용자 이름 및 비밀번호 정보를 입력하거나 익명 연결을 작성하는 대신 이 신임 정보를 사용할 수 있습니다. 기존 신임 정보를 사용하려면 **저장된 신임 정보**를 선택하고 **신임 정보 이름**을 입력하거나 찾아보십시오.

Cognos 네임스페이스는 IBM SPSS Collaboration and Deployment Services의 도메인을 통해 모델링됩니다.

네임스페이스 ID 서버에 로그인하는 데 사용되는 Cognos 보안 인증 제공자를 지정하십시오. 인증 제공자는 사용자, 그룹 및 역할을 정의 및 유지보수하고 인증 프로세스를 제어하는 데 사용됩니다. 이는 네임스페이스 이름이 아니라 네임스페이스 ID입니다(ID는 이름과 항상 동일하지는 않음).

사용자 이름 서버에 로그온하는 데 사용하는 Cognos 사용자 이름을 입력하십시오.

비밀번호 지정된 사용자 이름과 연관된 비밀번호를 입력하십시오.

기본값으로 저장 노드를 열 때마다 설정을 다시 입력하지 않아도 되도록 이 설정을 기본값으로 저장하려면 이 단추를 클릭하십시오.

⑤ Cognos 위치 선택

위치 지정 대화 상자에서는 데이터를 가져올 Cognos 패키지 또는 보고서를 가져올 패키지 또는 폴더를 선택할 수 있습니다.

공용 폴더. 데이터를 가져오는 경우 이는 선택된 서버에서 사용 가능한 패키지 및 폴더를 나열합니다. 사용할 패키지를 선택한 후 **확인**을 클릭하십시오. Cognos 소스 노드당 하나의 패키지만 선택할 수 있습니다.

보고서를 가져오는 경우 이는 선택된 서버에서 사용 가능한 보고서가 포함된 폴더 및 패키지를 나열합니다. 패키지 또는 보고서 폴더를 선택한 후 **확인**을 클릭하십시오. Cognos 소스 노드당 하나의 패키지 또는 보고서 폴더만 선택할 수 있지만 보고서 폴더는 다른 보고서 폴더 및 개별 보고서를 포함할 수 있습니다.

⑥ 데이터 또는 보고서에 대한 매개변수 지정

Cognos Analytics에 데이터 오브젝트 또는 보고서에 대한 매개변수가 정의된 경우에는 해당 오브젝트 또는 보고서를 가져오기 전에 이 매개변수에 대한 값을 지정할 수 있습니다. 보고서에 대한 매개변수의 예로는 보고서 컨텐츠의 시작 및 종료 날짜가 있습니다.

이름 Cognos 데이터베이스에 지정된 매개변수 이름입니다.

유형. 매개변수에 대한 설명입니다.

값 매개변수에 할당할 값입니다. 값을 입력하거나 편집하려면 테이블에서 해당 셀을 두 번 클릭하십시오. 값은 여기서 검증되지 않으므로 유효하지 않은 값은 런타임 시 발견됩니다.

테이블에서 유효하지 않은 매개변수 자동으로 제거. 이 옵션은 기본적으로 선택되며 데이터 오브젝트 또는 보고서에서 발견된 유효하지 않은 매개변수를 모두 제거합니다.

(10) IBM Cognos TM1 소스 노드

IBM Cognos TM1 소스 노드에서는 Cognos TM1 데이터를 데이터 마이닝 세션으로 가져올 수 있습니다. 이러한 방식으로 Cognos의 엔터프라이즈 계획 기능과 IBM® SPSS® Modeler의 예측 분석 기능을 결합할 수 있습니다. 다차원 OLAP 큐브 데이터의 플랫 버전을 가져올 수 있습니다.

> **참고**: TM1 사용자에게는 큐브 쓰기 권한, 차원 읽기 권한 및 차원 요소 쓰기 권한이 필요합니다. 또한, IBM Cognos TM1 10.2 수정팩 3 이상이 있어야 SPSS Modeler에서 Cognos TM1 데이터를 가져오고 내보낼 수 있습니다. 이전 버전을 기반으로 하는 기존 스트림은 여전히 작동합니다.
> 이 노드에는 관리자 신임 정보가 필요하지 않습니다. 이전 레거시 17.1 이전의 TM1 노드를 여전히 사용하는 경우 관리자 신임 정보가 필요합니다.
>
> SPSS Modeler는 IntegratedSecurityMode 1, 4, 5를 통한 Cognos TM1 작업만 지원합니다.

데이터를 가져오기 전에 TM1의 데이터를 수정해야 합니다. 가져올 데이터는 UTF-8 형식이어야 합니다.

IBM Cognos TM1 관리 호스트 연결에서 먼저 데이터를 가져올 TM1 서버를 선택합니다. 서버에는 하나 이상의 TM1 큐브가 있습니다. 그런 다음 필요한 큐브를 선택하고 해당 큐브 내에서 가져올 열 및 행을 선택합니다.

> **참고**: SPSS Modeler에서 TM1 소스 또는 내보내기 노드를 사용하려면 먼저 tm1s.cfg 파일에서 일부 설정을 유효화해야 합니다. 이 파일은 TM1 서버의 루트 디렉토리에 있는 TM1 서버 구성 파일입니다.
> - HTTPPortNumber - 유효한 포트 번호를 설정합니다. 일반적으로 1 - 65535입니다. 이 번호는 나중에 노드의 연결에 지정한 포트 번호가 아닙니다. 이 포트는 기본적으로 사용되지 않는 TM1에서 사용하는 내부 포트입니다. 필요하면 TM1 관리자에게 문의하여 이 포트의 올바른 설정을 확인하십시오.
> - UseSSL - *참*으로 설정하면 HTTPS가 전송 프로토콜로 사용됩니다. 이 경우 TM1 인증을 SPSS Modeler Server JRE로 가져와야 합니다.

① IBM Cognos TM1 데이터 가져오기

IBM Cognos TM1 데이터베이스에서 데이터를 가져오려면 IBM Cognos TM1 대화 상자의 **데이터** 탭에서 서버 연결 세부사항을 지정하고 큐브 및 데이터 세부사항을 선택하십시오.

> **참고**: 데이터를 가져오기 전에 TM1 내에서 몇 가지 전처리를 수행하여 데이터가 IBM® SPSS® Modeler에서 인식할 수 있는 형식인지 확인해야 합니다. 여기에는 보기가 가져오기에 올바른 크기 및 형태가 되도록 서브세트 편집기를 사용하여 데이터를 필터링하는 작업이 포함됩니다.

TM1에서 가져오는 영(0) 값은 "널" 값으로 취급됩니다(TM1은 공백과 0 값을 구별하지 않음). 또한 *정규 차원*의 비수치 데이터(또는 메타데이터)를 IBM SPSS Modeler로 가져올 수 있습니다. 그러나 비수치 *측도* 가져오기는 현재 지원되지 않습니다.

연결 유형. **관리 서버** 또는 **TM1 서버**를 선택하십시오. 관리 서버는 Planning Analytics on Cloud에서 제거되었으므로, 이전 관리 서버에 연결되는 이전 스트림이 있는 경우 해당 스트림이 Planning Analytics on Cloud를 대신 가리키도록 수정할 수 있습니다. 여기서 **관리 서버**를 선택할 경우 서버 URL(REST API의 호스트 이름)과 서버 이름을 입력해야 합니다. **TM1 서버**를 선택할 경우 다음 절로 이동하십시오.

TM1 서버 URL. 연결할 TM1 서버가 설치된 관리 호스트의 URL을 입력하십시오. 관리 호스트는 모든 TM1 서버에 대한 단일 URL로 정의됩니다. 이 URL에서, 사용하는 환경에 설치되어 실행 중인 모든 IBM Cognos TM1 서버를 검색하고 액세스할 수 있습니다. **로그인**을 클릭하십시오. 이전에 이 서버에 연결한 적이 없는 경우, **사용자 이름** 및 **비밀번호** 입력을 요구하는 프롬프트가 표시됩니다. 또는 이전에 입력하여 **저장된 신임 정보**로 저장한 로그인 세부사항을 검색할 수 있습니다.

가져올 TM1 큐브 보기 선택. 데이터를 가져올 수 있는 TM1 서버 내의 큐브 이름을 표시합니다. 큐브를 두 번 클릭하면 가져올 수 있는 보기 데이터가 표시됩니다.

> **참고**:
> 차원이 있는 큐브만 IBM SPSS Modeler로 가져올 수 있습니다.
>
> TM1 큐브 내의 요소에 대해 별명이 정의된 경우(예를 들어, 23277 값에 Sales라는 별명이 있는 경우), 별명이 아니라 값을 가져옵니다.

가져올 데이터를 선택하려면 보기를 선택하고 오른쪽 화살표를 클릭하여 **가져올 보기** 필드로 이동하십시오. 필요한 보기가 표시되지 않으면 큐브를 두 번 클릭하여 보기 목록을 펼치십시오. 공용 또는 개인용 보기를 선택할 수 있습니다.

행 차원. 가져오기로 선택한 데이터의 행 차원 이름을 나열합니다. 수준 목록에서 화면 이동하여 필요한 수준을 선택하십시오.

열 차원. 가져오기로 선택한 데이터의 열 차원 이름을 나열합니다. 수준 목록에서 화면 이동하여 필요한 수준을 선택하십시오.

컨텍스트 차원. 표시만 됩니다. 선택된 열 및 행과 관련된 컨텍스트 차원을 표시합니다.

(11) TWC 소스 노드

TWC 소스 노드는 The Weather Company, IBM Business에서 기상 데이터를 가져옵니다. 위치의 히스토리 기상 데이터 또는 예측 기상 데이터를 얻는 데 사용할 수 있습니다. 그러면 사용 가능한 가장 정확하고 정밀한 날짜 데이터를 사용하여 더 나은 의사결정을 내리기 위한 기상 중심 비즈니스 솔루션을 개발할 수 있습니다.

이 노드를 사용하여 latitude, longitude, time, day_ind(밤 또는 낮을 표시), temp, dewpt(이슬점), rh(상대 습도), feels_like 온도, heat_index, wc(풍속 냉각), wx_phrase(대체로 흐림, 부분 흐림 등), pressure, clds(구름), vis(가시성), wspd(풍속), gust, wdir(풍향), wdir_cardinal(NW, NNW, N 등), uv_index(자외선 지수) 등 uv_desc(낮음, 높음 등) 등의 기상 관련 데이터를 입력할 수 있습니다.

TWC 소스 노드는 다음 API를 사용합니다.
- 날씨 내역 데이터용 TWC Historical Observations Airport(http://goo.gl/DplOKj)
- 날씨 예측 데이터용 TWC Hourly Forecast(http://goo.gl/IJhhvZ)

위치

위도. 기상 데이터를 가져올 지역의 위도 값을 [-90.0~90.0] 형식으로 입력하십시오.

경도. 기상 데이터를 가져올 지역의 경도 값을 [-180.0~180.0] 형식으로 입력하십시오.

기타

라이센스 키. 라이센스 키는 필수입니다. The Weather Company에서 얻은 라이센스 키를 입력하십시오. 키가 없으면 관리자나 IBM 담당자에게 문의하십시오.

모든 사용자에게 키를 발행하는 대신 관리자가 IBM® SPSS® Modeler Server에서 새로운 config.cfg 파일에 키를 지정할 수 있습니다 이 경우 이 필드는 공백으로 둘 수 있습니다. 두 위치에 다 지정된 경우 이 대화 상자의 키에 우선순위가 있습니다. 관리자의 참고 사항: 서버에서 라이센스 키를 추가하려면 〈ModelerServerInstallation〉₩ext₩bin₩pasw.twcdata 위치에 LicenseKey=〈LICENCEKEY〉 컨텐츠(여기서 〈LICENSEKEY〉는 라이센스 키)가 있는 config.cfg 라는 새 파일을 작성하십시오.

단위. 사용할 측정 단위를 선택하십시오. **영어**, **메트릭** 또는 **하이브리드**. 기본값은 **메트릭**입니다.

시간 형식

UTC: 히스토리 기상 데이터를 가져오되 SPSS Modeler에서 TWC 시간별 예측 API에 액세스하지 않으려면 UTC를 선택하십시오. 이 옵션을 선택하는 경우, TWC 히스토리 관측값 공항 API에 액세스할 수 있는 권한만 사용권 키에 있으면 됩니다.

로컬: 시간을 UTC 시간에서 로컬 시간으로 변환하기 위해 SPSS Modeler에서 TWC 시간별 예측 API에 액세스해야 하는 경우, 로컬 시간 형식을 선택하십시오. 이 옵션을 선택하는 경우, 두 TWC API에 액세스할 수 있는 권한이 사용권 키에 필요합니다.

데이터 유형

히스토리. 히스토리 기상 데이터를 가져오려면 **히스토리**를 선택한 다음 시작 날짜와 종료 날짜를 YYYYMMDD 형식(예: 2012년 1월 1일의 경우 20120101)으로 지정하십시오.

예측. 예측 기상 데이터를 가져오려면 **예측**을 선택한 다음 예측할 시간을 지정하십시오.

(12) SAS 소스 노드

이 기능은 SPSS® Modeler Professional 및 SPSS Modeler Premium에서 사용 가능합니다.
SAS 소스 노드를 사용하면 SAS 데이터를 데이터 마이닝 세션으로 가져올 수 있습니다. 다음과 같은 네 가지 유형의 파일을 가져올 수 있습니다.

- SAS for Windows/OS2(.sd2)
- SAS for UNIX(.ssd)
- SAS 전송 파일(.tpt)
- SAS 버전 7/8/9(.sas7bdat)

데이터를 가져올 때 모든 변수가 유지되고 변수 유형이 변경되지 않습니다. 모든 케이스가 선택됩니다.

① SAS 소스 노드에 대한 옵션 설정

이 기능은 SPSS® Modeler Professional 및 SPSS Modeler Premium에서 사용 가능합니다.

가져오기. 전송할 SAS 파일의 유형을 선택합니다. **SAS for Windows/OS2(.sd2)**, **SAS for UNIX(.SSD)**, **SAS Transport File(.tpt)** 또는 **SAS 버전 7/8/9(.sas7bdat)**를 선택할 수 있습니다.

파일 가져오기. 파일의 이름을 지정합니다. 파일 이름을 입력하거나 생략 기호 단추(**...**)를 클릭하여 파일 위치를 찾아보십시오.

멤버. 위에서 선택한 SAS 전송 파일에서 가져올 멤버를 선택합니다. 멤버 이름을 입력하거나 **선택**을 클릭하여 파일 내의 모든 멤버를 찾아볼 수 있습니다.

SAS 데이터 파일에서 사용자 형식 읽기. 사용자 형식을 읽으려면 선택하십시오. SAS 파일은 데이터 및 데이터 형식(변수 레이블 등)을 다른 파일에 저장합니다. 대부분의 경우 형식도 함께 가져오려고 할 것입니다. 그러나 큰 데이터 세트가 있는 경우에는 이 옵션을 선택 취소하여 메모리를 절약할 수 있습니다.

형식 파일. 형식 파일이 필요하면 이 텍스트 상자가 활성화됩니다. 파일 이름을 입력하거나 생략기호 단추(**...**)를 클릭하여 파일 위치를 찾아보십시오.

변수 이름. SAS 파일에서 가져올 때 변수 이름 및 레이블을 처리하는 방법을 선택합니다. 여기에 포함시키려고 선택하는 메타데이터는 IBM® SPSS Modeler 내의 작업 전체에서 지속되며 SAS에서 사용하도록 다시 내보낼 수 있습니다.

- **이름 및 레이블 읽기**. 변수 이름 및 레이블을 둘 다 IBM SPSS Modeler에 읽어들일 수 있습니다. 기본적으로 이 옵션이 선택되고 변수 이름이 유형 노드에 표시됩니다. 레이블은 스트림 특성 대화 상자에서 지정된 옵션에 따라 표현식 작성기, 도표, 모델 브라우저 및 기타 유형의 출력에 표시될 수 있습니다.
- **이름으로 레이블 읽기**. SAS 파일에서 짧은 필드 이름이 아니라 설명 변수 레이블을 읽고 해당 레이블을 IBM SPSS Modeler에서 변수 이름으로 사용하려면 선택하십시오.

가. 멤버 선택

이 기능은 SPSS® Modeler Professional 및 SPSS Modeler Premium에서 사용 가능합니다.

일단 SAS 전송 파일을 지정했으면 IBM® SPSS Modeler로 읽어올 멤버를 선택해야 합니다. 이 대화 상자를 사용하면 전송 파일에서 사용 가능한 모든 멤버를 통해 찾아볼 수 있습니다.

(13) Excel 소스 노드

Excel 소스 노드를 사용하면 Microsoft Excel에서 데이터를 .xlsx 파일 형식으로 가져올 수 있습니다.

파일 유형. 가져오는 Excel 파일 유형을 선택하십시오.

파일 가져오기. 가져올 스프레드시트 파일의 이름 및 위치를 지정합니다.

이름 지정된 범위 사용. Excel 워크시트에서 정의된 대로 셀의 이름 지정된 범위를 지정할 수 있게 합니다. 생략 기호 단추(**...**)를 클릭하여 사용 가능한 범위 목록에서 선택하십시오. 이름 지정된 범위가 사용되는 경우에는 기타 워크시트 및 데이터 범위 설정은 더 이상 적용할 수 없으며 결과적으로 사용 안함으로 설정됩니다.

워크시트 선택. 가져올 워크시트를 지정합니다(인덱스별 또는 이름별).

- **인덱스별**. 가져올 워크시트에 대한 인덱스 값을 지정하십시오(첫 번째 워크시트에 대해 0으로 시작하여 두 번째 워크시트에 대해 1을 지정하는 등의 방식임).
- **이름별**. 가져올 워크시트의 이름을 지정하십시오. 생략 기호 단추(**...**)를 클릭하여 사용 가능한 워크시트 목록에서 선택하십시오.

워크시트의 범위. 첫 번째 비공백 행 또는 명시적 셀 범위로 시작하여 데이터를 가져올 수 있습니다.

- **첫 번째 비공백 행에서 범위 시작**. 첫 번째 비공백 셀을 찾아서 데이터 범위의 왼쪽 상단으로 사용합니다.
- **셀의 명시적 범위**. 행 및 열을 기준으로 명시적 범위를 지정할 수 있게 합니다. 예를 들어, Excel 범위 A1:D5를 지정하려면 첫 번째 필드에 A1을 입력하고 두 번째 필드에 D5를 입력하십시오(또는 R1C1과 R5C4). 공백 행을 포함하여 지정된 범위의 모든 행이 리턴됩니다.

공백 행에서. 둘 이상의 공백 행이 발생하는 경우에는 **읽기를 중지**할지 아니면 **공백행을 리턴**하여 공백 행을 포함한 모든 데이터를 워크시트의 끝까지 계속 읽을지 선택할 수 있습니다.

첫 번째 행에 열 이름이 있음. 지정된 범위의 첫 번째 행을 필드(열) 이름으로 사용해야 함을 나타냅니다. 선택되지 않은 경우에는 필드 이름이 자동으로 생성됩니다.

열과 유형에 대해 스캔할 행. IBM® SPSS® Modeler에서 Excel 데이터의 더 많은 행을 스캔하여 열 유형 및 저장 유형을 판별하도록 하려는 경우 이 값을 늘릴 수 있습니다. 기본값은 200개 행입니다. 이 설정은 성능에 영향을 미칠 수 있습니다.

필드 저장 공간 및 측정 수준

Excel에서 값을 읽어오는 경우에는 기본적으로 *연속형* 측정 수준으로 숫자 저장 공간을 가진 필드를 읽어오며 문자열 필드는 명목으로 읽어옵니다. 유형 탭에서 수동으로 측정 수준(연속형 대 명목)을 변경할 수 있지만 저장 공간은 자동으로 결정됩니다(필요한 경우에는 채움 노드 또는 파생 노드에서 to_integer 등의 변환 함수를 사용하여 변경할 수 있음). 자세한 정보는 필드 저장 공간 및 형식화 설정의 내용을 참조하십시오.

기본적으로 숫자 값과 문자열 값의 혼합을 가진 필드는 숫자로 읽어오며 이는 모든 문자열 값이 IBM SPSS Modeler에서 널(시스템 결측) 값으로 설정됨을 의미합니다. 이는 Excel과 달리 IBM SPSS Modeler는 필드 내에서 혼합 저장 유형을 허용하지 않기 때문에 발생합니다. 이를 방지하려면 Excel 스프레드시트에서 셀 형식을 텍스트로 수동으로 설정하십시오. 그러면 모든 값(숫자 포함)을 문자열로 읽어옵니다.

(14) XML 소스 노드

이 기능은 SPSS® Modeler Professional 및 SPSS Modeler Premium에서 사용 가능합니다.

XML 소스 노드를 사용하여 XML 형식 파일의 데이터를 IBM® SPSS Modeler 스트림으로 가져오십시오. XML은 데이터 교환을 위한 표준 언어이며 이러한 목적을 위해 다수의 조직에서 선택하는 형식입니다. 예를 들어, 국세청에서 온라인으로 제출된 XML 형식의 세금 환급 데이터를 분석하려 할 수 있습니다(http://www.w3.org/standards/xml/ 참조).

XML 데이터를 IBM SPSS Modeler 스트림으로 가져오면 소스에 대해 광범위한 예측 분석 함수를 수행할 수 있습니다. XML 데이터는 테이블 형식으로 구문 분석됩니다. 테이블 형식에서 열은 XML 요소 및 속성의 다양한 중첩 수준에 해당합니다. XML 항목은 XPath 형식으로 표시됩니다(http://www.w3.org/TR/xpath20/ 참조).

> **중요사항**: XML 소스 노드에서는 네임스페이스 선언을 고려하지 않습니다. 따라서 예를 들어 XML 파일의 name 태그에는 콜론(:) 문자가 포함될 수 없습니다. 포함하는 경우 실행 중에 올바르지 않은 문자에 대한 오류가 표시됩니다.

단일 파일 읽기. 기본적으로, SPSS Modeler는 **XML 데이터 소스** 필드에 지정하는 하나의 파일을 읽습니다.

디렉토리의 모든 XML 파일 읽기. 특정 디렉토리에 있는 모든 XML 파일을 읽으려면 이 옵션을 선택하십시오. 표시되는 **디렉토리** 필드에 위치를 지정하십시오. 지정된 디렉토리의 모든 서브디렉토리에 있는 파일을 추가로 읽으려면 **서브디렉토리 포함** 선택란을 선택하십시오.

XML 데이터 소스. 가져올 XML 소스 파일의 전체 경로 및 파일 이름을 입력하거나 찾아보기 단추를 사용하여 파일을 찾으십시오.

XML 스키마. (선택사항) XML 구조를 읽을 XSD 또는 DTD 파일의 전체 경로 및 파일 이름을 지정하거나 찾아보기 단추를 사용하여 이 파일을 찾으십시오. 이 필드를 비워 두면 XML 소스 파일에서 구조를 읽습니다. XSD 또는 DTD 파일은 둘 이상의 루트 요소를 포함할 수 있습니다. 이 경우, 초점을 다른 필드로 바꾸면 사용할 루트 요소를 선택하는 대화 상자가 표시됩니다. 자세한 정보는 여러 루트 요소에서 선택의 내용을 참조하십시오.

> **참고**: SPSS Modeler는 XSD 표시기를 무시합니다.

XML 구조. XML 소스 파일의 구조를 표시하는 계층 구조 트리(또는 **XML 스키마** 필드에 스키마를 지정한 경우 스키마)입니다. 레코드 경계를 정의하려면 요소를 선택하고 오른쪽 화살표 단추를 클릭하여 해당 항목을 **레코드** 필드로 복사하십시오.

속성 표시. **XML 구조** 필드에 있는 XML 요소의 속성을 표시하거나 숨깁니다.

레코드(XPath 표현식). XML 구조 필드에서 복사된 요소의 XPath 구문을 표시합니다. 그러면 이 요소가 XML 구조에서 강조표시되며 레코드 경계를 정의합니다. 소스 파일에서 이 요소를 만날 때마다 새 레코드가 작성됩니다. 이 필드가 비어 있으면 루트 아래의 첫 번째 하위 요소가 레코드 경계로 사용됩니다.

모든 데이터 읽기. 기본적으로 소스 파일의 모든 데이터를 스트림으로 읽습니다.

읽을 데이터 지정. 개별 요소나 속성 또는 둘 다를 가져오려면 이 옵션을 선택하십시오. 이 옵션을 선택하면 가져올 데이터를 지정할 수 있는 필드 테이블이 사용 가능하게 됩니다.

필드. 이 테이블은 **읽을 데이터 지정** 옵션을 선택한 경우 가져오기를 위해 선택한 요소 및 속성을 나열합니다. 요소 또는 속성의 XPath 구문을 XPath 열에 직접 입력하거나 XML 구조에서 요소 또는 속성을 선택하고 오른쪽 화살표 단추를 클릭하여 해당 항목을 테이블로 복사할 수 있습니다. 요소의 모든 하위 요소 및 속성을 복사하려면 XML 구조에서 해당 요소를 선택하고 이중 화살표 단추를 클릭하십시오.

- **XPath**. 가져올 항목의 XPath 구문입니다.
- **위치**. 가져올 항목의 XML 구조에서의 위치입니다. **고정 경로**는 XML 구조에서 강조표시된 요소(또는 강조표시된 요소가 없는 경우 루트 아래의 첫 번째 하위 요소)에 상대적인 항목의 경로를 표시합니다. **임의 위치**는 XML 구조의 임의 위치에 지정된 이름의 항목이 있음을 표시합니다. **사용자 정의**는 XPath 열에 직접 위치를 입력하는 경우에 표시됩니다.

① 여러 루트 요소에서 선택

이 기능은 SPSS® Modeler Professional 및 SPSS Modeler Premium에서 사용 가능합니다.

적절한 양식의 XML 파일은 하나의 루트 요소만 포함할 수 있는 반면, XSD 또는 DTD 파일은 여러 루트를 포함할 수 있습니다. 루트 중 하나가 XML 소스 파일의 루트와 일치하는 경우 해당 루트 요소가 사용되고, 그렇지 않은 경우에는 사용할 루트 요소를 선택해야 합니다.

표시할 루트 선택. 사용할 루트 요소를 선택하십시오. 기본값은 XSD 또는 DTD 구조의 첫 번째 루트 요소입니다.

② XML 소스 데이터에서 원하지 않는 공백 제거

이 기능은 SPSS® Modeler Professional 및 SPSS Modeler Premium에서 사용 가능합니다.

XML 소스 데이터의 행 바꾸기는 [CR][LF] 문자 조합으로 구현될 수 있습니다. 일부 경우에 행 바꾸기는 텍스트 문자열 가운데 올 수 있습니다.

```
<description>An in-depth look at creating applications[CR][LF]
with XML.</description>
```

파일이 일부 애플리케이션(예: 웹 브라우저)에서 열릴 때는 이러한 행 바꾸기가 표시되지 않을 수 있습니다. 그러나 XML 소스 노드를 통해 스트림으로 데이터를 읽으면 행 바꾸기가 일련의 공백 문자로 변환됩니다.

채움 노드를 사용하여 원하지 않는 공백을 제거함으로써 이러한 문제를 정정할 수 있습니다.

다음은 이러한 목적을 달성하기 위한 방법의 예입니다.

1. XML 소스 노드에 채움 노드를 연결하십시오.
2. 채움 노드를 열고 필드 선택기를 사용하여 원하지 않는 공백이 있는 필드를 선택하십시오.
3. **대체를 조건 기반**으로 설정하고 **조건**을 **true**로 설정하십시오.
4. **대체 대상** 필드에 replace(" ","",@FIELD)를 입력하고 확인을 클릭하십시오.
5. 테이블 노드를 채움 노드에 연결하고 스트림을 실행하십시오.

이제 테이블 노드 출력에서 텍스트가 추가 공백 없이 표시됩니다.

(15) 사용자 입력 노드

사용자 입력 노드는 처음부터 또는 기존 데이터를 변경하여 합성 데이터를 작성하는 쉬운 방법을 제공합니다. 이것은 예를 들어 모델링을 위한 검정 데이터 세트를 작성할 때 유용합니다.

처음부터 데이터 작성

사용자 입력 노드는 소스 팔레트에서 사용할 수 있으며 스트림 캔버스에 바로 추가할 수 있습니다.

1. 노드 팔레트의 **소스** 탭을 클릭하십시오.
2. 끌어서 놓기 또는 두 번 클릭으로 사용자 입력 노드를 스트림 캔버스에 추가하십시오.
3. 두 번 클릭하여 해당 대화 상자를 열고 필드 및 값을 지정하십시오.

참고: 소스 팔레트에서 선택한 사용자 입력 노드는 완전히 비어 있고 필드 및 데이터 정보가 없습니다. 이로 인해 완전히 처음부터 합성 데이터를 작성할 수 있습니다.

기존 데이터 소스에서 데이터 생성

스트림의 비터미널 노드에서 사용자 입력 노드를 생성할 수도 있습니다.

1. 스트림에서 노드를 대체할 위치를 결정하십시오.
2. 해당 데이터를 사용자 입력 노드에 공급할 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 **사용자 입력 노드 생성**을 선택하십시오.
3. 모든 다운스트림 프로세스가 연결된 사용자 입력 노드가 표시되어 데이터 스트림의 해당 위치에 있는 기존 노드를 대체합니다. 노드는 생성될 때 메타데이터에서 모든 데이터 구조 및 필드 유형 정보(사용 가능한 경우)를 상속합니다.

참고: 데이터가 스트림의 모든 노드를 거치지 않았다면 해당 노드들은 완전히 인스턴스화되지 않았으며, 이는 사용자 입력 노드로 대체할 때 저장 공간 및 데이터 값이 사용 불가능할 수도 있음을 의미합니다.

① 사용자 입력 노드의 옵션 설정

사용자 입력 노드에 대한 대화 상자에는 값을 입력하고 합성 데이터의 데이터 구조를 정의하는 데 사용할 수 있는 여러 도구가 있습니다. 생성된 노드의 경우, 데이터 탭의 테이블에 원래 데이터 소스의 필드 이름이 있습니다. 소스 팔레트에서 추가된 노드의 경우에는 테이블이 비어 있습니다. 테이블 옵션을 사용하여 다음 태스크를 수행할 수 있습니다.

- 테이블의 오른쪽에 있는 새 필드 추가 단추를 사용하여 새 필드를 추가합니다.
- 기존 필드의 이름을 바꿉니다.
- 각 필드의 데이터 저장 공간을 지정합니다.
- 값을 지정합니다.
- 표시 화면에서 필드 순서를 변경합니다.

데이터 입력

필드마다 테이블 오른쪽에 있는 값 선택도구 단추를 사용하여 원래 데이터 세트의 값을 삽입하거나 값을 지정할 수 있습니다. 값 지정에 대한 자세한 정보는 아래에 설명된 규칙을 참조하십시오. 필드를 비워 둘 수도 있습니다. 빈 필드는 시스템 널($null$)로 채워집니다.

문자열 값을 지정하려면 값 열에 문자열 값을 공백으로 분리하여 입력하면 됩니다.

```
Fred Ethel Martin
```

공백을 포함하는 문자열은 큰따옴표로 묶을 수 있습니다.

```
"Bill Smith" "Fred Martin" "Jack Jones"
```

숫자 필드의 경우, 여러 값을 동일한 방식으로(사이에 공백을 사용하여 나열) 입력할 수 있습니다.

```
10 12 14 16 18 20
```

또는 동일한 숫자 열을 해당 한계(10, 20)와 숫자 사이의 단계(2)를 설정하여 지정할 수 있습니다. 이 방법을 사용하는 경우 다음과 같이 입력합니다.

```
10,20,2
```

이 두 방법은 다음과 같이 하나를 다른 하나에 임베드하여 결합할 수 있습니다.

```
1 5 7 10,20,2 21 23
```

이와 같이 입력하는 경우 다음 값이 생성됩니다.

```
1 5 7 10 12 14 16 18 20 21 23
```

스트림 특성 대화 상자에서 선택된 현재 기본 형식을 사용하여 날짜 및 시간 값을 입력할 수 있습니다. 예를 들어, 다음과 같습니다.

```
11:04:00 11:05:00 11:06:00
```

```
2007-03-14 2007-03-15 2007-03-16
```

날짜 및 시간 구성요소를 모두 포함하는 시간소인 값의 경우 큰따옴표를 사용해야 합니다.

```
"2007-03-14 11:04:00" "2007-03-14 11:05:00" "2007-03-14 11:06:00"
```

추가 세부사항은 아래의 데이터 저장 공간에 대한 주석을 참조하십시오.

데이터 생성. 스트림을 실행할 때 레코드가 생성되는 방법을 지정할 수 있습니다.

- **모든 조합**. 가능한 모든 필드 값 조합을 포함하는 레코드를 생성하므로 각 필드 값이 여러 레코드에 표시됩니다. 가끔 필요한 양보다 많은 데이터를 생성할 수 있으므로 종종 이 노드 뒤에 표본 노드를 사용할 수도 있습니다.
- **순서대로**. 데이터 필드 값이 지정된 순서대로 레코드를 생성합니다. 각 필드 값이 하나의 레코드에만 표시됩니다. 총 레코드 수는 단일 필드의 값 최대 수와 같습니다. 필드에 값 최대 수보다 적은 수의 값이 있는 경우 정의되지 않은($null$) 값이 삽입됩니다.

예 표시

예를 들어, 다음과 같이 입력하면 다음 두 개의 표 예에 나열된 레코드를 생성합니다.

- **나이**. 30,60,10
- **BP**. 낮음
- **콜레스테롤**. 정상 높음
- **약제**. (비워 둠)

표 1. 데이터 생성 필드를 모든 조합으로 설정

나이	BP	콜레스테롤	약제
30	낮음	정상	$null$
30	낮음	높음	$null$
40	낮음	정상	$null$
40	낮음	높음	$null$
50	낮음	정상	$null$
50	낮음	높음	$null$
60	낮음	정상	$null$
60	낮음	높음	$null$

표 2. 데이터 생성 필드를 순서대로로 설정

나이	BP	콜레스테롤	약제
30	낮음	정상	$null$
40	$null$	높음	$null$
50	$null$	$null$	$null$
60	$null$	$null$	$null$

데이터 저장 공간

저장 공간은 데이터를 필드에 저장하는 방식을 설명합니다. 예를 들어, 값이 1 및 0인 필드는 정수 데이터를 저장합니다. 이는 데이터 사용에 대해 설명하고 저장 공간에 영향을 미치지 않는 측정 수준과 구별됩니다. 예를 들어, 값이 1 및 0인 정수 필드에 대한 측정 수준을 *플래그*로 설정할 수 있습니다. 이는 일반적으로 1 = *True* 및 0 = *False*임을 나타냅니다. 저장 공간은 소스에서 결정해야 하지만 측정 수준은 스트림의 어느 지점에서나 유형 노드를 사용하여 변경할 수 있습니다. 자세한 정보는 측정 수준의 내용을 참조하십시오.

사용 가능한 저장 유형은 다음과 같습니다.

- **문자열** 영숫자 데이터라고도 하는 숫자가 아닌 데이터가 포함된 필드에 사용됩니다. 문자열은 *fred*, *Class 2* 또는 *1234* 등의 문자열 시퀀스를 포함할 수 있습니다. 문자열의 숫자는 계산에서 사용할 수 없습니다.
- **정수** 값이 정수인 필드입니다.

- **실수** 값은 10진수를 포함할 수 있는 숫자입니다(정수로 제한되지 않음). 표시 형식은 스트림 특성 대화 상자에서 지정되며 유형 노드(형식 탭)에서 개별 필드에 대해 대체될 수 있습니다.
- **날짜** 연도, 월 및 일 등의 표준 형식으로 지정된 날짜 값입니다(예: 2007-09-26). 구체적인 형식은 스트림 특성 대화 상자에서 지정됩니다.
- **시간** 기간으로 측정된 시간입니다. 예를 들어, 1시간 26분 38초 동안 지속되는 서비스 호출은 스트림 특성 대화 상자에서 지정된 대로 현재 시간 형식에 따라 01:26:38로 표시될 수 있습니다.
- **시간소인** 스트림 특성 대화 상자의 현재 날짜 및 시간 형식에 따라 날짜와 시간 구성요소를 모두 포함하는 값입니다(예: 2007-09-26 09:04:00). 시간소인 값을 별도의 날짜 및 시간 값 대신 단일 값으로 해석하기 위해 시간소인 값을 큰따옴표로 묶어야 할 수 있습니다. (이는 예를 들어, 사용자 입력 노드에서 값을 입력하는 경우에 적용됩니다.)
- **목록** 지리 공간 및 콜렉션이라는 새로운 측정 수준과 함께 SPSS® Modeler 버전 17에서 도입된 목록 저장 공간 필드에는 단일 레코드에 대한 다중 값이 포함되어 있습니다. 모든 기타 저장 유형의 목록 버전이 있습니다.

표 3. 목록 저장 유형 아이콘

아이콘	저장유형
[A]	문자열 목록
[◇]	정수 목록
[#]	실수 목록
[🕒]	시간 목록
[📅]	날짜 목록
[📅]	시간소인 목록
[[]]	0(영)보다 큰 깊이를 가진 목록

또한 콜렉션 측정 수준과 함께 사용하기 위해 다음과 같은 측정 수준의 목록 버전이 있습니다.

표 4. 목록 측정 수준 아이콘

아이콘	측정수준
[📏]	연속형목록
[📊]	범주형목록
[🚦]	플래그목록
[●●●]	명목목록
[📶]	순서목록

목록은 세 가지 소스 노드(Analytic Server, 지리 공간 또는 가변파일) 중 하나에서 SPSS Modeler로 가져오거나 파생 또는 채움 필드 작업 노드를 사용하여 스트림 내에서 작성할 수 있습니다.

목록 및 해당 콜렉션 및 지리 공간 측정 수준과의 상호작용에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

저장 공간 변환. 채움 노드에서 to_string 및 to_integer 등의 다양한 변환 함수를 사용하여 필드에 대한 저장 공간을 변환할 수 있습니다. 자세한 정보는 채움 노드를 사용한 저장 공간 변환의 내용을 참조하십시오. 변환 함수(및 날짜 또는 시간 값 등의 특정 입력 유형이 필요한 기타 함수)는 스트림 특성 대화 상자에서 지정된 현재 형식에 따라 다릅니다. 예를 들어, 값이 *Jan 2018, Feb 2018* 등인 문자열 필드를 날짜 저장 공간으로 변환하려면 **MON YYYY**를 스트림의 기본 날짜 형식으로 선택하십시오. 파생 계산 중 임시 변환의 경우 파생 노드에서도 변환 함수를 사용할 수 있습니다. 파생 노드를 사용하여 범주형 값을 가진 문자열 필드 코딩 변경 등의 기타 조작을 수행할 수도 있습니다. 자세한 정보는 파생 노드를 사용하여 값 코딩변경의 내용을 참조하십시오.

혼합 데이터 읽기. 숫자 저장 공간(정수, 실수, 시간, 시간소인 또는 날짜)을 가진 필드에서 읽을 때 숫자가 아닌 값은 널값 또는 시스템 결측값으로 설정됩니다. 이는 일부 애플리케이션과는 달리 IBM® SPSS Modeler가 필드 내에서 혼합 저장 유형을 허용하지 않기 때문입니다. 이를 방지하기 위해 혼합 데이터가 포함된 필드는 필요에 따라 소스 노드 또는 외부 애플리케이션에서 저장 유형을 변경하여 문자열로 읽어야 합니다.

참고: 생성된 사용자 입력 노드는 인스턴스화된 경우 소스 노드에서 얻은 저장 공간 정보를 이미 포함할 수도 있습니다. 인스턴스화되지 않은 노드에는 저장 공간 또는 사용 유형 정보가 포함되지 않습니다.

값 지정 규칙

기호 필드의 경우 다음과 같이 값과 값 사이에 공백을 두어야 합니다.

HIGH MEDIUM LOW

숫자 필드의 경우, 여러 값을 동일한 방식으로(사이에 공백을 사용하여 나열) 입력할 수 있습니다.

10 12 14 16 18 20

또는 동일한 숫자 열을 해당 한계(10, 20)와 숫자 사이의 단계(2)를 설정하여 지정할 수 있습니다. 이 방법을 사용하는 경우 다음과 같이 입력합니다.

10,20,2

이 두 방법은 다음과 같이 하나를 다른 하나에 임베드하여 결합할 수 있습니다.

1 5 7 10,20,2 21 23

이와 같이 입력하는 경우 다음 값이 생성됩니다.

1 5 7 10 12 14 16 18 20 21 23

(16) 시뮬레이션 생성 노드

시뮬레이션 생성 노드는 히스토리 데이터 없이 사용자가 지정한 통계 분포를 사용하거나 기존 히스토리 데이터에 대해 시뮬레이션 적합 노드를 실행하여 얻은 분포를 자동으로 사용하여 시뮬레이션된 데이터를 생성하는 쉬운 방법을 제공합니다. 시뮬레이션된 데이터를 생성하는 것은 모델 입력에 불확실성이 존재하는 상황에서 예측 모델의 결과를 평가하기 원할 때 유용합니다.

히스토리 데이터 없이 데이터 작성

소스 팔레트에서 시뮬레이션 생성 노드를 사용할 수 있으며 스트림 캔버스에 직접 추가할 수 있습니다.

1. 노드 팔레트의 **소스** 탭을 클릭하십시오.
2. 시뮬레이션 생성 노드를 스트림 캔버스에 추가하려면 끌어서 놓거나 두 번 클릭하십시오.
3. 두 번 클릭하여 대화 상자를 열고 필드, 저장 유형, 통계 분포 및 분포모수를 지정하십시오.

참고: 소스 팔레트에서 선택된 시뮬레이션 생성 노드는 필드 및 분포 정보 없이 완전히 비어 있습니다. 따라서 사용자가 히스토리 데이터 없이 시뮬레이션 데이터를 작성할 수 있습니다.

기존 히스토리 데이터를 사용하여 시뮬레이션된 데이터 생성

시뮬레이션 생성 노드는 시뮬레이션 적합 터미널 노드를 실행하는 방법으로도 작성할 수 있습니다.

1. 시뮬레이션 적합 노드를 마우스 오른쪽 단추로 클릭하고 메뉴에서 **실행**을 선택하십시오.
2. 시뮬레이션 생성 노드는 시뮬레이션 적합 노드에 대한 업데이트 링크가 있는 스트림 캔버스에 표시됩니다.
3. 시뮬레이션 생성 노드는 생성될 때 모든 필드, 저장 유형 및 통계 분포 정보를 시뮬레이션 적합 노드에서 상속합니다.

시뮬레이션 적합 노드에 대한 업데이트 링크 정의

시뮬레이션 생성 노드 및 시뮬레이션 적합 노드 사이에 링크를 작성할 수 있습니다. 링크 작성은 히스토리 데이터에 대한 맞춤에 의해 판별된 가장 적합한 분포 정보가 있는 하나 이상의 필드를 업데이트할 때 유용합니다.

1. 시뮬레이션 생성 노드를 마우스 오른쪽 단추로 클릭하십시오.
2. 메뉴에서 **업데이트 링크 정의**를 선택하십시오. 커서가 링크 커서로 변경됩니다.
3. 또 다른 노드를 클릭하십시오. 이 노드가 시뮬레이션 적합 노드이면 링크가 작성된 것입니다. 이 노드가 시뮬레이션 적합 노드가 아니면 링크가 작성되지 않은 것이며 커서가 다시 일반 커서로 변경됩니다.

시뮬레이션 적합 노드의 필드가 시뮬레이션 생성 노드의 필드와 다른 경우, 차이가 있음을 알리는 메시지가 표시됩니다.

링크된 시뮬레이션 생성 노드를 업데이트하기 위해 시뮬레이션 적합 노드가 사용된 경우, 결과는 동일한 필드가 두 노드에 모두 존재하는지, 필드가 시뮬레이션 생성 노드에서 잠겨있지 않은지 여부에 따라 다릅니다. 시뮬레이션 적합 노드의 업데이트 결과는 다음 표에 표시됩니다.

표 1. 시뮬레이션 적합 노드의 업데이트 결과

	시뮬레이션의 필드	**적합노드**
시뮬레이션 생성 노드의 필드	**존재**	**결측값**
존재하며 잠겨있지 않음	필드를 덮어씁니다.	필드가 삭제됩니다.
결측	필드가 추가됩니다.	변경되지 않습니다.
존재하며 잠겨있음	필드의 분포를 덮어쓰지 않습니다. 세부사항 적합 대화 상자의 정보 및 상관관계가 업데이트됩니다.	필드를 덮어쓰지 않습니다. 상관계수는 0으로 설정됩니다.
재적합 시 최소/최대 지우지 않음 선택란이 선택됩니다.	최소, 최대 열의 값과 별개로	필드를 덮어씁니다.
재적합 시 상관계수를 재계산하지 않음 선택란이 선택됩니다.	필드가 잠겨있지 않으면 덮어씁니다.	상관계수는 덮어쓰지 않습니다.

시뮬레이션 적합 노드에 대한 업데이트 링크 제거

다음 단계를 수행하여 시뮬레이션 생성 노드 및 시뮬레이션 적합 노드 사이에 링크를 제거할 수 있습니다.

1. 시뮬레이션 생성 노드를 마우스 오른쪽 단추로 클릭하십시오.
2. 메뉴에서 **업데이트 링크 제거**를 선택하십시오. 링크가 제거됩니다.

① 시뮬레이션 생성 노드에 대한 옵션 설정

시뮬레이션 생성 노드 대화 상자의 데이터 탭에 대한 옵션을 사용하여 다음을 수행할 수 있습니다.

- 필드에 대한 통계 분포 정보를 보고 지정하고 편집할 수 있습니다.
- 필드 사이의 상관관계를 보고 지정하고 편집할 수 있습니다.
- 시뮬레이션할 반복계산 및 케이스 수를 지정합니다.

항목 선택. 시뮬레이션 생성 노드의 세 가지 뷰(시뮬레이션한 필드, 상관관계 및 고급 옵션) 사이에서 전환할 수 있습니다.

시뮬레이션한 필드 보기

시뮬레이션 생성 노드가 히스토리 데이터가 있는 시뮬레이션 적합 노드에서 생성되거나 업데이트된 경우, 시뮬레이션한 필드 보기에서 각 필드에 대한 통계 분포 정보를 보고 편집할 수 있습니다. 각 필드에 대한 다음 정보가 시뮬레이션 적합 노드로부터 시뮬레이션 생성 노드의 유형 탭에 복사됩니다.

- 측정 수준
- 값
- 결측값
- 검사
- 역할

히스토리 데이터가 없으면 저장 유형 및 분포 유형을 선택하고 필수 모수를 If you do not have historical data, you can define fields and specify their 입력하여 필드를 정의하고 분포를 지정할 수 있습니다. 이런 방법으로 데이터를 생성하면 **유형** 탭 또는 유형 노드 등에서 데이터가 인스턴스화될 때까지 각 필드의 측정 수준에 대한 정보를 사용할 수 없습니다.

시뮬레이션한 필드 보기는 다음 태스크를 수행하는 데 사용할 수 있는 여러 가지 도구를 포함합니다.

- 필드 추가 및 제거
- 표시 화면에서 필드 순서를 변경합니다.
- 각 필드에 대한 저장 유형 지정
- 각 필드에 대한 통계 분포 지정
- 각 필드의 통계 분포에 대한 모수 값 지정

시뮬레이션한 필드. 소스 팔레트에서 스트림 캔버스로 시뮬레이션 생성 노드가 추가된 경우, 이 표에 한 개의 빈 행이 추가됩니다. 이 행을 편집하면 비어 있는 새 행이 표의 아래쪽에 추가됩니다. 시뮬레이션 생성 노드가 시뮬레이션 적합 노드로부터 작성된 경우, 이 표에 히스토리 데이터의 각 필드에 대해 한 행이 포함됩니다. **새 필드 추가** 아이콘을 클릭하여 표에 추가 행을 추가할 수 있습니다.
시뮬레이션한 필드 표는 다음 열로 구성됩니다.

- **필드**. 필드의 이름을 포함합니다. 필드 이름은 셀에 입력하여 편집할 수 있습니다.
- **저장 공간** 이 열의 셀은 저장 유형의 드롭 다운 목록을 포함합니다. 사용가능한 저장 유형은 **문자열**, **정수**, **실수**, **시간**, **날짜** 및 **시간소인**입니다. 저장 유형을 선택하면 분포 열에서 사용 가능한 분포가 결정됩니다. 시뮬레이션 생성 노드가 시뮬레이션 적합 노드로부터 작성된 경우, 시뮬레이션 적합 노드로부터 저장 유형이 복사됩니다.

> ⓘ **참고**: 날짜/시간 저장 유형이 있는 필드의 경우, 분포모수를 정수로 지정해야 합니다. 예를 들어, 1970년 1월 1일의 평균 날짜를 지정하려면 정수 0을 사용하십시오. 부호 있는 정수는 1970년 1월 1일 자정 이후 또는 이전의 초 수를 나타냅니다.

- **상태**. 상태 열의 아이콘은 각 필드에 대한 적합 상태를 표시합니다.

	필드에 대해 분포가 지정되지 않았거나 하나 이상의 분포모수가 결측값입니다. 시뮬레이션을 실행하려면 이 필드에 대한 분포를 지정하고 모수에 대한 유효한 값을 입력해야 합니다.
	필드가 가장 가까운 적합 분포로 설정됩니다. ⓘ **참고**: 이 아이콘은 시뮬레이션 생성 노드가 시뮬레이션 적합 노드로부터 생성된 경우에만 표시될 수 있습니다.
	가장 가까운 적합 분포가 세부사항 적합 하위 대화 상자의 대체 분포로 대체되었습니다. 자세한 정보는 세부사항 적합의 내용을 참조하십시오.
	분포가 수동으로 지정되거나 편집되었으며 두 개 이상의 수준에서 지정된 모수를 포함할 수 있습니다.

- **잠겨있음**. 잠금 아이콘이 있는 열에서 선택란을 선택하여 시뮬레이션한 필드를 잠그면 링크된 시뮬레이션 적합 노드에 의한 자동 업데이트에서 필드가 제외됩니다. 이 방법은 수동으로 분포를 지정하고 링크된 시뮬레이션 적합 노드가 실행될 때 자동 분포 적합에 의해 영향을 받지 않도록 할 때 가장 유용합니다.
- **분포**. 이 열의 셀은 통계 분포의 드롭 다운 목록을 포함합니다. 저장 유형을 선택하면 지정된 필드에 대한 해당 열에서 사용 가능한 분포가 결정됩니다. 자세한 정보는 분포의 내용을 참조하십시오.

> ⓘ **참고**: 모든 필드에 대해 고정 분포를 지정할 수는 없습니다. 생성된 데이터의 모든 필드가 고정되도록 하려면 뒤에 균형 노드가 있는 사용자 입력 노드를 사용하십시오.

- **모수** 맞춰진 각 분포와 연관된 분포모수가 이 열에 표시됩니다. 모수의 다중 값은 콤마로 구분됩니다. 모수에 대한 다중 값을 지정하면 시뮬레이션에 대한 다중 반복이 생성됩니다. 자세한 정보는 반복의 내용을 참조하십시오. 모수가 결측값이면 상태 열에 표시되는 아이콘에서 반영됩니다. 모수에 대한 값을 지정하려면 관심 있는 필드에 대한 행에서 이 열을 클릭하고 목록에서 **지정**을 선택하십시오. 그러면 매개변수 지정 하위 대화 상자가 열립니다. 자세한 정보는 매개변수 지정의 내용을 참조하십시오. 분포 열에서 경험적 분포가 선택되면 이 열을 사용하지 않습니다.
- **최소**, **최대**. 일부 분포의 경우, 이 열에서 시뮬레이션한 데이터에 대한 최소값, 최대값 또는 둘 다를 지정할 수 있습니다. 최소값보다 작고 최대값보다 큰 시뮬레이션한 데이터는 지정된 분포에 유효한 경우라도 거부됩니다. 최소값 및 최대값을 지정하려면 관심 있는 필드에 해당되는 행에서 이 열을 클릭하고 목록에서 지정을 선택하십시오. 그러면 매개변수 지정 하위 대화 상자가 열립니다. 자세한 정보는 매개변수 **지정**의 내용을 참조하십시오. 분포 열에서 경험적 분포가 선택되면 이 열을 사용하지 않습니다.

가장 가까운 적합 사용. 시뮬레이션 생성 노드가 히스토리 데이터가 있는 시뮬레이션 적합 노드에서 자동으로 생성되고 시뮬레이션한 필드 표에서 단일 행이 선택된 경우에만 사용 가능합니다. 선택된 행의 필드에 대한 정보를 필드에 대한 가장 가까운 적합 분포의 정보로 바꾸십시오. 선택된 행의 정보를 편집한 경우, 이 단추를 누르면 시뮬레이션 적합 노드에서 판별된 가장 가까운 적합 분포로 정보가 다시 설정됩니다.

세부사항 적합. 시뮬레이션 생성 노드가 시뮬레이션 적합 노드에서 자동으로 생성된 경우에만 사용 가능합니다. 그러면 세부사항 적합 하위 대화 상자가 열립니다. 자세한 정보는 세부사항 적합의 내용을 참조하십시오.

시뮬레이션한 필드 보기의 오른쪽에 있는 아이콘을 사용하여 여러 가지 유용한 태스크를 수행할 수 있습니다. 이러한 아이콘에 대해서는 다음 표에서 설명합니다.

표 1. 시뮬레이션한 필드 보기의 아이콘

아이콘	도구팁	설명
	분포 모수 편집	시뮬레이션한 필드 표에서 단일 행이 선택된 경우에만 사용 가능합니다. 선택된 행에 대한 모수 지정 하위 대화 상자가 열립니다. 자세한 정보는 매개변수 지정의 내용을 참조하십시오.
	새 필드 추가	시뮬레이션한 필드 표에서 단일 행이 선택된 경우에만 사용 가능합니다. 시뮬레이션한 필드 표의 아래쪽까지 비어 있는 새 행을 추가합니다.
	다중 사본 생성	시뮬레이션한 필드 표에서 단일 행이 선택된 경우에만 사용 가능합니다. 그러면 필드 복제 하위 대화 상자가 열립니다. 자세한 정보는 복제 필드 주제를 참조하십시오.
	선택된 필드 삭제	시뮬레이션한 필드 표에서 선택된 행이 삭제됩니다.
	맨 위로 이동	선택된 행이 시뮬레이션한 필드 표의 맨 위 행이 아닌 위치에 있는 경우에만 사용 가능합니다. 선택된 행을 시뮬레이션한 필드 표의 맨 위쪽으로 이동합니다. 이 조치는 시뮬레이션한 데이터의 필드 순서에 영향을 미칩니다.
	위로 이동	선택된 행이 시뮬레이션한 필드 표의 맨 위 행이 아닌 위치에 있는 경우에만 사용 가능합니다. 선택된 행을 시뮬레이션한 필드 표의 한 위치 위쪽으로 이동합니다. 이 조치는 시뮬레이션한 데이터의 필드 순서에 영향을 미칩니다.
	아래로 이동	선택된 행이 시뮬레이션한 필드 표의 맨 아래 행이 아닌 위치에 있는 경우에만 사용 가능합니다. 선택된 행을 시뮬레이션한 필드 표의 한 위치 아래쪽으로 이동합니다. 이 조치는 시뮬레이션한 데이터의 필드 순서에 영향을 미칩니다.
	맨 아래로 이동	선택된 행이 시뮬레이션한 필드 표의 맨 아래 행이 아닌 위치에 있는 경우에만 사용 가능합니다. 선택된 행을 시뮬레이션한 필드 표의 맨 아래쪽으로 이동합니다. 이 조치는 시뮬레이션한 데이터의 필드 순서에 영향을 미칩니다.

재적합 시 최소 및 최대 지우지 않음. 이 옵션을 선택하면 분포가 연결된 시뮬레이션 적합 노드를 실행하여 업데이트될 때 최소값 및 최대값을 덮어쓰지 않습니다.

상관관계 보기

예측 모형에 대한 입력 필드는 종종 상관관계가 있는 것으로 알려져 있습니다. 예를 들어, 높이 및 가중치가 있습니다. 시뮬레이션한 값이 이러한 상관관계를 유지하기 위해서는 시뮬레이션되는 필드 사이의 상관관계를 고려해야 합니다.

시뮬레이션 생성 노드가 히스토리 데이터가 있는 시뮬레이션 적합 노드에서 생성되거나 업데이트된 경우, 상관관계 보기에서 필드 쌍 사이의 계산된 상관관계를 보고 편집할 수 있습니다. 히스토리 데이터가 없으면 필드가 상관된 방법에 대한 지식을 기반으로 하여 수동으로 상관관계를 지정할 수 있습니다.

> **참고**: 데이터가 생성되기 전에 상관행렬이 자동으로 positive semi-definite인지 여부를 확인하므로 도치될 수 있습니다. 행렬이 선형 독립변수이면 도치될 수 있습니다. 상관행렬을 도치할 수 없으면 자동으로 도치될 수 있도록 조정됩니다.

행렬 또는 목록 형식으로 상관관계를 표시할 수 있습니다.

상관행렬. 행렬에서 필드 쌍 사이의 상관관계를 표시합니다. 필드 이름이 행렬의 맨 위에서 아래 왼쪽으로 문자순으로 나열됩니다. 대각선 아래의 셀만 편집할 수 있습니다. -1.000 이상이며 1.000 이하인 값을 입력해야 합니다. 초점이 대각선 아래의 미러링된 셀에서 변경될 때 대각선 위 셀이 업데이트되고 두 셀 모두 동일한 값을 표시합니다. 대각선 셀은 항상 사용하지 않으며 항상 상관관계가 1.000입니다. 모든 기타 셀에 대한 기본값은 0.000입니다. 0.000 값은 연관된 필드 쌍 사이에 상관관계가 없음을 표시합니다. 연속형 및 순서 필드만 행렬에 포함됩니다. 고정 분포에 지정된 명목형, 범주형 및 플래그 필드는 표에 표시되지 않습니다.

상관관계 목록. 표에서 필드 쌍 사이의 상관관계를 표시합니다. 표의 각 행은 필드 쌍 사이의 상관관계를 표시합니다. 행은 추가되거나 삭제될 수 없습니다. 머리말 필드 1 및 필드 2가 있는 열은 편집할 수 없는 필드 이름을 포함합니다. 상관관계 열은 편집할 수 있는 상관관계를 포함하며 -1.000 이상이며 1.000 이하인 값을 입력해야 합니다. 모든 셀에 대한 기본값은 0.000입니다. 연속형 및 순서 필드만 목록에 포함됩니다. 고정 분포에 지정된 명목형, 범주형 및 플래그 필드는 목록에 표시되지 않습니다.

상관관계 재설정. 상관관계 재설정 대화 상자가 열립니다. 히스토리 데이터를 사용할 수 있으면 다음 세 가지 옵션 중 하나를 선택할 수 있습니다.
- **맞춰짐**. 현재 상관관계를 히스토리 데이터를 사용하여 계산한 상관관계로 바꿉니다.
- **0값**. 현재 상관관계를 0으로 바꿉니다.
- **취소**. 대화 상자를 닫습니다. 상관관계가 변경되지 않습니다.

히스토리 데이터를 사용할 수 없으나 상관관계를 변경한 경우에는 현재 상관관계를 0값으로 대체하거나 취소할 수 있습니다.
표시 형식. 상관관계를 행렬로 표시하려면 표를 선택하십시오. 상관관계를 목록으로 표시하려면 목록을 선택하십시오.

재적합 시 상관관계를 재계산하지 않음. 수동으로 상관관계를 지정하고 시뮬레이션 적합 노드 및 히스토리 데이터를 사용하여 자동으로 분포를 맞출 때 덮어쓰지 않도록 하려면 이 옵션을 선택하십시오.

범주형 분포가 있는 입력에 대해 맞춰진 다방향 분할표 사용. 기본적으로 범주형 분포가 있는 모든 필드는 분할표(또는 범주형 분포가 있는 필드의 수에 따라 다방향 분할표)에 포함됩니다. 분할표는 상관관계와 같이 시뮬레이션 적합 노드가 실행될 때 구성됩니다. 분할표는 볼 수 없습니다. 이 옵션을 선택하면 범주형 분포가 있는 필드가 분할표의 실제 퍼센트를 사용하여 시뮬레이션됩니다. 즉, 명목 필드 사이의 모든 연관이 시뮬레이션한 새 데이터에서 다시 작성됩니다. 이 옵션을 선택 취소하면 범주형 분포가 있는 필드가 분할표의 예측 퍼센트를 사용하여 시뮬레이션됩니다. 필드를 수정하면 분할표에서 필드가 제거됩니다.

고급 옵션 보기

시뮬레이션할 케이스 수. 시뮬레이션할 케이스 수 및 반복의 이름을 설정하는 방법을 지정할 수 있는 옵션이 표시됩니다.

- **최대 케이스 수**. 생성할 시뮬레이션한 데이터 및 연관된 목표 값의 최대수를 지정합니다. 기본값은 100,00, 최소값은 1000, 최대값은 2,147,483,647입니다.
- **반복**. 이 수는 자동으로 계산되며 편집할 수 없습니다. 반복은 분포 모수에 다중 값이 지정될 때마다 자동으로 계산됩니다.
- **전체 행**. 반복계산 수가 1보다 큰 경우에만 사용 가능합니다. 수는 표시된 방정식을 사용하여 자동으로 계산되며 편집할 수 없습니다.
- **반복 필드 생성**. 반복계산 수가 1보다 큰 경우에만 사용 가능합니다. 선택하면 이름 필드를 사용할 수 있습니다. 자세한 정보는 반복의 내용을 참조하십시오.
- **이름 반복 필드 생성** 선택란이 선택되었으며 반복계산 수가 1보다 큰 경우에만 사용 가능합니다. 이 텍스트 필드를 입력하여 반복계산 수를 편집하십시오. 자세한 정보는 반복 주제를 참조하십시오.

난수 시드. 난수 시드를 설정하면 시뮬레이션을 복제할 수 있습니다.

- **결과 복제**. 선택하면 **생성** 단추 및 **난수 시드** 필드를 사용할 수 있습니다.
- **난수 시드**. **복제 결과** 선택란을 선택한 경우에만 사용 가능합니다. 이 필드에서 난수 시드로 사용할 정수를 지정할 수 있습니다. 기본값은 629111597입니다.
- **생성**. **복제 결과** 선택란을 선택한 경우에만 사용 가능합니다. **난수 시드** 필드에 1 이상 999999999 이하의 의사 난수 정수를 생성합니다.

② 복제 필드

복제 필드 대화 상자에서는 선택한 필드의 사본 작성 수 및 각 사본의 이름을 결정하는 방식을 지정할 수 있습니다. 복합 효과를 조사할 때 필드의 다중 사본이 있는 경우에 유용합니다. 예를 들어, 수많은 연속된 시간 주기에 걸친 이율 또는 성장 비율 등입니다.

대화 상자의 제목 표시줄에는 선택된 필드의 이름이 포함됩니다.

작성할 사본 수. 작성할 필드의 사본 수를 포함합니다. 작성할 사본 수를 선택하려면 화살표를 클릭하십시오. 최소 사본 수는 1이며 최대값은 512입니다. 사본 수는 처음에 10으로 설정됩니다.

사본 접미문자. 각 사본에 대한 필드 이름의 끝에 추가되는 문자를 포함합니다. 이러한 문자는 필드 이름과 사본 수를 구분합니다. 이 필드를 입력하여 접미문자를 편집할 수 있습니다. 이 필드는 비어 있는 상태로 둘 수 있습니다. 이 경우, 필드 이름과 사본 수 사이에 문자가 없습니다. 기본 문자는 밑줄입니다.

초기 사본 번호. 첫 번째 사본에 대한 접미문자를 포함합니다. 초기 사본 수를 선택하려면 화살표를 클릭하십시오. 최소 초기 사본 수는 1이며 최대값은 1000입니다. 기본 초기 사본 수는 1입니다.

사본 수 단계. 접미문자 수에 대한 증분을 포함합니다. 증분을 선택하려면 화살표를 클릭하십시오. 최소 증분은 1이며 최대값은 255입니다. 증분은 초기에 1로 설정됩니다.

필드. 사본에 대한 필드 이름 미리보기를 포함합니다. 이는 복제 필드 대화 상자의 필드 중 임의의 필드가 편집될 때 업데이트됩니다. 이 텍스트는 자동으로 생성되고 편집할 수 없습니다.

확인. 대화 상자에서 지정된 대로 모든 사본을 생성합니다. 시뮬레이션 생성 노드 대화 상자의 시뮬레이션한 필드 표에서 복사된 필드를 포함하는 행 바로 아래에 사본이 추가됩니다.

취소. 대화 상자를 닫습니다. 작성된 변경사항을 삭제합니다.

③ 세부사항 적합

세부사항 적합 대화 상자는 시뮬레이션 적합 노드를 실행하여 시뮬레이션 생성 노드가 작성되었거나 업데이트된 경우에만 사용 가능합니다. 이 대화 상자는 선택된 필드에 대한 자동 분포 적합의 결과를 표시합니다. 분포는 적합도에 의해 순서가 지정되며 가장 가까운 적합 분포가 첫 번째로 나열됩니다. 이 대화 상자에서는 다음 태스크를 수행할 수 있습니다.

- 히스토리 데이터에 맞춰진 분포 검사
- 맞춰진 분포 중 한 개 선택

필드. 선택된 필드의 이름을 포함합니다. 이 텍스트는 편집할 수 없습니다.

다음으로 처리(측도). 선택된 필드의 측정 유형을 표시합니다. 시뮬레이션 생성 노드 대화 상자의 시뮬레이션한 필드 표에서 가져옵니다. 측정 유형은 화살표를 클릭하여 변경할 수 있으며 드롭 다운 목록에서 측정 유형을 선택할 수 있습니다. **연속형**, **명목형** 및 **순서**라는 세 가지 옵션이 있습니다.

분포. 분포 표는 측정 유형에 적합한 모든 분포를 표시합니다. 히스토리 데이터에 맞춰진 분포는 가장 적합한 것부터 가장 적합하지 않은 것까지 적합도에 의해 순서가 지정됩니다. 적합도는 시뮬레이션 적합 노드에서 선택된 적합 통계량에 의해 판별됩니다. 히스토리 데이터에 맞춰지지 않은 분포는 표에서 맞춰진 분포 아래에 문자순으로 나열됩니다.

분포 테이블에는 다음 열이 포함됩니다.

- **사용**. 선택된 단일 선택 단추는 현재 필드에 대해 선택된 분포를 표시합니다. 사용 열에서 원하는 분포에 대한 단일 선택 단추를 선택하여 가장 가까운 적합 분포를 대체할 수 있습니다. 또한 사용 열에서 단일 선택 단추를 선택하면 선택한 필드에 대한 히스토리 데이터의 히스토그램 또는 막대형 차트에 겹쳐진 분포의 도표가 표시됩니다. 한 번에 하나의 분포만 선택할 수 있습니다.
- **분포**. 분포의 이름을 포함합니다. 이 열은 편집할 수 없습니다.
- **통계량 적합**. 분포에 대해 계산되는 적합 통계량을 포함합니다. 이 열은 편집할 수 없습니다. 셀의 컨텐츠는 필드의 측정 유형에 따라 결정됩니다.
  - **연속형**. Anderson-Darling 및 Kolmogorov-Smirnoff 검정의 결과를 포함합니다. 검정과 연관된 p-값도 표시됩니다. 시뮬레이션 적합 노드에서 적합도 기준으로 선택된 적합 통계량이 첫 번째로 표시되고 분포 순서 지정에 사용됩니다. Anderson-Darling 통계량은 A=aval P=pval로 표시됩니다. Kolmogorov-Smirnoff 통계량은 K=kval P=pval로 표시됩니다. 통계를 계산할 수 없으면 숫자 대신 점이 표시됩니다.
  - **명목형 및 순서**. 카이제곱 검정의 결과를 포함합니다. 검정과 연관된 p-값도 표시됩니다. 통계량은 카이제곱=val P=pval로 표시됩니다. 분포가 맞춰지지 않으면 맞춰지지 않음이 표시됩니다. 분포가 수학적으로 맞춰지지 않으면 적합할 수 없음이 표시됩니다.

  참고: 경험적 분포에 대한 셀은 항상 비어 있습니다.

- **모수**. 맞춰진 각 분포와 연관된 분포모수를 포함합니다. 모수는 parameter_name = parameter_value로 표시되고 단일 공백으로 구분됩니다. 범주형 분포의 경우, 모수 이름은 범주형이며 모수값은 확률과 연관됩니다. 분포가 히스토리 데이터에 맞춰지지 않으면 셀이 비어 있습니다. 이 열은 편집할 수 없습니다.

히스토그램 썸네일. 선택한 필드에 대한 히스토리 데이터의 히스토그램에 겹쳐진 선택된 분포의 도표가 표시됩니다.

분포 썸네일. 선택된 분포에 대한 설명을 표시합니다.

확인. 대화 상자를 닫고 선택된 필드에 대한 시뮬레이션한 필드 표의 측정, 분포, 모수 및 최소, 최대 열의 값을 선택한 분포의 정보로 업데이트합니다. 상태 열의 아이콘도 선택된 분포가 데이터에 가장 가까운 적합이 있는 분포인지 여부를 반영하도록 업데이트됩니다.

취소. 대화 상자를 닫습니다. 작성된 변경사항을 삭제합니다.

④ 매개변수 지정

매개변수 지정 대화 상자에서 선택한 필드의 분포에 대한 모수값을 수동으로 지정할 수 있습니다. 또한 선택한 필드에 대해 다른 분포를 선택할 수도 있습니다.

매개변수 지정 대화 상자는 세 가지 방법으로 열 수 있습니다.

- 시뮬레이션 생성 노드 대화 상자의 시뮬레이션한 필드 표에서 필드 이름을 두 번 클릭하십시오.
- 시뮬레이션한 필드 표의 모수 또는 최소, 최대 열을 클릭하고 목록에서 지정을 선택하십시오.
- 시뮬레이션한 필드 표에서 행을 선택한 다음 분포모수 편집 아이콘을 클릭하십시오.

필드. 선택된 필드의 이름을 포함합니다. 이 텍스트는 편집할 수 없습니다.

분포. 선택된 필드의 분포를 포함합니다. 이는 시뮬레이션한 필드 표에서 가져옵니다. 분포는 화살표를 클릭하여 변경할 수 있으며 드롭 다운 목록에서 분포를 선택할 수 있습니다. 사용가능한 분포는 선택된 필드의 저장 유형에 따라 다릅니다.

방향. 이 옵션은 **분포** 필드에서 Dice 분포를 선택한 경우에만 사용 가능합니다. 필드를 분할할 방향 또는 범주의 수를 선택하려면 화살표를 클릭하십시오. 최소 방향 수는 2이며 최대수는 20입니다. 방향 수는 처음에 6으로 설정됩니다.

분포모수. 분포모수 표에는 선택한 분포의 각 모수에 대해 한 행이 포함됩니다.

> ⓘ **참고**: 분포가 형상 모수 $\alpha = k$이며 역 척도 모수 $\beta = 1/\theta$인 비율 모수를 사용합니다.

이 테이블에는 두 개의 열이 있습니다.
- **모수**. 모수의 이름을 포함합니다. 이 열은 편집할 수 없습니다.
- **값**. 모수의 값을 포함합니다. 시뮬레이션 생성 노드가 시뮬레이션 적합 노드로부터 작성되거나 업데이트된 경우, 분포를 히스토리 데이터에 맞춰서 결정된 모수값이 이 열의 셀에 포함됩니다. 소스 노드 팔레트에서 스트림 캔버스로 시뮬레이션 생성 노드가 추가된 경우, 이 열의 셀이 비어 있습니다. 값은 셀에 입력하여 편집할 수 있습니다. 각 분포에 필요한 모수 및 허용 가능한 모수값에 대한 자세한 정보는 분포의 내용을 참조하십시오.

모수에 대한 다중 값은 콤마로 구분되어야 합니다. 모수에 대한 다중 값을 지정하면 시뮬레이션의 다중 반복이 정의됩니다. 한 모수에 대해서만 다중 값을 지정할 수 있습니다.

> ⓘ **참고**: 날짜/시간 저장 유형이 있는 필드의 경우, 분포모수를 정수로 지정해야 합니다. 예를 들어, 1970년 1월 1일의 평균 날짜를 지정하려면 정수 0을 사용하십시오.

> ⓘ **참고**: Dice 분포를 선택하는 경우, 분포모수 표가 약간 다릅니다. 표에 각 방향 또는 범주에 대해 한 행이 포함됩니다. 표에 값 열 및 확률 열이 포함됩니다. 값 열은 각 범주에 대한 레이블을 포함합니다. 레이블의 기본값은 1-N 정수이며 여기서, N은 방향의 수입니다. 레이블은 셀에 입력하여 편집할 수 있습니다. 셀에 임의의 값을 입력할 수 있습니다. 숫자가 아닌 값을 사용하려는 경우, 저장 유형이 아직 문자열로 설정되지 않았으면 데이터 필드의 저장 유형을 문자열로 변경해야 합니다. 확률 열은 각 범주의 확률을 포함합니다. 확률은 편집할 수 없으며 1/N로 계산됩니다.

미리보기. 지정된 모수를 기준으로 하여 분포의 표본 도표를 표시합니다. 한 모수에 대해 두 개 이상의 값을 지정하면 모수의 각 값에 대한 표본 도표가 표시됩니다. 선택한 필드에 대해 히스토리 데이터를 사용할 수 있으면 분포의 도표가 히스토리 데이터의 히스토그램에 겹쳐집니다.

선택적 설정. 이 옵션을 사용하여 시뮬레이션한 데이터에 대한 최소값, 최대값 또는 둘 다를 지정할 수 있습니다. 최소값보다 작고 최대값보다 큰 시뮬레이션한 데이터는 지정된 분포에 유효한 경우라도 거부됩니다.

- 최소값 지정. **아래 값 거부** 필드를 사용하려면 선택하십시오. 경험적 분포를 선택하면 이 선택란을 사용할 수 없습니다.
- **아래 값 거부**. **최소값 지정**이 선택된 경우에만 사용 가능합니다. 시뮬레이션한 데이터에 대한 최소값을 입력하십시오. 이 값 보다 작은 시뮬레이션한 모든 값이 거부됩니다.
- **최대값 지정**. **위 값 거부** 필드를 사용하려면 선택하십시오. 경험적 분포를 선택하면 이 선택란을 사용할 수 없습니다.
- **위 값 거부**. **최대값 지정**이 선택된 경우에만 사용 가능합니다. 시뮬레이션한 데이터에 대한 최대값을 입력하십시오. 이 값 보다 큰 시뮬레이션한 모든 값이 거부됩니다.

확인. 대화 상자를 닫고 선택된 필드에 대한 시뮬레이션한 필드 표의 분포, 모수 및 최소, 최대 열의 값을 업데이트합니다. 상태 열의 아이콘도 선택된 분포를 반영하도록 업데이트됩니다.

취소. 대화 상자를 닫습니다. 작성된 변경사항을 삭제합니다.

가. 반복

고정 필드 또는 분포모수에 대해 둘 이상의 값을 지정한 경우, 지정된 각 값에 대해 별도의 시뮬레이션에 대해 효과적인 시뮬레이션된 케이스의 독립변수 집합이 생성됩니다. 그러면 필드 또는 모수 변형의 효과를 조사할 수 있습니다. 각 시뮬레이션된 케이스 집합을 *반복*이라고 합니다. 시뮬레이션된 데이터에서 반복이 수직누적됩니다.

시뮬레이션 생성 노드 대화 상자의 고급 옵션 보기에서 **반복 작성 필드** 선택란을 선택하면 반복 필드가 시뮬레이션된 데이터에 숫자 저장 공간이 있는 명목 필드로 추가됩니다. 이 필드의 이름은 고급 옵션 보기의 **이름** 필드에 입력하여 편집할 수 있습니다. 이 필드에는 시뮬레이션된 각 케이스가 속한 반복을 표시하는 레이블이 포함됩니다. 레이블의 양식은 반복의 유형에 따라 다릅니다.

- **고정 필드 반복**. 레이블이 필드의 이름이며 뒤에 등호 부호가 표시되고 그 뒤에 해당 반복에 대한 필드의 값이 표시됩니다. 다음과 같습니다.
  field_name = field_value

- **분포모수 반복**. 레이블이 필드의 이름이며 뒤에 콜론이 오고 그 뒤에 반복 모수의 이름이 오고 그 뒤에 등호 부호가 표시되고 그 뒤에 해당 반복에 대한 필드의 값이 표시됩니다. 다음과 같습니다.
  field_name:parameter_name = parameter_value

- **범주형 또는 범위 분포에 대한 분포모수 반복**. 레이블이 필드의 이름이며 뒤에 콜론이 표시되고 그 뒤에 "Iteration"이 표시되고 그 뒤에 반복 수가 표시됩니다. 다음과 같습니다.
  field_name: Iteration *iteration_number*

⑤ 분포

해당 필드에 대한 모수 지정 대화 상자를 열고 **분포** 목록에서 원하는 분포를 선택하고 **분포모수** 표에 분포모수를 입력하여 임의의 필드에 대한 확률 분포를 지정할 수 있습니다. 다음은 특정 분포에 대한 몇 가지 참고사항입니다.

- **범주형**. 범주형 분포는 범주라고 하는 고정 숫자의 숫자 값이 있는 입력 필드를 설명합니다. 각 범주에는 연관된 확률이 있으며 모든 범주의 확률 합계가 1이 됩니다.

  ⓘ **참고**: 합계가 1이 되지 않는, 범주에 대한 확률을 지정하면 경고가 수신될 수 있습니다.

- **음이항 - 실패**. 지정된 수의 성공이 관측되기 전에 시도 시퀀스의 실패 수의 분포를 설명합니다. Threshold 모수는 지정된 수의 성공이며 Probability 모수는 지정된 시도 내의 성공 확률입니다.
- **음이항 - 시행**. 지정된 수의 성공이 관측되기 전에 요구되는 시도 수의 분포를 설명합니다. Threshold 모수는 지정된 수의 성공이며 Probability 모수는 지정된 시도 내의 성공 확률입니다.
- **범위**. 이 분포는 각 구간에 확률이 지정된 구간 집합으로 구성되며 전체 구간에 걸친 확률의 값이 1이 됩니다. 지정된 구간 내의 값은 해당 구간에 대해 정의된 균일 분포에서 파생됩니다. 구간은 최소값, 최대값 및 연관된 확률을 입력하여 지정됩니다.
  예를 들어, 원자재의 비용이 단위당 $10 - $15 범위에 해당될 가능성이 40%이며 단위당 $15 - $20 범위에 해당될 가능성이 60%라고 가정해 봅시다. 그러면 첫 번째 구간과 연관된 확률을 0.4로 설정하고 두 번째 구간과 연관된 확률을 0.6으로 설정하여 두 개의 구간 [10 -

15] 및 [15 - 20]으로 구성되는 범위 분포로 비용을 모델링할 수 있습니다. 구간이 연속될 필요가 없으며 겹칠 수도 있습니다. 예를 들어, $10 - $15 및 $20 - $25 또는 $10 - $15 및 $13 - $16의 구간으로 지정할 수도 있습니다.

- **와이블**. *Location* 모수가 선택적 위치 모수이며 분포의 원점 위치를 지정합니다.

다음은 사용자 정의 분포 맞춤에 대해 사용할 수 있는 분포 및 모수에 허용 가능한 값을 표시하는 표입니다. 이러한 분포 중 일부는 시뮬레이션 적합 노드에 의해 저장 유형에 자동으로 맞춰지지 않는 경우에도 특정 저장 유형에 대해 사용자 정의 맞춤을 수행하는 데 사용 가능합니다.

표 1. 사용자 정의 맞춤에 대해 사용 가능한 분포

분포	사용자 정의 맞춤에 대해 지원되는 저장 유형	모수	모수한계	참고
베르누이	정수, 실수, 날짜/시간	확률	0 ≤ Probability ≤ 1	
베타	정수, 실수, 날짜/시간	모양 1 모양 2 최소값 최대값	≥ 0 ≥ 0 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.
이항	정수, 실수, 날짜/시간	시행 수(n) 확률 최소값 최대값	> 0, integer 0 ≤ Probability ≤ 1 < Maximum > Minimum	시행 수는 정수여야 합니다. 최소값 및 최대값은 선택사항입니다.
범주형	정수, 실수, 날짜/시간, 문자열	범주 이름 (또는 레이블)	0 ≤ Value ≤ 1	값은 범주의 확률입니다. 값의 합계가 1이 되어야 합니다. 그렇지 않으면 경고가 생성됩니다.
Dice	정수, 문자열	방향	2 ≤ Sides ≤ 20	각 범주(방향)의 확률이 1/N로 계산되며 N은 방향의 수입니다. 확률을 편집할 수 없습니다.
경험적 분포	정수, 실수, 날짜/시간			경험적 분포를 편집하거나 유형으로 선택할 수 없습니다. 경험적 분포는 히스토리 데이터가 있는 경우에만 사용할 수 있습니다.
지수	정수, 실수, 날짜/시간	척도 최소값 최대값	> 0 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.

분포	사용자 정의 맞춤에 대해 지원되는 저장 유형	모수	모수한계	참고
고정	정수, 실수, 날짜/시간, 문자열	값		모든 필드에 대해 고정 분포를 지정할 수는 없습니다. 생성된 데이터의 모든 필드가 고정되도록 하려면 뒤에 균형 노드가 있는 사용자 입력 노드를 사용하십시오.
감마	정수, 실수, 날짜/시간	형태 척도 최소값 최대값	≥ 0 ≥ 0 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다. 분포가 형상 모수 α = k이며 역 척도 모수 β = 1/θ인 비율 모수를 사용합니다.
로그정규	정수, 실수, 날짜/시간	모양 1 모양 2 최소값 최대값	≥ 0 ≥ 0 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.
음이항 - 실패	정수, 실수, 날짜/시간	임계값 확률 최소값 최대값	≥ 0 0 ≤ Probability ≤ 1 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.
음이항 - 시행	정수, 실수, 날짜/시간	임계값 확률 최소값 최대값	≥ 0 0 ≤ Probability ≤ 1 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.
정규	정수, 실수, 날짜/시간	평균 표준 편차 최소값 최대값	≥ 0 > 0 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.
포아송	정수, 실수, 날짜/시간	평균 최소값 최대값	≥ 0 < Maximum > Minimum	최소값 및 최대값은 선택사항입니다.
범위	정수, 실수, 날짜/시간	시작(X) 끝(X) 확률(X)	0 ≤ Value ≤ 1	X는 각 구간의 지수입니다. 확률 값의 합계가 1이 되어야 합니다.
삼각	정수, 실수, 날짜/시간	모드 최소값 최대값	Minimum ≤ Value ≤ Maximum < Maximum > Minimum	

분포	사용자 정의 맞춤에 대해 지원되는 저장 유형	모수	모수한계	참고
균일	정수, 실수, 날짜/시간	최소값 최대값	〈 Maximum 〉 Minimum	
와이블	정수, 실수, 날짜/시간	비율 척도 위치 최소값 최대값	〉0 〉0 ≥ 0 〈 Maximum 〉 Minimum	위치, 최대값 및 최소값은 선택사항입니다.

(17) 확장 가져오기 노드

확장 가져오기 노드를 사용하면 R 또는 Python for Spark 스크립트를 실행하여 데이터를 가져올 수 있습니다.

① 확장 가져오기 노드 - 명령문 탭

구문 유형(**R** 또는 **Python for Spark**)을 선택하십시오. 데이터를 가져오기 위한 사용자 정의 스크립트를 입력하거나 붙여넣으십시오. 명령문이 준비되면 **실행**을 클릭하여 확장 가져오기 노드를 실행할 수 있습니다.

② 확장 가져오기 노드 - 콘솔 출력 탭

콘솔 출력 탭에는 명령문 탭에서 R 스크립트 또는 Python for Spark 스크립트가 실행될 때 수신된 모든 출력이 포함됩니다. 예를 들어, R 스크립트를 사용하는 경우, **명령문** 탭의 **R 명령문** 필드의 R 스크립트가 실행될 때 R 콘솔에서 수신된 출력을 표시합니다. 이 출력에는 R 또는 Python 스크립트가 실행될 때 생성되는 R 또는 Python 오류 메시지 또는 경고가 포함됩니다. 출력은 주로 스크립트를 디버그하는 데 사용될 수 있습니다. **콘솔 출력** 탭에는 **R 명령문** 또는 **Python 명령문** 필드의 스크립트도 포함됩니다.

확장 가져오기 스크립트가 실행될 때마다 R 콘솔 또는 Python for Spark에서 수신된 출력이 **콘솔 출력** 탭의 내용을 덮어씁니다. 출력은 편집할 수 없습니다.

③ 필드 필터링 또는 이름 바꾸기

스트림의 어느 지점에서나 필드의 이름을 바꾸고 필드를 제외할 수 있습니다. 예를 들어, 의학 연구자로서 환자(레코드 수준 데이터)의 칼륨 수준(필드 수준 데이터)에 대해 관심이 없을 수 있으므로 K(칼륨) 필드를 필터링할 수 있습니다. 이는 소스 또는 출력 노드의 필터 탭을 사용하거나 별도의 필터 노드를 사용하여 수행할 수 있습니다. 액세스되는 노드에 관계없이 기능은 동일합니다.

- 가변파일, 고정 파일, Statistics 파일, XML 또는 확장 가져오기 등의 소스 노드에서 데이터를 IBM® SPSS® Modeler로 읽어올 때 필드의 이름을 바꾸거나 필드를 필터링할 수 있습니다.
- 필터 노드를 사용하면 스트림의 어느 지점에서나 필드의 이름을 바꾸거나 필드를 필터링할 수 있습니다.
- Statistics 내보내기, Statistics 변환, Statistics 모델 및 Statistics 출력 노드에서 IBM SPSS Statistics 이름 지정 표준을 준수하도록 필드를 필터링하거나 필드의 이름을 바꿀 수 있습니다. 자세한 정보는 IBM SPSS Statistics에 대한 필드 이름 변경 또는 필터링 주제를 참조하십시오.
- 위 노드의 필터 탭을 사용하여 다중 응답 세트를 정의하거나 편집할 수 있습니다. 자세한 정보는 다중 응답 세트 편집 주제를 참조하십시오.
- 최종적으로 필터 노드를 사용하여 한 소스 노드에서 다른 소스 노드로 필드를 맵핑할 수 있습니다.

④ 유형에 대한 정보 보기 및 설정

다양한 소스 노드 및 유형 노드에서 IBM® SPSS® Modeler의 모델링 및 기타 작업에 대해 유용한 필드 메타데이터 및 특성을 지정할 수 있습니다. 이러한 특성에는 다음이 포함됩니다.

- 데이터 세트 내의 각 필드에 대해 범위, 변수군, 정렬된 변수군 또는 플래그 등의 사용 유형을 지정합니다.
- 결측값 및 시스템 널값을 처리하기 위한 옵션을 설정합니다.
- 모델링 목적으로 필드 역할을 설정합니다.
- 데이터 세트에서 자동으로 값을 읽기 위해 사용되는 옵션 및 필드에 대한 값을 지정합니다.
- 필드 및 값 레이블을 지정합니다.

아래 목록에서 사용자의 상황에 맞는 도움말을 선택하십시오.

(18) 지리 공간적 소스 노드

지리 공간적 소스 노드를 사용하여 맵 또는 지리 공간적 데이터를 데이터 마이닝 세션으로 가져옵니다. 다음 두 방법 중 하나로 데이터를 가져올 수 있습니다.

- 모양 파일(.shp)로
- 맵 파일을 포함하는 계층 구조 파일 시스템이 포함된 ESRI 서버에 연결하여

> ⓘ **참고**: 공용 맵 서비스에만 연결할 수 있습니다.

STP(Spatio-Temporal Prediction) 모델은 예측에 맵 또는 공간 요소를 포함할 수 있습니다. 추가 정보는 STP(Spatio-Temporal Prediction) 모델링 노드의 내용을 참조하십시오.

① 지리 공간적 소스 노드에 대한 옵션 설정

데이터 소스 유형 모양 파일(.shp)에서 데이터를 가져오거나 **맵 서비스**에 연결할 수 있습니다.

모양 파일을 사용하는 경우에는 파일 이름 및 파일 경로를 입력하거나 파일을 찾아서 선택하십시오. 파일은 로컬 디렉토리에 있거나 맵핑된 드라이브에서 액세스해야 합니다. UNC(Uniform Naming Convention) 경로를 사용하여 파일에 액세스할 수는 없습니다.

> ⓘ **참고**: 모양 데이터에는 .shp 파일과 .dbf 파일이 모두 필요합니다. 두 파일은 동일한 이름을 가지고 동일한 폴더에 있어야 합니다. .shp 파일을 선택하면 .dbf 파일을 자동으로 가져옵니다. 또한 모양 데이터에 대한 좌표계를 지정하는 .prj 파일이 있을 수 있습니다.

맵 서비스를 사용하는 경우에는 서비스에 대한 URL을 입력한 후 **연결**을 클릭하십시오. 서비스에 연결하고 나면 해당 서비스 내 레이어가 **사용 가능한 맵** 분할창의 트리 구조에서 대화 상자의 맨 아래에 표시됩니다. 트리를 펼쳐서 필요한 레이어를 선택하십시오.

> ⓘ **참고**: 공용 맵 서비스에만 연결할 수 있습니다.

지리 공간적 데이터의 자동 정의

기본적으로 SPSS® Modeler는 가능한 경우 올바른 메타데이터를 가진 소스 노드에서 지리 공간적 데이터 필드를 자동으로 정의합니다. 메타데이터는 점 또는 다각형 등의 지리 공간적 필드의 측정 수준과 필드에서 사용하는 좌표계(원점(예: 위도 0, 경도 0) 및 측정 단위 등의 세부사항 포함)를 포함할 수 있습니다. 측정 수준에 대한 자세한 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.

모양 파일을 구성하는 .shp 및 .dbf 파일에는 키로 사용되는 공통 식별자 필드가 포함되어 있습니다. 예를 들어, .shp 파일은 국가(식별자로 사용되는 국가명 필드 포함)를 포함하고 .dbf 파일은 식별자로도 사용되는 국가명이 포함된 해당 국가에 대한 정보를 포함할 수 있습니다.

> ⓘ **참고**: 좌표계가 기본 SPSS Modeler 좌표계와 동일하지 않은 경우에는 필요한 좌표계를 사용하도록 데이터를 재투영해야 할 수 있습니다. 추가 정보는 재투영 노드의 내용을 참조하십시오.

(19) JSON 소스 노드

JSON 소스 노드에서는 UTF-8 인코딩을 사용하여 JSON 파일의 데이터를 SPSS® Modeler 스트림으로 가져올 수 있습니다. JSON 파일의 데이터는 *오브젝트*, *배열* 또는 값 형식일 수 있습니다. 이 JSON 소스 노드는 오브젝트 *배열* 읽기만 지원합니다. 오브젝트는 중첩할 수 없습니다.

JSON 데이터 예:

```
[
        {
                "After": 122762,
                "Promotion": 1467,
                "Cost": 23.99,
                "Class": "Confection",
                "Before": 114957
        },
        {
                "After": 137097,
                "Promotion": 1745,
                "Cost": 79.29,
                "Class": "Drink",
                "Before": 123378
        }
]
```

SPSS Modeler가 JSON 파일에서 데이터를 읽어들일 때 다음과 같은 변환이 수행됩니다.

표 1. JSON 데이터 저장 공간 변환

JSON값	SPSS Modeler 데이터 저장 공간
문자열	문자열
number(int)	정수
number(real)	실수
TRUE	1(정수)
FALSE	0(정수)
null	결측값

JSON 소스 노드 대화 상자에서는 다음 옵션을 사용할 수 있습니다.

JSON 데이터 소스: 가져올 JSON 파일을 선택하십시오.

JSON 문자열 형식: JSON 문자열의 형식을 지정하십시오. JSON 파일이 이름 및 값 쌍 콜렉션일 경우 **레코드**를 선택하십시오. JSON 소스 노드가 이름을 SPSS Modeler의 필드 이름으로 가져옵니다. JSON 데이터가 (이름 없이) 값만 사용하는 경우 **값**을 선택하십시오.

(20) 공통 소스 노드 탭

다음은 해당되는 탭을 클릭하여 모든 소스 노드에 지정할 수 있는 옵션입니다.

- **데이터 탭**. 기본 저장 유형을 변경하는 데 사용됩니다.
- **필터 탭**. 데이터 필드를 제거하거나 이름을 변경하는 데 사용됩니다. 이 탭은 필터 노드와 같은 기능을 제공합니다.자세한 정보는 필터링 옵션 설정 주제를 참조하십시오.
- **유형 탭**. 측정 수준을 설정하는 데 사용됩니다. 이 탭은 유형 노드와 같은 기능을 제공합니다.
- **주석 탭**. 이 탭은 모든 노드에 사용되며 노드의 이름을 바꾸고 사용자 맞춤 도구팁을 제공하며 긴 주석을 저장하기 위한 옵션을 제공합니다.

① 소스 노드에서 측정 수준 설정

필드 특성은 소스 노드 또는 별도의 유형 노드에서 지정할 수 있습니다. 두 노드 모두에서 기능은 비슷합니다. 사용 가능한 특성은 다음과 같습니다.

- **필드** 필드 이름을 두 번 클릭하여 IBM® SPSS® Modeler에서 데이터에 대한 값 및 필드 레이블을 지정하십시오. 예를 들어, IBM SPSS Statistics에서 가져온 필드 메타데이터를 여기서 보거나 수정할 수 있습니다. 마찬가지로 필드 및 해당 값에 대한 새 레이블을 작성할 수 있습니다. 여기서 지정하는 레이블은 스트림 특성 대화 상자에서 작성하는 선택사항에 따라 IBM SPSS Modeler 전체에 표시됩니다.
- **측정** 지정된 필드의 데이터 특성을 설명하는 데 사용되는 측정 수준입니다. 필드의 모든 세부사항이 알려져 있는 경우 이를 **완전히 인스턴스화**되어 있다고 합니다. 추가 정보는 측정 수준의 내용을 참조하십시오.

> ⓘ **참고**: 필드의 측정 수준은 데이터가 문자열, 정수, 실수, 날짜, 시간소인, 목록 중 어느 것으로 저장되는지를 표시하는 해당 저장 유형과 다릅니다.

- **값** 이 열에서는 데이터 세트에서 데이터 값을 읽어오기 위한 옵션을 지정하거나 **지정** 옵션을 사용하여 별도의 대화 상자에서 측정 수준 및 값을 지정할 수 있습니다. 해당 값을 읽지 않고 필드를 통과하도록 선택할 수도 있습니다. 추가 정보는 데이터 값의 내용을 참조하십시오.

> ⓘ **참고**: 해당 **필드** 항목에 목록이 포함되어 있으면 이 열에서 셀을 수정할 수 없습니다.

- **결측** 필드의 결측값이 처리되는 방식을 지정하는 데 사용됩니다. 추가 정보는 결측값 정의의 내용을 참조하십시오.

> ⓘ **참고**: 해당 **필드** 항목에 목록이 포함되어 있으면 이 열에서 셀을 수정할 수 없습니다.

- **확인** 이 열에서는 필드 값이 지정된 값 또는 범위를 준수하는지 확인하는 옵션을 설정할 수 있습니다. 추가 정보는 유형 값 검사의 내용을 참조하십시오.

> ⓘ **참고**: 해당 **필드** 항목에 목록이 포함되어 있으면 이 열에서 셀을 수정할 수 없습니다.

- **역할** 필드가 시스템 학습 프로세스에 대해 **입력**(예측자 필드)인지 아니면 **목표**(예측 필드)인지를 모델링 노드에 알리는 데 사용됩니다. 레코드를 훈련, 검정 및 검증을 위한 별도의 표본으로 파티셔닝하는 데 사용되는 필드를 표시하는 **파티션**과 함께 **모두** 및 **없음** 역할도 사용할 수 있습니다. **분할** 값은 필드의 가능한 각각의 값에 대해 별도의 모델이 작성되도록 지정합니다. 추가 정보는 필드 역할 설정의 내용을 참조하십시오.

자세한 정보는 유형 노드 주제를 참조하십시오.

가. 소스 노드에서 인스턴스화할 시기

두 가지 방법으로 필드의 값 및 데이터 저장 공간에 대해 학습할 수 있습니다. 이 인스턴스화는 처음으로 데이터를 IBM® SPSS® Modeler로 가져올 때 소스 노드에서 발생하거나 유형 노드를 데이터 스트림에 삽입하여 발생할 수 있습니다.

소스 노드에서 인스턴스화는 다음 경우에 유용합니다.

- 데이터 세트가 작은 경우
- 표현식 작성기를 사용하여 새 필드를 파생시키려고 계획하는 경우(인스턴스화를 수행하면 표현식 작성기에서 필드 값을 사용할 수 있게 됨)

일반적으로 데이터 세트가 그다지 크지 않고 나중에 스트림에서 필드를 추가하지 않을 계획인 경우에는 소스 노드에서 인스턴스화가 가장 편리한 방법입니다.

> ⓘ **참고**: 데이터베이스 내보내기 노드에서 데이터를 내보내는 경우 데이터를 완전히 인스턴스화해야 합니다.

② 소스 노드에서 필드 필터링

소스 노드 대화 상자의 필터 탭을 사용하면 데이터의 초기 검사를 기준으로 하여 다운스트림 작업에서 필드를 제외할 수 있습니다. 예를 들어, 데이터에 중복 필드가 있거나 관련이 없는 필드를 제외할 수 있을 정도로 데이터에 익숙한 경우에 유용합니다. 또는 나중에 스트림에 별도의 필터 노드를 추가할 수도 있습니다. 기능은 두 경우 모두 유사합니다. 자세한 정보는 필터링 옵션 설정 주제를 참조하십시오.

2) 레코드 작업 노드

(1) 레코드 작업 개요

레코드 작업 노드는 레코드 수준에서 데이터를 변경하는 데 사용됩니다. 이러한 작업은 데이터 마이닝의 **데이터 이해** 및 **데이터 준비** 단계에서 중요합니다. 사용자가 데이터를 사용자의 특정 비즈니스 요구에 맞출 수 있기 때문입니다.

예를 들어, 데이터 검토 노드(출력 팔레트)를 사용하여 수행된 데이터 검토의 결과를 기준으로 하여 지난 삼 개월 동안의 고객 구매 레코드를 합치도록 결정할 수 있습니다. 합치기 노드를 사용하여 *고객 ID* 등의 키 필드 값을 기준으로 하여 레코드를 합칠 수 있습니다. 또는 웹 사이트 방문에 대한 정보를 포함하는 데이터베이스의 레코드가 백만 개 이상이 되면 관리할 수 없다는 것을 발견하게 될 수도 있습니다. 표본 노드를 사용하여 모델링에 사용할 데이터 서브세트를 선택할 수 있습니다.

레코드 작업 팔레트에는 다음 노드가 포함됩니다.

	선택 노드는 특정 조건을 기반으로 데이터 스트림에서 레코드의 서브세트를 선택 또는 삭제합니다. 예를 들어, 특정 영업 지역에 관련된 레코드를 선택할 수 있습니다.
	표본 노드는 레코드의 서브세트를 선택합니다. 층화, 수평배열, 비임의(구조화) 표본을 포함하여 다양한 표본 유형이 지원됩니다. 표본추출은 성능을 개선하고 분석을 위해 관련 레코드나 트랜잭션 집단을 선택하는 데 유용할 수 있습니다.
	균형 노드는 데이터 세트의 불균형을 정정하므로, 데이터 세트가 지정된 조건을 준수합니다. 균형 지시문이 조건이 지정된 요인만큼 참인 레코드 비율을 조정합니다.
	통합 노드는 입력 레코드의 시퀀스를 요약되고 통합된 출력 레코드로 대체합니다.
	RFM(Recency, Frequency, Monetary) 통합 노드를 사용하면 고객의 히스토리 트랜잭션 데이터를 취하고 모든 사용하지 않은 데이터를 제거하고 마지막으로 다룬 시기, 작성한 트랜잭션 수, 해당 트랜잭션의 구매총액을 나열하는 단일 행으로 모든 나머지 트랜잭션 데이터를 결합할 수 있습니다.
	정렬 노드는 하나 이상의 필드의 값을 기반으로 레코드를 내림차순 또는 오름차순으로 정렬합니다.
	병합 노드는 다중 입력 레코드를 취하고 입력 필드의 일부 또는 모두를 포함하는 단일 출력 레코드를 작성합니다. 내부 고객 데이터 및 구매한 인구통계학적 데이터 같은 상이한 소스의 데이터 병합에 유용합니다.

	붙여쓰기 노드는 레코드 세트를 연결합니다. 비슷한 구조를 갖지만 상이한 데이터를 갖는 데이터 세트 결합에 유용합니다.
	고유 노드는 첫 번째 고유 레코드를 데이터 스트림으로 전달하거나 첫 번째 레코드를 삭제하고 대신 모든 중복을 데이터 스트림으로 전달하여 중복 레코드를 제거합니다.
	스트리밍 시계열 노드는 한 단계로 시계열 모델을 작성하고 스코어링합니다. 로컬 또는 분산 환경의 데이터와 함께 노드를 사용할 수 있으며 분산 환경에서는 IBM® SPSS® Analytic Server의 기능을 이용할 수 있습니다.
	Spectral Clustering© 알고리즘은 여러 고유 벡터를 사용하여 차원 수가 적은 공간에 데이터를 표시합니다. 그런 다음 새 공간에 k-평균 군집 알고리즘을 적용하여 데이터를 군집으로 분리합니다. 여러 필드가 있는 작은 레코드의 경우 상당히 빠르지만, 대형 데이터 세트의 경우 컴퓨팅 리소스가 많이 소비됩니다. SPSS Modeler의 스펙트럼 군집화 노드에는 스펙트럼 군집화 라이브러리의 핵심 기능과 일반적으로 사용되는 매개변수가 표시됩니다. 이 노드는 Python으로 구현됩니다.
	STB(Space-Time-Box)는 Geohash 공간 위치의 확장입니다. 보다 상세하게 설명하자면 STB는 공간 및 시간을 정기적으로 모양으로 표시하는 알파뉴메릭 문자열입니다.
	스트리밍 TCM 노드는 한 단계로 임시 원인 모델을 작성하고 스코어링합니다.
	CPLEX 최적화 노드에서는 OPL(Optimization Programming Language) 모델 파일을 통해 복잡한 수학(CPLEX) 기반 최적화를 수행할 수 있습니다. 이 기능은 더 이상 지원되지 않는 IBM Analytical Decision Management 제품에서 사용 가능합니다. 그러나 IBM Analytical Decision Management 없이도 SPSS Modeler에서 CPLEX 노드도 사용할 수 있습니다.

레코드 작업 팔레트 내의 많은 노드에 대해 사용자가 CLEM 표현식을 사용해야 합니다. CLEM에 익숙한 경우에는 필드에 표현식을 입력할 수 있습니다. 단, 모든 표현식 필드가 CLEM 표현식 작성기를 여는 단추를 제공하므로 이를 사용하여 해당 표현식을 자동으로 작성하는 것이 도움이 됩니다.

그림 1. 표현식 작성기 단추

(2) 선택 노드

선택 노드를 사용하여 BP(혈압) = "높음"과 같이 특정 조건을 기준으로 하여 데이터 스트림에서 레코드의 서브세트를 선택 또는 삭제할 수 있습니다.

모드. 조건을 충족하는 레코드를 데이터 스트림에 대해 포함할 것인지 제외할 것인지 지정합니다.

- **포함**. 선택 조건을 충족하는 레코드를 포함하려면 선택하십시오.
- **삭제**. 선택 조건을 충족하는 레코드를 제외하려면 선택하십시오.

조건. CLEM 표현식을 사용하여 지정하는 각 레코드를 검정하기 위해 사용할 선택 조건을 표시합니다. 창에 표현식을 입력하거나 창의 오른쪽에 있는 계산기(표현식 작성기) 단추를 클릭하여 표현식 작성기를 사용하십시오.

다음과 같이 조건을 기준으로 하여 레코드를 삭제하도록 선택합니다.

```
(var1='value1' and var2='value2')
```

기본적으로 선택 노드는 모든 선택 필드에 대해 널값을 가진 레코드도 삭제합니다. 이런 현상을 피하려면 다음 조건을 원본에 추가하십시오.

```
and not(@NULL(var1) and @NULL(var2))
```

레코드의 비율을 선택하는 데도 선택 노드가 사용됩니다. 일반적으로 이 작업에 대해서는 다른 노드인 표본 노드를 사용합니다. 그러나 지정할 조건이 제공된 모수보다 복잡한 경우, 선택 노드를 사용하여 직접 조건을 작성할 수 있습니다. 예를 들어 다음과 같은 조건을 작성할 수 있습니다.

BP = "HIGH" and random(10) <= 4

그러면 높은 혈압을 나타내는 40% 정도의 레코드가 선택되고 해당 레코드가 추가 분석을 위해 다운스트림에 전달됩니다.

(3) 표본 노드

표본 노드를 사용하여 분석할 레코드 서브세트를 선택하거나 삭제할 레코드 비율을 지정할 수 있습니다. 층화, 수평배열, 비임의(구조화) 표본을 포함하여 다양한 표본 유형이 지원됩니다. 표본추출은 다음과 같은 여러 가지 이유로 사용될 수 있습니다.

- 데이터 서브세트에서 모형을 추정하여 성능을 개선하기 위해서 사용됩니다. 표본에서 추정된 모형은 종종 전체 데이터 세트에서 파생된 모형만큼 정확하며 이전에는 불가능했던 다양한 방법을 개선된 성능을 사용하여 시도함으로써 더 정확한 결과를 얻을 수도 있습니다.

- 분석할 관련 레코드 또는 트랜잭션 그룹을 선택하기 위해서 사용됩니다. 예를 들어, 온라인 장바구니에서 모든 항목을 선택하거나 특정 인접 항목에서 모든 특성을 선택하는 것이 있습니다.
- 품질 보장, 사기 방지 또는 보안 등의 목적으로 임의 검사를 수행하기 위해 단위 또는 케이스를 식별하기 위해서 사용됩니다.

참고: 검증 목적으로 데이터를 학습 및 검정 표본으로 파티션만 하면 되는 경우에는 파티션 노드를 대신 사용하십시오. 자세한 정보는 파티션 노드 주제를 참조하십시오.

표본 유형

군집 표본. 개별 단위가 아니라 표본 그룹 또는 군집입니다. 예를 들어, 한 학생당 하나의 레코드가 있는 데이터 파일이 있다고 가정합니다. 학교 기준으로 군집화하고 표본 크기가 50%인 경우, 50%의 학교가 선택되고 선택된 각 학교의 모든 학생이 선택됩니다. 선택되지 않은 학교의 학생은 유효하지 않습니다. 평균적으로 약 50%의 학생이 선택될 것으로 예상하나 학교 크기가 다양하므로 퍼센트가 정확하지 않을 수 있습니다. 이와 유사하게, 장바구니 항목을 트랜잭션 ID 기준으로 군집화하여 선택된 트랜잭션의 모든 항목이 유지되는지 확인할 수 있습니다. 마을 기준으로 군집화되는 예를 보려면 *complexsample_property.str* 샘플 스트림을 참조하십시오.

층화 표본. 모그룹 밀도 또는 계층의 겹치지 않는 하위 그룹 내에서 표본을 독립적으로 선택합니다. 예를 들어, 여성 및 남성이 동일한 비율로 선택되도록 하거나 도시 인구의 모든 지역 또는 사회 경제적 그룹이 표시되도록 할 수 있습니다. 또한 각 계층별로 다른 표본 크기를 지정할 수 있습니다. 예를 들어, 한 그룹이 원 데이터에서 적게 표시된다고 생각할 수 있습니다. 지역 기준으로 특성을 층화하는 예를 보려면 *complexsample_property.str* 샘플 스트림을 참조하십시오.

계통 또는 n중1 표본추출. 임의선택이 어려운 경우 계통적(고정된 간격) 또는 순차적으로 단위를 표본추출할 수 있습니다.
표본추출 가중치. 표본추출 가중치는 복합 표본을 그릴 때 자동으로 계산되며 각 표본 단위가 원 데이터에서 나타내는 "빈도"와 거의 일치합니다. 따라서 표본에 대한 가중치 합계가 원 데이터의 크기를 추정해야 합니다.

표본추출 프레임

표본추출 프레임은 표본 또는 연구에 포함되는 케이스의 잠재적 소스를 정의합니다. 어떤 경우에는 모집단의 각각의 모든 멤버를 식별하고 그 각각을 표본에 포함시킬 수 있습니다. 예를 들어, 생산 공정에서 만들어내는 항목을 표본추출하는 경우입니다. 모든 가능한 케이스에 액세스할 수 없는 경우가 더 많습니다. 예를 들어, 선거가 시행될 때까지는 누가 선거에서 투표할 것인지 확인할 수 없습니다. 이 경우, 등록한 사람 중 일부가 투표하지 않으며 사용자가 등록을 확인하는 시점에 나열되지 않았음에도 투표하는 사람이 있더라도 선거 등록을 표본추출 프레임으로 사용할 수 있습니다. 표본추출 프레임 내에 없는 사람은 표본추출에 포함될 가능성이 없습니다. 표본추출 프레임이 평가하려고 시도하는 모집단의 성격에 충분히 근접한지 여부는 실생활의 각 케이스에서 반드시 다루어야 하는 안건입니다.

① **표본 노드 옵션**

요구 사항에 따라 **단순** 또는 **복합** 방법을 선택할 수 있습니다.

단순 표본추출 옵션

단순 방법을 사용하면 레코드를 임의의 비율로 선택하거나 연속된 레코드를 선택하거나 모든 *n 번째* 레코드를 선택할 수 있습니다.

모드 다음 방법에 대해 레코드를 전달(포함)하거나 삭제(제외)할 수 있습니다.

- **표본 포함** 선택된 레코드를 데이터 스트림에 포함시키고 모든 기타를 삭제합니다. 예를 들어, 모드를 **표본 포함**으로 설정하고 **n중1** 옵션을 5로 설정하면 매 다섯 번째 레코드가 포함되고 대략 원래 크기의 1/5인 데이터 세트가 생성됩니다. 이 모드가 데이터 표본추출의 기본 모드이며 복합 방법을 사용할 때의 유일한 모드입니다.
- **표본 삭제** 선택된 레코드를 제외하고 모든 기타를 포함합니다. 예를 들어, 모드를 **표본 삭제**로 설정하고 **n중1** 옵션을 5로 설정하면 매 다섯 번째 레코드가 삭제됩니다. 이 모드는 단순 방법에서만 사용할 수 있습니다.

표본 다음 옵션 중에서 표본추출 방법을 선택합니다.

- **처음** 연속된 데이터 표본추출을 사용하도록 선택합니다. 예를 들어, 최대 표본 크기가 10000으로 설정되면 처음 10,000개의 레코드가 선택됩니다.
- **n 중 1** 매 *n*번째 레코드를 전달하거나 삭제하여 데이터를 표본추출하도록 선택합니다. 예를 들어, *n*이 5로 설정되면 매 다섯 번째 레코드가 선택됩니다.
- **임의 %** 데이터의 임의 퍼센트를 표본 추출하도록 선택합니다. 예를 들어, 퍼센트를 20으로 설정하면 선택한 모드에 따라 데이터의 20%가 데이터 스트림에 전달되거나 삭제됩니다. 표본 추출 퍼센트를 지정하려면 필드를 사용하십시오. 또한 **난수 시드 설정** 제어를 사용하여 시드 값을 지정할 수 있습니다.
  블록 수준 표본추출(In-Database 전용)을 사용하십시오. 이 옵션은 사용자가 Oracle 또는 IBM Db2 데이터베이스에서 In-Database 마이닝을 수행할 때 임의 퍼센트 표본추출을 선택한 경우에만 사용 가능합니다. 이러한 상황에서는 블록 수준 표본추출이 더 효율적일 수 있습니다.

> ⓘ **참고**: 동일한 임의 표본 설정을 실행할 때마다 정확히 동일한 수의 행이 리턴되지는 않습니다. 그 이유는 각 입력 레코드의 확률이 표본에 포함되는 N/100이며(여기서, N은 사용자가 노드에서 지정한 **임의 %**임) 확률은 독립변수입니다. 따라서 결과가 정확히 N%가 아닙니다.

최대 표본 크기 표본에 포함될 최대 레코드 수를 지정합니다. 이 옵션은 중복이므로 **First** 및 **Include**가 선택될 때는 사용하지 않습니다. 또한 **임의 %** 옵션과 함께 사용되는 경우에는 이 설정이 특정 레코드가 선택되는 것을 방지할 수 있습니다. 예를 들어, 데이터 세트에 천만 개의 레코드가 있으며 최대 표본 크기가 삼백만 레코드인 50%의 레코드를 선택한 경우, 첫 번째 육백만 개의 레코드의 50%가 선택되며 나머지 사백만 개의 레코드는 선택될 기회가 없습니다. 이 제한을 방지하기 위해 **복합** 표본추출 방법을 선택하고 군집 또는 층화 변수를 지정하지 말고 삼백만 개의 레코드의 임의 표본을 요청하십시오.

복합 표본추출 옵션

복합 표본 옵션을 사용하면 기타 옵션과 함께 군집, 층화 및 가중치 표본 등을 사용하여 표본을 더 미세하게 제어할 수 있습니다.

군집 및 층화 필요에 따라 군집, 층화 및 입력 가중 필드를 지정할 수 있습니다. 자세한 정보는 군집 및 층화 설정의 내용을 참조하십시오.

모델 유형.

- **임의** 각 계층 내에서 군집 또는 레코드를 임의로 선택합니다.
- **계통** 고정된 간격으로 레코드를 선택합니다. 이 옵션은 *n중1* 방법과 동일하게 작동하나 난수 시드에 따라 첫 번째 레코드의 위치가 변경된다는 점만 다릅니다. *n*의 값은 표본 크기 또는 비율을 기준으로 하여 자동으로 결정됩니다.

표본 단위 기본 표본 단위로 비율 또는 빈도를 선택할 수 있습니다.

표본 크기 여러 가지 방법으로 표본 크기를 지정할 수 있습니다.

- **고정됨** 개수 또는 비율로 표본의 전체 크기를 지정할 수 있습니다.
- **사용자 정의** 각 하위 그룹 또는 계층의 표본 크기를 지정할 수 있습니다. 이 옵션은 군집 및 층화 하위 대화 상자에서 층화 필드가 지정된 경우에만 사용 가능합니다.
- **변수** 각 하위 그룹 또는 계층의 표본 크기를 정의하는 필드를 사용자가 선택할 수 있도록 허용합니다. 이 필드는 특정 계층 내의 각 레코드에 대해 동일한 값을 가져야 합니다. 예를 들어, 표본이 지역별로 계층화된 경우, *county = Surrey* 내의 모든 레코드가 동일한 값을 가져야 합니다. 필드는 숫자여야 하며 해당 값이 선택된 표본 단위와 일치해야 합니다. 비율의 경우, 값이 0보다 크고 1보다 작아야 하며 개수의 경우 최소값이 1입니다.

계층 당 최소 표본 최소 레코드 수를 지정합니다. 군집 필드가 지정된 경우에는 최소 군집 수가 지정됩니다.

계층 당 최대 표본 레코드 또는 군집의 최대 수를 지정합니다. 군집 또는 층화 필드를 지정하지 않고 이 옵션을 선택하면 지정된 크기의 임의 또는 계통 표본이 선택됩니다.

난수 시드 설정. 임의 퍼센트를 기반으로 사용하여 표본추출 레코드 또는 파티셔닝을 수행하는 경우에 이 옵션을 사용하면 다른 세션에서 동일한 결과를 복제할 수 있습니다. 난수 생성기에서 사용하는 시작값을 지정하면 노드를 실행할 때마다 동일한 레코드를 지정하도록 보장할 수 있습니다. 원하는 시드 값을 입력하거나 **생성** 단추를 클릭하여 자동으로 임의 값을 생성하십시오. 이 옵션을 선택하지 않으면 노드를 실행할 때마다 다른 표본이 생성됩니다.

> **참고**: 데이터베이스에서 읽은 레코드에서 **난수 시드 설정** 옵션을 사용하는 경우 노드를 실행할 때마다 동일한 결과를 보장하려면 표본 추출 전에 정렬 노드가 필요할 수도 있습니다. 난수 시드는 레코드 순서에 의존하여 관계형 데이터베이스에서는 동일하게 보장되지 않기 때문입니다. 자세한 정보는 정렬 노드의 내용을 참조하십시오.

② 군집 및 층화 설정

군집 및 층화 대화 상자를 사용하면 복합 표본을 그릴 때 군집, 층화 및 가중 필드를 선택할 수 있습니다.

군집. 레코드를 군집화하는 데 사용할 범주형 필드를 지정합니다. 레코드는 일부 군집은 포함되고 다른 일부는 제외된 형태로 소속군집을 기준으로 하여 표본화됩니다. 단, 지정된 군집의 어느 한 레코드가 포함되면 모든 레코드가 포함됩니다. 예를 들어, 장바구니와 관련하여 제품을 분석할 때 항목을 트랜잭션 ID 기준으로 군집화하여 선택된 트랜잭션의 모든 항목이 유지되는지 확인할 수 있습니다. 함께 판매된 항목에 대한 정보를 영구 삭제하는 표본추출 레코드 대신 트랜잭션에 대한 표본추출을 수행하여 선택된 트랜잭션에 대한 모든 레코드를 유지할 수 있습니다.

계층화 기준. 모그룹 밀도 또는 계층의 겹치지 않는 하위 그룹 내에서 표본이 독립적으로 선택되도록 레코드를 층화하는 데 사용되는 범주형 필드를 지정합니다. 성별 기준으로 층화된 50% 표본을 선택하면 남성에 대해 하나, 여성에 대해 하나의 두 개의 50% 표본이 선택됩니다. 예를 들어, 계층은 사회 경제적인 그룹, 작업 범주, 연령 그룹 또는 인종 그룹일 수 있으며 사용자가 관심 있는 부그룹에 대한 적절한 표본 크기를 선택할 수 있도록 해줍니다. 원래 데이터 세트에 남성보다 여성이 세 배 많은 경우, 이 비율은 각 그룹에서 별도로 표본추출할 때 유지됩니다. 다중 층화 필드도 지정할 수 있습니다. 예를 들어, 지역 내의 제품군 표본추출 또는 그 반대의 경우도 가능합니다.

참고: 결측값(널 또는 시스템 결측값, 비어 있는 문자열, 공백 또는 사용자 정의 결측값)이 있는 필드를 기준으로 하여 계층화하는 경우, 계층에 대한 고객 표본 크기를 지정할 수 없습니다. 결측값 또는 공백값이 있는 필드를 기준으로 하여 계층화하는 경우에 사용자 정의 표본 크기를 사용하려면 해당 값을 채워야 합니다.

입력 가중치 사용. 표본추출 전에 레코드를 가중하는 데 사용되는 필드를 지정합니다. 예를 들어, 가중 필드에 1에서 5 범위의 값이 있는 경우, 5로 가중된 레코드는 5배 많이 선택되는 경향이 있습니다. 노드에 의해 생성된 최종 출력 가중값이 이 필드의 값을 덮어씁니다. 다음 단락을 참조하십시오.

새 출력 가중치. 입력 가중 필드가 지정되지 않은 경우에 최종 가중치가 작성되는 필드의 이름을 지정합니다. (입력 가중 필드가 지정되면 위에서 설명한 대로 그 값이 최종 가중값에 의해 교체되고 별도의 출력 가중 필드가 작성되지 않습니다.) 출력 가중값은 원 데이터 내의 표본추출된 각 레코드가 나타내는 레코드 수를 표시합니다. 가중값의 합계는 표본 크기에 대한 추정값을 제공합니다. 예를 들어, 10%의 임의 표본이 사용되면 출력 가중치가 모든 레코드에 대해 10이 됩니다. 즉, 표본추출된 각 레코드가 원 데이터에서의 약 10개의 레코드를 나타냅니다. 층화 또는 가중된 표본에서 출력 가중값은 각 계층에 대한 표본 비율에 따라 다릅니다.

설명

- 수평배열 표본추출은 표본추출할 모그룹 분포의 전체 목록을 가져올 수 없으나 특정 그룹 또는 군집에 대한 전체 목록은 가져올 수 있는 경우에 유용합니다. 임의 표본이 접촉하기에 비실용적인 검정 개체를 생성하는 경우에도 유용합니다. 예를 들어, 국가 내의 모든 지역에 흩어져 있는 농부를 선택하기 보다 한 지역 내의 모든 농부를 방문하는 것이 더 쉬울 것입니다.
- 각 계층 내에서 독립적으로 군집을 표본추출하기 위해 군집 및 층화 필드를 둘 다 지정할 수 있습니다. 예를 들어, 지역별로 계층화된 특성 값을 표본추출하고 각 지역 내의 마을별로 군집화할 수 있습니다. 그러면 마을의 독립적 표본이 각 지역 내에 그려집니다. 일부 마을은 포함될 것이며 기타 마을은 포함되지 않을 것이나 포함되는 각 마을에 대해 마을 내의 모든 특성이 포함될 것입니다.
- 각 군집 내에서 임의 표본 단위를 선택하기 위해 두 표본 노드를 함께 연결할 수 있습니다. 예를 들어, 먼저 위에서 설명한 대로 지역 기준으로 계층화된 마을을 표본추출할 수 있습니다. 그런 다음 두 번째 표본 노드를 첨부하고 층화 필드로 마을을 선택하여 각 마을의 레코드의 비율을 표본추출할 수 있습니다.
- 군집을 고유하게 식별하기 위해 필드의 조합이 필요한 경우, 파생 노드를 사용하여 새 필드가 생성될 수 있습니다. 예를 들어, 여러 가게가 트랜잭션에 대해 동일한 번호 매기기 시스템을 사용하는 경우, 가게와 트랜잭션 ID를 연결하는 새 필드를 파생시킬 수 있습니다.

③ 계층에 대한 표본 크기

층화 표본을 그릴 때 기본 옵션은 각 계층의 레코드 또는 군집에서 동일한 비율로 표본을 추출하는 것입니다. 예를 들어, 한 그룹이 다른 그룹보다 그 수가 세 배 많은 경우, 일반적으로 표본에서도 동일한 비율을 유지하려고 할 것입니다. 그러나 이런 케이스가 아니라면 각 계층에 대해 별도로 표본 크기를 지정할 수 있습니다.

계층에 대한 표본 크기 대화 상자에는 층화 필드의 각 값이 나열되어 있어서 사용자가 해당 계층의 기본값을 대체할 수 있습니다. 다중 층화 필드를 선택하면 가능한 모든 값 조합이 나열되어 사용자가 각 도시 내의 각 인종 그룹에 대한 크기 또는 각 지역 내의 각 마을에 대한 크기 등을 지정할 수 있습니다. 크기는 표본 노드의 현재 설정에 의해 결정되는 비율 또는 빈도로 지정됩니다.

계층에 대한 표본 크기를 지정하려면 다음을 수행하십시오.

1. 표본 노드에서 **복합**을 선택하고 하나 이상의 층화 필드를 선택하십시오. 자세한 정보는 군집 및 층화 설정의 내용을 참조하십시오.
2. **사용자 정의**를 선택하고 **크기 지정**을 선택하십시오.
3. 계층에 대한 표본 크기 대화 상자의 왼쪽 하단에 있는 **값 읽기** 단추를 클릭하여 표시를 채우십시오. 필요한 경우 업스트림 소스 또는 유형 노드에서 값을 인스턴스화할 수 있습니다. 자세한 정보는 인스턴스화란? 주제를 참조하십시오.
4. 임의의 행을 클릭하여 해당 계층에 대한 기본 크기를 대체하십시오.

표본 크기에 대한 참고

다양한 계층에 다양한 분산이 있는 경우, 예를 들어, 표준 편차에 비례하는 표본 크기를 작성하는 경우에 사용자 정의 표본 크기가 유용합니다. (계층 내의 케이스가 더 다양한 경우에는 더 많은 표본을 추출하여 대표 표본을 얻어야 합니다.) 계층이 작은 경우에는 높은 표본 비율을 사용하여 관측값의 최소수가 포함되도록 할 수 있습니다.

참고: 결측값(널 또는 시스템 결측값, 비어 있는 문자열, 공백 또는 사용자 정의 결측값)이 있는 필드를 기준으로 하여 계층화하는 경우, 계층에 대한 고객 표본 크기를 지정할 수 없습니다. 결측값 또는 공백값이 있는 필드를 기준으로 하여 계층화하는 경우에 사용자 정의 표본 크기를 사용하려면 해당 값을 채워야 합니다.

(4) 균형 노드

균형 노드를 사용하면 지정된 테스트 기준을 준수하도록 데이터 세트에서 불균형을 정정할 수 있습니다. 예를 들어, 데이터 세트에 *low* 또는 *high*의 두 값만 있으며 케이스의 10%만 *high*일 때 케이스의 90%가 *low*라고 가정하십시오. 여러 모델링 기술은 거의 드물지만 *low* 결과만 학습하고 *high* 결과는 무시하는 경향이 있으므로 여러 모델링 기술을 사용할 때 이러한 편향 데이터에 문제가 있습니다. 데이터가 대략적으로 동일한 수의 *low* 및 *high* 결과와 올바른 균형을 유지하는 경우 모델은 두 그룹을 구별하는 패턴을 찾을 가능성이 높습니다. 이 경우 균형 노드는 *low* 결과의 케이스를 줄이는 균형 지시문을 작성하는 데 유용합니다.

사용자가 지정하는 조건에 기반하여 레코드를 복제한 후 버려서 균형이 수행됩니다. 어떤 조건도 없는 레코드는 항상 전달됩니다. 레코드를 복제하거나 버려서 이 프로세스가 작동되므로 원래 데이터 시퀀스는 다운스트림 작업에서 유실됩니다. 데이터 스트림에 균형 노드를 추가하기 전에 시퀀스 관련 값을 파생시켜야 합니다.

참고: 분포 차트 및 히스토그램에서 균형 노드가 자동으로 생성될 수 있습니다. 예를 들어, 분포 도표에 표시되어 있는 바와 같이 범주형 필드의 모든 범주에 동일한 비율을 표시하도록 데이터의 균형을 유지할 수 있습니다.

예. 이전 마케팅 캠페인에 긍정적으로 반응한 최근 고객을 식별하기 위해 RFM 스트림을 작성하는 경우 판매 회사의 마케팅 부서는 균형 노드를 사용하여 데이터에 있는 true 반응과 false 반응의 차이에 대한 균형을 유지합니다.

① 균형 노드의 옵션 설정

레코드 균형 지시문. 현재 균형 지시문을 나열합니다. 각 지시문에는 "조건이 true인 경우 계수를 지정하여 레코드 비율을 증가시키도록" 소프트웨어에 지시하는 조건과 계수가 포함됩니다. 1.0 미만의 계수는 표시된 레코드의 비율이 감소됨을 의미합니다. 예를 들어, Y 약물이 치료 약물인 레코드 수를 줄이려면 계수 0.7과 조건 Drug = "drugY"를 사용하여 균형 지시문을 작성할 수 있습니다. 이 지시문은 Y 약물이 치료 약물인 레코드 수가 모든 다운스트림 작업에 대해 70%로 감소됨을 의미합니다.

참고: 감소를 위한 균형 계수가 4자리의 소수 자리로 지정될 수 있습니다. 계수가 0.0001 미만으로 설정된 경우 결과가 올바르게 계산되지 않으므로 오류가 발생합니다.

- 텍스트 필드 오른쪽의 단추를 클릭하여 **조건 작성**을 수행하십시오. 이 경우 새 조건을 입력할 비어 있는 행이 삽입됩니다. 조건으로 CLEM 표현식을 작성하려면 표현식 작성기 단추를 클릭하십시오.
- 빨간색 삭제 단추를 사용하여 **지시문 삭제**를 수행하십시오.
- 위로 및 아래로 화살표 단추를 사용하여 **지시문 정렬**을 수행하십시오.

균형 학습 데이터만. 파티션 필드가 스트림에 있는 경우 이 옵션은 학습 파티션의 데이터 균형만 유지합니다. 특히 조정된 성향 스코어를 생성하는 경우에 유용할 수 있으며 불균형 테스트 또는 검증 파티션이 필요합니다. 파티션 필드가 스트림에 없는 경우(또는 여러 파티션 필드가 지정된 경우) 이 옵션은 무시되고 모든 데이터의 균형이 유지됩니다.

(5) 통합 노드

통합은 데이터 세트의 크기를 줄이기 위해 자주 사용되는 데이터 준비 작업입니다. 통합을 계속하기 전에 오래 걸려도 데이터를 정리해야 하며 특히 결측값에 집중해야 합니다. 통합했으면 결측값과 관련된 잠재적으로 유용한 정보가 유실될 수 있습니다.

통합 노드를 사용하여 입력 레코드 시퀀스를 요약, 통합 출력 레코드로 대체할 수 있습니다. 예를 들어, 다음 표와 같은 입력 판매 레코드 세트가 있을 수 있습니다.

표 1. 판매 레코드 입력 예

나이	성별	지역	지점	판매
23	M	S	8	4
45	M	S	16	4
37	M	S	8	5
30	M	S	5	7
44	M	N	4	9
25	M	N	2	11
29	F	S	16	6
41	F	N	4	8
23	F	N	6	2
45	F	N	4	5
33	F	N	6	10

성별 및 *지역*을 키 필드로 사용하여 이러한 레코드를 통합할 수 있습니다. 그런 다음 *나이*를 모드 **평균**과 통합하고 *판매*를 모드 **합계**와 통합하도록 선택하십시오. 통합 노드 대화 상자에서 **필드에 레코드 수 포함**을 선택하면 통합 출력은 다음 표와 같습니다.

표 2. 통합 레코드 예

나이(평균)	성별	지역	판매(합계)	레코드수
35.5	F	N	25	4
29	F	S	6	1
34.5	M	N	20	2
33.75	M	S	20	4

이를 통해 예를 들어, 북부 지역에 있는 네 명의 여성 판매 담당자의 평균 나이가 35.5이며 총 판매 합계가 25 단위임을 알 수 있습니다.

참고: 통합 모드가 지정되지 않은 경우 *지점*과 같은 필드를 자동으로 버립니다.

① 통합 노드의 옵션 설정

통합 노드에서 다음을 지정하십시오.

- 통합의 범주로 사용할 하나 이상의 키 필드
- 통합 값을 계산할 하나 이상의 통합 필드
- 각 통합 필드에 대해 출력할 하나 이상의 통합 모드(통합 유형)

새로 추가된 필드에 사용할 기본 통합 모드를 지정하고 통합을 분류하기 위한 표현식(공식과 비슷함)을 사용할 수 있습니다.

성능 향상을 위해 병렬 처리를 사용하면 통합 작업에 유용할 수 있다는 점을 참고하십시오.

키 필드. 통합의 범주로 사용할 수 있는 필드를 나열합니다. 연속형(숫자) 및 범주형 필드를 모두 키로 사용할 수 있습니다. 두 개 이상의 키 필드를 선택하는 경우 값을 결합하여 레코드를 통합하기 위한 키 값을 생성합니다. 각 고유 키 필드마다 하나의 통합 레코드가 생성됩니다. 예를 들어 Sex 및 Region이 키 필드인 경우, N과 S를 갖는 M과 F의 각각의 고유한 조합(4가지 고유 조합)이 통합 레코드를 갖습니다. 키 필드를 추가하려면 창 오른쪽의 필드 선택기 단추를 사용하십시오.

대화 상자의 나머지는 두 개의 기본 영역인 **기본 통합** 및 **통합 표현식**으로 구분됩니다.

기본 통합

통합 필드. 값이 통합되는 필드와 선택된 통합 모드를 나열합니다. 이 목록에 필드를 추가하려면 오른쪽의 필드 선택기 단추를 사용하십시오. 다음 통합 모드를 사용할 수 있습니다.

> ⓘ **참고**: 일부 모드는 숫자가 아닌 필드에 적용할 수 없습니다(예: 날짜/시간 필드의 **합계**). 선택된 통합 필드와 함께 사용할 수 없는 모드가 사용 안함으로 설정되어 있습니다.

- **합계** 각 키 필드 조합의 합계 값을 리턴하려면 이를 선택하십시오. 합계는 결측값이 있는 모든 케이스에서 값의 총계입니다.
- **평균** 각 키 필드 조합의 평균 값을 리턴하려면 이를 선택하십시오. 평균은 중심 경향의 측도이며 산술 평균(합계를 케이스 수로 나눈 값)입니다.
- **최소** 각 키 필드 조합의 최소값을 리턴하려면 이를 선택하십시오.
- **최대** 각 키 필드 조합의 최대값을 리턴하려면 이를 선택하십시오.
- **SDev**. 각 키 필드 조합의 표준 편차를 리턴하려면 이를 선택하십시오. 표준 편차는 평균 주변의 산포도이며 분산 측정의 제곱근입니다.
- **중앙값** 각 키 필드 조합의 중앙값을 리턴하려면 이를 선택하십시오. 중앙값은 벗어난 값의 영향을 받지 않는 중심 경향 측도이며 평균과 달리 상한 극단값 또는 하한 극단값의 영향을 받을 수 있습니다. 50번째 백분위수 또는 두 번째 사분위수라고도 합니다.

- **개수** 각 키 필드 조합의 널이 아닌 값 수를 리턴하려면 이를 선택하십시오.
- **분산**. 각 키 필드 조합의 분산 값을 리턴하려면 이를 선택하십시오. 분산은 평균 주변의 산포도이며 평균을 케이스 수에서 1을 뺀 값으로 나눈 값의 제곱 편차 합계와 같습니다.
- **첫 번째 사분위수**. 각 키 필드 조합의 첫 번째 사분위수(25번째 백분위수) 값을 리턴하려면 이를 선택하십시오.
- **세 번째 사분위수**. 각 키 필드 조합의 세 번째 사분위수(75번째 백분위수) 값을 리턴하려면 이를 선택하십시오.

> ⓘ **참고**: 통합 노드가 있는 스트림을 실행하는 경우 SQL을 Oracle 데이터베이스에 푸시백할 때 첫 번째 및 세 번째 사분위수에 대해 리턴되는 값이 원시 모드에서 리턴되는 값과 다를 수 있습니다.

기본 모드. 새로 추가된 필드에 대해 사용할 기본 통합 모드를 지정하십시오. 동일한 통합을 자주 사용하는 경우 여기에서 하나 이상의 모드를 선택하고 오른쪽의 모두에 적용 단추를 사용하여 위에 나열된 모든 필드에 선택된 모드를 적용하십시오.

새 필드 이름 확장자. 통합 필드를 복제하기 위해 접미부 또는 접두부(예: **1** 또는 **new**)를 추가하려면 이를 선택하십시오. 예를 들어, 접미부 옵션을 선택하고 확장자로 **1**을 지정한 경우 **연령** 필드에서 최소값 통합의 결과로 Age_Min_1이라는 필드 이름이 생성됩니다. _Min 또는 Max_와 같은 통합 확장자는 새 필드에 자동으로 추가되어 수행된 통합의 유형을 표시합니다. 선호하는 확장자 스타일을 표시하려면 **접미어** 또는 **접두어**를 선택하십시오.

필드에 레코드 개수 포함 기본적으로 Record_Count라는 각 출력 레코드에 추가 필드를 포함하려면 이를 선택하십시오. 이 필드는 각 통합 레코드를 형성하기 위해 통합된 입력 레코드 수를 표시합니다. 편집 필드에 입력하여 이 필드의 사용자 정의 이름을 작성하십시오.

> ⓘ **참고**: 통합이 계산될 때 시스템 널값이 제외되지만 레코드 수에는 포함됩니다. 반면, 공백 값은 통합 및 레코드 수에 모두 포함됩니다. 공백 값을 제외하려면 채움 노드를 사용하여 공백을 널값으로 대체할 수 있습니다. 선택 노드를 사용하여 공백을 제거할 수도 있습니다.

통합 표현식

표현식은 값, 필드 이름, 연산자, 함수에서 작성되는 공식과 비슷합니다. 한 번에 하나의 레코드에서 작동되는 함수와 달리 통합 표현식은 레코드의 그룹, 세트 또는 콜렉션에서 작동됩니다.

> ⓘ **참고**: 스트림에 데이터베이스 연결이 포함된 경우에만 통합 표현식을 작성할 수 있습니다(데이터베이스 소스 노드 사용).

새 표현식이 파생 필드로 작성됩니다. 표현식을 작성하려면 표현식 작성기에서 사용 가능한 *데이터베이스* 통합 함수를 사용하십시오.

추가 정보는 표현식 작성기의 내용을 참조하십시오.

통합 표현식이 키 필드에 따라 그룹화되어 있으므로 **키 필드**와 통합 표현식 간 연결이 있다는 점을 참고하십시오.

유효한 통합 표현식은 결과를 통합하기 위해 평가되는 표현식입니다. 유효한 통합 표현식에 대한 두 개의 예와 이를 제어하는 규칙은 다음과 같습니다.

- 스칼라 함수를 사용하면 여러 통합 함수를 함께 결합하여 단일 통합 결과를 생성할 수 있습니다. 예:

```
max(C01) - min(C01)
```

- 통합 함수는 여러 스칼라 함수의 결과에서 작동될 수 있습니다. 예:

```
sum (C01*C01)
```

② 최적화 설정 통합

최적화 탭에서 다음을 지정하십시오.

키가 인접 키임. 동일한 키 값을 갖는 모든 레코드가 입력에서 함께 그룹화됨(예를 들어, 입력이 키 필드에서 정렬됨)을 아는 경우 이 옵션을 선택하십시오. 그렇게 하면 성능이 개선될 수 있습니다.

중앙값 및 사분위수의 근사치 허용. Analytic Server에서 데이터를 처리할 때 주문 통계(중앙값, 첫 번째 사분위수, 세 번째 사분위수)는 현재 지원되지 않습니다. Analytic Server를 사용하는 경우 데이터를 구간화한 후 구간에 대한 분포에 기반하여 통계의 추정값을 계산하여 계산되는 통계 대신 이러한 통계에 대한 근사 값을 사용하려면 이 선택란을 선택하십시오. 기본적으로 이 옵션은 선택되어 있지 않습니다.

구간 수. **중앙값 및 사분위수의 근사치 허용** 선택란을 선택하는 경우에만 사용할 수 있습니다. 통계를 추정할 때 사용할 구간 수를 선택하십시오. 구간 수는 **최대 오차 %**에 영향을 줍니다. 기본적으로 구간 수는 1000이며 범위의 0.1 퍼센트의 최대 오차에 해당합니다.

(6) RFM 통합 노드

RFM(최근, 빈도, 구매총액) 통합 노드를 사용하면 고객의 히스토리 트랜잭션 데이터를 사용하고 모든 사용하지 않은 데이터를 제거하고 고유 고객 ID를 키로 사용하여 마지막으로 다룬 시기(최근), 작성한 트랜잭션 수(빈도), 해당 트랜잭션의 총 값(구매총액)을 나열하는 단일 행으로 모든 나머지 트랜잭션 데이터를 결합할 수 있습니다.

통합을 진행하기 전에 특히 결측값에 집중하여 데이터를 정리하는 시간이 필요합니다.

일단 RFM 통합 노드를 사용하여 데이터를 식별하고 변환한 후에는 RFM 분석 노드를 사용하여 추가 분석을 수행할 수 있습니다. 자세한 정보는 RFM 분석 노드 주제를 참조하십시오.

일단 RFM 통합 노드를 통해 데이터 파일이 실행된 후에는 어떠한 목표 값도 갖지 않습니다. 따라서 이를 C5.0 또는 CHAID 등의 모델링 노드를 사용하는 추가 예측 분석에 대한 입력으로 사용하려면 다른 고객 데이터와 합쳐야 합니다. 예를 들어, 고객 ID 일치 등의 방법이 있습니다. 자세한 정보는 합치기 노드 주제를 참조하십시오.

IBM® SPSS® Modeler의 RFM 통합 및 RFM 분석 노드는 독립적 구간화를 사용하기 위해 설정됩니다. 즉, 해당 값 또는 다른 두 측도에 관계없이 RFM(Recency, Frequency, Monetary) 값의 각 측도에 대한 데이터를 순위화하고 구간화합니다.

① RFM 통합 노드에 대한 옵션 설정

RFM 통합 노드의 설정 탭은 다음 필드를 포함합니다.

최근성 계산 기준 날짜 트랜잭션의 최근성이 계산되는 날짜를 지정하십시오. 사용자가 입력하는 고정 날짜 또는 시스템에서 설정된 **오늘 날짜**일 수 있습니다. **오늘 날짜**는 기본적으로 입력되며 노드가 실행될 때 자동으로 업데이트됩니다.

> **참고**: **고정 날짜**는 로케일에 따라 다를 수 있습니다. 예를 들어, 스트림에 2007-8-10 값이 Fri Aug 10 00:00:00 CST 2007로 저장된 경우 'UTC+8' 시간대의 시간과 날짜입니다. 그러나 'UTC-8' 시간대의 Thu Aug 9 12:00:00 EDT 2007로 표시됩니다.

연속 ID ID가 동일한 모든 레코드를 데이터 스트림에서 함께 표시하도록 데이터를 사전 정렬한 경우 이 옵션을 선택하여 처리 속도를 높입니다. 데이터가 사전 정렬되지 않았거나 정렬 여부가 확실하지 않은 경우 이 옵션을 선택하지 않은 상태로 두면 노드가 데이터를 자동으로 정렬합니다.

ID 고객 및 고객의 트랜잭션을 식별하기 위해 사용할 필드를 선택하십시오. 선택할 수 있는 필드를 표시하려면 오른쪽의 필드 선택기 단추를 사용하십시오.

날짜 최근성을 계산하기 위해 사용할 날짜 필드를 선택하십시오. 선택할 수 있는 필드를 표시하려면 오른쪽의 필드 선택기 단추를 사용하십시오.

입력으로 사용하려면 적절한 형식의 날짜 또는 시간소인 저장 공간이 있는 필드가 필요합니다. 예를 들어, 값이 *Jan 2007, Feb 2007* 등인 문자열 필드가 있는 경우 이를 채움 노드 및 to_date() 함수를 사용하여 날짜 필드로 변환할 수 있습니다. 자세한 정보는 채움 노드를 사용한 저장 공간 변환의 내용을 참조하십시오.

값 고객 트랜잭션의 구매총액 값을 계산하는 데 사용할 필드를 선택하십시오. 선택할 수 있는 필드를 표시하려면 오른쪽의 필드 선택기 단추를 사용하십시오. *참고:* 숫자 값이어야 합니다.

새 필드 이름 확장 접미문자 또는 접두문자를 첨부할 것인지 선택하십시오. 예를 들어, 새로 생성된 최근, 빈도 및 구매총액 필드에 "12_month"를 첨부할 수 있습니다. 선호하는 확장자 유형을 표시하려면 **접미문자** 또는 **접두문자**를 선택하십시오. 예를 들어, 여러 시간 주기를 검사할 때 유용합니다.

아래 값의 레코드 삭제 필요한 경우, RFM 총계를 계산할 때 그 아래의 모든 트랜잭션 세부사항을 사용하지 않는 최소값을 지정할 수 있습니다. 값의 단위는 선택된 **Value** 필드와 관계가 있습니다.

최근 트랜잭션만 포함 큰 데이터베이스를 분석하는 경우, 최근 레코드만 사용되도록 지정할 수 있습니다. 특정 날짜 이후 또는 최근 주기 내에 기록된 데이터만 사용하도록 선택할 수 있습니다.

- **이후의 트랜잭션 날짜** 그 이후의 레코드가 분석에 포함될 트랜잭션 날짜를 지정하십시오.
- **지난 기간 동안의 트랜잭션** 그 이후의 레코드가 분석에 포함될 날짜에 **상대적인 최근 계산**으로부터의 기간(일, 주, 월 또는 년)의 유형과 수를 지정하십시오.

두 번째로 최신인 트랜잭션 날짜 저장 각 고객에 대한 두 번째 최근 트랜잭션의 날짜를 알려면 이 선택란을 선택하십시오. 또한 **세 번째 가장 최근 트랜잭션의 날짜** 선택란도 선택할 수 있습니다. 예를 들어, 상당한 정도의 기간 이전에는 많은 트랜잭션을 수행했으나 최근 트랜잭션은 한 번뿐인 고객을 식별하는 데 도움을 줍니다.

(7) 정렬 노드

정렬 노드를 사용하여 하나 이상의 필드의 값을 기반으로 레코드를 내림차순 또는 오름차순으로 정렬할 수 있습니다. 예를 들어, 가장 일반적인 데이터 값을 가진 레코드를 보고 선택하기 위해 정렬 노드를 사용하는 경우가 가장 빈번합니다. 일반적으로 먼저 통합 노드를 사용하여 데이터를 통합한 다음 정렬 노드를 사용하여 통합된 데이터를 레코드 개수의 내림차순으로 정렬합니다. 표에 이런 결과를 표시하면 데이터를 탐색하여 최고의 고객 열 명의 레코드를 선택하는 것 등을 포함하여 의사결정을 내릴 수 있습니다.

정렬 노드의 설정 탭은 다음 필드를 포함합니다.

정렬 기준. 정렬 키로 사용하기 위해 선택한 모든 필드가 표에 표시됩니다. 키 필드는 숫자 필드일 때 정렬이 가장 효과적입니다.

- 오른쪽의 필드 선택기 단추를 사용하여 이 목록에 **필드를 추가**하십시오.
- 표의 **정렬** 열에서 **오름차순** 또는 **내림차순** 화살표를 클릭하여 순서를 선택하십시오.

- 빨간색 삭제 단추를 사용하여 **필드를 삭제**하십시오.
- 위로 및 아래로 화살표 단추를 사용하여 **지시문을 정렬**하십시오.

기본 정렬 순서. 새 필드가 추가될 때 기본 정렬 순서로 사용할 **오름차순** 또는 **내림차순**을 선택하십시오.

> ⓘ **참고**: 모델 스트림 아래로 고유 노드가 있으면 정렬 노드가 적용되지 않습니다. 고유 노드에 대한 자세한 정보는 고유 노드의 내용을 참조하십시오.

① 정렬 최적화 설정

일부 키 필드를 기준으로 하여 이미 정렬된 데이터를 사용하여 작업하는 경우, 이미 정렬된 필드를 지정하여 시스템이 데이터의 나머지를 더 효율적으로 정렬하도록 설정할 수 있습니다. 예를 들어, *연령*(내림차순) 및 *약물*(오름차순)로 정렬하려고 하나 데이터가 이미 *연령*(내림차순)으로 정렬되었음을 알고 있습니다.

데이터가 사전 정렬됨. 데이터가 이미 하나 이상의 필드에 의해 정렬되는지 여부를 지정합니다.

기존 정렬 순서 지정. 이미 정렬된 필드를 지정하십시오. 필드 선택 대화 상자를 사용하여 목록에 필드를 추가하십시오. *순서* 열에서 각 필드가 오름차순 또는 내림차순으로 정렬되는지 여부를 지정하십시오. 여러 필드를 지정하는 경우에는 올바른 정렬 순서로 필드를 나열하는지 확인하십시오. 목록 오른쪽의 화살표를 사용하여 올바른 순서로 필드를 배열하십시오. 올바른 기존 정렬 순서를 지정하지 못하면 스트림을 실행할 때 오류가 표시되고 사용자가 지정한 정렬과 일치하지 않는 레코드 수가 표시됩니다.

참고: 병렬 처리를 사용할 경우 정렬 속도가 개선될 수 있습니다.

(8) 합치기 노드

병합 노드의 기능은 다중 입력 레코드를 가져와서 입력 필드 중 일부 또는 전부가 포함된 단일 출력 레코드를 작성하는 것입니다. 이는 내부 고객 데이터 및 구매한 인구 통계 데이터 등의 다양한 소스의 데이터를 병합하려고 할 때 유용한 조작입니다. 다음의 방법으로 데이터를 병합할 수 있습니다.

- **순서**별 병합은 가장 작은 데이터 소스가 고갈될 때까지 입력 순서로 모든 소스의 해당 레코드를 연결합니다. 이는 정렬 노드를 사용하여 데이터를 정렬한 이 옵션을 사용하는 경우 중요합니다.
- 한 데이터 소스의 레코드를 다른 데이터 소스의 레코드와 일치시키는 방법을 지정하기 위해 **키** 필드(예: *고객 ID*)를 사용하여 병합. 내부 결합, 전체 외부 결합, 부분 외부 결합, 안티 결합을 포함한 여러 유형의 결합을 사용할 수 있습니다. 자세한 정보는 결합의 유형 주제를 참조하십시오.

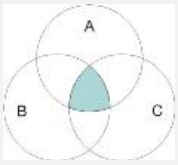
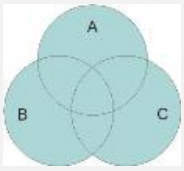
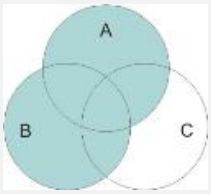
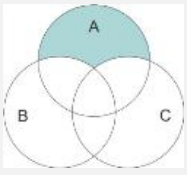
- **조건**별 병합은 병합을 수행하기 위해 충족할 조건을 지정할 수 있음을 의미합니다. 노드에서 작업 조건을 지정하거나 표현식 작성기를 사용하여 조건을 작성할 수 있습니다.
- **순위가 매겨진 조건**별 병합은 병합을 수행하기 위해 충족할 조건과 오름차순으로 정렬하는 순위화 표현식을 지정하는 왼쪽 외부 결합입니다. 지리 공간적 데이터를 병합하는 데 가장 자주 사용되며 노드에서 직접 조건을 지정하거나 표현식 작성기를 사용하여 조건을 작성할 수 있습니다.

① 결합의 유형

데이터 병합을 위해 키 필드를 사용하는 경우 제외할 레코드와 포함할 레코드에 대해 잠시 생각해 보는 것이 도움이 됩니다. 다양한 결합이 있으며 이러한 결합에 대해서는 아래에서 자세하게 다룹니다.

결합의 두 가지 기본적인 유형을 내부 결합과 외부 결합이라고 합니다. 이 방법은 *고객 ID* 등의 키 필드의 공통 값을 기반으로 관련 데이터 세트의 테이블을 병합하는 데 자주 사용됩니다. 내부 결합에서는 완전한 레코드만 포함하는 출력 데이터 세트와 정리 병합을 고려합니다. 외부 결합은 병합된 데이터의 완전한 데이터도 포함하지만 하나 이상의 입력 테이블의 고유 데이터를 포함할 수도 있게 합니다.

허용되는 결합의 유형에 대해서는 아래에 더 자세히 설명되어 있습니다.

	내부결합은 키 필드의 값이 모든 입력 테이블에 대해 공통인 레코드만 포함합니다. 즉, 일치하지 않는 레코드는 출력 데이터 세트에 포함되지 않습니다.
	전체외부결합은 입력 테이블의 모든 레코드(일치하는 레코드와 일치하지 않는 레코드 모두)를 포함합니다. 왼쪽 및 오른쪽 외부 결합은 부분 외부 결합이라고 하며 아래에 설명되어 있습니다.
	부분외부결합은 키 필드를 사용하는 일치된 모든 레코드와 지정된 테이블의 일치하지 않는 레코드를 포함합니다. (즉, 일부 테이블의 모든 레코드와 다른 테이블의 일치하는 레코드만). 병합 탭의 선택 단추를 사용하여 외부 결합에 포함할 테이블(예: 여기에 표시된 A 및 B)을 선택할 수 있습니다. 부분 결합은 두 개의 테이블만 병합 중인 경우 왼쪽 또는 오른쪽 외부 결합이라고도 합니다. IBM® SPSS® Modeler는 셋 이상의 테이블의 병합을 허용하므로 이를 부분 외부 결합이라고 합니다.
	안티결합은 첫 번째 입력 테이블(여기에 표시된 테이블 A)에 대해 일치하지 않는 레코드만 포함합니다. 이 유형의 결합은 내부 결합과 반대이며 출력 데이터 세트에 완전한 레코드를 포함하지 않습니다.

예를 들어, 한 데이터 세트에 농장에 대한 정보가 있고 다른 데이터 세트에 농장 관련 보험 클레임에 대한 정보가 있는 경우 병합 옵션을 사용하여 첫 번째 소스의 레코드를 두 번째 소스에 일치시킬 수 있습니다.

농장 표본에 있는 고객이 보험 클레임을 제기했는지 판별하려면 내부 결합 옵션을 사용하여 두 표본에서 모든 ID가 일치하는 위치를 표시하는 목록을 리턴하십시오.

그림 1. 내부 결합 병합의 표본 출력

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0...
2	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0...
3	id620	name620	north	880.000	74.000	6.000	426988.000	rapeseed	arable_de	118885.0...

전체 외부 결합 옵션을 사용하면 입력 테이블에서 일치하는 레코드와 일치하지 않는 레코드가 모두 리턴됩니다. 불완전한 값에는 시스템 결측값($null$)이 사용됩니다.

그림 2. 전체 외부 결합 병합의 표본 출력

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalu
1	id601	$null$	$null$	$null$	$null$	$null$	$null$	$null$	decomm...	74703.1(
2	id602	name602	north	1780.000	42.000	9.000	734118.000	maize	$null$	$nul
3	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0
4	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0
5	id606	$null$	$null$	$null$	$null$	$null$	$null$	$null$	arable_de	122135.0

부분 외부 결합은 키 필드를 사용하는 일치된 모든 레코드와 지정된 테이블의 일치하지 않는 레코드를 포함합니다. 테이블에는 ID 필드에서 일치된 모든 레코드와 첫 번째 데이터 세트에서 일치된 레코드가 표시됩니다.

그림 3. 부분 외부 결합 병합의 표본 출력

	id	claimtype	claimvalue	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	$null$	$null$	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id604	decomm...	281082.0...	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes
3	id605	decomm...	122006.0...	name605	north	1700.000	46.000	8.000	621148.000	wheat
4	id607	$null$	$null$	name607	southeast	1820.000	29.000	6.000	211605.000	maize
5	id608	$null$	$null$	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
6	id609	$null$	$null$	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
7	id615	$null$	$null$	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
8	id618	$null$	$null$	name618	southeast	1180.000	98.000	3.000	368646.000	maize

안티 결합 옵션을 사용하는 경우 테이블은 첫 번째 입력 테이블에 대해 일치하지 않는 레코드만 리턴합니다.

그림 4. 안티 결합 병합의 표본 출력

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id607	name607	southeast	1820.000	29.000	6.000	211605.000	maize
3	id608	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
4	id609	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
5	id615	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
6	id618	name618	southeast	1180.000	98.000	3.000	368646.000	maize
7	id619	name619	north	840.000	64.000	8.000	457552.000	potatoes

② 병합 방법 및 키 지정

병합 노드의 병합 탭에는 다음과 같은 필드가 포함되어 있습니다.

병합 방법 레코드 병합에 사용할 방법을 선택하십시오. **키** 또는 **조건**을 선택하면 대화 상자의 아래쪽 절반이 활성화됩니다.

- **순서** 각 입력의 *n*번째 레코드가 병합되어 n번째 출력 레코드를 생성하도록 순서별로 레코드를 병합합니다. 레코드에 일치하는 입력 레코드가 부족해지면 더 이상 출력 레코드가 생성되지 않습니다. 이는 작성되는 레코드 수는 가장 작은 데이터 세트의 레코드 수임을 의미합니다. 이 옵션을 사용하는 경우에는 정렬 노드를 사용하여 데이터를 정렬했다는 점이 중요합니다.
- **키** *트랜잭션 ID* 등의 키 필드를 사용하여 레코드를 키 필드의 동일한 값과 병합합니다. 이는 데이터베이스 "일치 결합"과 동등합니다. 키 값이 두 번 이상 발생하는 경우에는 가능한 모든 조합이 리턴됩니다. 예를 들어, 동일한 키 필드 값 *A*를 가진 레코드의 다른 필드에 다른 값 *B*, *C* 및 *D*가 포함되어 있는 경우 병합된 필드는 *B* 값을 가진 *A*, *C* 값을 가진 *A* 및 *D* 값을 가진 *A*라는 각각의 조합에 대해 별도의 레코드를 생성합니다.
  참고: 널값은 키별 병합 방법에서 동일한 것으로 간주되지 않으므로 결합되지 않습니다.

- **조건** 병합에 대한 조건을 지정하려면 이 옵션을 사용하십시오. 추가 정보는 병합을 위한 조건 지정의 내용을 참조하십시오.
- **순위화된 조건** 1차 및 모든 2차 데이터 세트에 있는 각각의 행 쌍을 병합해야 하는지 여부를 지정하려면 이 옵션을 사용하십시오. 순위화 표현식을 사용하여 다중 일치를 오름차순으로 정렬하십시오. 추가 정보는 병합을 위한 순위화된 조건 지정의 내용을 참조하십시오.

가능한 키 모든 입력 데이터 소스에서 정확하게 일치하는 필드 이름을 가진 필드만 나열합니다. 이 목록에서 필드를 선택한 후 화살표 단추를 사용하여 레코드 병합에 사용되는 키 필드로 추가하십시오. 둘 이상의 키 필드를 사용할 수 있습니다. 필터 노드 또는 소스 노드의 필터 탭을 사용하여 비일치 입력 필드의 이름을 바꿀 수 있습니다.

병합을 위한 키 키 필드의 값을 기반으로 모든 입력 데이터 소스의 레코드를 병합하는 데 사용되는 모든 필드를 나열합니다. 목록에서 키를 제거하려면 하나의 키를 선택한 후 화살표 단추를 사용하여 가능한 키 목록에 리턴하십시오. 둘 이상의 키 필드가 선택되면 아래의 옵션을 사용할 수 있습니다.

중복 키 필드 결합 위에서 둘 이상의 키 필드가 선택되는 경우 이 옵션을 사용하면 해당 이름의 출력 필드가 하나만 있게 됩니다. 이 옵션은 IBM® SPSS® Modeler의 이전 버전에서 스트림을 가져온 경우를 제외하고 기본적으로 사용으로 설정됩니다. 이 옵션이 사용 안함으로 설정되는 경우에는 합치기 노드 대화 상자의 필터 탭을 사용하여 중복 키 필드의 이름을 바꾸거나 중복 키 필드를 제외해야 합니다.

일치 레코드만 포함(내부 결합) 완전한 레코드만 병합하려면 선택하십시오.

일치 및 비일치 레코드 포함(전체 외부 결합) "전체 외부 결합"을 수행하려면 선택하십시오. 이는 키 필드의 값이 모든 입력 테이블에 없는 경우 불완전한 레코드가 계속 보존됨을 의미합니다. 정의되지 않은 값($null$)이 키 필드에 추가되고 출력 레코드에 포함됩니다.

일치 및 선택된 비일치 레코드 포함(부분 외부 결합) 하위 대화 상자에서 선택하는 테이블의 "부분 외부 결합"을 수행하려면 선택하십시오. 선택을 클릭하여 병합 시 불완전한 레코드를 보존할 테이블을 지정하십시오.

다른 레코와 일치하지 않는 첫 번째 데이터 세트의 레코드 포함(안티 결합) 첫 번째 데이터 세트의 일치하지 않는 레코드만 다운스트림으로 전달되는 "안티 결합"의 한 유형을 수행하려면 선택하십시오. 입력 탭의 화살표를 사용하여 입력 데이터 세트의 순서를 지정할 수 있습니다. 이 유형의 결합에서는 완전한 레코드를 출력 데이터 세트에 포함하지 않습니다. 추가 정보는 결합의 유형의 내용을 참조하십시오.

③ 부분 결합에 대한 데이터 선택

부분 외부 결합의 경우 불완전한 레코드를 보유할 테이블을 선택해야 합니다. 예를 들어, 주택 담보 대출 테이블의 일치된 레코드만 보유하면서 고객 테이블의 모든 레코드를 보유하길 원할 수 있습니다.

부분 결합에 대한 데이터 세트를 선택하려면 다음을 수행하십시오.

1. 병합 노드의 **병합** 탭을 클릭하십시오.
2. **일치하는 레코드 및 선택한 일치하지 않는 레코드 포함(부분 외부 결합)**을 선택하십시오.
3. **선택** 단추를 클릭하여 부분 결합에 대한 데이터 세트를 지정할 수 있는 대화 상자를 여십시오.

외부 결합 열. *외부 결합* 열에서 포함할 데이터 세트를 전부 선택하십시오. 부분 결합의 경우 겹치는 레코드뿐만 아니라 여기서 선택한 데이터 세트에 대한 불완전한 레코드도 보유됩니다. 자세한 정보는 결합의 유형 주제를 참조하십시오.

④ 병합을 위한 조건 지정

병합 방법을 **조건**으로 설정하여 병합을 수행하기 위해 충족해야 하는 하나 이상의 조건을 지정할 수 있습니다.

조건 필드에 직접 조건을 입력하거나 필드 오른쪽의 계산기 기호를 클릭하여 표현식 작성기를 통해 조건을 작성할 수 있습니다.

병합 충돌을 피하기 위해 중복 필드 이름 태그 추가 병합할 둘 이상의 데이터 세트에 동일한 필드 이름이 포함되어 있는 경우에는 이 선택란을 선택하여 필드 열 헤더의 시작 부분에 다른 접두부 태그를 추가하십시오. 예를 들어, *Name*이라는 필드가 두 개 있으면 병합 결과에는 *1_Name*과 *2_Name*이 포함됩니다. 데이터 소스에서 태그의 이름이 바뀐 경우에는 번호가 지정된 접두부 태그 대신 새 이름이 사용됩니다. 이 선택란을 선택하지 않은 경우 데이터에 중복 이름이 있으면 선택란 오른쪽에 경고가 표시됩니다.

⑤ 병합을 위한 순위화된 조건 지정

순위화된 조건 병합은 조건별 왼쪽 외부 결합 병합으로 간주될 수 있습니다. 병합의 왼쪽은 각각의 레코드가 이벤트인 1차 데이터 세트입니다. 예를 들어, 범죄 데이터에서 패턴을 찾는 데 사용되는 모델에서 1차 데이터 세트의 각 레코드는 범죄 및 해당 연관된 정보(위치, 유형 등)가 됩니다. 이 예제에서 오른쪽에는 관련 지리 공간적 데이터 세트가 포함되어 있습니다.

이 병합에서는 병합 조건과 순위화 표현식을 모두 사용합니다. 병합 조건은 within 또는 close_to 등의 지리 공간적 함수를 사용할 수 있습니다. 병합을 수행하는 동안 오른쪽 데이터 세트의 모든 필드가 왼쪽 데이터 세트에 추가되지만 다중 일치가 있으면 목록 필드가 생성됩니다. 예:

- 왼쪽: Crime data
- 오른쪽: Counties 데이터 세트 및 roads 데이터 세트
- 병합 조건: Crime data within counties and close_to roads, along with a definition of what counts as close_to.

이 예제에서 범죄가 3개 도로의 요구된 close_to 거리 내에서 발생한 경우(또한 리턴될 일치 수가 3 이상으로 설정된 경우) 3개 도로 모두 목록 항목으로 리턴됩니다.

병합 방법을 **순위화된 조건**으로 설정하여 병합을 수행하기 위해 충족할 하나 이상의 조건을 지정할 수 있습니다.

1차 데이터 세트 병합할 1차 데이터 세트를 선택하십시오. 선택하는 데이터 세트에 다른 모든 데이터 세트의 필드가 추가됩니다. 이것은 외부 결합 병합의 왼쪽으로 간주될 수 있습니다.

1차 데이터 세트를 선택하면 병합 노드에 연결되는 다른 모든 입력 데이터 세트가 자동으로 **병합** 테이블에 나열됩니다.

병합 충돌을 피하기 위해 중복 필드 이름에 태그 추가 병합할 둘 이상의 데이터 세트에 동일한 필드 이름이 포함되어 있는 경우에는 이 선택란을 선택하여 필드 열 헤더의 시작 부분에 다른 접두부 태그를 추가하십시오. 예를 들어, Name이라는 필드가 두 개 있으면 병합 결과에는 *1_Name*과 *2_Name*이 포함됩니다. 데이터 소스에서 태그의 이름이 바뀌는 경우에는 번호가 지

정된 접두부 태그 대신 새 이름이 사용됩니다. 이 선택란을 선택하지 않은 경우 데이터에 중복 이름이 있으면 선택란 오른쪽에 경고가 표시됩니다.

병합

데이터 세트

병합 노드에 대해 입력으로 연결되는 2차 데이터 세트의 이름을 표시합니다. 기본적으로 둘 이상의 2차 데이터 세트가 있으면 병합 노드에 연결된 순서대로 해당 데이터 세트가 나열됩니다.

병합 조건

1차 데이터 세트를 가진 테이블의 각 데이터 세트를 병합하는 데 필요한 고유 조건을 입력하십시오. 셀에 직접 조건을 입력하거나 셀 오른쪽에 있는 계산기 기호를 클릭하여 표현식 작성기를 통해 조건을 작성할 수 있습니다. 예를 들어, 지리 공간적 술어를 사용하여 다른 데이터 세트의 구/군 데이터 내에 한 데이터 세트의 범죄 데이터를 배치하는 병합 조건을 작성할 수 있습니다. 기본 병합 조건은 아래 목록에 표시된 대로 지리 공간적 측정 수준에 따라 다릅니다.

- 점, 선 스트링, 다중 점, 다중 선 스트링 - close_to의 기본 조건입니다.
- 다각형, 다중 다각형 - within의 기본 조건입니다.

이 수준에 대한 자세한 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.

데이터 세트에 다양한 유형의 여러 지리 공간적 필드가 포함되어 있으면 사용되는 기본 조건은 다음 내림차순으로 데이터에서 발견되는 첫 번째 측정 수준에 의해 결정됩니다.

- 점
- 선 문자열
- 다각형

> **참고**: 기본값은 2차 데이터베이스에 지리 공간적 데이터 필드가 있는 경우에만 사용할 수 있습니다.

순위화 표현식

데이터 세트의 병합을 순위화하는 표현식을 지정하십시오. 이 표현식은 순위화 기준을 기반으로 하는 순서로 다중 일치를 정렬하는 데 사용됩니다. 셀에 직접 조건을 입력하거나 셀 오른쪽에 있는 계산기 기호를 클릭하여 표현식 작성기를 통해 조건을 작성할 수 있습니다.

거리 및 영역의 기본 순위화 표현식이 표현식 작성기에서 제공되며 둘 다 오름차순으로 순위화됩니다(예를 들어, 거리에 대한 맨 위 일치가 가장 작은 값임을 의미함). 거리별 순위화의 예제는 1차 데이터 세트에 범죄 및 해당 연관된 위치가 포함되어 있고 각각의 다른 데이터

세트에 위치를 가진 범죄 대상이 포함되어 있는 경우입니다. 이 경우 범죄와 범죄 대상 사이의 거리를 순위화 기준으로 사용할 수 있습니다. 기본 순위화 표현식은 아래 목록에 표시된 대로 지리 공간적 측정 수준에 따라 다릅니다.

- 점, 선 스트링, 다중 점, 다중 선 스트링 - 기본 표현식은 distance입니다.
- 다각형, 다중 다각형 - 기본 표현식은 area입니다.

참고: 기본값은 2차 데이터베이스에 지리 공간적 데이터 필드가 있는 경우에만 사용할 수 있습니다.

일치 수

조건 및 순위화 표현식을 기반으로 리턴되는 일치의 수를 지정하십시오. 기본 일치 수는 아래 목록에 표시된 대로 2차 데이터 세트의 지리 공간적 측정 수준에 따라 다릅니다. 하지만 셀을 두 번 클릭하여 최대 100까지 자체 값을 입력할 수 있습니다.
- 점, 선 스트링, 다중 점, 다중 선 스트링 - 기본값 3
- 다각형, 다중 다각형 - 기본값 1
- 지리 공간적 필드가 포함되어 있지 않은 데이터 세트 - 기본값 1

예를 들어, **병합 조건** close_to 및 **순위화 표현식** distance를 기반으로 하는 병합을 설정하면 1차 데이터 세트의 각 레코드에 대한 2차 데이터 세트의 상위 3개(가장 가까운) 일치가 결과 목록 필드에 값으로 리턴됩니다.

⑥ 병합 노드의 필드 필터링

병합 노드에는 여러 데이터 소스를 병합한 결과로 중복 필드를 필터링하거나 이름을 바꾸는 편리한 방법이 포함되어 있습니다. 대화 상자의 **필터** 탭을 클릭하여 필터링 옵션을 선택하십시오.

여기서 제공되는 옵션은 필터 노드의 옵션과 거의 동일합니다. 하지만 필터 메뉴에서는 여기서 다루지 않은 추가적인 옵션을 사용할 수 있습니다. 자세한 정보는 필드 필터링 또는 이름 바꾸기 주제를 참조하십시오.

필드. 현재 연결된 데이터 소스의 입력 필드를 표시합니다.

태그. 데이터 소스 링크와 연관된 태그 이름(또는 번호)을 나열합니다. **입력** 탭을 클릭하여 이 병합 노드에 대한 활성 링크를 변경하십시오.

소스 노드. 데이터를 병합 중인 소스 노드를 표시합니다.

연결된 노드. 병합 노드에 연결된 노드의 노드 이름을 표시합니다. 복잡한 데이터 마이닝을 사용하려면 동일한 소스 노드를 포함할 수 있는 여러 병합 또는 추가 조작이 필요할 수 있습니다. 연결된 노드 이름은 이를 구분하는 방법을 제공합니다.

필터. 입력 필드와 출력 필드 간 현재 연결을 표시합니다. 활성 연결은 중단되지 않은 화살표를 표시합니다. 빨간색 X가 있는 연결은 필터링된 필드를 표시합니다.

필드. 병합 또는 추가 후 출력 필드를 나열합니다. 중복 필드는 빨간색으로 표시됩니다. 위의 필터 필드를 클릭하여 중복 필드를 사용 안함으로 설정하십시오.

현재 필드 보기. 키 필드로 사용하기 위해 선택한 필드에 대한 정보를 보려면 선택하십시오.

사용하지 않은 필드 설정 보기. 현재 사용되고 있지 않은 필드에 대한 정보를 보려면 선택하십시오.

⑦ 입력 순서 및 태그 지정 설정

합치기 및 붙여쓰기 노드 대화 상자에서 입력 탭을 사용하면 입력 데이터 소스의 순서를 지정하고 각 소스에 대한 태그 이름을 변경할 수 있습니다.

태그 및 입력 데이터 세트의 순서. 완전한 레코드만 합치거나 붙여쓰려면 선택하십시오.

- **태그**. 각 입력 데이터 소스에 대한 현재 태그 이름을 나열합니다. 태그 이름 또는 tags는 합치기 또는 붙여쓰기 작업에 대한 데이터 링크를 고유하게 식별하는 방법입니다. 예를 들어, 여러 파이프의 물이 한 지점에서 합쳐져서 단일 파이프를 통해 흘러가는 것을 상상해 보십시오. IBM® SPSS® Modeler의 데이터도 유사하게 흘러가고 합치는 포인트는 종종 다양한 데이터 소스 사이의 복합 상호작용이 됩니다. 태그는 노드가 저장되거나 연결 해제될 때 링크를 유지하고 쉽게 인식할 수 있도록 합치기 또는 붙여쓰기 노드에 대한 입력("파이프")을 관리하는 방법을 제공합니다.

추가 데이터 소스를 합치기 또는 붙여쓰기 노드에 연결할 때 사용자가 노드를 연결한 순서를 표시하기 위해 자동으로 숫자를 사용하여 기본 태그가 작성됩니다. 이 순서는 입력 또는 출력 데이터 세트의 필드 순서와는 연관이 없습니다. *태그* 열에 새 이름을 입력하여 기본값을 변경할 수 있습니다.

- **소스 노드**. 데이터가 결합되는 소스 노드를 표시합니다.
- **연결된 노드**. 합치기 또는 붙여쓰기 노드에 연결된 노드의 노드 이름을 표시합니다. 복합 데이터 마이닝의 경우, 동일한 소스 노드를 포함하는 여러 합치기 작업이 필요한 경우가 종종 있습니다. 연결된 노드 이름은 이를 구분하는 방법을 제공합니다.
- **필드**. 각 데이터 소스 내의 필드 수를 나열합니다.

현재 태그 보기. 합치기 또는 붙여쓰기 노드에 의해 활성 상태로 사용되는 태그를 보려면 선택하십시오. 즉, 현재 태그는 데이터가 플로우되는 노드에 대한 링크를 식별합니다. 파이프 비유를 사용하자면, 현재 태그는 기존 물이 흘러가고 있는 파이프와 유사합니다.

사용되지 않은 태그 설정 보기. 이전에는 합치기 또는 붙여쓰기 노드에 연결되는 데 사용되었으나 현재는 데이터 소스와 연결되어 있지 않은 태그 또는 링크를 보려면 선택하십시오. 이는 배수 시스템 내에 아직 그대로 있는 비어 있는 파이프와 유사합니다. 이러한 "파이프"를 새 소스에 연결하거나 제거할 수 있습니다. 노드에서 사용되지 않은 태그를 제거하려면 **지우기**를 클릭하십시오. 그러면 사용되지 않은 모든 태그가 한 번에 선택 취소됩니다.

⑧ 병합 최적화 설정

시스템은 특정 상황에서 더 효율적으로 데이터를 병합할 수 있는 두 가지 옵션을 제공합니다. 이 옵션을 사용하면 한 입력 데이터 세트가 다른 데이터 세트보다 상당히 크거나 병합에 사용하는 키 필드 중 일부 또는 전부를 기준으로 데이터가 이미 정렬되어 있는 경우 병합을 최적화할 수 있습니다.

> **참고**: 이 탭에서의 최적화는 IBM® SPSS® Modeler 원시 노드 실행에만 적용됩니다. 즉, 병합 노드가 SQL로 푸시백되지 않습니다. 최적화 설정은 SQL 생성에 영향을 미치지 않습니다.

한 입력 데이터 세트가 상대적으로 큼. 입력 데이터 세트 중 하나가 다른 입력 데이터 세트보다 훨씬 큼을 나타내려면 선택하십시오. 시스템은 메모리에서 더 작은 데이터 세트를 캐싱한 후 큰 데이터 세트는 캐싱 또는 정렬하지 않고 처리하여 병합을 수행합니다. 공유 데이터의 큰 중심 테이블이 있는 경우(예: 트랜잭션 데이터에) 스타 스키마 또는 비슷한 디자인을 사용하여 설계된 이 유형의 데이터 결합을 일반적으로 사용합니다. 이 옵션을 선택하는 경우에는 **선택**을 클릭하여 큰 데이터 세트를 지정하십시오. 큰 데이터 세트는 *하나*만 선택할 수 있습니다. 다음 표에는 이 방법을 사용하여 최적화할 수 있는 결합이 요약되어 있습니다.

표 1. 결합 최적화 요약

결합유형	큰 입력 데이터 세트에 대해 최적화할 수 있는지 여부
내부	예
부분	예(큰 데이터 세트에 불완전한 레코드가 없는 경우)
전체	아니오
안티 결합	예(큰 데이터 세트가 첫 번째 입력인 경우)

모든 입력이 이미 키 필드별로 정렬되어 있음. 병합을 위해 사용하는 키 필드 중 하나 이상을 기준으로 입력 데이터가 이미 정렬되어 있음을 나타내려면 선택하십시오. *모든* 입력 데이터 세트가 정렬되어 있는지 확인하십시오.

기존 정렬 순서 지정. 이미 정렬된 필드를 지정하십시오. 필드 선택 대화 상자를 사용하여 목록에 필드를 추가하십시오. 병합 탭에서 지정된 병합에 사용 중인 키 필드 중에서만 선택할 수 있습니다. *순서* 열에서 각 필드가 오름차순 또는 내림차순으로 정렬되는지 여부를 지정하십시오. 여러 필드를 지정하는 경우에는 올바른 정렬 순서로 필드를 나열하는지 확인하십시오. 목록 오른쪽의 화살표를 사용하여 올바른 순서로 필드를 배열하십시오. 올바른 기존 정렬 순서를 지정하지 못하면 스트림을 실행할 때 오류가 표시되고 사용자가 지정한 정렬과 일치하지 않는 레코드 수가 표시됩니다.

데이터베이스에서 사용하는 데이터 정렬 방법의 대소문자 구분 여부에 따라 하나 이상의 입력이 데이터베이스에 의해 정렬된 경우 최적화가 올바르게 작동하지 않을 수 있습니다. 예를 들어, 두 개의 입력이 있는데 하나는 대소문자를 구분하고 다른 하나는 대소문자를 구분하지 않는 경우에는 정렬 결과가 다를 수 있습니다. 병합 최적화를 수행하면 정렬된 순서를 사용하여 레코드가 처리됩니다. 결과적으로 서로 다른 데이터 정렬 방법을 사용하여 입력을 정렬하면 병합 노드에서 오류를 보고하고 정렬이 불일치하는 레코드 번호를 표시합니다. 모든 입력이 한 소스에서 제공되거나 서로 포함하는 데이터 배열을 사용하여 정렬되는 경우에는 레코드가 성공적으로 병합될 수 있습니다.

참고: 병렬 처리를 사용하면 병합 속도가 향상될 수 있습니다.

(9) 붙여쓰기 노드

붙여쓰기 노드를 사용하면 레코드 세트를 연결할 수 있습니다. 다른 소스의 레코드를 함께 결합하는 병합 노드와 달리 붙여쓰기 노드는 한 소스에서 모든 레코드를 더 이상 없을 때까지 다운스트림으로 읽고 전달합니다. 이후에 다음 소스의 레코드는 첫 번째 또는 기본 입력과 동일한 데이터 구조(레코드 수, 필드 수 등)를 사용하여 읽습니다. 기본 소스에 다른 입력 소스보다 많은 필드가 있는 경우 시스템 널 문자열($null$)이 불완전한 값에 사용됩니다.

붙여쓰기 노드는 구조가 비슷하지만 다른 데이터가 있는 데이터 세트를 결합하는 데 유용합니다. 예를 들어, 3월의 판매 데이터 파일 및 4월의 개별 파일과 같이 다른 시간 중 다른 파일에 트랜잭션 데이터를 저장했을 수 있습니다. 구조가 동일하다고 가정할 때(동일한 순서의 동일한 필드) 붙여쓰기 노드는 이러한 파일을 분석할 수 있도록 하나의 큰 파일에 결합할 수 있습니다.

참고: 파일을 붙여쓰려면 필드 측정 수준이 비슷해야 합니다. 예를 들어, *명목* 필드는 측정 수준이 *연속형*인 필드와 함께 붙여쓸 수 없습니다.

① 붙여쓰기 옵션 설정

필드 일치 기준. 매치하는 필드를 붙여쓸 때 사용할 방법을 선택하십시오.

- **위치**. 기본 데이터 소스에서 필드의 위치에 기반하여 데이터 세트를 추가하려면 이를 선택하십시오. 이 방법을 사용할 때 적절하게 붙여쓰기 위해 데이터를 정렬해야 합니다.
- **이름**. 입력 데이터 세트에서 필드의 이름에 기반하여 데이터 세트를 붙여쓰려면 이를 선택하십시오. 필드 이름을 일치시킬 때 대소문자를 감지하도록 하려면 **대소문자 구분**도 선택하십시오.

출력 필드. 붙여쓰기 노드에 연결되는 소스 노드를 나열합니다. 목록의 첫 번째 노드는 기본 입력 소스입니다. 열 표제를 클릭하여 화면에서 필드를 정렬할 수 있습니다. 이 정렬은 데이터 세트에서 필드를 실제로 다시 정렬하지 않습니다.

필드 포함. 기본 데이터 세트의 필드에 기반하여 출력 필드를 생성하려면 **기본 데이터 세트**만을 선택하십시오. 기본 데이터 세트는 입력 탭에 지정된 첫 번째 입력입니다. 모든 입력 데이터 세트에서 일치하는 필드가 있는지 여부에 관계없이 모든 데이터 세트에서 모든 필드의 출력 필드를 생성하려면 **모든 데이터 세트**를 선택하십시오.

필드에 소스 데이터 세트를 포함하여 레코드 태그 지정. 해당 값이 각 레코드의 소스 데이터 세트를 표시하는 출력 파일에 추거 필드를 추가하려면 선택하십시오. 텍스트 필드에 이름을 지정하십시오. 기본 필드 이름은 *Input*입니다.

(10) 고유 노드

데이터 마이닝을 시작하려면 먼저 데이터 세트의 중복 레코드를 제거해야 합니다. 예를 들어, 마케팅 데이터베이스에서 주소 또는 회사 정보가 다른 개인이 여러 번 표시될 수 있습니다. 고유 노드를 사용하여 데이터에서 중복 레코드를 찾거나 제거하거나 중복 레코드 그룹에서 하나의 복합 레코드를 작성할 수 있습니다.

고유 노드를 사용하려면 먼저 두 레코드가 중복으로 간주되는 경우를 판별하는 키 필드 세트를 정의해야 합니다.

모든 필드를 키 필드로 선택하지 않는 경우 두 개의 "중복" 레코드는 나머지 필드의 값에 여전히 차이가 있을 수 있기 때문에 진정으로 동일할 수 없습니다. 이 경우에는 각각의 중복 레코드 그룹 내에서 적용되는 정렬 순서도 정의할 수 있습니다. 이 정렬 순서를 통해 그룹에서 첫 번째로 처리되는 레코드를 미세 제어할 수 있습니다. 그렇지 않으면 모든 중복이 교환 가능한 것으로 간주되어 모든 레코드를 선택할 수 있습니다. 레코드의 수신 순서는 고려되지 않으므로 업스트림 정렬 노드를 사용해도 도움이 되지 않습니다(아래의 "고유 노드 내에서 레코드 정렬" 참조).

모드. 복합 레코드를 작성할지 아니면 첫 번째 레코드를 포함 또는 제외(삭제)할지 지정하십시오.

- **각 그룹에 대해 복합 레코드 작성**. 숫자가 아닌 필드를 통합하는 방법을 제공합니다. 이 옵션을 선택하면 복합 레코드 작성 방법을 지정하는 복합 탭을 사용할 수 있습니다. 자세한 정보는 고유 복합 설정을 참조하십시오.
- **각 그룹의 첫 번째 레코드만 포함**. 중복 레코드 그룹 각각의 첫 번째 레코드를 선택하고 나머지는 삭제합니다. *첫 번째* 레코드는 레코드의 수신 순서가 아니라 아래에 정의된 정렬 순서에 의해 결정됩니다.
- **각 그룹의 첫 번째 레코드만 삭제**. 중복 레코드 그룹 각각의 첫 번째 레코드를 삭제하는 대신 나머지는 선택합니다. *첫 번째* 레코드는 레코드의 수신 순서가 아니라 아래에 정의된 정렬 순서에 의해 결정됩니다. 이 옵션은 스트림에서 나중에 중복을 검사할 수 있도록 데이터에서 중복을 *찾는* 경우 유용합니다.

그룹화를 위한 키 필드. 레코드가 동일한지 판별하는 데 사용되는 필드를 나열합니다. 다음을 수행할 수 있습니다.

- 오른쪽의 필드 선택도구 단추를 사용하여 이 목록에 필드를 추가하십시오.
- 빨간색 X(제거) 단추를 사용하여 목록에서 필드를 삭제하십시오.

그룹 내에서 레코드 정렬 기준. 각 중복 그룹 내에서 레코드가 정렬되는 방식과 해당 레코드가 오름차순과 내림차순 중 어느 순서로 정렬되는지를 결정하는 데 사용되는 필드를 나열합니다. 다음을 수행할 수 있습니다.

- 오른쪽의 필드 선택도구 단추를 사용하여 이 목록에 필드를 추가하십시오.
- 빨간색 X(제거) 단추를 사용하여 목록에서 필드를 삭제하십시오.
- 위로 또는 아래로 단추를 사용하여 필드를 이동하십시오(둘 이상의 필드를 기준으로 정렬하는 경우).

각 그룹의 첫 번째 레코드를 포함하거나 제외하도록 선택한 경우 첫 번째로 처리되는 레코드가 사용자에게 중요하면 정렬 순서를 지정해야 합니다.

기본 정렬 순서. 기본적으로 레코드가 정렬 키 값의 **오름차순**과 **내림차순** 중 어느 순서로 정렬되는지를 지정하십시오.

고유 노드 내에서 레코드 정렬

중복 그룹 내 레코드의 순서가 사용자에게 중요한 경우에는 고유 노드에서 **그룹 내, 레코드 정렬 기준** 옵션을 사용하여 순서를 지정해야 합니다. 업스트림 정렬 노드에 의존하지 마십시오. 레코드의 수신 순서는 고려되지 않으며 노드 내에서 지정된 순서만 고려된다는 점을 기억하십시오.

정렬 필드를 지정하지 않는 경우(또는 충분하지 않은 정렬 필드를 지정하는 경우) 각 중복 그룹 내 레코드는 정렬되지 않거나 불완전하게 정렬되므로 결과를 예측할 수 없습니다.

예를 들어, 다수의 머신에 관한 매우 큰 로그 레코드 세트가 있다고 가정해 봅니다. 이 로그에는 다음과 같은 데이터가 포함되어 있습니다.

표 1. 머신 로그 데이터

시간 소인	Machine	온도
17:00:22	Machine A	31
13:11:30	Machine B	26
16:49:59	Machine A	30
18:06:30	Machine X	32
16:17:33	Machine A	29
19:59:04	Machine C	35
19:20:55	Machine Y	34
15:36:14	Machine X	28
12:30:41	Machine Y	25
14:45:49	Machine C	27
19:42:00	Machine B	34
20:51:09	Machine Y	36
19:07:23	Machine X	33

레코드 수를 각 머신에 대한 최신 레코드까지 줄이려면 Machine을 키 필드로 사용하고 Timestamp를 정렬 필드(내림차순)로 사용하십시오. 정렬 선택사항은 지정된 머신에 대한 다수의 행 중 리턴될 행을 지정하므로 입력 순서는 결과에 영향을 미치지 않으며 최종 데이터 출력은 다음과 같습니다.

표 2. 정렬된 머신 로그 데이터

시간소인	Machine	온도
17:00:22	Machine A	31
19:42:00	Machine B	34
19:59:04	Machine C	35
19:07:23	Machine X	33
20:51:09	Machine Y	36

① 고유 최적화 설정

작업 중인 데이터가 적은 수의 레코드만 포함하고 있거나 이미 정렬된 경우에는 IBM® SPSS® Modeler가 데이터를 더 효율적으로 처리할 수 있도록 처리 방식을 최적화할 수 있습니다.

참고: **입력 데이터 세트에 적은 수의 고유 키가 있음**을 선택하거나 노드에 대해 SQL 생성을 사용하는 경우 고유 키 값 내의 행이 리턴될 수 있습니다. 고유 키 내부에 리턴되는 행을 제어하려면 설정 탭의 **그룹 내, 레코드 정렬 기준** 필드를 사용하여 정렬 순서를 지정해야 합니다. 설정 탭에서 정렬 순서를 지정한 경우에는 최적화 옵션이 고유 노드에 의한 결과 출력에 영향을 미치지 않습니다.

입력 데이터 세트에 적은 수의 고유 키가 있음. 적은 수의 레코드, 적은 수의 키 필드 고유 값 또는 둘 다를 가진 경우 이 옵션을 선택하십시오. 그렇게 하면 성능이 개선될 수 있습니다.

설정 탭에서 필드를 정렬하거나 필드를 그룹화하여 입력 데이터 세트가 이미 정렬되어 있음. 설정 탭의 **그룹 내, 레코드 정렬 기준**에 나열되는 모든 필드를 기준으로 데이터가 이미 정렬되어 있거나 데이터의 오름차순 정렬 순서와 내림차순 정렬 순서가 동일한 경우에만 이 옵션을 선택하십시오. 그렇게 하면 성능이 개선될 수 있습니다.

SQL 생성 사용 안함. 노드에 대한 SQL 생성을 사용 안함으로 설정하려면 이 옵션을 선택하십시오.

② 고유 복합 설정

작업 중인 데이터에 예를 들어, 동일한 사용자에 대한 다중 레코드가 포함되어 있는 경우 처리할 하나의 복합(또는 통합) 레코드를 작성하여 데이터 처리 방식을 최적화할 수 있습니다.

> ⓘ **참고**: 이 탭은 설정 탭에서 **각 그룹에 대해 복합 레코드 작성**을 선택하는 경우에만 사용할 수 있습니다.

복합 탭에 대한 옵션 설정

필드. 이 열에는 데이터 모델의 키 필드를 제외한 모든 필드가 기본 정렬 순서로 표시됩니다. 노드가 연결되어 있지 않으면 필드가 표시되지 않습니다. 필드 이름을 기준으로 알파벳순으로 행을 정렬하려면 열 헤더를 클릭하십시오. Shift+클릭 또는 Ctrl+클릭을 사용하여 둘 이상의 행을 선택할 수 있습니다. 또한 필드를 마우스 오른쪽 단추로 클릭하면 모든 행을 표시하거나 오름차순 또는 내림차순 필드 이름 또는 값을 기준으로 행을 정렬하거나 측도 또는 저장 유형별로 필드를 선택하거나 값을 선택하여 동일한 **다음을 기반으로 값 채우기** 항목을 선택된 모든 행에 자동으로 추가하도록 선택할 수 있는 메뉴가 표시됩니다.

다음을 기반으로 값 채우기. 필드에 대한 복합 레코드에 사용할 값 유형을 선택하십시오. 사용 가능한 옵션은 필드 유형에 따라 다릅니다.
- 숫자 범위 필드의 경우 다음 중에서 선택할 수 있습니다.
  - 그룹에서 첫 번째 레코드
  - 그룹에서 마지막 레코드
  - 총계
  - 평균
  - 최소값
  - 최대값
  - Custom

- 시간 또는 날짜 필드의 경우 다음 중에서 선택할 수 있습니다.
  - 그룹에서 첫 번째 레코드
  - 그룹에서 마지막 레코드
  - 최초
  - 최근
  - Custom

- 문자열 또는 유형 없는 필드의 경우 다음 중에서 선택할 수 있습니다.
  - 그룹에서 첫 번째 레코드
  - 그룹에서 마지막 레코드
  - 첫 번째 영숫자
  - 마지막 영숫자
  - Custom

각각의 경우에 **사용자 정의** 옵션을 사용하여 복합 레코드를 채우는 데 사용되는 값에 대한 제어를 향상시킬 수 있습니다. 자세한 정보는 고유 복합 - 사용자 정의 탭을 참조하십시오.
필드에 레코드 개수 포함 각각의 출력 레코드에 추가 필드를 포함하려면 이 옵션을 선택하십시오(기본적으로 Record_Count라고 함). 이 필드는 각각의 통합 레코드를 형성하기 위해 통합된 입력 레코드 수를 표시합니다. 이 필드에 대해 사용자 정의 이름을 작성하려면 편집 필드에서 항목을 입력하십시오.

가. 고유 복합 - 사용자 정의 탭

사용자 정의 채우기 대화 상자는 새 복합 레코드를 완료하는 데 사용되는 값에 대한 추가적인 제어를 제공합니다. 복합 탭에서 단일 필드 행만 사용자 정의하는 경우에는 이 옵션을 사용하기 전에 먼저 데이터를 인스턴스화해야 합니다.

> ⓘ **참고**: 이 대화 상자는 복합 탭의 **다음을 기반으로 값 채우기** 열에서 사용자 정의 값을 선택하는 경우에만 사용할 수 있습니다.

필드 유형에 따라 다음 옵션 중 하나에서 선택할 수 있습니다.

- 빈도별 선택. 데이터 레코드에서 발생하는 빈도를 기반으로 값을 선택하십시오.

> ⓘ **참고**: 연속형, 유형 없음 또는 날짜/시간 유형을 가진 필드에는 사용할 수 없습니다.

  - 사용. 최대 빈도와 최소 빈도 중에서 선택하십시오.
  - 등순위. 발생 빈도가 동일한 레코드가 둘 이상 있는 경우에는 필요한 레코드를 선택하는 방법을 지정하십시오. 첫 번째 사용, 마지막 사용, 최저 사용 또는 최고 사용이라는 네 가지 옵션 중 하나에서 선택할 수 있습니다.

- **값(T/F) 포함**. 필드를 그룹의 레코드가 지정된 값을 가지고 있는지 식별하는 플래그로 변환하려면 선택하십시오. 그런 다음 선택된 필드에 대한 목록에서 **값**을 선택할 수 있습니다.

> ⓘ **참고**: 복합 탭에서 둘 이상의 필드 행을 선택하는 경우에는 사용할 수 없습니다.

- **목록에서 첫 번째 일치**. 복합 레코드에 제공할 값의 우선 순위를 지정하려면 선택하십시오. 그런 다음 선택된 필드에 대한 목록에서 **항목** 중 하나를 선택할 수 있습니다.

> ⓘ **참고**: 복합 탭에서 둘 이상의 필드 행을 선택하는 경우에는 사용할 수 없습니다.

- **값 연결**. 그룹의 모든 값을 문자열로 연결하여 유지하려면 선택하십시오. 각각의 값 사이에서 사용할 구분자를 지정해야 합니다.

> ⓘ **참고**: 이는 연속형, 유형 없음 또는 날짜/시간 유형을 가진 하나 이상의 필드 행을 선택하는 경우 사용 가능한 유일한 옵션입니다.

- **구분자 사용**. **공백** 또는 **쉼표**를 연결된 문자열의 구분자 값으로 사용하도록 선택할 수 있습니다. 또는 **기타** 필드에서 자체 구분자 값 문자를 입력할 수 있습니다.

> ⓘ **참고**: **값 연결** 옵션을 선택하는 경우에만 사용할 수 있습니다.

(11) 스트리밍 시계열 노드

스트리밍 시계열 노드를 사용하여 한 단계로 시계열 모델을 작성하고 스코어링합니다. 목표 필드마다 개별 시계열 모델이 작성되지만, 생성된 모델 팔레트에 모델 너깃이 추가되지 않으며 모델 정보를 찾아볼 수 없습니다.

시계열 데이터를 모델링하는 방법에서는 각 측정 사이에 균일한 구간이 필요합니다(빈 행으로 결측값을 표시함). 데이터가 이미 이 요구사항을 만족하지 않으면 필요에 따라 값을 변환해야 합니다.

시계열 노드와 관련하여 또 다른 주의 사항은 다음과 같습니다.

- 필드는 숫자여야 합니다.
- 날짜 필드를 입력으로 사용할 수 없습니다.
- 파티션은 무시됩니다.

스트리밍 시계열 노드는 시계열에 대한 지수평활, 일변량 ARIMA(Autoregressive Integrated Moving Average), 다변량 ARIMA(또는 전이 함수) 모델을 추정하고 시계열 데이터에 기반하여 예측을 생성합니다. 하나 이상의 목표 필드에 가장 적합한 ARIMA 또는 지수평활 모델을 자동으로 식별하고 추정하는 자동 Modeler도 사용 가능합니다.

시계열 모델링에 대한 자세한 정보는 시계열 데이터의 내용을 참조하십시오.

IBM® SPSS® Collaboration and Deployment Services 스코어링 서비스를 사용하여 IBM SPSS Modeler Solution Publisher를 통해 스트리밍 배포 환경에서 스트리밍 시계열 노드를 사용할 수 있습니다.

① 스트리밍 시계열 노드 - 필드 옵션

사전 정의된 역할 사용 이 옵션은 업스트림 유형 노드(또는 업스트림 소스 노드의 유형 탭)에서 역할 설정(목표, 예측자 등)을 사용합니다.

사용자 정의 필드 할당 사용: 수동으로 대상, 예측자 및 기타 역할을 할당하려면 이 옵션을 선택하십시오.

> ⓘ **참고**: 데이터를 파티셔닝한 경우 **사전 정의된 역할 사용**을 선택하면 파티션이 고려되지만 **사용자 정의 필드 할당**을 선택하면 고려되지 않습니다.

필드. 화살표 단추를 사용하여 수동으로 이 목록의 항목을 화면 오른쪽의 다양한 역할 필드에 할당하십시오. 아이콘은 각 역할 필드에 대한 유효한 측정 수준을 나타냅니다.

목록의 모든 필드를 선택하려면 **모두** 단추를 클릭하거나 개별 측정 수준 단추를 클릭하여 이 측정 수준의 모든 필드를 선택하십시오.

대사 하나 이상의 필드를 예측의 대상으로 선택하십시오.

후보 입력 예측에 대한 입력으로 하나 이상의 필드를 선택합니다.

이벤트 및 개입 - 이 영역을 사용하여 특정 입력 필드를 이벤트 또는 개입 필드로 지정할 수 있습니다. 이와 같이 지정하면 이벤트(판매 프로모션과 같은 예측 가능한 반복 상황) 또는 개입(정전 또는 직원 파업과 같은 일회성 사건)의 영향을 받을 수 있는 시계열 데이터를 포함하는 항목으로 필드를 식별합니다.

② 스트리밍 시계열 노드 - 데이터 지정 사항 옵션

데이터 지정 사항 탭에서는 모델에 포함될 데이터에 대한 모든 옵션을 설정할 수 있습니다. **날짜/시간 필드**와 **시간 간격**을 모두 지정하는 경우에만 **실행** 단추를 클릭하여 모든 기본 옵션으로 모델을 작성할 수도 있지만, 보통 사용자는 고유한 목적을 위해 작성을 사용자 정의하려고 합니다.

탭은 모델에 특정한 사용자 정의를 설정하는 여러 창을 포함합니다.

가. 스트리밍 시계열 노드 - 관측값

이 분할창의 설정을 사용하여 관측값을 정의하는 필드를 지정할 수 있습니다.

날짜/시간 필드로 지정된 관측값

관측값이 날짜, 시간 또는 시간소인 필드를 통해 정의되도록 지정할 수 있습니다. 관측값을 정의하는 필드 외에, 관측값을 설명하는 적절한 시간 간격을 선택하십시오. 지정된 시간 간격에 따라, 관측값(증분) 사이의 구간이나 주당 일 수와 같은 다른 설정을 지정할 수도 있습니다. 다음 고려사항은 시간 간격에 적용됩니다.

- 관측값이 시간에서 비정규적으로 간격이 있는 경우(판매 순서가 처리되는 시간과 같이), **비정규** 값을 사용하십시오. **비정규**가 선택될 때, 데이터 지정 사항 탭의 **시간 간격** 설정에서 분석에 사용되는 시간 간격을 지정해야 합니다.
- 관측값이 날짜와 시간을 나타내고 시간 간격이 시, 분 또는 초인 경우 **하루 중 시간(시)**, **하루 중 시간(분)** 또는 **하루 중 시간(초)**을 사용하십시오. 관측값이 날짜에 대한 참조 없이 시간(기간)을 나타내고 시간 간격이 시, 분 또는 초일 경우, **시(비주기적)**, **분(비주기적)** 또는 **초(비주기적)**를 사용하십시오.
- 선택된 시간 간격을 기초로, 프로시저는 결측 관측값을 발견할 수 있습니다. 프로시저에서는 모든 관측값이 시간에서 동일하게 간격을 두고 결측 관측값이 없다고 정하므로, 결측 관측값을 발견해야 합니다. 예를 들어, 시간 간격이 일(Days)이고 날짜 2015-10-27 뒤에 2015-10-29가 있는 경우, 2015-10-28에 대해 결측 관측값이 있습니다. 결측 관측값에 대해 값이 대체됩니다. 데이터 지정 사항 탭의 **결측값 처리** 영역에서 결측값 처리 설정을 지정하십시오.
- 지정된 시간 간격은 프로시저가 함께 통합해야 하는 동일한 시간 간격의 여러 관측값을 발견하고 관측값에 동일하게 간격이 있도록 월의 첫 번째와 같은 구간 경계에 관측값을 맞출 수 있도록 합니다. 예를 들어, 시간 간격이 월일 경우, 동일 월에 있는 여러 날짜가 함께 통합됩니다. 이 유형의 통합을 *그룹화*라고 합니다. 기본적으로, 관측값은 그룹화될 때 합산됩니다. 데이터 지정 사항 탭의 **통합 및 분포** 설정에서, 그룹화에 다른 방법(예: 관측값의 평균)을 지정할 수 있습니다.
- 일부 시간 간격의 경우, 추가 설정은 동일하게 간격이 있는 정규 구간에서 중단을 설정할 수 있습니다. 예를 들어, 시간 간격이 일(Days)이지만 평일만 유효한 경우, 주에 5일이 있고 주는 월요일에 시작함을 지정할 수 있습니다.

관측값이 주기 또는 순환 주기로 정의됨

관측값은 임의의 순환 수준 수까지, 주기 또는 반복 주기 순환을 나타내는 하나 이상의 정수 필드로 정의할 수 있습니다. 이 구조를 사용할 경우 표준 시간 간격 중 하나에 맞지 않은 관측값 계열을 기술할 수 있습니다. 예를 들어, 10개월만 있는 회계연도는 연도를 나타내는 순환 필드와, 월을 나타내는 주기 필드로 설명할 수 있습니다. 여기서 하나의 주기 길이는 10입니다.

순환 주기를 지정하는 필드는 주기적 수준의 계층 구조를 정의합니다. 가장 낮은 수준은 **주기** 필드에 의해 정의됩니다. 다음 최상위 수준은 수준이 1인 순환 필드에 의해 지정되고, 그 다음은 수준 2의 순환 필드로 지정되며 뒤로도 마찬가지입니다. 가장 높은 수준을 제외하고, 각 수준의 필드 값은 다음 최상위 수준에 관하여 주기적이어야 합니다. 최상위 수준의 값은 주기적이 될 수 없습니다. 예를 들어, 10달 회계연도의 경우 월은 연도 내에서 주기적이며 연도는 주기적이 아닙니다.

- 특정 수준에 있는 순환의 길이는 다음으로 가장 낮은 수준의 주기성입니다. 회계연도 예의 경우, 단 하나 순환 수준이 있고 순환 길이는 10입니다. 다음으로 가장 낮은 수준이 월을 나타내고 지정된 회계 연도에 10달이 있기 때문입니다.
- 주기적 필드의 시작 값(1부터 시작하지 않음)을 지정하십시오. 이 설정은 결측값을 발견하는데 필요합니다. 예를 들어, 주기적 필드는 2에서 시작하지만 시작 값은 1로 지정되는 경우, 프로시저는 해당 필드의 각 순환에 있는 첫 번째 주기에 대해 결측값이 있다고 가정합니다.

나. 스트리밍 시계열 노드 - 분석 시간 구간

분석에 사용할 시간 구간은 관측값의 시간 구간과 다를 수 있습니다. 예를 들어, 관측값의 시간 구간이 일(Days)일 경우, 분석의 시간 구간으로는 월을 선택할 수 있습니다. 그런 다음 모델이 작성되기 전에 매일 데이터에서 매월 데이터까지 데이터가 통합됩니다. 또한 데이터를 장기 시간 구간에서 단기 시간 구간으로 분포할 것을 선택할 수도 있습니다. 예를 들어, 관측값이 분기별인 경우, 데이터를 분기별에서 월별 데이터로 분포할 수 있습니다.

이 분할창의 설정을 사용하여 분석 시간 구간을 지정할 수 있습니다. 데이터가 통합되거나 분포되는 방법은 데이터 지정 사항 탭의 **통합 및 분포** 설정에서 지정됩니다.

분석이 행해지는 시간 구간에 대해 사용 가능한 선택은 해당 관측값 정의 방법과 관측값의 시간 구간에 따라 다릅니다. 특히, 관측값이 순환 주기로 정의될 경우 통합만 지원됩니다. 그러한 경우, 분석의 시간 구간은 관측값의 시간 구간보다 크거나 같아야 합니다.

다. 스트리밍 시계열 노드 - 통합 및 분포 옵션

이 분할창의 설정을 사용하여 관측값의 시간 구간과 관련한 입력 데이터 통합 또는 분포 설정을 지정할 수 있습니다.

집계 함수

분석에 사용되는 시간 구간이 관측에 사용되는 시간 구간보다 길 경우, 입력 데이터는 통합됩니다. 예를 들어, 관측값의 시간 구간이 일(Days)이고 분석의 시간 구간이 월일 경우 통합이 수행됩니다. mean, sum, mode, min 또는 max 집계 함수를 사용할 수 있습니다.

분포 함수

분석에 사용되는 시간 구간이 관측의 시간 구간보다 짧을 경우, 입력 데이터는 분포됩니다. 예를 들어, 관측값의 시간 구간이 분기이고 분석의 시간 구간이 월일 경우 분포가 수행됩니다. mean 또는 sum 분포 함수를 사용할 수 있습니다.

그룹화 함수

그룹화는 관측값이 날짜/시간에 의해 정의되고 여러 관측값이 동일 시간 구간에 발생하는 경우에 적용됩니다. 예를 들어, 관측값의 시간 구간이 월일 경우, 동일 월에 있는 여러 날짜가 그룹화되어 날짜가 발생하는 월과 연관됩니다. mean, sum, mode, min 또는 max와 같은 그룹화 함수를 사용할 수 있습니다. 다음 그룹화는 항상 관측값이 날짜/시간에 의해 정의되고 관측값의 시간 간격이 비정규로 지정된 경우에 수행됩니다

> **참고**: 그룹화가 통합 양식이어도, 그룹화는 결측값 처리 이전에 수행됩니다(정상 통합은 결측값 처리 이후에 수행됩니다). 관측값의 시간 구간이 비정규로 지정되는 경우, 통합은 그룹화 함수로만 수행됩니다.

교차-일 관측값을 이전 일로 통합

1일 경계를 교차하는 시간을 사용하는 관측값이 전날의 값에 통합되는지 여부를 지정합니다. 예를 들어, 20:00시에 시작하는 8시간 노동의 시간별 관측값의 경우, 이 설정은 00:00 및 04:00 사이의 관측값이 전날 통합 결과에 포함되는지 여부를 지정합니다. 이 설정은 관측값의 시간 구간이 하루 중 시간(시), 하루 중 시간(분) 또는 하루 중 시간(초)이고 분석의 시간 구간이 일(Days)인 경우에만 적용됩니다.

지정된 필드에 대한 사용자 정의 설정

필드 기준으로 필드에 통합, 분포 및 그룹화 함수를 지정할 수 있습니다. 이 설정은 통합, 분포 및 그룹화 함수에 대한 기본 설정을 대체합니다.

라. 스트리밍 시계열 노드 - 결측값 옵션

이 분할창의 설정을 사용하여 입력 데이터의 결측값을 대체값으로 바꾸는 방법을 지정할 수 있습니다. 다음 방법으로 바꿀 수 있습니다.

선형 보간법

선형 보간법을 사용하여 결측값을 바꿉니다. 결측값 이전의 마지막 유효한 값과 결측값 이후의 첫 번째 유효한 값이 보간법에 사용됩니다. 계열에서 첫 번째 또는 마지막 관측값에 결측값이 있는 경우, 계열의 시작 또는 종료에서 두 개의 가장 근접한 비결측 값이 사용됩니다.
선형 보간법을 사용하여 결측값을 바꿉니다.

- 비계절 데이터의 경우 결측값 이전의 마지막 유효한 값과 결측값 이후의 첫 번째 유효한 값이 보간법에 사용됩니다. 결측값이 시계열의 시작 또는 끝에 있는 경우 가장 근접한 두 개의 유효 값을 기준으로 선형 외삽법이 사용됩니다.
- 계절 데이터의 경우 결측값 이전 동일 기간의 마지막 유효한 값과 결측값 이후 동일 기간의 첫 번째 유효한 값을 사용하여 결측값이 선형 보간됩니다. 결측값에서 동일한 기간의 두 값 중 하나를 찾을 수 없는 경우 데이터는 비계절 데이터로 간주되어 비계절 데이터의 선형 보간을 사용하여 결측값을 대치합니다.

계열 평균

결측값을 전체 계열에 대한 평균으로 바꿉니다.

근접한 값들의 평균

결측값을 유효한 근접 값의 평균으로 바꿉니다. 근접한 값들의 계산너비는 평균을 계산하는데 사용되는 결측값 전후의 유효값 수입니다.

근접한 값들의 중앙값

결측값을 근접한 유효한 값의 중앙값으로 바꿉니다. 근접한 값들의 계산너비는 평균을 계산하는데 사용되는 결측값 전후의 유효값 수입니다.

선형 추세

이 옵션은 단순 선형 회귀 모델에 맞게 계열의 모든 비결측 관측값을 사용한 후 이 모델을 사용하여 결측값을 대체합니다.

기타 설정:

최저 데이터 품질 스코어(%)

각 시계열에 해당하는 입력 데이터 및 시간 변수에 대한 데이터 품질 측도를 계산합니다. 데이터 품질 스코어가 이 임계값보다 낮을 경우 해당 시계열은 삭제됩니다.

마. 스트리밍 시계열 노드 - 추정 기간

추정 기간 분할창에서 모델 추정에 사용될 레코드의 범위를 지정할 수 있습니다. 기본적으로 추정 기간은 모든 계열에 걸쳐 최초 관측값 시간에 시작되고 최근 관측값 시간에 종료됩니다.

시작 및 종료 시간 기준

추정 기간의 시작 및 종료 둘 다를 지정하거나 시작 또는 종료만 지정할 수 있습니다. 추정 기간의 시작 또는 종료를 생략하는 경우, 기본값이 사용됩니다.

- 날짜/시간 필드에 의해 관측값이 정의된 경우, 날짜/시간 필드에 사용되는 것과 동일한 형식으로 시작 및 종료 값을 입력하십시오.
- 순환 주기에 의해 정의된 관측값의 경우, 순환 주기 필드마다 값을 지정하십시오. 각 필드는 별도의 열에 표시됩니다.

최근이거나 최초의 시간 구간(L)

선택적 오프셋으로, 데이터의 최초 시간 구간에 시작하거나 최근 시간 구간에 종료하는, 지정된 시간 구간 수로 추정 기간을 정의합니다. 이 컨텍스트에서, 시간 구간은 분석의 시간 구간을 가리킵니다. 예를 들어, 관측값이 매월 단위이지만 분석의 시간 구간은 분기일 수 있습니다. **최근**과 **시간 구간 수**로 24 값을 지정하면 최근 24개 분기를 의미합니다.

선택적으로, 지정된 시간 구간 수를 제외할 수 있습니다. 예를 들어, 최근 24 시간 구간을 지정하고 제외할 수로 1를 지정하면, 추정 기간은 마지막 구간 앞에 있는 24개 구간으로 구성됩니다.

③ 스트리밍 시계열 노드 - 작성 옵션

작성 옵션 탭은 모델을 작성하기 위한 모든 옵션을 설정하는 위치입니다. 물론, **실행** 단추를 클릭하여 모든 기본 옵션으로 모델을 작성할 수도 있지만, 보통 사용자는 고유한 목적을 위해 작성을 사용자 정의하려고 합니다.

이 탭은 해당 모델에만 적용되는 사용자 정의를 설정하는 두 개의 분할창으로 구성됩니다.

가. 스트리밍 시계열 노드 - 일반 작성 옵션

이 분할창에서 사용 가능한 옵션은 **방법** 목록에서 선택한 다음 세 가지 설정에 따라 다릅니다.

- **자동 모델 생성기**. 자동 모델 생성기를 사용하려면 이 옵션을 선택합니다. 그러면 각 종속 계열에 대한 최적 적합 모델을 자동으로 찾습니다.
- **지수평활**. 이 옵션을 사용하여 사용자 정의 지수평활 모델을 지정합니다.
- **ARIMA**. 이 옵션을 사용하여 사용자 정의 ARIMA 모델을 지정합니다.

자동 모델 생성기

모델 유형에서 작성하려는 모델의 유형을 선택하십시오.

- **모든 모델**. 자동 모델 생성기에서 ARIMA 및 지수평활 모델을 모두 고려합니다.
- **지수평활 모델만**. 자동 모델 생성기에서 지수평활 모델만 고려합니다.
- **ARIMA 모델만**. 자동 모델 생성기에서 ARIMA 모델만 고려합니다.

자동 모델 생성기에서 계절 모델 고려. 이 옵션은 활성 데이터 세트에 대해 주기성이 정의된 경우에만 사용할 수 있습니다.이 옵션을 선택하면 자동 모델 생성기는 계절 및 비계절 모델을 모두 고려합니다. 이 옵션을 선택하지 않은 경우 자동 모델 생성기에서 비계절 모델만 고려합니다.

자동 모델 생성기에서 정교한 지수평활 모델 고려. 이 옵션을 선택하면 자동 모델 생성기가 총 13개의 지수평활 모델(7개는 원래 시계열 노드에 있고 6개는 버전 18.1에서 추가됨)을 검색합니다. 이 옵션을 선택하지 않으면 자동 모델 생성기가 원래 7개의 지수평활 모델만 검색합니다.
이상치에서 다음 옵션 중 하나를 선택하십시오.

자동으로 이상값 발견. 기본적으로 이상값 자동 발견을 수행하지 않습니다. 이상값 자동 발견을 수행하려면 이 옵션을 선택하고 원하는 이상값 유형을 선택합니다.

추가 정보는 이상값의 내용을 참조하십시오.

입력 필드는 이 목록에 포함되기 전에 플래그, 명목 또는 순서와 같은 측정 수준이 있어야 하며 숫자여야 합니다(예: 플래그 필드의 경우 참/거짓이 아닌 1/0).

추가 정보는 펄스 및 단계의 내용을 참조하십시오.

자동 모델 생성기는 **필드** 탭의 이벤트 또는 개입 필드로 식별된 입력에 대해 임의 전이 함수가 아닌 단순 회귀분석만 고려합니다.

지수평활

모델 유형. 지수평활 모델은 계절 또는 비계절로 분류됩니다.[1] 계절 모델은 데이터 지정 사항 탭의 시간 구간 분할창을 사용하여 정의된 주기성이 계절인 경우에만 사용할 수 있습니다. 계절 주기성은 다음과 같습니다. 주기적 기간, 연도, 분기, 월, 한 주의 요일, 하루의 시간, 하루의 분, 하루의 초. 다음 모델 유형을 사용할 수 있습니다.

- **단순**. 이 모델은 추세나 계절성이 없는 계열에 적합합니다. 유일하게 관련된 평활 모수는 수준입니다. 단순 지수평활은 자동 선형회귀 차수가 0, 차이 차수가 1, 이동 평균 차수가 1, 및 상수 없음인 ARIMA와 가장 비슷합니다.
- **Holt의 선형 추세**. 이 모델은 선형 추세가 있고 계절성이 없는 계열에 적합합니다. 관련 평활 모수는 수준 및 추세이며, 이 모델에서는 서로의 값으로 제한되지 않습니다. Holt 모델은 Brown 모델보다 일반적이지만 대형 계열의 추정값 계산에는 더 오래 걸립니다. Holt 지수평활은 자동 선형회귀 차수가 0, 차이 차수가 2, 이동 평균 차수가 2인 ARIMA와 가장 비슷합니다.
- **진폭감소 추세**. 이 모델은 점점 소멸되는 선형 추세가 있고 계절성이 없는 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 진폭감소 추세입니다. 진폭감소 지수평활은 자동 선형회귀 차수가 1, 차이 차수가 1, 이동 평균 차수가 2인 ARIMA와 가장 비슷합니다.
- **승법 추세**. 이 모델은 계열의 규모가 변경되는 추세가 있고 계절성이 없는 계열에 적합합니다. 관련된 평활 모수는 수준과 추세입니다. 승법 추세 지수평활은 ARIMA 모델과 다릅니다.

---

1) Gardner, E. S. 1985. Exponential smoothing: The state of the art. Journal of Forecasting, 4, 1-28.

- **Brown의 선형 추세**. 이 모델은 선형 추세가 있고 계절성이 없는 계열에 적합합니다. 관련 평활 모수는 수준 및 추세지만, 이 모델에서는 동일하다고 가정합니다. 따라서 Brown의 모형은 Holt 모형의 특별한 케이스입니다. Brown의 지수평활은 자동 선형회귀 차수가 0, 차이 차수가 2, 이동 평균 차수가 2이며, 이동 평균의 두 번째 차수에 대한 계수가 첫 번째 차수에 대한 계수의 절반과 같은 ARIMA와 가장 비슷합니다.
- **단순 계절**. 이 모델은 추세와 계절 효과가 없고 시간에 따라 일정한 계열에 적합합니다. 관련된 평활 모수는 수준과 계절입니다. 계절 지수평활은 자동 선형회귀 차수가 0, 차이 차수가 1, 계절 차이 차수가 1, 이동 평균의 경우 차수 1, p, p+1인 ARIMA와 가장 유사합니다. 여기서 p는 계절 간격에서 기간 수입니다. 월별 데이터의 경우 p = 12입니다.
- **Winters의 가법**. 이 모델은 선형 추세와 계절 효과가 있고 시간에 따라 일정한 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 계절입니다. Winters의 가법 지수평활은 자동 선형회귀 차수가 0, 차이 차수가 1, 계절 차이 차수가 1, 이동 평균의 경우 차수가 p+1인 ARIMA와 가장 유사합니다. 여기서 p는 계절 간격에서 기간 수입니다. 월별 데이터의 경우 p = 12입니다.
- **가법 계절이 있는 진폭감소 추세**. 이 모델은 감소되는 선형 추세와 계절 효과가 있고 시간에 따라 일정한 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 진폭감소 추세 및 계절입니다. 진폭감소 추세 및 가법 계절 지수평활은 ARIMA 모델과 다릅니다.
- **가법 계절이 있는 승법 추세**. 이 모델은 계열의 규모가 변경되는 추세와 계절 효과가 있고 시간에 따라 일정한 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 계절입니다. 승법 추세 및 가법 계절 지수평활은 ARIMA 모델과 다릅니다.
- **승법 계절성**. 이 모델은 선형 추세와 계절 효과가 없고 계절 규모가 변경되는 계열에 적합합니다. 관련된 평활 모수는 수준과 계절입니다. 승법 계절 지수평활은 ARIMA 모델과 다릅니다.
- **Winters의 승법**. 이 모델은 선형 추세와 계절 효과가 있고 계절 규모가 변경되는 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 계절입니다. Winters의 승법 지수평활은 ARIMA 모델과 다릅니다.
- **승법 계절이 포함된 진폭감소 추세**. 이 모델은 감소되는 선형 추세와 계절 효과가 있고 계절 규모가 변경되는 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 진폭감소 추세 및 계절입니다. 진폭감소 추세 및 승법 계절 지수평활은 ARIMA 모델과 다릅니다.
- **승법 계절이 있는 승법 추세**. 이 모델은 추세와 계절 효과가 있고 둘 다 계절 규모와 함께 변경되는 계열에 적합합니다. 관련된 평활 모수는 수준, 추세, 계절입니다. 승법 추세 및 승법 계절 지수평활은 ARIMA 모델과 다릅니다.

목표 변환. 모델링하기 전에 각 종속 변수에서 수행할 변환을 지정할 수 있습니다.

추가 정보는 계열 변환의 내용을 참조하십시오.

- **없음**. 변환을 수행하지 않습니다.
- **제곱근**. 제곱근 변환을 수행합니다.
- **자연로그**. 자연로그 변환을 수행합니다.

ARIMA

사용자 정의 ARIMA 모델의 구조를 지정하십시오.

ARIMA 차수. 눈금의 해당 셀에 모델의 다양한 ARIMA 구성요소 값을 입력합니다. 모든 값은 0 또는 음이 아닌 정수여야 합니다. 자기회귀 및 이동 평균 구성요소의 경우 값은 최대 차수를 나타냅니다. 더 낮은 양의 차수가 모두 모델에 포함됩니다. 예를 들어, 2를 지정하면 모델에 차수 2와 1이 포함됩니다. 계절 열의 셀은 활성 데이터 세트에 대해 주기성이 정의된 경우에만 사용할 수 있습니다.

- **자기회귀(p)**. 모델의 자기회귀 차수 수입니다. 자기회귀 차수는 계열의 이전 값 중 현재 값 예측에 사용될 값을 지정합니다. 예를 들어, 자기회귀 차수 2는 과거 2개 시간 주기의 계열 값을 현재 값 예측에 사용하도록 지정합니다.
- **차이(d)**. 모델을 추정하기 전 계열에 적용할 차이 차수를 지정합니다. 추세가 존재하며(추세가 있는 계열은 일반적으로 비정상이며 ARIMA 모델링은 정상성을 가정) 해당 효과 제거를 위해 사용되는 경우 차이가 필요합니다. 차이 차수는 계열 추세 수준에 해당합니다. 1차 차이는 선형 추세, 2차 차이는 2차 추세 등을 나타냅니다.
- **이동 평균(q)**. 모델의 이동 평균 차수 수입니다. 이동 평균 차수는 이전 값에 대한 계열 평균 편차를 사용하여 현재 값을 예측하는 방법을 지정합니다. 예를 들어, 이동 평균 차수 1과 2는 계열의 현재 값을 예측하는 경우 지난 2개 시간 주기 각각의 계열 평균값 편차를 고려하도록 지정합니다.

계절. 계절 자기회귀, 이동 평균, 차이 구성요소는 해당 비계절 구성요소와 동일한 역할을 합니다. 그러나 계절 차수의 경우 현재 계열 값이 한 개 이상의 계절 주기에 의해 구분된 이전 계열 값의 영향을 받습니다. 예를 들어, 월별 데이터(계절 주기 12)의 경우 계절 차수 1은 현재 계열 값이 현재 계열 이전의 계열 값 12 주기의 영향을 받는다는 것을 의미합니다. 월별 데이터의 경우 계절 차수 1은 비계절 차수 12를 지정하는 것과 동일합니다.

자동으로 이상값 발견. 이상값 자동 발견을 수행하려면 이 옵션을 선택하고 사용 가능한 하나 이상의 이상값 유형을 선택합니다.

발견할 이상값 유형. 발견할 이상값 유형을 선택합니다. 지원되는 유형은 다음과 같습니다.

- 가법(기본값)
- 수준 이동(기본값)
- 혁신적
- 일시적
- 계절 가법
- 국소적 추세
- 가법 수정

전이 함수 차수 및 변환. 변환을 지정하고 ARIMA 모델의 일부 또는 전체 입력 모델에 대해 전이 함수를 정의하려면 설정을 클릭하십시오. 전이 및 변환 세부사항을 입력할 수 있는 별도의 대화 상자가 표시됩니다.

모델에 상수 포함. 전체 평균 계열 값이 0인지 잘 모르는 경우 상수를 포함하는 것이 표준입니다. 차이를 적용하는 경우에는 상수를 제외하는 것이 좋습니다.

추가 세부사항

- 이상치 유형에 대한 자세한 정보는 이상값의 내용을 참조하십시오.
- 전송 및 변환 기능에 대한 자세한 정보는 전이 및 변환 함수의 내용을 참조하십시오.

ㄱ. 전이 및 변환 함수

전이 함수 차수 및 변환 대화 상자에서는 변환을 지정하고 ARIMA 모델의 일부 또는 전체 입력 모델에 대해 전이 함수를 정의할 수 있습니다.

목표 변환. 이 분할창에서는 모델링하기 전에 각 목표변수에 대해 수행할 변환을 지정할 수 있습니다.

- **없음**. 변환을 수행하지 않습니다.
- **제곱근**. 제곱근 변환을 수행합니다.
- **자연로그**. 자연로그 변환을 수행합니다.

추가 정보는 계열 변환의 내용을 참조하십시오.

후보 입력 전이 함수 및 변환. 전이 함수를 사용하여 목표 계열의 미래 값을 예측하는 데 입력 필드의 과거 값을 사용하는 방식을 지정할 수 있습니다. 분할창 왼쪽에 있는 목록에는 모든 입력 필드가 표시됩니다. 이 분할창의 나머지 정보는 선택한 입력 필드에 따라 다릅니다.

전이 함수 차수. 구조 눈금의 해당 셀에 전이 함수의 다양한 구성요소 값을 입력합니다. 모든 값은 0또는 음이 아닌 정수여야 합니다. 분자 및 분모 구성요소의 경우 값은 최대 차수를 나타냅니다. 더 낮은 양의 차수가 모두 모델에 포함됩니다. 또한 분자 구성요소에 대해 차수 0은 항상 포함됩니다. 예를 들어, 분자로 2를 지정하면 모델에 차수 2, 1, 0이 포함됩니다. 분모로 3을 지정하면 모델에 차수 3, 2, 1이 포함됩니다. 계절 열의 셀은 활성 데이터 세트에 대해 주기성이 정의된 경우에만 사용할 수 있습니다.

분자. 전이 함수의 분자 차수는 종속 계열의 현재 값을 예측하기 위해 사용되는 선택된 독립(예측자) 계열의 이전 값을 지정합니다. 예를 들어, 분자 차수 1은 각 종속 계열의 현재 값 예측에 과거 1개 시간 주기의 독립 계열 값과 독립 계열의 현재 값을 사용하도록 지정합니다.

분모. 전이 함수의 분모 차수는 종속 계열의 현재 값을 예측하기 위해 사용되는 선택된 독립(예측자) 계열의 이전 값에 대한 계열 평균의 편차를 지정합니다. 예를 들어 분모 차수 1은 각 종속 계열의 현재 값을 예측하는 경우 과거 1개 시간 주기의 독립 계열 평균 값 편차를 고려하도록 지정합니다.

차이. 모델을 추정하기 전 선택된 독립(예측자) 계열에 적용할 차이 차수를 지정합니다. 추세가 있으며 해당 효과 제거를 위해 사용되는 경우 차이가 필요합니다.
계절. 계절 분자, 분모 및 차이 구성요소는 해당 비계절 구성요소와 동일한 역할을 합니다. 그러나 계절 차수의 경우 현재 계열 값이 한 개 이상의 계절 주기에 의해 구분된 이전 계열 값의 영향을 받습니다. 예를 들어, 월별 데이터(계절 주기 12)의 경우 계절 차수 1은 현재 계열 값이 현재 계열 이전의 계열 값 12 주기의 영향을 받는다는 것을 의미합니다. 월별 데이터의 경우 계절 차수 1은 비계절 차수 12를 지정하는 것과 동일합니다.

지연. 지연을 설정하면 지정된 구간 수만큼 입력 필드의 영향력이 지연됩니다. 예를 들어, 지연이 5로 설정된 경우 시간 t의 입력 필드 값은 다섯 구간이 경과할 때까지 예측에 영향을 주지 않습니다(t + 5).

변환. 독립 변수 세트에서 전이 함수를 지정하면 해당 변수에서 수행할 선택적 변환도 포함됩니다.

- **없음**. 변환을 수행하지 않습니다.
- **제곱근**. 제곱근 변환을 수행합니다.
- **자연로그**. 자연로그 변환을 수행합니다.

④ 스트리밍 시계열 노드 - 모델 옵션

신뢰구간 너비(%). 모델 예측 및 잔차 자기상관에 대해 신뢰구간을 계산합니다. 100보다 작은 양수 값을 지정할 수 있습니다. 기본적으로, 95% 신뢰구간이 사용됩니다.

레코드를 미래로 확장 옵션은 추정 기간이 끝난 이후에 예측할 시간 구간 수를 설정합니다. 이 경우의 시간 구간은 데이터 지정 사항 탭에 지정한 분석의 시간 구간입니다. 예측이 요청되면 물론 목표가 아닌 입력 계열에서 자기회귀분석 모델이 자동으로 작성됩니다. 그런 다음, 이 모델을 사용하여 예측 기간에 해당 입력 계열의 값을 생성합니다. 이 설정에 최대 한계는 없습니다.

예측에 사용할 미래 값

- **입력의 미래 값 계산** 이 옵션을 선택하면 예측자의 예측 값, 잡음 예측, 분산 추정 및 미래 값이 자동으로 계산됩니다. 예측이 요청되면 물론 목표가 아닌 입력 계열에서 자기회귀분석 모델이 자동으로 작성됩니다. 그런 다음, 이 모델을 사용하여 예측 기간에 해당 입력 계열의 값을 생성합니다.

- **데이터에 추가할 값이 있는 필드 선택**. 예측자 필드를 사용하는 경우(역할이 입력으로 설정됨), 예측하려는 레코드(검증용 제외)마다 각 예측자에 대한 예측 기간 동안의 추정값을 지정할 수 있습니다. 값을 수동으로 지정하거나 목록에서 선택할 수 있습니다.
  - **필드**. 필드 선택기 단추를 클릭하고 예측자로 사용할 수 있는 필드를 선택하십시오. 여기서 선택한 필드가 모델링에 사용되거나 사용되지 않을 수 있습니다. 실제로 필드를 예측자로 사용하려면 다운스트림 모델링 노드에서 해당 필드를 선택해야 합니다. 이 대화 상자에서는 단지 각 노드에서 개별적으로 미래 값을 지정하지 않고도 여러 다운스트림 모델링 노드에서 값을 공유할 수 있도록 미래 값을 지정하는 편리한 장소를 제공합니다. 작성 옵션 탭의 선택사항을 통해 사용 가능한 필드의 목록을 제한할 수도 있습니다.
    삭제되었거나 작성 옵션 탭에서 선택한 사항을 업데이트했기 때문에 더 이상 스트림에서 사용할 수 없는 필드에 미래 값이 지정된 경우 필드는 빨간색으로 표시됩니다.

  - **값**. 필드마다, 함수 목록에서 선택하거나 **지정**을 클릭하여 수동으로 값을 입력하거나 또는 사전 정의된 값 목록에서 선택할 수 있습니다. 예측자 필드가 제어 가능한 항목 또는 미리 알 수 있는 항목과 관련이 있으면 값을 수동으로 입력해야 합니다. 예를 들어, 룸 예약 수를 기반으로 호텔의 다음 달 수입을 예측하는 경우, 해당 기간 동안의 실제 예약 수를 지정할 수 있습니다. 반면, 예측자 필드가 제어 불가능한 항목(예: 주식 가격)과 관련이 있으면 most recent value 또는 mean of recent points와 같은 함수를 사용할 수 있습니다.
    사용 가능한 함수는 필드의 측정 수준에 따라 다릅니다.

표 1. 측정 수준에 사용 가능한 함수

측정수준	함수
연속형 또는 명목 필드	공백 최신 점의 평균 최신 값 지정
플래그 필드	공백 최신 값 TRUE FALSE 지정

최신 점의 평균은 마지막 세 개의 데이터 점 평균에서 미래 값을 계산합니다.

최신 값은 미래 값을 최신 데이터 점의 값으로 설정합니다.

True/False는 플래그 필드의 미래 값을 지정된 대로 True나 False로 설정합니다.

지정은 수동으로 미래 값을 지정하거나 사전 정의된 목록에서 선택하여 대화 상자를 엽니다.

점수에 사용 가능

모델 너깃에 대한 대화 상자에 표시될 스코어링 옵션에 대한 기본값을 여기서 설정할 수 있습니다.

- **신뢰구간 상한 및 하한 계산**. 선택할 경우 이 옵션은 각 목표 필드에서 하한 및 하한 신뢰구간에 대한 새 필드(기본 접두문자: $TSLCI- 및 $TSUCI-)를 작성합니다.
- **잡음 잔차 계산**. 이 옵션을 선택한 경우 각 목표 필드에서 모형 잔차와 이러한 값의 총계에 대한 새 필드(기본 접두문자가 $TSResidual-임)를 작성합니다.

모델 설정

출력에 표시할 최대 모델 수. 출력에 포함될 최대 모델 수를 지정하십시오. 작성된 모델 수가 이 임계값을 초과할 경우 모델이 출력에 표시되지는 않지만 스코어링에 계속 사용할 수 있습니다. 기본값은 10입니다. 더 큰 모델 수를 표시하면 성능이나 안정성이 저하될 수 있습니다.

(12) SMOTE 노드

SMOTE(Synthetic Minority Over-sampling Technique) 노드는 불균형 데이터 세트를 처리하기 위해 초과 표본추출 알고리즘을 제공합니다. 또한 데이터 균형을 조정하기 위한 고급 방법을 제공합니다. SMOTE 프로세스 노드는 Python으로 구현되며, imbalanced-learn© Python 라이브러리가 필요합니다. imbalanced-learn 라이브러리에 대한 자세한 내용은 https://imbalanced-learn.org/stable/[2]을 참조하십시오.

노드 팔레트의 Python 탭은 SMOTE 노드와 다른 Python 노드로 구성됩니다.

① SMOTE 노드 설정

SMOTE 노드의 **설정** 탭에서 다음 설정을 정의하십시오.

목표 설정

대상 필드. 대상 필드를 선택하십시오. 플래그, 명목, 순서, 및 이산형 측정 유형이 모두 지원됩니다. 파티션 섹션에서 **파티션된 데이터 사용** 옵션을 선택한 경우 학습 데이터만 초과 표본추출됩니다.

---

2) Lemaître, Nogueira, Aridas. "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning." Journal of Machine Learning Research, vol. 18, no. 17, 2017, pp. 1-5. (http://jmlr.org/papers/v18/16-365.html)

초과 표본 비율

초과 표본 비율을 자동으로 선택하려면 **자동**을 선택하고, 사용자 정의 비율 값을 설정하려면 **비율 설정(소수/다수)**을 선택하십시오. 비율은 다수 클래스의 표본 수 대비 소수 클래스의 표본 수입니다. 값은 **0**보다 크고 **1**보다 작거나 같아야 합니다.

난수 시드

난수 시드 설정. 난수 생성기에서 사용한 시드를 생성하려면 이 옵션을 선택하고 **생성**을 클릭하십시오.

방법

알고리즘 종류 사용할 SMOTE 알고리즘의 유형을 선택하십시오.

표본 규칙

K 이웃 합성 표본을 생성하는 데 사용할 최근접 이웃의 수를 지정하십시오.

M 이웃 소수 표본이 위험한지 판별하는 데 사용할 최근접 이웃의 수를 지정하십시오. 이 옵션은 **Borderline1** 또는 **Borderline2** SMOTE 알고리즘 유형을 선택한 경우에만 사용됩니다.

파티션

파티션된 데이터 사용. 학습 데이터를 초과 표본추출하려는 경우에만 이 옵션을 선택하십시오.

SMOTE 노드에는 imbalanced-learn© Python 라이브러리가 필요합니다. 다음 표는 SPSS® Modeler SMOTE 노드 대화 상자의 설정과 Python 알고리즘 간의 관계를 보여줍니다.

표 1. Python 라이브러리 모수에 맵핑되는 노드 특성

SPSS Modeler 설정	스크립트 이름(특성 이름)	PythonAPI 매개변수이름
초과 표본 비율(숫자 입력 제어)	sample_ratio_value	ratio
난수 시드	random_seed	random_state
K_Neighbours	k_neighbours	k
M_Neighbours	m_neighbours	m
알고리즘 종류	algorithm_kind	kind

(13) 확장 변환 노드

확장 변환 노드를 사용하면 IBM® SPSS® Modeler 스트림에서 데이터를 가져와서 R 스크립팅 또는 Python for Spark 스크립팅을 사용하여 변환을 데이터에 적용할 수 있습니다. 데이터가 수정되면 추가 처리, 모델 작성 및 모델 스코어링을 위해 스트림에 리턴됩니다. 확장 변환 노드를 사용하면 R 또는 Python for Spark로 작성된 알고리즘을 사용하여 데이터를 변환할 수 있으며 사용자가 특정 문제점에 적합한 데이터 변환 방법을 개발할 수 있습니다.

이 노드를 R과 함께 사용하려면 IBM SPSS Modeler - Essentials for R을 설치해야 합니다. *IBM SPSS Modeler - Essentials for R: 설치 지시사항*에서 설치 지시사항 및 호환성 정보를 참조하십시오. 또한 R의 호환 가능한 버전이 컴퓨터에 설치되어 있어야 합니다.

① 확장 변환 노드 - 명령문 탭

구문 유형(**R** 또는 **Python for Spark**)을 선택하십시오. 자세한 정보는 다음 섹션을 참조하십시오. 명령문이 준비되면 **실행**을 클릭하여 확장 변환 노드를 실행할 수 있습니다.

R 구문

R 구문. 데이터 분석을 위해 R 스크립트 구문을 이 필드에 입력, 붙여넣기 또는 사용자 정의할 수 있습니다.

플래그 필드 변환. 플래그 필드를 처리하는 방법을 지정합니다. **문자열에서 요인으로**, **정수 및 실수에서 double로** 및 **논리 값(True, False)**이라는 두 가지 옵션이 있습니다. **논리 값(True, False)**을 선택하면 플래그 필드의 원래 값이 손실됩니다. 예를 들어, 필드에 남성 및 여성 값이 있는 경우, 해당 값이 True 및 False로 변경됩니다.

결측값을 R '사용할 수 없음' 값(NA)으로 변환. 선택하면 모든 결측값이 R NA 값으로 변환됩니다. NA 값은 결측값을 식별하기 위해 R에서 사용됩니다. 사용하는 일부 R 함수에는 데이터에 NA가 포함된 경우에 함수가 작동하는 방식을 제어하는 데 사용되는 인수가 있습니다. 예를 들어, 함수에서 NA를 포함하는 레코드를 자동으로 제외하도록 선택할 수 있습니다. 이 옵션을 선택하지 않으면 모든 결측값이 변경되지 않은 상태로 R에 전달되고 R 스크립트가 실행될 때 오류가 발생할 수 있습니다.

날짜/시간 필드를 시간대의 특수 제어가 있는 R 클래스로 변환. 이 옵션을 선택하면 날짜 또는 날짜/시간 형식의 변수가 R 날짜/시간 개체로 변환됩니다.
다음 옵션 중 하나를 선택해야 합니다.

- **R POSIXct**. 날짜 또는 날짜/시간 형식의 변수가 R POSIXct 개체로 변환됩니다.
- **R POSIXlt (목록)**. 날짜 또는 날짜/시간 형식의 변수가 R POSIXlt 개체로 변환됩니다.

ⓘ **참고**: POSIX 형식은 고급 옵션입니다. R 스크립트에서 날짜/시간 필드가 해당 형식이 필요한 방식으로 처리되도록 지정된 경우에만 이 옵션을 사용하십시오. POSIX 형식은 시간 형식이 있는 변수에 적용되지 않습니다.

Python 구문(S)

Python 구문. 이 필드에 데이터 분석을 위한 Python 스크립팅 구문을 입력하거나 붙여넣거나 사용자 정의할 수 있습니다. Python for Spark에 대한 자세한 정보는 Python for Spark 및 Python for Spark로 스크립팅의 내용을 참조하십시오.

② 확장 변환 노드 - 콘솔 출력 탭

콘솔 출력 탭에는 명령문 탭에서 R 스크립트 또는 Python for Spark 스크립트가 실행될 때 수신된 모든 출력이 포함됩니다. 예를 들어, R 스크립트를 사용하는 경우, **명령문** 탭의 **R 명령문** 필드의 R 스크립트가 실행될 때 R 콘솔에서 수신된 출력을 표시합니다. 이 출력에는 R 또는 Python 스크립트가 실행될 때 생성되는 R 또는 Python 오류 메시지 또는 경고가 포함됩니다. 출력은 주로 스크립트를 디버그하는 데 사용될 수 있습니다. **콘솔 출력** 탭에는 **R 명령문** 또는 **Python 명령문** 필드의 스크립트도 포함됩니다.

확장 변환 스크립트가 실행될 때마다 R 콘솔 또는 Python for Spark에서 수신된 출력이 **콘솔 출력** 탭의 내용을 덮어씁니다. 출력은 편집할 수 없습니다.

(14) Space-Time-Box 노드

STB(Space-Time-Box)는 Geohash 공간 위치의 확장입니다. 보다 상세하게 설명하자면 STB는 공간 및 시간을 정기적으로 모양으로 표시하는 알파뉴메릭 문자열입니다.

예를 들어, STB **dr5ru7|2013-01-01 00:00:00|2013-01-01 00:15:00**은 다음과 같은 세 부분으로 구성됩니다.

- Geohash **dr5ru7**
- 시작 시간소인 **2013-01-01 00:00:00**
- 종료 시간소인 **2013-01-01 00:15:00**

예를 들어, 공간 및 시간 정보를 사용하면 두 개의 엔티티가 동시에 동일한 공간에 시각적으로 표시되므로 동일함에 대한 신뢰도를 높일 수 있습니다. 또는 두 엔티티가 공간 및 시간의 근접성으로 인해 관련됨을 표시함으로써 관계 식별의 정확도를 높일 수 있습니다.

사용자의 요구 사항에 맞게 **개별 레코드** 또는 **행아웃(Hangout)** 모드를 선택할 수 있습니다. 두 모드 모두 다음과 같은 동일한 기본 세부사항이 필요합니다.

위도 필드. WGS84 좌표계에서 위도를 식별하는 필드를 선택하십시오.

경도 필드. WGS84 좌표계에서 경도를 식별하는 필드를 선택하십시오.

시간소인 필드. 시간 또는 날짜를 식별하는 필드를 선택하십시오.

개별 레코드 옵션

지정된 시간의 위치를 식별하기 위해 레코드에 추가 필드를 추가하려면 이 모드를 사용하십시오.

파생. 새 필드를 파생시킬 공간 및 시간의 밀도를 하나 이상 선택하십시오. 자세한 정보는 Space-Time-Box 밀도 정의의 내용을 참조하십시오.

필드 이름 확장. 새 필드 이름에 추가할 확장자를 입력하십시오. 이 확장자를 **접미어** 또는 **접두어**로 추가할 수 있습니다.

행아웃 옵션

행아웃은 엔티티가 지속적으로 또는 반복적으로 발견될 수 있는 위치 및/또는 시간으로 생각할 수 있습니다. 예를 들어, 정기적인 운송 작업을 하는 차량을 식별하고 규범으로부터의 일탈을 식별하기 위해 사용할 수 있습니다.

행아웃 감지기는 엔티티의 움직임을 모니터링하고 엔티티가 영역에서 "행아웃"으로 관찰되는 조건에 플래그를 지정합니다. 행아웃 감지기는 플래그가 지정된 각 행아웃을 하나 이상의 STB에 자동으로 지정하고 인메모리 엔티티 및 이벤트 추적을 사용하여 가장 효율적인 행아웃을 감지합니다.

STB 밀도. 새 필드를 파생시킬 공간 및 시간의 밀도를 선택하십시오. 예를 들어, **STB_GH4_10MINS** 값은 크기가 약 20km x 20km이며 시간이 10분인 4문자 Geohash 상자 창에 해당됩니다. 자세한 정보는 Space-Time-Box 밀도 정의의 내용을 참조하십시오.

엔티티 ID 필드. 행아웃 식별자로 사용할 엔티티를 선택하십시오. 이 ID 필드는 이벤트를 식별합니다.

최소 이벤트 수. 이벤트는 데이터의 행입니다. 행아웃으로 간주하기 위한 엔티티에 대한 이벤트 최소 발생 수를 선택하십시오. 행아웃은 **최소 체류 시간** 필드 기준도 충족해야 합니다.

최소 체류 시간. 엔티티가 동일한 위치에 체류해야 하는 최소 기간을 지정합니다. 예를 들어, 차량이 신호등 때문에 대기하는 것이 행아웃으로 간주되는 경우를 제외할 수 있습니다. 행아웃은 앞에서 설명한 **최소 이벤트 수** 필드 기준도 충족해야 합니다.

다음은 행아웃을 규정하는 것에 대한 더 자세한 정보입니다.

e_1, ..., e_n은 지속 기간(t_1, t_n) 동안 지정된 이벤트 ID에서 수신된 시간 순으로 정렬된 모든 이벤트를 표시합니다. 이러한 이벤트는 다음 경우에 행아웃으로 규정됩니다.
- n >= 최소 이벤트 수
- tn - t1 >= 최소 체류 시간
- 모든 이벤트 e_1, ..., e_n이 동일한 STB에서 발생

행아웃 범위가 STB 경계를 포함하도록 허용. 이 옵션이 선택되면 행아웃의 정의가 덜 엄격해집니다. 예를 들어, 둘 이상의 Space-Time-Box에 행아웃하는 엔티티를 포함할 수 있습니다. STB가 전체 시간으로 정의된 경우, 이 옵션을 선택하면 한 시간이 자정 전 30분과 자정 후 30분으로 구성되더라도 유효한 한 시간 동안 행아웃하는 엔티티로 인식합니다. 이 옵션을 선택하지 않으면 100%의 행아웃 시간이 단일 Space-Time-Box 내에 있어야 합니다.

시간상자 규정의 최소 이벤트 비율(%). 행아웃 범위가 STB 경계를 포함하도록 허용이 선택된 경우에만 사용 가능합니다. 한 STB에서 보고되는 행아웃이 실제로 다른 STB에 겹칠 수 있는 정도를 제어하는 데 사용됩니다. 행아웃을 식별하기 위해 단일 STB 내에서 발생해야 하는 최소 이벤트 비율을 선택하십시오. 25%로 설정한 상태에서 이벤트의 비율이 26%이면 이는 행아웃으로 규정됩니다.

예를 들어, 4바이트 Geohash 공간 상자 및 10분 시간 상자(STB_NAME = STB_GH4_10MINS) 내에 두 개 이상의 이벤트(최소 이벤트 수 = 2) 및 2분 이상의 연속 체류 시간을 요구하도록 행아웃 감지기를 구성한 경우를 가정해 보십시오. 행아웃이 발견될 때 4:57pm 및 5:07pm 사이의 10분의 시간 범위 내에서 4:58pm, 5:01pm 및 5:03pm에 세 개의 규정 이벤트가 발생하는 동안 엔티티가 동일한 4바이트 Geohash 공간 상자에 체류한다고 말할 수 있습니다. 규정 시간 상자 퍼센트 값은 STB가 행아웃으로 간주되기 위한 퍼센트 값을 다음과 같이 지정합니다.

- **100%**. 행아웃이 5:00 - 5:10pm 시간 상자에서는 보고되고 4:50 - 5:00pm 시간 상자에서는 보고되지 않습니다(5:01pm 및 5:03pm의 이벤트는 행아웃을 규정하는 데 필요한 모든 조건을 충족하며 이러한 이벤트의 100%가 5:00 - 5:10 시간 상자에서 발생했습니다).
- **50%**. 두 시간 상자 모두에서 행아웃이 보고됩니다(5:01pm 및 5:03pm의 이벤트는 행아웃을 규정하는 데 필요한 모든 조건을 충족하며 이러한 이벤트 중 50% 이상이 4:50 - 5:00 시간 상자에서 발생하고 이러한 이벤트 중 50% 이상이 5:00 - 5:10 시간 상자에서 발생했습니다).
- **0%**. 두 시간 상자 모두에서 행아웃이 보고됩니다.

0%를 지정하면 행아웃 보고서에 규정 기간이 속한 모든 시간 상자를 나타내는 STB가 포함됩니다. 규정 기간은 STB 내의 시간 상자의 해당 기간 이하여야 합니다. 즉, 10분 STB가 20분의 규정 기간과 함께 구성될 수 없습니다.

행아웃은 규정 조건이 충족되자 마자 보고되며 STB당 두 번 이상 보고되지 않습니다. 세 개의 이벤트가 행아웃을 규정하고 총 열 개의 이벤트가 모두 동일한 STB 내의 규정 기간 동안 발생했다고 가정해 보십시오. 이 경우, 세 번째 규정 이벤트가 발생할 때 행아웃이 보고됩니다. 추가 일곱 개의 이벤트는 행아웃 보고서를 트리거하지 않습니다.

참고:
- 행아웃 감지기의 인메모리 이벤트 데이터는 프로세스에 걸쳐 공유되지 않습니다. 따라서 특정 엔티티가 특정 행아웃 감지기 노드와 연관관계를 갖습니다. 즉, 엔티티에 대해 수신되는 이동 데이터는 항상 해당 엔티티를 추적하는 행아웃 감지기 노드(일반적으로 실행 전체에서 동일 노드임)에 일관적으로 전달되어야 합니다.
- 행아웃 감지기의 인메모리 이벤트 데이터는 일시적입니다. 행아웃 감지기가 종료되고 다시 시작되면 모든 진행 중인 행아웃이 손실됩니다. 즉, 프로세스를 중지하고 다시 시작하면 시스템이 실제 행아웃을 손실하게 됩니다. 잠재적인 해결 방법은 히스토리 이동 데이터의 일부를 재생하는 것입니다. 예를 들어, 48시간 뒤로 이동하여 재시작된 임의의 노드에 적용 가능한 이동 레코드를 재생하는 것입니다.
- 시간 순서로 행아웃 감지기에 데이터를 공급해야 합니다.

① Space-Time-Box 밀도 정의

각각 포함할 실제 영역 및 경과 시간을 지정하여 STB(Space-Time-Box)의 크기(밀도)를 지정할 수 있습니다.

지역 밀도. 각 STB에 포함할 영역의 크기를 선택하십시오.

시간 구간. 각 STB에 포함할 시간 수를 선택하십시오.

필드 이름. STB라는 접두문자가 사용되고 앞의 두 필드의 선택사항을 기준으로 자동으로 완료됩니다.

(15) 스트리밍 TCM 노드

스트리밍 TCM 노드는 임시 원인 모델을 한 번에 작성하고 스코어링하는 데 사용할 수 있습니다.

시간 인과 모델링에 대한 자세한 정보는 시간 인과 모델의 내용을 참조하십시오.

① 스트리밍 TCM 노드 - 시계열 옵션

필드 탭에서, 모델 시스템에서 포함할 계열을 지정하려면 **시계열** 설정을 사용하십시오.

데이터에 적용되는 데이터 구조에 맞는 옵션을 선택하십시오. 다차원 데이터의 경우, 차원 필드를 지정하기 위해 **차원 선택**을 클릭하십시오. 차원 필드의 지정된 순서는 모두 연속 대화 상자 및 출력에 차원이 나타나는 순서를 정의합니다. 차원 필드를 다시 정렬하려면 차원 선택 하위 대화 상자에서 위로 및 아래로 화살표 단추를 사용하십시오.

열 기반 데이터에 대해 *계열* 용어의 의미는 *필드* 용어 의미와 같습니다. 다차원 데이터의 경우, 시계열을 포함하는 필드는 메트릭 필드로 언급됩니다. 다차원 데이터에 대해 시계열은 차원 필드 각각에 대한 메트릭 필드 및 값으로 정의됩니다. 다음 고려사항은 열 기반 및 다차원 데이터에 적용됩니다.

- 후보 입력으로, 또는 목표 및 입력 둘 다로 지정되는 계열은 각 목표에 대해 모델에서 포함을 위해 고려됩니다. 각 목표에 대한 모델에는 항상 목표 자체의 시차 값이 포함됩니다.
- 강제 입력으로 지정되는 계열은 항상 각 목표에 대해 모델에 포함됩니다.
- 최소 하나의 계열을 목표로 또는 목표 및 입력 둘 다로 지정해야 합니다.
- **사전 정의된 역할 사용**이 선택되면, 입력 역할을 가지고 있는 필드가 후보 입력으로 설정됩니다. 사전 정의된 어떤 규칙도 강제 입력에 맵핑하지 않습니다. 자세한 정보는 역할 주제를 참조하십시오.

다차원 데이터

다차원 데이터의 경우, 눈금에서 메트릭 필드 및 연관된 역할을 지정합니다. 눈금의 각 행은 단일 메트릭 및 역할을 지정합니다. 기본적으로, 모델 시스템에는 눈금의 각 행에 대한 차원 필드의 모든 조합에 대한 계열이 포함됩니다. 예를 들어, *region* 및 *brand*에 대한 차원이 있는 경우, 기본적으로 목표로 메트릭 *sales*를 지정하면, 이는 각각의 *region* 및 *brand* 조합에 대해 별도의 sales 목표 계열이 있음을 의미합니다.

눈금의 각 행에 대해, 차원의 생략 기호 단추를 클릭하여 차원 필드에 대한 값 세트를 사용자 정의할 수 있습니다. 이 동작은 차원 값 선택 하위 대화 상자를 엽니다. 또한 눈금 행을 추가, 삭제 또는 복사할 수도 있습니다.

계열 개수 열은 현재 연관 메트릭에 대해 지정된 차원 값 세트 수를 표시합니다. 표시된 값은 실제 계열 수(세트당 하나의 계열)보다 클 수 있습니다. 이 조건은 지정된 차원 값 조합 중 일부가 연관 메트릭에 의해 포함된 계열에 해당되지 않는 경우에 발생합니다.

가. 스트리밍 TCM 노드 - 차원 값 선택

다차원 데이터의 경우, 특정 역할이 있는 특정 메트릭 필드에 적용되는 차원 값을 지정하여 분석을 사용자 정의할 수 있습니다. 예를 들어, *sales*가 메트릭 필드이고 *channel*이 'retail' 및 'web' 값을 가지고 있는 차원인 경우, 'web' 판매가 입력이고 'retail' 판매가 목표임을 지정할 수 있습니다. 또한 분석에 사용되는 모든 메트릭 필드에 적용되는 차원 서브세트를 지정할 수도 있습니다. 예를 들어, *region*이 지리적 지역을 나타내는 차원 필드인 경우 분석을 특정 지역으로 제한할 수 있습니다.

모든 값

현재 차원 필드의 모든 값이 포함됨을 지정합니다. 이 옵션은 기본값입니다.

포함하거나 제외할 값 선택

현재 차원 필드의 값 세트를 지정하려면 이 옵션을 사용하십시오. **모드**에 대해 **포함**이 선택되는 경우, **선택된 값** 목록에 지정되는 값만 포함됩니다. **모드**에 대해 **제외**가 선택되는 경우, **선택된 값** 목록에 지정된 값이 아닌 다른 모든 값이 포함됩니다.

선택할 값 세트를 필터링할 수 있습니다. 필터 조건에 충족하는 값은 **매치됨** 탭에 나타나고, 필터 조건과 일치하지 않는 값은 **선택되지 않은 값** 목록의 **매치하지 않음** 탭에 나타납니다. **모두** 탭은 필터 조건과 관계없이 선택되지 않은 모든 값을 나열합니다.

- 필터를 지정할 때 와일드카드 문자를 표시하기 위해 별표(*)를 사용할 수 있습니다.
- 현재 필터를 지우려면, 표시된 값 필터링 대화 상자에 검색어로 비어 있는 값을 지정하십시오.

② 스트리밍 TCM 노드 - 관측 옵션

필드 탭에서, 관측값을 정의하는 필드를 지정하려면 **관측값** 설정을 사용하십시오.

날짜/시간에 의해 정의되는 관측값

관측값이 날짜, 시간 또는 시간소인 필드에 의해 정의됨을 지정할 수 있습니다. 관측값을 정의하는 필드 외에, 관측값을 설명하는 적절한 시간 구간을 선택하십시오. 지정된 시간 구간에 따라, 관측값(증분) 사이의 구간이나 주당 일 수와 같은 다른 설정을 지정할 수도 있습니다. 다음 고려사항은 시간 간격에 적용됩니다.

- 관측값이 시간에서 비정규적으로 간격이 있는 경우(판매 순서가 처리되는 시간과 같이), **비정규** 값을 사용하십시오. **비정규**가 선택될 때, 데이터 지정 사항 탭의 **시간 간격** 설정에서 분석에 사용되는 시간 구간을 지정해야 합니다.
- 관측값이 날짜와 시간을 나타내고 시간 구간이 시, 분 또는 초인 경우 **하루 중 시간(시)**, **하루 중 시간(분)** 또는 **하루 중 시간(초)**을 사용하십시오. 관측값이 날짜에 대한 참조 없이 시간(기간)을 나타내고 시간 구간이 시, 분 또는 초일 경우, **시(비주기적)**, **분(비주기적)** 또는 초**(비주기적)**를 사용하십시오.
- 선택된 시간 간격을 기초로, 프로시저는 결측 관측값을 발견할 수 있습니다. 프로시저에서는 모든 관측값이 시간에서 동일하게 간격을 두고 결측 관측값이 없다고 가정하므로, 결측

관측값을 발견해야 합니다. 예를 들어, 시간 구간이 일(Days)이고 날짜 2014-10-27 뒤에 2014-10-29가 있는 경우, 2014-10-28에 대해 결측 관측값이 있습니다. 값은 결측 관측값에 대해 대체됩니다. 결측값 처리에 대한 설정은 데이터 지정 사항 탭으로부터 지정할 수 있습니다.

- 지정된 시간 구간은 프로시저가 함께 통합해야 하는 동일한 시간 구간의 여러 관측값을 발견하고 관측값에 동일하게 간격이 있도록 월의 첫 번째와 같은 구간 경계에 관측값을 맞출 수 있도록 합니다. 예를 들어, 시간 구간이 월일 경우, 동일 월에 있는 여러 날짜가 함께 통합됩니다. 이 유형의 통합을 그룹화라고 합니다. 기본적으로, 관측값은 그룹화될 때 합산됩니다. 데이터 지정 사항 탭의 **통합 및 분포** 설정에서, 그룹화에 다른 방법(예: 관측값의 평균)을 지정할 수 있습니다.
- 일부 시간 구간의 경우, 추가 설정은 동일하게 간격이 있는 정규 구간에서 중단을 설정할 수 있습니다. 예를 들어, 시간 구간이 일(Days)이지만 평일만 유효한 경우, 주에 5일이 있고 주는 월요일에 시작함을 지정할 수 있습니다.

관측값이 주기 또는 순환 주기로 정의됨

관측값은 임의의 순환 수준 수까지, 주기 또는 반복 주기 순환을 나타내는 하나 이상의 정수 필드로 정의할 수 있습니다. 이 구조에서, 표준 시간 구간 중 하나에 맞지 않은 관측값 계열을 설정할 수 있습니다. 예를 들어, 10개월만 있는 회계연도는 연도를 나타내는 순환 필드와, 월을 나타내는 주기 필드로 설명할 수 있습니다. 여기서 하나의 주기 길이는 10입니다.

순환 주기를 지정하는 필드는 주기적 수준의 계층 구조를 정의합니다. 가장 낮은 수준은 **주기** 필드에 의해 정의됩니다. 다음 최상위 수준은 수준이 1인 순환 필드에 의해 지정되고, 그 다음은 수준 2의 순환 필드로 지정되며 뒤로도 마찬가지입니다. 가장 높은 수준을 제외하고, 각 수준의 필드 값은 다음 최상위 수준에 관하여 주기적이어야 합니다. 최상위 수준의 값은 주기적이 될 수 없습니다. 예를 들어, 10달 회계연도의 경우 월은 연도 내에서 주기적이며 연도는 주기적이 아닙니다.

- 특정 수준에 있는 순환의 길이는 다음으로 가장 낮은 수준의 주기성입니다. 회계연도 예의 경우, 단 하나 순환 수준이 있고 순환 길이는 10입니다. 다음으로 가장 낮은 수준이 월을 나타내고 지정된 회계 연도에 10달이 있기 때문입니다.
- 1에서 시작하지 않은 주기적 필드의 시작 값을 지정하십시오. 이 설정은 결측값을 발견하는 데 필요합니다. 예를 들어, 주기적 필드는 2에서 시작하지만 시작 값은 1로 지정되는 경우, 프로시저는 해당 필드의 각 순환에 있는 첫 번째 주기에 대해 결측값이 있다고 가정합니다.

③ 스트리밍 TCM 노드 - 시간 구간 옵션

분석에 사용되는 시간 구간은 관측값의 시간 구간과 다를 수 있습니다. 예를 들어, 관측값의 시간 구간이 일(Days)일 경우, 분석의 시간 구간으로는 월을 선택할 수 있습니다. 데이터는 모델이 작성되기 전에 매일 데이터에서 매월 데이터까지 통합됩니다. 또한 데이터를 장기 시간 구간에서 단기 시간 구간으로 분포할 것을 선택할 수도 있습니다. 예를 들어, 관측값이 분기별인 경우, 데이터를 분기별에서 월별 데이터로 분포할 수 있습니다.

분석이 행해지는 시간 구간에 대해 사용 가능한 선택은 해당 관측값 정의 방법과 관측값의 시간 구간에 따라 다릅니다. 특히, 관측값이 순환 주기로 정의될 경우, 통합만 지원됩니다. 그러한 경우, 분석의 시간 구간은 관측값의 시간 구간보다 크거나 같아야 합니다.

분석 시간 구간은 데이터 지정 사항 탭의 **시간 구간** 설정에서 지정됩니다. 데이터가 통합되거나 분포되는 방법은 데이터 지정 사항 탭의 **통합 및 분포** 설정에서 지정됩니다.

④ 스트리밍 TCM 노드 - 통합 및 분포 옵션

집계 함수

분석에 사용되는 시간 구간이 관측에 사용되는 시간 구간보다 길 경우, 입력 데이터는 통합됩니다. 예를 들어, 관측값의 시간 구간이 일(Days)이고 분석의 시간 구간이 월일 경우 통합이 수행됩니다. mean, sum, mode, min 또는 max 집계 함수를 사용할 수 있습니다.

분포 함수

분석에 사용되는 시간 구간이 관측의 시간 구간보다 짧을 경우, 입력 데이터는 분포됩니다. 예를 들어, 관측값의 시간 구간이 분기이고 분석의 시간 구간이 월일 경우 분포가 수행됩니다. mean 또는 sum 분포 함수를 사용할 수 있습니다.

그룹화 함수

그룹화는 관측값이 날짜/시간에 의해 정의되고 여러 관측값이 동일 시간 구간에 발생하는 경우에 적용됩니다. 예를 들어, 관측값의 시간 간격이 월일 경우, 동일 월에 있는 여러 날짜가 그룹화되어 날짜가 발생하는 월과 연관됩니다. mean, sum, mode, min 또는 max와 같은 그룹화 함수를 사용할 수 있습니다. 다음 그룹화는 항상 관측값이 날짜/시간에 의해 정의되고 관측값의 시간 간격이 비정규로 지정된 경우에 수행됩니다.

> ⓘ **참고**: 그룹화가 통합 양식이어도, 그룹화는 결측값 처리 이전에 수행됩니다(정상 통합은 결측값 처리 이후에 수행됩니다). 관측값의 시간 구간이 비정규로 지정되는 경우, 통합은 그룹화 함수로만 수행됩니다.

교차-일 관측값을 이전 일로 통합

1일 경계를 교차하는 시간을 사용하는 관측값이 전날의 값에 통합되는지 여부를 지정합니다. 예를 들어, 20:00시에 시작하는 8시간 노동의 시간별 관측값의 경우, 이 설정은 00:00 및 04:00 사이의 관측값이 전날 통합 결과에 포함되는지 여부를 지정합니다. 이 설정은 관측값의 시간 구간이 하루 중 시간(시), 하루 중 시간(분) 또는 하루 중 시간(초)이고 분석의 시간 구간이 일(Days)인 경우에만 적용됩니다.

지정된 필드에 대한 사용자 정의 설정

필드 기준으로 필드에 통합, 분포 및 그룹화 함수를 지정할 수 있습니다. 이 설정은 통합, 분포 및 그룹화 함수에 대한 기본 설정을 대체합니다.

⑤ 스트리밍 TCM 노드 - 결측값 옵션

입력 데이터의 결측값은 대치된 값으로 바뀝니다. 다음 방법으로 바꿀 수 있습니다.

선형 보간법

선형 보간법을 사용하여 결측값을 바꿉니다. 결측값 이전의 마지막 유효한 값과 결측값 이후의 첫 번째 유효한 값이 보간법에 사용됩니다. 계열에서 첫 번째 또는 마지막 관측값에 결측값이 있는 경우, 계열의 시작 또는 종료에서 두 개의 가장 근접한 비결측 값이 사용됩니다.

계열 평균

결측값을 전체 계열에 대한 평균으로 바꿉니다.

근접한 값들의 평균

결측값을 유효한 근접 값의 평균으로 바꿉니다. 근접한 값들의 계산너비는 평균을 계산하는데 사용되는 결측값 전후의 유효값 수입니다.

근접한 값들의 중앙값

결측값을 근접한 유효한 값의 중앙값으로 바꿉니다. 근접한 값들의 계산너비는 평균을 계산하는데 사용되는 결측값 전후의 유효값 수입니다.

선형 추세

이 옵션은 단순 선형 회귀 모형을 적합시키기 위해 계열에서 모든 비결측 관측값을 사용합니다. 이 모델은 결측값을 대치하기 위해 사용됩니다.

기타 설정:

결측값의 최대 퍼센트(%)

어떤 계열에 대해서도 허용되는 최대 결측값 퍼센트를 지정합니다. 지정된 최대값보다 많은 결측값이 있는 계열은 분석에서 제외됩니다.

⑥ 스트리밍 TCM 노드 - 일반 데이터 옵션

차원 필드당 최대 고유 값 수

이 설정은 다차원 데이터에 적용되며 하나의 차원 필드에 대해 허용되는 최대 고유 값 수를 지정합니다. 기본적으로, 이 한계는 10000으로 설정되지만, 임의로 큰 숫자로 증가될 수 있습니다.

⑦ 스트리밍 TCM 노드 - 일반 작성 옵션

신뢰구간 너비(%)

이 설정은 예측 및 모델 모수 둘 다의 신뢰구간을 제어합니다. 100보다 작은 양수 값을 지정할 수 있습니다. 기본적으로, 95% 신뢰구간이 사용됩니다.

각 목표에 대한 최대 입력 수

이 설정은 각 목표에 대한 모델에서 허용되는 최대 입력 수를 지정합니다. 1 - 20 범위의 정수를 지정할 수 있습니다. 각 목표에 대한 모델에는 항상 자체의 시차 값이 포함되므로, 이 값을 1로 설정하면 입력만 목표 자체가 됩니다.

모델 허용 한도

이 설정은 각 목표에 대한 최상의 입력 세트를 판별하기 위해 사용되는 반복 프로세스를 제어합니다. 0보다 큰 값을 지정할 수 있습니다. 기본값은 0.001입니다. 모델 공차는 예측자 선택에 대한 중단 기준입니다. 이는 최종 모델에 포함되는 예측자 수에 영향을 미칠 수 있습니다. 단, 목표가 스스로 매우 잘 예측할 수 있는 경우, 기타 예측자가 최종 모델에 포함되지 않을 수 있습니다. 일부 시행 및 오차가 필요할 수 있습니다. 예를 들어, 이 값을 높게 설정한 경우, 이를 더 작은 값으로 설정하여 기타 예측자가 포함될 수 있는지 여부를 알 수 있습니다.

이상치 임계값(%)

모델에서 계산된 확률(이상치)이 이 임계값을 초과하는 경우 관측값은 이상치로 플래그가 붙습니다. 50 - 100 범위의 값을 지정할 수 있습니다.

각 입력의 시차 수

이 설정은 각 목표에 대한 모델에서 각 입력의 시차 항 수를 지정합니다. 기본적으로, 시차 항 수는 분석에서 사용되는 시간 구간에서 자동으로 결정됩니다. 예를 들어, 시간 구간이 월(증분: 한 달)인 경우 시차 수는 12입니다. 선택적으로, 시차 수를 명시적으로 지정할 수 있습니다. 지정된 값은 1 - 20 범위의 정수여야 합니다.

기존 모델을 사용하여 추정 계속

이미 시간 인과 모델을 생성한 경우, 새 모델을 작성하기 보다는 해당 모델에 대해 지정된 기준 설정을 재사용하려면 이 옵션을 선택하십시오. 이 방식에서는 이전(그러나, 최근의 데이터)과 동일한 모델 설정을 기반으로 하는 새 예측을 다시 추정하고 생성하여 시간을 절약할 수 있습니다.

⑧ 스트리밍 TCM 노드 - 추정 기간 옵션

기본적으로 추정 기간은 모든 계열에 걸쳐 최초 관측값 시간에 시작되고 최근 관측값 시간에 종료됩니다.

시작 및 종료 시간 기준

추정 기간의 시작 및 종료 둘 다를 지정하거나 시작 또는 종료만 지정할 수 있습니다. 추정 기간의 시작 또는 종료를 생략하는 경우, 기본값이 사용됩니다.

- 날짜/시간 필드에 의해 관측값이 정의된 경우, 날짜/시간 필드에 사용되는 동일한 형식으로 시작 및 종료 값을 입력하십시오.
- 순환 주기에 의해 정의된 관측값의 경우, 순환 주기 필드마다 값을 지정하십시오. 각 필드는 별도의 열에 표시됩니다.

최근이거나 최초의 시간 구간(L)

선택적 오프셋으로, 데이터의 최초 시간 구간에 시작하거나 최근 시간 구간에 종료하는, 지정된 시간 구간 수로 추정 기간을 정의합니다. 이 컨텍스트에서, 시간 구간은 분석의 시간 구간을 가리킵니다. 예를 들어, 관측값이 매월 단위이지만 분석의 시간 구간은 분기일 수 있습니다. **최근**과 **시간 구간 수**로 24 값을 지정하면 최근 24개 분기를 의미합니다.

선택적으로, 지정된 시간 구간 수를 제외할 수 있습니다. 예를 들어, 최근 24 시간 구간을 지정하고 제외할 수로 1를 지정하면, 추정 기간은 마지막 구간 앞에 있는 24개 구간으로 구성됩니다.

⑨ 스트리밍 TCM 노드 - 모델 옵션

모델 이름

모델에 대한 사용자 정의 이름을 지정하거나 자동으로 생성되는 이름(TCM)을 승인할 수 있습니다.

예측

레코드를 미래로 확장 옵션은 추정 기간이 끝난 이후에 예측할 시간 구간 수를 설정합니다. 이 경우의 시간 구간은 데이터 지정 사항 탭에 지정된 분석의 시간 구간입니다. 예측이 요청되면 물론 목표가 아닌 입력 계열에서 자기회귀분석 모델이 자동으로 작성됩니다. 그런 다음, 이 모델을 사용하여 예측 기간에 해당 입력 계열의 값을 생성합니다. 이 설정에 최대 한계는 없습니다.

(16) CPLEX 최적화 노드

CPLEX 최적화 노드에서는 OPL(Optimization Programming Language) 모델 파일을 통해 복잡한 수학(CPLEX) 기반 최적화를 수행할 수 있습니다. 이 기능은 더 이상 지원되지 않는 IBM® Analytical Decision Management 제품에서 제공하지만, 이제 IBM Analytical Decision Management 없이도 SPSS® Modeler에서 CPLEX 노드를 사용할 수 있습니다.

CPLEX 최적화 및 OPL에 대한 자세한 정보는 IBM ILOG CPLEX Optimization Studio 문서를 참조하십시오.

CPLEX 최적화 노드는 다중 데이터 소스 또는 다차원 수신 데이터를 지원합니다. CPLEX 최적화 노드에 여러 노드를 연결하고, 이전 노드 각각을 활용하여 OPL 모델 컴퓨팅에 대한 데이터를 제공합니다. 이는 각 필드 맵핑에서 개별 튜플 변수군으로 설정됩니다.

CPLEX 최적화 노드에서 생성된 데이터를 출력할 때, 데이터 소스의 원래 데이터는 단일 지수로서 또는 결과의 다차원 지수로서 함께 출력될 수 있습니다.

참고:

- IBM SPSS Modeler Server에서 CPLEX 최적화 노드를 포함하는 스트림을 실행하는 경우, 기본적으로 임베드된 커뮤니티 에디션 CPLEX 라이브러리가 사용됩니다. 여기에는 1000개의 변수 및 1000개의 제약조건이라는 제한이 있습니다. IBM ILOG CPLEX의 전체 에디션을 설치한 경우에 이러한 제한이 없는 전체 에디션의 CPLEX 엔진을 대신 사용하려면 플랫폼별로 다음 단계를 완료하십시오.
  - Windows의 경우, options.cfg를 편집하고 OPL 라이브러리 경로를 추가하십시오. 예:

    ```
    cplex_opl_lib_path="<CPLEX_path>₩opl₩bin₩<Platform_dir>"
    ```

    여기서, <CPLEX_path>는 C:₩Program Files₩IBM₩ILOG₩CPLEX_Studio127과 같은 CPLEX 설치 디렉토리이며 <Platform_dir>은 x64_win64와 같은 플랫폼별 디렉토리입니다.
  - Linux의 경우, modelersrv.sh를 편집하고 OPL 라이브러리 경로를 추가하십시오. 예:

    ```
    CPLEX_OPL_LIB_PATH=<CPLEX_path>/opl/bin/<Platform_dir>
    ```

    여기서, <CPLEX_path>는 /root/Libs_127_FullEdition/Linux_x86_64와 같은 CPLEX 설치 디렉토리이며 <Platform_dir>은 x86-64_linux와 같은 플랫폼별 디렉토리입니다.

참고:

- **SPSS Modeler Solution Publisher**에서 CPLEX 최적화 노드를 포함하는 스트림을 실행하는 경우, 기본적으로 임베드된 커뮤니티 에디션 CPLEX 라이브러리가 사용됩니다. 여기에는 1000개의 변수 및 1000개의 제약조건이라는 제한이 있습니다. IBM ILOG CPLEX의 전체 에디션을 설치한 경우에 이러한 제한이 없는 전체 에디션의 CPLEX 엔진을 대신 사용하려면 플랫폼별로 다음 단계를 완료하십시오.
  - Windows의 경우, modelerrun.exe에 대한 명령행 인수로 OPL 라이브러리 경로를 추가하십시오. 예:

    ```
    -o cplex_opl_lib_path="<CPLEX_path>₩opl₩bin₩<Platform_dir>"
    ```

    여기서, <CPLEX_path>는 C:₩Program Files₩IBM₩ILOG₩CPLEX_Studio127과 같은 CPLEX 설치 디렉토리이며 <Platform_dir>은 x64_win64와 같은 플랫폼별 디렉토리입니다.
  - Linux의 경우, modelerrun를 편집하고 OPL 라이브러리 경로를 추가하십시오. 예:

    ```
    CPLEX_OPL_LIB_PATH=<CPLEX_path>/opl/bin/<Platform_dir>
    ```

    여기서, <CPLEX_path>는 /root/Libs_127_FullEdition/Linux_x86_64와 같은 CPLEX 설치 디렉토리이며 <Platform_dir>은 x86-64_linux와 같은 플랫폼별 디렉토리입니다.
- CPLEX는 macOS에서 지원되지 않습니다. 노드를 사용할 수 있지만(스트림에 추가, 속성 편집 등) 실행할 수는 없습니다.

① CPLEX 최적화 노드에 대한 옵션 설정

CPLEX 최적화 노드의 옵션 탭은 다음 필드로 구성됩니다.

OPL 모델 파일. OPL(Optimization Programming Language) 모델 파일을 선택하십시오.

OPL 모델. OPL 모델을 선택하면 컨텐츠가 여기에 표시됩니다.

입력 데이터

입력 데이터 탭에서 **데이터 소스** 드롭 다운 목록은 현재 CPLEX 최적화 노드에 연결된 모든 데이터 소스(사전확률 노드)를 나열합니다. 드롭 다운에서 데이터 소스를 선택하면 아래의 **입력 맵핑** 섹션이 새로 고쳐집니다. **모든 필드 적용**을 클릭하여 선택된 데이터 소스에 대한 모든 필드 맵핑을 자동으로 생성하십시오. **입력 맵핑** 테이블이 자동으로 채워집니다.

수신 데이터에 해당하는 OPL 모델의 튜플 세트 이름을 입력하십시오. 그런 다음 필요한 경우 모든 튜플 필드가 튜플 정의에 설정된 순서에 따라 데이터 입력 필드에 맵핑되었는지 확인하십시오.

데이터 소스에 대한 입력 맵핑을 설정한 다음 드롭 다운에서 다른 데이터 소스를 선택하고 프로세스를 반복할 수 있습니다. 이전 데이터 소스 맵핑이 자동으로 저장됩니다. 완료되면 **적용** 또는 **확인**을 클릭하십시오.

기타 데이터

최적화에 대해 다른 데이터를 지정해야 하는 경우 기타 데이터 탭에서 **OPL 데이터** 섹션을 사용하십시오.

출력

출력이 의사결정 변수인 경우, 사전확률 데이터 소스(수신 데이터)를 지수로 사용해야 하며 지수가 입력 데이터 탭의 **입력 맵핑** 섹션에서 사전정의되어야 합니다. 의사결정 변수의 기타 유형은 현재 지원되지 않습니다. 의사결정 변수는 단일 지수 또는 다중 지수를 가질 수 있습니다. SPSS® Modeler에서는 원 수신 데이터의 일부 또는 전부와 함께 CPLEX 결과를 출력합니다. 이는 기타 SPSS Modeler 노드와 일관성이 있습니다. 참조되는 해당 지수는 아래에서 설명하는 **출력 튜플** 필드에서 지정되어야 합니다.

출력 탭에서 출력 모드(**원시 출력** 또는 **의사결정 변수**)를 선택하고 다른 옵션을 적절하게 지정하십시오. 원시 출력 옵션은 이름에 관계없이 목적 함수 값을 직접 출력합니다.

OPL의 목적 함수 값 변수 이름. 이 필드는 **의사결정 변수** 출력 모드를 선택한 경우에만 사용으로 설정됩니다. OPL 모델의 목적 함수 값 변수 이름을 입력하십시오.

출력을 위한 목적 함수 값 필드 이름 - 출력에 사용할 필드 이름을 입력하십시오. 기본값은 _OBJECTIVE입니다.

출력 튜플. 수신 데이터의 사전정의된 튜플의 이름을 입력하십시오. 이는 의사결정 변수의 지수 역할을 수행하며 변수 출력이 있는 출력인 것으로 예측됩니다. 출력 튜플은 OPL 내의 의사결정 변수 정의와 일관성이 있어야 합니다. 다중 지수인 경우, 튜플 이름이 쉼표(,)로 결합되어야 합니다.

변수 출력 - 출력에 포함될 변수를 한 개 이상 추가하십시오.

3) 필드 조작 노드

(1) 필드 조작 개요

초기 데이터 탐색 후에는 분석을 준비하기 위해 데이터를 선택하거나 정리하거나 구성해야 할 수 있습니다. 필드 조작 팔레트에는 이 변환 및 준비를 위해 유용한 다수의 노드가 포함되어 있습니다.

예를 들어, 파생 노드를 사용하면 데이터에서 현재 표시되지 않는 속성을 작성할 수 있습니다. 또는 구간화 노드를 사용하여 목표로 지정된 분석을 위해 자동으로 필드 값의 코딩을 변경할 수 있습니다. 유형 노드를 자주 사용하고 있음을 알 수 있으며 이를 통해 데이터 세트의 각 필드에 대한 측정 수준 값 및 모델링 역할을 지정할 수 있습니다. 해당 조작은 결측값 및 다운스트림 모델링 처리를 위해 유용합니다.

필드 조작 팔레트에는 다음과 같은 노드가 포함되어 있습니다.

자동 데이터 준비(ADP) 노드는 데이터를 분석하고 수정사항을 식별하고, 문제가 있거나 유용할 것 같지 않은 필드를 제외시키고, 적절한 경우 새 속성을 파생시키고, 지능형 선별 및 표본추출 기법을 통해 성능을 개선할 수 있습니다. 완전 자동화된 방식으로 노드를 사용하여 노드가 수정사항을 선택하고 적용할 수 있게 하거나, 변경사항이 작성 및 승인되기 전에 변경을 미리보거나, 거부 또는 원하는 대로 개정할 수 있습니다.

유형 노드는 필드 메타데이터 및 특성을 지정합니다. 예를 들어 각 필드에 대한 측정 수준(연속형, 명목형, 순서형 또는 플래그)을 지정하고, 결측값 및 시스템 널 처리를 위한 옵션을 설정하고, 모델링 목적으로 필드의 역할을 설정하고, 필드와 값 레이블을 지정하고, 필드의 값을 지정할 수 있습니다.

필터 노드는 필드를 필터링(삭제)하고, 필드 이름을 변경하고 한 소스에서 다른 소스로 필드를 맵핑합니다.

파생 노드는 데이터 값을 수정하거나 하나 이상의 기존 필드로부터 새 필드를 작성합니다. 수식, 플래그, 명목형, 상태, 개수, 조건부 유형의 필드를 작성합니다.

앙상블 노드는 둘 이상의 모델 너깃을 결합하여 임의의 한 모델에서 얻을 수 있는 것보다 정확한 예측을 얻습니다.

채움 노드는 필드 값을 대체하고 저장 공간을 변경합니다. @BLANK(@FIELD) 같은 CLEM 조건을 기반으로 값을 대체할 수 있습니다. 또는 모든 공백 또는 널값을 특정 값으로 대체할 것을 선택할 수 있습니다. 채움 노드는 종종 유형 노드와 함께 사용하여 결측값을 대체합니다.

값 익명화 노드는 필드 이름 및 값이 다운스트림으로 표시되는 방법을 변환하여 원 데이터를 위장합니다. 이것은 다른 사용자가 고객 이름이나 기타 세부사항 같은 민감한 데이터를 사용하여 모델을 작성하도록 허용하려는 경우에 유용할 수 있습니다.

재분류 노드는 한 세트의 범주형 값을 다른 값으로 변환합니다. 재분류는 분석을 위해 범주를 접거나 데이터를 재그룹화하는 데 유용합니다.

구간화 노드는 하나 이상의 기존 연속형(숫자 범위) 필드의 값을 기반으로 새 명목형(세트) 필드를 자동으로 작성합니다. 예를 들어, 연속형 수입 필드를 평균값에서의 편차로서 수입 그룹을 포함하는 새 범주형 필드로 변환할 수 있습니다. 새 필드에 대한 구간을 작성한 후에는 절단점을 기반으로 파생 노드를 생성할 수 있습니다.

RFM(Recency, Frequency, Monetary) 분석 노드를 사용하면 얼마나 최근에 사용자로부터 구매했는지(최근성), 얼마나 자주 구매했는지(빈도) 및 모든 트랜잭션에서 얼마나 소비했는지(구매총액)를 조사하여 최고의 고객이 될 수 있는 고객을 정량적으로 판별할 수 있습니다.

파티션 노드는 파티션 필드를 생성하는데, 이 필드는 모델 작성의 학습, 검정, 검증 단계를 위한 별개의 서브세트로 데이터를 분할합니다.

플래그로 설정 노드는 하나 이상의 명목 필드에 대해 정의된 범주형 값을 바탕으로 다중 플래그 필드를 파생시킵니다.

구조변환 노드는 명목 또는 플래그 필드를 아직 또 다른 필드의 값으로 채워질 수 있는 필드 그룹으로 변환합니다. 예를 들어, payment type이라는 이름의 필드와 credit, cash, debit의 값이 주어진 경우, 각각이 실제 이루어진 지불의 값을 포함할 수 있는 세 개의 새 필드(credit, cash, debit)가 작성됩니다.

전치 노드는 행 및 열의 데이터를 바꿔서 레코드가 필드가 되고 필드가 레코드가 되게 합니다.

구간을 지정하고 추정 또는 시계열 분석을 위한 새 시간 필드를 파생하려면 시간 간격 노드를 사용하십시오. 초부터 년까지, 모든 범위의 시간 간격이 지원됩니다.

히스토리 노드는 이전 레코드의 필드에 있는 데이터를 포함하는 새 필드를 작성합니다. 히스토리 노드는 시계열 데이터 같은 순차 데이터에 가장 자주 사용합니다. 히스토리 노드를 사용하기 전에 정렬 노드를 사용하여 데이터를 정렬할 수 있습니다.

필드 재정렬 노드는 필드를 다운스트림으로 표시하는 데 사용하는 기본 순서를 정의합니다. 이 순서는 테이블, 목록 및 필드 선택기 같은 다양한 장소에서 필드의 표시에 영향을 줍니다. 이 작업은 관심있는 필드를 더 잘 보이게 만들기 위해 넓은 데이터 세트에 대해 작업할 때 유용합니다.

SPSS® Modeler 내에서 표현식 작성기 공간 함수, STP(Spatio-Temporal Prediction) 노드, 맵 시각화 노드 같은 항목이 투영된 좌표계를 사용합니다. 지리적 좌표계를 사용하고 사용자가 가져오는 임의의 데이터의 좌표계를 변경하려면 재투영 노드를 사용하십시오.

이 노드 중 몇몇 노드는 데이터 검토 노드에 의해 작성된 감사 보고서에서 직접 생성될 수 있습니다. 자세한 정보는 데이터 준비를 위해 기타 노드 생성의 내용을 참조하십시오.

(2) 자동 데이터 준비

분석을 위한 데이터 준비는 모든 프로젝트에서 가장 중요한 단계 중 하나이며 일반적으로 가장 많은 시간이 소요되는 단계 중 하나입니다. 자동 데이터 준비(ADP)는 데이터 분석, 수정사항 식

별, 문제가 있거나 유용할 것 같지 않은 필드 필터링, 적절한 경우 새 속성 파생 및 지능형 선별 기술을 통한 성능 개선 등의 작업을 자동으로 처리합니다. 완전히 **자동화**된 방식으로 알고리즘을 사용하여 알고리즘이 수정사항을 선택하고 적용할 수 있도록 하거나, **대화식** 방식으로 알고리즘을 사용하여 변경을 수행하기 전에 변경사항을 미리보고 원하는 바에 따라 변경사항을 수락하거나 거부할 수 있습니다.

ADP를 사용하면 관련된 통계 개념에 대한 사전 지식 없이 모델 작성을 위한 데이터를 쉽고 빠르게 준비할 수 있습니다. 모델은 더 빨리 작성되고 스코어링되는 경향이 있습니다. 이에 더하여 ADP를 사용하면 자동화된 모델링 프로세스(예: 모델 새로 고치기 및 챔피언/도전자)의 강력함이 더욱 향상됩니다.

> ⓘ **참고**: ADP는 분석을 위해 필드를 준비할 때 이전 필드의 기존 값 및 특성을 대체하는 대신 조정 또는 변환을 포함하는 새 필드를 작성합니다. 이전 필드는 추가 분석에서 사용되지 않습니다(해당 역할이 없음으로 설정됨).

예제. 자택 소유자의 보험 청구를 조사하기 위한 제한된 자원을 가진 보험 회사가 사기일 가능성이 높은 의심스러운 보험 청구를 플래그 지정하는 모델을 작성하려 합니다. 모델을 작성하기 전에 보험 회사는 자동 데이터 준비를 사용하여 모델링을 위한 데이터를 준비합니다. 변환이 적용되기 전에 제안된 변환을 검토할 수 있기를 원하므로 대화식 모드에서 자동 데이터 준비를 사용합니다.

자동차 산업 그룹은 다양한 개인용 자동차의 판매량을 추적합니다. 과성능 모델과 성능 미달 모델을 식별하기 위한 노력으로 자동차 판매량과 자동차 특성 사이의 관계를 설정하려 합니다. 자동 데이터 준비를 사용하여 분석을 위한 데이터를 준비하고 결과가 어떻게 다른지 확인하기 위해 준비 "전"과 "후"의 데이터를 사용하여 모델을 작성합니다.

사용자의 목표 자동 데이터 준비는 다른 알고리즘이 모델을 작성하는 속도에 영향을 주고 그러한 모델의 예측력을 향상시키는 데이터 준비 단계를 권장합니다. 여기에는 기능 변환, 생성 및 선택이 포함됩니다. 목표도 변환할 수 있습니다. 데이터 준비 프로세스가 집중해야 하는 모델 작성 우선순위를 지정할 수 있습니다.

- **속도와 정확도의 균형**. 이 옵션은 모델 작성 알고리즘이 데이터를 처리하는 속도와 예측의 정확도 둘 다에 동일한 우선순위를 부여하도록 데이터를 준비합니다.
- **속도 최적화**. 이 옵션은 모델 작성 알고리즘이 데이터를 처리하는 속도에 우선순위를 부여하도록 데이터를 준비합니다. 매우 큰 데이터 세트를 사용하여 작업하거나 빠른 해답을 찾으려면 이 옵션을 선택하십시오.
- **정확도 최적화**. 이 옵션은 모델 작성 알고리즘이 생성하는 예측의 정확도에 우선순위를 부여하도록 데이터를 준비합니다.
- **사용자 정의 분석**. 설정 탭에서 알고리즘을 수동으로 변경하려면 이 옵션을 선택하십시오. 이후 설정 탭에서 다른 목적 중 하나와 호환되지 않는 옵션을 변경하면 이 설정이 자동으로 선택됩니다.

노드 학습

ADP 노드는 프로세스 노드로 구현되고 유형 노드와 유사한 방식으로 작동합니다. ADP 노드를 **훈련**하는 것은 유형 노드를 인스턴스화하는 것에 해당합니다. 일단 분석이 수행되면, 업스트림 데이터 모델이 변경되지 않는 한 추가 분석 없이 지정된 변환이 데이터에 적용됩니다. 유형 및 필터 노드와 마찬가지로, ADP 노드는 다시 연결될 때 다시 훈련하지 않아도 되도록 연결이 끊어질 때 데이터 모델 및 변환을 기억합니다. 이로 인해 일반 데이터 서브세트를 대상으로 ADP 노드를 훈련한 후 ADP 노드를 복사하거나 배포하여 필요한 만큼 자주 실시간 데이터를 대상으로 ADP 노드를 사용할 수 있습니다.

도구 모음 사용

도구 모음을 사용하여 데이터 분석을 실행하고 표시를 업데이트하며 원래 데이터와 관련하여 사용할 수 있는 노드를 생성할 수 있습니다.

- **생성** 이 메뉴에서 필터 또는 파생 노드를 생성할 수 있습니다. 이 메뉴는 분석 탭에 분석이 표시된 경우에만 사용할 수 있습니다.
  필터 노드는 변환된 입력 필드를 제거합니다. 원래 입력 필드를 데이터 세트에 남겨 두도록 ADP 노드를 구성하면 원래 입력 세트가 복원되고, 이로 인해 입력의 관점에서 스코어 필드를 해석할 수 있습니다. 예를 들어, 이는 다양한 입력에 대한 스코어 필드 그래프를 생성하려는 경우에 유용할 수 있습니다.

  파생 노드는 원래 데이터 세트 및 목표 단위를 복원할 수 있습니다. ADP 노드에 범위 목표의 척도를 조정하는 분석이 있는 경우(즉, 입력 및 출력 준비 패널에서 Box-Cox 척도 조정을 선택한 경우)에만 파생 노드를 생성할 수 있습니다. 목표가 범위가 아니거나 Box-Cox 척도 조정을 선택하지 않은 경우에는 파생 노드를 생성할 수 없습니다. 자세한 정보는 파생 노드 생성의 내용을 참조하십시오.

- **보기**에는 분석 탭에 표시되는 내용을 제어하는 옵션이 있습니다. 여기에는 그래프 편집 제어와 기본 패널 및 링크된 보기 둘 다에 대한 표시 선택사항이 포함됩니다.
- **미리보기**: 입력 데이터에 적용될 변환 표본을 표시합니다.
- **데이터 분석**: 현재 설정을 사용하여 분석을 시작하고 분석 탭에 결과를 표시합니다.
- **분석 지우기**: 기존 분석을 지웁니다(현재 분석이 있는 경우에만 사용 가능).

노드 상태

IBM® SPSS® Modeler 캔버스에서 ADP 노드의 상태는 분석이 수행되었는지 여부를 나타내는 아이콘 위의 화살표 또는 체크로 표시됩니다.

자동화된 데이터 준비 노드를 사용하여 수행된 계산에 대한 자세한 정보는 *IBM SPSS Modeler 알고리즘 안내서*의 *자동화된 데이터 준비 알고리즘* 절을 참조하십시오. 이 안내서는 제품 다운로드의 일부로 설치 디스크의 ₩Documentation 디렉토리나 웹에서 PDF 형식으로 사용 가능합니다.

① 필드 탭(자동 데이터 준비)

모델을 작성하려면 먼저 목표 및 입력으로 사용할 필드를 지정해야 합니다. 몇 가지 예외가 있지만, 모든 모델링 노드는 업스트림 유형 노드에서 필드 정보를 사용합니다. 유형 노드를 사용하여 입력 및 대상 필드를 선택하는 경우 이 탭의 내용을 변경하지 않아도 됩니다.

유형 노드 설정 사용. 이 옵션은 업스트림 유형 노드의 필드 정보를 사용하도록 노드에 지시합니다. 이는 기본값입니다.

사용자 정의 설정 사용. 이 옵션은 업스트림 유형 노드에 제공된 정보 대신 여기에 지정된 필드 정보를 사용하도록 노드에 알립니다. 이 옵션을 선택한 후 필요하면 아래 필드를 지정합니다.

목표: 하나 이상의 대상 필드가 필요한 모델의 경우 하나 이상의 대상 필드를 선택합니다. 유형 노드에서 필드 역할을 *목표*로 설정하는 것과 유사합니다.

입력 하나 이상의 입력 필드를 선택합니다. 유형 노드에서 필드 역할을 *입력*으로 설정하는 것과 유사합니다.

② 설정 탭(자동 데이터 준비)

설정 탭에는 서로 다른 여러 설정 그룹이 포함되며, 이러한 설정을 수정하여 알고리즘이 데이터를 처리하는 방법을 미세 조정할 수 있습니다. 다른 목적과 호환되지 않는 기본 설정을 변경하면 목적 탭이 자동으로 업데이트되어 **분석 사용자 정의** 옵션을 선택합니다.

가. 필드 설정

빈도 필드 사용: 이 옵션에서는 빈도 가중치로 필드를 선택할 수 있습니다. 학습 데이터의 레코드가 각각 둘 이상의 단위를 나타내는 경우(예: 통합 데이터를 사용하는 경우) 이 옵션을 사용하십시오. 필드 값은 각 레코드로 나타낸 노드 수여야 합니다.

가중 필드 사용: 이 옵션에서는 케이스 가중치로 필드를 선택할 수 있습니다. 케이스 가중치는 출력 필드의 수준에서 분산의 차이를 설명하는 데 사용됩니다.

모델링에서 제외된 필드 처리 방법. 제외된 필드에 적용되는 조치를 지정하십시오. 제외된 필드를 데이터에서 필터링하거나 단순히 해당 역할을 **없음**으로 설정할 수 있습니다.

ⓘ **참고**: 이 조치는 목표가 변환되는 경우 목표에도 적용됩니다. 예를 들어, **목표**의 새 파생된 버전이 목표 필드로 사용되는 경우 원래 목표는 필터링되거나 **없음**으로 설정됩니다.

수신 필드가 기존 분석과 일치하지 않는 경우. 학습된 ADP 노드를 실행할 때 수신 데이터 세트에서 하나 이상의 필수 입력 필드가 누락된 경우에 발생하는 상황을 지정하십시오.

- **실행을 중지하고 기존 분석 유지**. 실행 프로세스를 중지하고 현재 분석 정보를 유지하며 오류를 표시합니다.
- **기존 분석을 지우고 새 데이터 분석**. 기존 분석을 지우고 수신 데이터를 분석하며 해당 데이터에 권장되는 변환을 적용합니다.

나. 날짜 및 시간 준비(자동 데이터 준비)

다수의 모델링 알고리즘은 날짜 및 시간 세부사항을 직접 처리할 수 없습니다. 이러한 설정을 사용하면 기존 데이터의 날짜 및 시간에서 모델 입력으로 사용할 수 있는 새 지속 기간 데이터를 파생시킬 수 있습니다. 날짜 및 시간을 포함하는 필드는 날짜 또는 시간 저장 유형으로 사전 정의해야 합니다. 원래 날짜 및 시간 필드는 자동 데이터 준비를 따르는 모델 입력으로 권장되지 않습니다.

모델링을 위한 날짜 및 시간 준비. 이 옵션을 선택 취소하면 기타 모든 날짜 및 시간 준비 제어가 사용되지 않습니다(단, 선택항목은 유지보수됨).

참조 날짜까지의 경과 시간 계산. 날짜를 포함하는 각 변수의 참조 날짜 이후 경과한 년/월/일 수를 생성합니다.

- **참조 날짜**. 입력 데이터의 날짜 정보와 관련하여 지속 기간을 계산할 시작 날짜를 지정하십시오. **오늘의 날짜**를 선택하면 ADP를 실행할 때 현재 시스템 날짜가 항상 사용됩니다. 특정 날짜를 선택하려면 **고정 날짜**를 선택하고 필요한 날짜를 입력하십시오. 노드를 처음 작성할 때는 **고정 날짜** 필드에 현재 날짜가 자동으로 입력됩니다.
- **날짜 지속 기간 단위**. ADP가 날짜 지속 기간 단위를 자동으로 결정하는지 여부를 지정하거나 **고정 단위** 년, 월 및 일 중에서 선택하십시오.

참조 시간까지의 경과 시간 계산. 시간을 포함하는 각 변수의 참조 시간 이후 경과한 시/분/초 수를 생성합니다.

- **참조 시간**. 입력 데이터의 시간 정보와 관련하여 지속 기간을 계산할 시작 시간을 지정하십시오. **현재 시간**을 선택하면 ADP를 실행할 때 현재 시스템 시간이 항상 사용됩니다. 특정 시간을 선택하려면 **고정 시간**을 선택하고 필요한 세부사항을 입력하십시오. 노드를 처음 작성할 때는 **고정 시간** 필드에 현재 시간이 자동으로 입력됩니다.
- **시간 지속 기간 단위**. ADP가 시간 지속 기간 단위를 자동으로 결정하는지 여부를 지정하거나 **고정 단위** 시, 분 및 초 중에서 선택하십시오.

순환 시간 요소 추출. 이러한 설정을 사용하여 단일 날짜 또는 시간 필드를 하나 이상의 필드로 분할하십시오. 예를 들어, 날짜 선택란을 세 개 모두 선택하는 경우, 입력 날짜 필드 "1954-05-23"은 세 개의 필드 1954, 5 및 23으로 분할되고 각 필드는 **필드 이름** 패널에 정의된 접미부를 사용하며 원래 날짜 필드는 무시됩니다.

- **날짜에서 추출**. 모든 날짜 입력에 대해 년, 월, 일 또는 임의 조합을 추출할지 여부를 지정하십시오.
- **시간에서 추출**. 모든 시간 입력에 대해 시, 분, 초 또는 임의 조합을 추출할지 여부를 지정하십시오.

다. 필드 제외(자동 데이터 준비)

품질이 낮은 데이터는 예측 정확도에 영향을 미칠 수 있습니다. 따라서 입력 기능에 대해 허용 가능한 품질 수준을 지정할 수 있습니다. 상수이거나 100% 결측값을 갖는 모든 필드는 자동으로 제외됩니다.

저품질 입력 필드 제외. 이 옵션을 선택 취소하면 기타 모든 필드 제외 제어가 사용되지 않습니다(단, 선택항목은 유지보수됨).

결측값이 너무 많은 필드 제외. 결측값이 지정된 백분율보다 더 많은 필드는 추가 분석에서 제거됩니다. 0(이 옵션을 선택 취소하는 것과 동등함)보다 크거나 같고 100(결측값이 100%인 필드는 자동으로 제외됨)보다 작거나 같은 값을 지정하십시오. 기본값은 50입니다.

고유 범주가 너무 많은 명목 필드 제외. 범주가 지정된 수보다 더 많은 명목 필드는 추가 분석에서 제거됩니다. 양수를 지정하십시오. 기본은 100입니다. 이 옵션은 모델링에서 ID, 주소, 이름 등의 레코드 고유 정보를 포함하는 필드를 자동으로 제거하는 데 유용합니다.

단일 범주에 너무 많은 값이 있는 범주형 필드 제외. 범주에 지정된 백분율보다 더 많은 레코드가 있는 순서 및 명목 필드는 추가 분석에서 제거됩니다. 0(이 옵션을 선택 취소하는 것과 동등함)보다 크거나 같고 100(상수 필드는 자동으로 제외됨)보다 작거나 같은 값을 지정하십시오. 기본값은 95입니다.

라. 입력 및 목표 준비

처리하기에 완벽한 상태인 데이터는 없으므로 분석을 실행하기 전에 일부 설정을 조정하려 할 수 있습니다. 예를 들어, 이러한 조정에 이상값 제거, 결측값 처리 방법 지정 또는 유형 조정 등이 포함될 수 있습니다.

> ⓘ **참고**: 이 패널의 값을 변경하는 경우, **목적** 탭이 자동으로 업데이트되어 **사용자 정의 분석** 옵션이 선택됩니다.

모델링을 위한 입력 및 대상 필드 준비. 패널의 모든 필드를 켜거나 끕니다.

유형 조정 및 데이터 품질 개선. 입력 및 목표에 대해 몇 가지 데이터 변환을 개별적으로 지정할 수 있습니다(목표의 값을 변경하지 않으려는 이유에서). 예를 들어, 달러 단위의 수입 예측이 log(dollars)로 측정된 예측보다 더 의미가 있습니다. 또한, 목표에 결측값이 있는 경우 결측값을 채워도 예측에 아무런 이점이 없지만, 입력의 결측값을 채우는 경우에는 일부 알고리즘이 결측값을 채우지 않은 경우에 유실될 수 있는 정보를 처리할 수도 있습니다.

이러한 변환을 위한 추가 설정(예: 이상값 절사 값)은 목표 및 입력 모두에 공통됩니다.

입력, 목표 또는 둘 다에 대해 다음 설정을 선택할 수 있습니다.

- **숫자 필드의 유형 조정**. 측정 수준이 *순서*인 숫자 필드를 *연속형*으로 또는 그 반대로 변환할 수 있는지 결정하려면 이를 선택하십시오. 변환을 제어하는 최소 및 최대 임계값을 지정할 수 있습니다.
- **명목 필드 재정렬**. 명목(세트) 필드를 가장 작은 범주에서 가장 큰 범주로 순서대로 정렬하려면 이를 선택하십시오.
- **연속형 필드의 이상값 대체**. 이상값을 대체할지 여부를 지정하십시오. 이 옵션을 아래의 **이상값 대체 방법** 옵션과 함께 사용하십시오.
- **연속형 필드: 결측값을 평균으로 대체**. 연속형(범위) 기능의 결측값을 대체하려면 이를 선택하십시오.
- **명목 필드: 결측값을 모드로 대체**. 명목(세트) 기능의 결측값을 대체하려면 이를 선택하십시오.
- **순서 필드: 결측값을 중앙값으로 대체**. 순서(순서형 세트) 기능의 결측값을 대체하려면 이를 선택하십시오.

순서 필드의 값 최대 수. 순서(순서형 세트) 필드를 연속형(범위)으로 재정의하기 위한 임계값을 지정하십시오. 기본값은 10입니다. 따라서, 순서 필드에 10개가 넘는 범주가 있는 경우 이 필드는 연속형(범위)으로 재정의됩니다.

연속형 필드의 값 최소 수. 척도 또는 연속형(범위) 필드를 순서(순서형 세트)로 재정의하기 위한 임계값을 지정하십시오. 기본값은 5입니다. 따라서, 연속형 필드에 5개 미만의 값이 있으면 이 필드는 순서(순서형 세트)로 재정의됩니다.

이상값 절사 값. 표준 편차에서 측정되는 이상값 절사 기준을 지정하십시오. 기본값은 3입니다.

이상값 대체 방법. 이상값을 자르기(강제 적용)를 통해 절사 값으로 대체할지 또는 이상값을 삭제하고 결측값으로 설정할지 여부를 선택하십시오. 결측값으로 설정된 이상값은 위에서 선택한 결측값 처리 설정을 따릅니다.

모든 연속형 입력 필드를 일반 척도에 둠. 연속형 입력 필드를 정규화하려면 이 선택란을 선택하고 정규화 방법을 선택하십시오. 기본값은 **z-스코어 변환**입니다. 이 방법에서는 기본값이 0인

최종 평균과 기본값이 1인 **최종 표준 편차**를 지정할 수 있습니다. 또는 **최소/최대 변환**을 사용하도록 선택하고 최소 및 최대 값(기본값은 각각 0과 100임)을 지정할 수 있습니다.

이 필드는 특히 기능 생성 및 선택 패널에서 **기능 생성 수행**을 선택할 때 유용합니다.

Box-Cox 변환으로 연속형 대상 척도 조정. 연속형(척도 또는 범위) 대상 필드를 정규화하려면 이 선택란을 선택하십시오. Box-Cox 변환에서 **최종 평균**의 기본값은 0이고 **최종 표준 편차**의 기본값은 1입니다.

> ⓘ **참고**: 목표를 정규화하도록 선택한 경우 목표의 차원이 변환됩니다. 이 경우, 추가 처리를 위해 변환된 단위를 다시 인식 가능한 형식으로 되돌리기 위해 역변환을 적용할 파생 노드를 생성해야 할 수도 있습니다. 자세한 정보는 파생 노드 생성의 내용을 참조하십시오.

마. 생성 및 필드선택

데이터의 예측력을 향상시키기 위해 입력 필드를 변환하거나 기존 필드를 기반으로 새 입력 필드를 생성할 수 있습니다.

> ⓘ **참고**: 이 패널의 값을 변경하는 경우, **목적** 탭이 자동으로 업데이트되어 **사용자 정의 분석** 옵션이 선택됩니다.

예측력 향상을 위해 입력 필드 변환, 생성 및 선택. 패널의 모든 필드를 켜거나 끕니다.

성긴 범주를 병합하여 목표와의 연관 최대화. 목표와 연관하여 처리할 변수의 수를 줄여 보다 경제적인 모델을 작성하려면 이 옵션을 선택하십시오. 필요한 경우, 확률값을 기본값 0.05에서 변경하십시오.

모든 범주가 하나로 병합되는 경우, 필드의 원래 버전과 파생된 버전은 예측변수로서 값이 없기 때문에 제외됩니다.

목표가 없는 경우 개수를 기반으로 성긴 범주 병합. 목표가 없는 데이터를 처리하는 경우, 순서(순서형 세트) 기능, 명목(세트) 기능 또는 둘 다의 성긴 범주를 병합하도록 선택할 수 있습니다. 병합할 범주를 식별하는 데이터에 케이스(또는 레코드)의 최소 백분율을 지정하십시오. 기본값은 10입니다.

다음 규칙에 따라 범주가 병합됩니다.

- 2진 필드에 대해서는 병합이 수행되지 않습니다.
- 병합 동안 범주가 두 개뿐인 경우에는 병합이 중지합니다.
- 원래 범주가 없고 병합 중에 범주가 작성되지 않으며 지정된 케이스 최소 백분율보다 적으면 병합이 중지합니다.

예측력을 유지하면서 연속형 필드 구간화. 범주형 대상이 포함된 데이터가 있는 경우, 처리 성능을 향상시키기 위해 강력한 연관이 있는 연속형 입력을 구간화할 수 있습니다. 필요한 경우, 동종 서브세트의 확률값을 기본값 0.05에서 변경하십시오.

구간화 조작으로 특정 필드의 단일 구간이 생성되는 경우, 필드의 원래 버전과 구간화된 버전은 예측변수로서 값이 없기 때문에 제외됩니다.

> **참고**: ADP에서의 구간화는 IBM® SPSS® Modeler의 다른 파트에서 사용되는 최적 구간화와 다릅니다. 최적 구간화는 엔트로피 정보를 사용하여 연속형 변수를 범주형 변수로 변환합니다. 이 구간화는 데이터를 정렬하고 모두 메모리에 저장해야 합니다. ADP는 동종 서브세트를 사용하여 연속형 변수를 구간화합니다. 이는 ADP 구간화는 데이터를 정렬하지 않아도 되며 모든 데이터를 메모리에 저장하지 않음을 의미합니다. 동종 서브세트를 사용하여 연속형 변수를 구간화하는 방법은 구간화 후의 범주 수가 항상 목표 범주 수 이하임을 의미합니다.

변수 선택 수행. 상관 계수가 낮은 기능을 제거하려면 이 옵션을 선택하십시오. 필요한 경우, 확률값을 기본값 0.05에서 변경하십시오.

이 옵션은 목표가 연속형인 연속형 입력 기능과 범주형 입력 기능에만 적용됩니다.

기능 생성 수행. 기존의 여러 기능의 조합(이후 모델링에서 삭제됨)에서 새 기능을 파생시키려면 이 옵션을 선택하십시오.

이 옵션은 목표가 연속형이거나 목표가 없는 연속형 입력 기능에만 적용됩니다.

바. 필드 이름(자동 데이터 준비)

변환된 새 기능을 쉽게 식별하기 위해 ADP는 기본 새 이름, 접두부 또는 접미부를 작성하고 적용합니다. 이러한 이름을 사용자 요구 및 데이터와 연관성이 더 큰 이름으로 수정할 수 있습니다. 다른 레이블을 지정하려면 다운스트림 유형 노드에서 이를 수행해야 합니다.

변환 및 생성 필드. 변환된 목표 및 입력 필드에 적용할 이름 확장자를 지정하십시오.

ADP 노드에서 문자열 필드를 아무것도 포함하지 않도록 설정하면 사용하지 않는 필드 처리를 위해 선택한 방법에 따라 오류가 발생할 수 있습니다. 설정 탭의 필드 설정 패널에서 **모델링에서 제외된 필드 처리 방법**을 **사용하지 않는 필드 필터링**으로 설정한 경우, 입력 및 목표의 이름 확장자를 빈 값으로 설정할 수 있습니다. 원래 필드는 필터링되고 변환된 필드가 그 대신 저장됩니다. 이 경우 변환된 새 필드는 원래 필드와 동일한 이름을 가집니다.

사용하지 않는 필드의 방향을 '없음'으로 설정을 선택한 경우, 목표 및 입력에 빈(또는 널) 이름 확장자를 사용하면 중복 필드 이름을 작성하게 되므로 오류가 발생합니다.

또한, 선택 및 생성 설정을 통해 생성되는 기능에 적용할 접두부 이름을 지정하십시오. 이 접두부 루트 이름에 숫자 접미부를 첨부하여 새 이름이 작성됩니다. 숫자의 형식은 파생되는 새 기능의 수에 따라 다릅니다. 예를 들면, 다음과 같습니다.

- 1-9개의 생성된 기능은 feature1 - feature9으로 이름이 지정됩니다.
- 10-99개의 생성된 기능은 feature01 - feature99으로 이름이 지정됩니다.
- 100-999개의 생성된 기능은 feature001 - feature999으로 이름이 지정되며, 계속해서 이와 같이 반복됩니다.

이 방식을 사용하면 생성된 기능을 기능의 수와 상관없이 합리적인 순서로 정렬할 수 있습니다.

날짜 및 시간에서 계산된 지속 기간. 날짜 및 시간 둘 다에서 계산된 지속 기간에 적용할 이름 확장자를 지정하십시오.

날짜 및 시간에서 추출된 순환 요소. 날짜 및 시간 둘 다에서 추출된 순환 요소에 적용할 이름 확장자를 지정하십시오.

③ 분석 탭(자동 데이터 준비)

1. 목적, 필드 및 설정 탭에서 수행된 변경을 포함하여 ADP 설정에 만족하는 경우 **데이터 분석**을 클릭하십시오. 알고리즘이 해당 설정을 데이터 입력에 적용하고 분석 탭에 결과를 표시합니다.

분석 탭은 데이터 처리를 요약하는 테이블 및 그래프 형식 출력을 모두 포함하며 스코어링을 위한 데이터 수정 또는 개선 방법과 관련하여 권장사항을 표시합니다. 그러면 그러한 권장사항을 검토하여 수락하거나 거부할 수 있습니다.

분석 탭은 두 개의 패널(왼쪽에 있는 기본 보기와 오른쪽에 있는 링크된(또는 보조) 보기)로 구성됩니다. 세 개의 기본 보기가 있습니다.

- 필드 처리 요약(기본값). 자세한 정보는 필드 처리 요약(자동 데이터 준비)의 내용을 참조하십시오.
- 필드. 자세한 정보는 필드(자동 데이터 준비)의 내용을 참조하십시오.
- 조치 요약. 자세한 정보는 동작 요약(자동 데이터 준비)의 내용을 참조하십시오.

4개의 링크/보조 보기가 있습니다.

- 예측력(기본값). 자세한 정보는 예측 력(자동 데이터 준비)의 내용을 참조하십시오.
- 필드 테이블. 자세한 정보는 필드 테이블(자동 데이터 준비)의 내용을 참조하십시오.
- 필드 세부사항. 자세한 정보는 필드 세부사항(자동 데이터 준비)의 내용을 참조하십시오.
- 조치 세부사항. 자세한 정보는 조치 세부사항(자동 데이터 준비)의 내용을 참조하십시오.

보기 사이의 링크

기본 보기 내에서 테이블의 밑줄 친 텍스트는 링크된 보기에서의 표시를 제어합니다. 텍스트를 클릭하면 특정 필드, 필드 세트 또는 처리 단계에 대한 세부사항을 볼 수 있습니다. 마지막으로 선택한 링크가 더 어두운 색으로 표시됩니다. 이는 두 보기 패널에 있는 컨텐츠 사이의 연결을 식별하는 데 도움이 됩니다.

보기 재설정

원래의 분석 권장사항을 다시 표시하고 분석 보기에 대해 수행한 모든 변경을 중단하려면 기본 보기 패널 맨 아래에 있는 **재설정**을 클릭하십시오.

가. 필드 처리 요약(자동 데이터 준비)

필드 처리 요약 테이블은 생성된 기능의 수 및 기능의 상태 변경을 포함하여 처리가 미치는 전반적인 영향에 대한 추정 스냅샷을 제공합니다.

모델이 실제로 작성되지는 않으므로 데이터 준비 전과 후의 전체 예측력 변화에 대한 측도나 그래프가 없습니다.

이 테이블에는 다음 정보가 표시됩니다.

- 대상 필드 수
- 원래(입력) 예측변수 수
- 분석 및 모델링에서의 사용이 권장되는 예측변수. 여기에는 권장되는 총 필드 수, 권장되는 변환되지 않은 원래 필드 수, 권장되는 변환된 필드 수(필드의 중간 버전, 날짜/시간 예측변수에서 파생된 필드 및 생성된 예측변수 제외), 날짜/시간 필드에서 파생된 권장되는 필드 수 및 권장되는 생성된 예측변수 수가 포함됩니다.
- 원래 양식이든, 파생된 필드로서든 또는 생성된 예측변수에 대한 입력으로서든 어떤 양식으로도 사용이 권장되지 않는 입력 예측변수 수

필드 정보에 밑줄이 있는 경우, 클릭하면 링크된 보기에 자세한 내용이 표시됩니다. **목표**, **입력 기능** 및 **사용되지 않는 입력 기능**의 세부사항이 필드 테이블 링크된 보기에 표시됩니다. 자세한 정보는 필드 테이블(자동 데이터 준비) 주제를 참조하십시오. **분석에서 사용이 권장되는 기능**은 예측력 링크된 보기에 표시됩니다. 자세한 정보는 예측 력(자동 데이터 준비)의 내용을 참조하십시오.

나. 필드(자동 데이터 준비)

필드 기본 보기는 처리된 필드를 표시하고 또한 ADP가 다운스트림 모델에서 처리된 필드의 사용을 권장하는지 여부를 표시합니다. 필드에 대한 권장을 대체할 수 있습니다(예를 들어, 생성된 기능을 제외시키거나 ADP가 제외시키도록 권장하는 기능을 포함시키기 위해). 필드가 변환된 경우, 제안된 변환을 수락할지 또는 원래의 버전을 사용할지 여부를 결정할 수 있습니다.

필드 보기는 두 개의 테이블(목표에 대한 테이블과 처리되었거나 작성된 예측변수에 대한 테이블)로 구성됩니다.

목표 테이블

목표 테이블은 데이터에 목표가 정의된 경우에만 표시됩니다.

이 테이블에는 두 개의 열이 있습니다.

- **이름** 대상 필드의 이름 또는 레이블입니다. 필드가 변환된 경우에도 원래 이름이 항상 사용됩니다.
- **측정 수준**. 측정 수준을 나타내는 아이콘을 표시합니다. 아이콘 위에 마우스를 놓으면 데이터를 설명하는 레이블(연속형, 순서, 명목 등)이 표시됩니다.
  목표가 변환된 경우 **측정 수준** 열은 변환된 최종 버전을 반영합니다. *참고*: 목표에 대해 변환을 끌 수 없습니다.

예측변수 테이블

예측변수 테이블은 항상 표시됩니다. 테이블의 각 행은 필드를 나타냅니다. 기본적으로 행은 예측력의 내림차순으로 정렬됩니다.

일반 기능의 경우 항상 원래 이름이 행 이름으로 사용됩니다. 날짜/시간 필드의 원래 및 파생 버전 모두 테이블에 표시됩니다(개별 행에). 테이블에는 생성된 예측변수도 포함됩니다.

테이블에 표시된 필드의 변환된 버전은 항상 최종 버전을 나타냅니다.

기본적으로 권장되는 필드만 예측변수 테이블에 표시됩니다. 나머지 필드를 표시하려면 테이블 위에 있는 **테이블에 권장되지 않는 필드 포함** 선택란을 선택하십시오. 그러면 나머지 필드가 테이블 맨 아래에 표시됩니다.

이 테이블에는 다음 열이 포함됩니다.

- **사용할 버전**. 필드가 다운스트림에서 사용될지 여부와 제안된 변환을 사용할지 여부를 제어하는 드롭 다운 목록을 표시합니다. 기본적으로, 이 드롭 다운 목록은 권장사항을 반영합니다.

  변환된 일반 예측변수의 경우 드롭 다운 목록에는 세 가지 선택사항인 **변환**, **원래** 및 **사용하지 않음**이 있습니다.

  변환되지 않은 일반 예측변수의 경우 선택사항은 **원래** 및 **사용하지 않음**입니다.

  파생된 날짜/시간 필드 및 생성된 예측변수의 경우 선택사항은 **변환** 및 **사용하지 않음**입니다.

  원래 날짜 필드의 경우 드롭 다운 목록이 사용되지 않고 **사용하지 않음**으로 설정됩니다.

  > ⓘ **참고**: 원래 및 변환된 버전이 모두 있는 예측변수의 경우, **원래** 버전과 **변환**된 버전 간에 변경하면 해당 기능의 **측정 수준**과 **예측력** 설정이 자동으로 업데이트됩니다.

- **이름** 각 필드의 이름은 하나의 링크입니다. 이름을 클릭하면 링크된 보기에 해당 필드에 대한 자세한 정보가 표시됩니다. 자세한 정보는 필드 세부사항(자동 데이터 준비)의 내용을 참조하십시오.
- **측정 수준**. 데이터 유형을 나타내는 아이콘을 표시합니다. 아이콘 위에 마우스를 놓으면 데이터를 설명하는 레이블(연속형, 순서, 명목 등)이 표시됩니다.
- **예측력**. 예측력은 ADP가 권장하는 필드에 대해서만 표시됩니다. 목표가 정의되지 않은 경우에는 이 열이 표시되지 않습니다. 예측력 범위는 0 - 1이며 값이 클수록 "더 좋은" 예측변수입니다. 일반적으로 예측력은 ADP 분석 내에서 예측변수를 비교하는 데 유용하지만, 분석들 간에 예측력 값을 비교해서는 안 됩니다.

다. 동작 요약(자동 데이터 준비)

자동 데이터 준비가 조치를 수행할 때마다 입력 예측변수가 변환 및/또는 필터링되며, 하나의 조치 이후에 남은 필드는 그 다음 조치에서 사용됩니다. 마지막 단계까지 남아 있는 필드는 모델링에서 사용하도록 권장되는 반면, 변환되고 생성된 예측변수에 대한 입력은 필터링됩니다.

조치 요약은 ADP가 수행한 처리 조치를 나열하는 단순한 테이블입니다. **조치**에 밑줄이 있는 경우, 클릭하면 링크된 보기에 수행된 조치에 대한 자세한 내용이 표시됩니다. 자세한 정보는 조치 세부사항(자동 데이터 준비)의 내용을 참조하십시오.

> ⓘ **참고**: 각 필드의 원래 버전과 최종 변환 버전만 표시되고 분석 중에 사용된 중간 버전은 표시되지 않습니다.

라. 예측력(자동 데이터 준비)

분석을 처음 실행할 때 기본적으로 표시되거나 필드 처리 요약 기본 보기에서 분**석에서 사용이 권장되는 예측변수**를 선택할 때 표시되는 차트에는 권장되는 예측변수의 예측력이 표시됩니다. 필드는 예측력을 기준으로 정렬되며 가장 높은 값을 갖는 필드가 맨 위에 표시됩니다.

일반 예측변수의 변환된 버전인 경우, 필드 이름은 설정 탭의 필드 이름 패널에서 선택한 접미부를 반영합니다(예: *_transformed*).

개별 필드 이름 뒤에는 측정 수준이 표시됩니다.

권장되는 각 예측변수의 예측력은 목표가 연속형인지 또는 범주형인지에 따라 선형 회귀 또는 naïve Bayes 모델에서 계산됩니다.

마. 필드 테이블(자동 데이터 준비)

필드 처리 요약 기본 보기에서 **목표**, **예측변수** 또는 **사용되지 않는 예측변수**를 클릭하면 표시되는 필드 테이블 보기에는 관련 기능을 나열하는 단순한 테이블이 표시됩니다.

이 테이블에는 두 개의 열이 있습니다.

- **이름** 예측변수 이름입니다.
  목표의 경우, 목표가 변환된 경우에도 필드의 원래 이름 또는 레이블이 사용됩니다.

  일반 예측변수의 변환된 버전인 경우, 이름은 설정 탭의 필드 이름 패널에서 선택한 접미부를 반영합니다(예: _transformed).

  날짜 및 시간에서 파생된 필드의 경우, 변환된 최종 버전의 이름이 사용됩니다(예: bdate_years).

  생성된 예측변수의 경우, 생성된 예측변수의 이름이 사용됩니다(예: Predictor1).

- **측정 수준**. 데이터 유형을 나타내는 아이콘을 표시합니다.
  목표의 경우, **측정 수준**은 항상 변환된 버전을 반영합니다(목표가 변환된 경우). 순서(순서형 세트)에서 연속형(범위, 척도) 또는 그 반대로의 변경을 예로 들 수 있습니다.

바. 필드 세부사항(자동 데이터 준비)

필드 기본 보기에서 아무 **이름**을 클릭하면 표시되는 필드 세부사항 보기에는 선택된 필드의 분포, 결측값 및 예측력 차트(적용 가능한 경우)가 있습니다. 또한 필드의 처리 히스토리와 변환된 필드의 이름도 표시됩니다(적용 가능한 경우).

차트 세트마다 두 개의 버전이 나란히 표시되어 변환이 적용된 필드와 변환이 적용되지 않은 필드를 비교할 수 있습니다. 필드의 변환된 버전이 없는 경우에는 원래 버전에 대한 차트만 표시됩니다. 파생된 날짜 또는 시간 필드와 생성된 예측변수의 경우 새 예측변수에 대한 차트만 표시됩니다.

> ⓘ **참고**: 너무 많은 범주를 포함해서 필드가 제외되는 경우에는 처리 히스토리만 표시됩니다.

분포 차트

연속형 필드 분포는 평균값에 대한 수직 참조선이 있고 정상 곡선이 오버레이된 히스토그램으로 표시됩니다. 범주형 필드는 막대형 차트로 표시됩니다.

히스토그램은 표준 편차 및 왜도를 표시하는 레이블이 지정됩니다. 그러나, 값 수가 2보다 적거나 최초 필드의 분산이 10-20 미만인 경우에는 왜도가 표시되지 않습니다.

히스토그램의 평균을 표시하거나 막대형 차트에서 범주의 수 및 총 레코드 수의 백분율을 표시하려면 차트 위에 마우스를 놓으십시오.

결측값 차트

원형 차트는 변환이 적용된 결측값 백분율과 변환이 적용되지 않은 결측값 백분율을 비교합니다. 차트 레이블은 백분율을 표시합니다.

ADP가 결측값 처리를 수행한 경우, 변환 후 원형 차트에는 대체값(즉, 결측값 대신 사용된 값)도 레이블로 포함됩니다.

차트 위에 마우스를 놓으면 결측값 수와 총 레코드 수의 백분율이 표시됩니다.

예측력 차트

권장되는 필드의 경우, 막대형 차트는 변환 전후의 예측력을 표시합니다. 목표가 변환된 경우, 계산된 예측력은 변환된 목표와 관련이 있습니다.

> ⓘ **참고**: 목표가 정의되지 않았거나 기본 보기 패널에서 목표를 클릭한 경우에는 예측력 차트가 표시되지 않습니다.

차트 위에 마우스를 놓으면 예측력 값이 표시됩니다.

처리 히스토리 테이블

이 테이블은 필드의 변환된 버전이 파생된 방법을 보여줍니다. ADP가 수행한 조치가 수행된 순서대로 나열됩니다. 그러나 특정 단계에서는 특정 필드에 대해 여러 조치가 수행되었을 수도 있습니다.

> ⓘ **참고:** 변환되지 않은 필드의 경우 이 테이블이 표시되지 않습니다.

테이블의 정보는 두 개 또는 세 개의 열로 분류됩니다.

- **조치**. 조치의 이름입니다. 예를 들어, 연속형 예측변수입니다. 자세한 정보는 조치 세부사항(자동 데이터 준비)의 내용을 참조하십시오.
- **세부사항**. 수행된 처리 목록입니다. 예를 들어, 표준 단위로의 변환입니다.
- **함수**. 생성된 예측변수에 대해서만 표시되며, 입력 필드의 선형 조합(예: .06*age + 1.21*height)을 표시합니다.

사. 조치 세부사항(자동 데이터 준비)

조치 요약 기본 보기에서 밑줄이 있는 **조치**를 선택할 때 표시되는 조치 세부사항 링크된 보기에는 수행된 각 처리 단계에 대한 조치별 정보와 공통 정보가 표시됩니다. 조치별 세부사항이 먼저 표시됩니다.

조치마다 링크된 보기의 맨 위에 설명이 제목으로 사용됩니다. 조치별 세부사항은 제목 아래에 표시되며 여기에는 생성되거나 제외된 예측변수, 병합되거나 다시 정렬된 범주, 목표 변환, 리캐스트된 필드 및 파생된 예측변수의 수 세부사항이 포함될 수 있습니다.

각 조치가 처리됨에 따라 처리에서 사용되는 예측변수의 수가 변경될 수도 있습니다(예를 들어, 예측변수가 제외되거나 병합됨에 따라).

> ⓘ **참고**: 조치가 꺼져 있거나 목표를 지정하지 않은 경우에 조치 요약 기본 보기에서 조치를 클릭하면 조치 세부사항 위치에 오류 메시지가 표시됩니다.

9개의 가능한 조치가 있지만 분석할 때마다 9개 조치 모두 활성 상태일 필요는 없습니다.

텍스트 필드 테이블

이 테이블에는 다음이 표시됩니다.

- 잘린 후미 공백 값 수
- 분석에서 제외된 예측변수 수

날짜 및 시간 예측변수 테이블

이 테이블에는 다음이 표시됩니다.

- 날짜 및 시간 예측변수에서 파생된 지속 기간 수
- 날짜 및 시간 요소 수
- 파생된 날짜 및 시간 예측변수 총계

날짜 지속 기간이 계산된 경우 참조 날짜 또는 시간이 각주로 표시됩니다.

예측변수 선별 테이블

이 테이블에는 처리에서 제외된 다음 예측변수의 수를 표시합니다.

- 상수
- 결측값이 너무 많은 예측변수
- 단일 범주에 케이스가 너무 많은 예측변수
- 범주가 너무 많은 명목 필드(세트)
- 제외된 예측변수 총계

측정 수준 확인 테이블

이 테이블은 리캐스된 필드 수를 표시하며 다음과 같이 분류됩니다.

- 연속형 필드로 리캐스트된 순서 필드(순서형 세트)
- 순서 필드로 리캐스트된 연속형 필드
- 리캐스트된 총 수

연속형 또는 순서 입력 필드(목표 또는 예측변수)가 없는 경우에는 각주가 표시됩니다.

이상치 테이블

이 테이블은 예상값이 처리된 필드 수를 표시합니다.

- 설정 탭의 입력 및 목표 준비 패널에 지정된 설정에 따라, 이상값이 발견되고 잘린 연속형 필드 수 또는 이상값이 발견되고 결측으로 설정된 연속형 필드 수
- 이상값 처리 후 상수여서 제외된 연속형 필드 수

하나의 각주는 이상값 절사 값을 표시하는 반면, 다른 각주는 연속형 입력 필드(목표 또는 예측변수)가 없는 경우에 표시됩니다.

결측값 테이블

이 테이블은 결측값이 대체된 필드의 수를 표시하며 다음과 같이 분류됩니다.

- 목표. 목표를 지정하지 않은 경우에는 이 행이 표시되지 않습니다.
- 예측변수. 다시 명목(세트), 순서(순서형 세트) 및 연속형의 수로 분류됩니다.
- 대체된 결측값 총계

목표 테이블

이 테이블은 목표가 변환되었는지 여부를 다음과 같이 표시합니다.

- 정규로의 Box-Cox 변환. 이는 다시 지정된 기준(평균 및 표준 편차) 및 람다를 표시하는 열로 분류됩니다.
- 목표 범주가 안정성 향상을 위해 다시 정렬됨

범주형 예측변수 테이블

테이블에 다음과 같은 범주형 예측변수의 수가 표시됩니다.

- 안정성을 향상시키기 위해 해당 범주가 가장 낮은 값에서 가장 높은 값 순으로 다시 정렬된 범주형 예측변수
- 목표와의 연관을 최대화하기 위해 해당 범주가 병합된 범주형 예측변수
- 성긴 범주를 처리하기 위해 해당 범주가 병합된 범주형 예측변수
- 목표와의 연관이 낮아 제외된 범주형 예측변수
- 병합 후 상수여서 제외된 범주형 예측변수

범주형 예측변수가 없는 경우에는 각주가 표시됩니다.

연속형 예측변수 테이블

두 개의 테이블이 있습니다. 첫 번째 테이블에는 다음과 같은 변환의 수 중 하나가 표시됩니다.

- 표준 단위로 변환된 예측변수 값. 또한, 변환된 예측변수 수, 지정된 평균값 및 표준 편차를 표시합니다.
- 공통 범위로 맵핑된 예측변수 값. 또한, 지정된 최소 및 최대 값과 함께 최소-최대 변환을 사용하여 변환된 예측변수 수를 표시합니다.
- 구간화된 예측변수 값 및 구간화된 예측변수 수

두 번째 테이블에는 다음과 같은 예측변수의 수로 표시되는 예측변수 공간 생성 세부사항이 표시됩니다.

- 생성된 예측변수
- 목표와의 연관이 낮아 제외된 예측변수
- 구간화 후 상수여서 제외된 예측변수
- 생성 후 상수여서 제외된 예측변수

연속형 예측변수가 입력되지 않은 경우에는 각주가 표시됩니다.

④ 파생 노드 생성

파생 노드를 생성하는 경우, 파생 노드는 스코어 필드에 역 목표 변환을 적용합니다. 기본적으로, 이 노드는 자동 모델 작성기(예: Auto Classifier 또는 Auto Numeric) 또는 앙상블 노드에 의해 생성될 스코어 필드의 이름을 입력합니다. 척도(범위) 목표가 변환된 경우 스코어 필드는 변환된 단위로 표시됩니다(예: $ 대신 log($)). 결과를 해석하고 사용하려면 예측값을 다시 원래 척도로 변환해야 합니다.

> ⓘ **참고**: ADP 노드에 범위 목표의 척도를 조정하는 분석이 있는 경우(즉, 입력 및 출력 준비 패널에서 Box-Cox 척도 조정을 선택한 경우)에만 파생 노드를 생성할 수 있습니다. 목표가 범위가 아니거나 Box-Cox 척도 조정을 선택하지 않은 경우에는 파생 노드를 생성할 수 없습니다.

파생 노드는 다중 모드에서 작성되고 표현식에서 @FIELD를 사용하므로 필요한 경우 변환된 목표를 추가할 수 있습니다. 예를 들어, 다음과 같은 세부사항을 사용합니다.

- 대상 필드 이름: response
- 변환된 대상 필드 이름: response_transformed
- 스코어 필드 이름: $XR-response_transformed

파생 노드는 새 필드 $XR-response_transformed_inverse를 작성합니다.

> ⓘ **참고**: 자동 모델 작성기 또는 앙상블 노드를 사용하지 않는 경우에는 모델에 알맞은 스코어 필드를 변환하도록 파생 노드를 편집해야 합니다.

정규화된 연속형 대상

기본적으로, 입력 및 목표 준비 패널에서 **Box-Cox 변환으로 연속형 대상 척도 조정** 선택란을 선택하면 목표가 변환되고 사용자는 모델 작성의 목표가 될 새 필드를 작성합니다. 예를 들어, 원래의 목표가 *response*였던 경우 새 목표는 *response_transformed*가 됩니다.

그러나 이 방법은 원래 목표에 따라 여러 가지 문제점이 발생할 수 있습니다. 예를 들어, 목표가 *Age*였던 경우, 새 목표의 값은 *Years*가 아니라 *Years*의 변환된 버전이 됩니다. 이는 스코어가 인식 가능한 단위가 아니기 때문에 스코어를 보고 해석할 수 없음을 의미합니다. 이 경우, 변환된 단위를 원래 의도한 것으로 되돌리는 역변환을 적용할 수 있습니다. 이를 위해서는 다음을 수행하십시오.

1. **데이터 분석**을 클릭하여 ADP 분석을 실행한 후 생성 메뉴에서 *파생 노드*를 선택하십시오.
2. 모델 캔버스에서 사용하는 너깃 다음에 파생 노드를 배치하십시오.

파생 노드는 예측값이 원래의 *Years* 값이 되도록 스코어 필드를 원래 차원으로 복원합니다.

기본적으로 파생 노드는 자동 모델 작성기 또는 앙상블 모형으로 생성된 스코어 필드를 변환합니다. 개별 모델을 작성하는 경우 실제 스코어 필드에서 파생되도록 파생 노드를 편집해야 합니다. 모델을 평가하려면 파생 노드에서 **파생 소스** 필드에 변환된 목표를 추가해야 합니다. 그러면 목표에 동일한 역변환이 적용되며, 다운스트림 평가 또는 분석 노드는 메타데이터 대신 필드 이름을 사용하도록 전환할 경우 변환된 데이터를 올바르게 사용합니다.

원래 이름을 복원하려면 필터 노드를 사용하여 원래 대상 필드(아직 거기에 있는 경우)를 제거하고 목표 및 스코어 필드의 이름을 바꾸십시오.

(3) 유형 노드

필드 특성은 소스 노드 또는 별도의 유형 노드에서 지정할 수 있습니다. 두 노드 모두에서 기능은 비슷합니다. 사용 가능한 특성은 다음과 같습니다.

- **필드** 필드 이름을 두 번 클릭하여 IBM® SPSS® Modeler에서 데이터에 대한 값 및 필드 레이블을 지정하십시오. 예를 들어, IBM SPSS Statistics에서 가져온 필드 메타데이터를 여기서 보거나 수정할 수 있습니다. 마찬가지로 필드 및 해당 값에 대한 새 레이블을 작성할 수 있습니다. 여기서 지정하는 레이블은 스트림 특성 대화 상자에서 작성하는 선택사항에 따라 IBM SPSS Modeler 전체에 표시됩니다.
- **측정** 지정된 필드의 데이터 특성을 설명하는 데 사용되는 측정 수준입니다. 필드의 모든 세부사항이 알려져 있는 경우 이를 **완전히 인스턴스화**되어 있다고 합니다. 추가 정보는 측정 수준의 내용을 참조하십시오.

> ⓘ **참고**: 필드의 측정 수준은 데이터가 문자열, 정수, 실수, 날짜, 시간소인, 목록 중 어느 것으로 저장되는지를 표시하는 해당 저장 유형과 다릅니다.

- **값** 이 열에서는 데이터 세트에서 데이터 값을 읽어오기 위한 옵션을 지정하거나 **지정** 옵션을 사용하여 별도의 대화 상자에서 측정 수준 및 값을 지정할 수 있습니다. 해당 값을 읽지 않고 필드를 통과하도록 선택할 수도 있습니다. 추가 정보는 데이터 값의 내용을 참조하십시오.

> ⓘ **참고**: 해당 **필드** 항목에 목록이 포함되어 있으면 이 열에서 셀을 수정할 수 없습니다.

- **결측** 필드의 결측값이 처리되는 방식을 지정하는 데 사용됩니다. 추가 정보는 결측값 정의의 내용을 참조하십시오.

> ⓘ **참고**: 해당 **필드** 항목에 목록이 포함되어 있으면 이 열에서 셀을 수정할 수 없습니다.

- **확인** 이 열에서는 필드 값이 지정된 값 또는 범위를 준수하는지 확인하는 옵션을 설정할 수 있습니다. 추가 정보는 유형 값 검사의 내용을 참조하십시오.

> ⓘ **참고**: 해당 **필드** 항목에 목록이 포함되어 있으면 이 열에서 셀을 수정할 수 없습니다.

- **역할** 필드가 시스템 학습 프로세스에 대해 **입력**(예측자 필드)인지 아니면 **목표**(예측 필드)인지를 모델링 노드에 알리는 데 사용됩니다. 레코드를 훈련, 검정 및 검증을 위한 별도의 표본으로 파티셔닝하는 데 사용되는 필드를 표시하는 **파티션**과 함께 **모두** 및 **없음** 역할도 사용할 수 있습니다. **분할** 값은 필드의 가능한 각각의 값에 대해 별도의 모델이 작성되도록 지정합니다. 추가 정보는 필드 역할 설정의 내용을 참조하십시오.

유형 노드 창을 사용하여 기타 여러 옵션을 지정할 수 있습니다.

- 도구 메뉴 단추를 사용하여 유형 노드가 인스턴스화(지정, 값 읽기 또는 스트림 실행을 통해)된 후 고유 필드를 무시하도록 선택할 수 있습니다. 고유 필드를 무시하는 경우 값이 하나뿐인 필드가 자동으로 무시됩니다.
- 도구 메뉴 단추를 사용하여 유형 노드가 인스턴스화된 후 **대형 세트를 무시**하도록 선택할 수 있습니다. 대형 세트를 무시하는 경우 다수의 멤버를 포함하는 세트가 자동으로 무시됩니다.
- 도구 메뉴 단추를 사용하여 유형 노드가 인스턴스화된 후 **연속형 정수를 순서로 변환**하도록 선택할 수 있습니다. 자세한 정보는 연속형 데이터 변환의 내용을 참조하십시오.
- 도구 메뉴 단추를 사용하여 선택된 필드를 버리는 필터 노드를 생성할 수 있습니다.
- 선글라스 토글 단추를 사용하여 모든 필드의 기본값을 읽기 또는 패스로 설정할 수 있습니다. 소스 노드의 유형 탭은 기본적으로 필드를 패스하는 반면, 유형 노드 자체는 기본적으로 값을 읽습니다.
- **값 지우기** 단추를 사용하여 이 노드에서 작성된 필드 값 변경사항(비상속 값)을 지우고 업스트림 조작에서 값을 다시 읽을 수 있습니다. 이 옵션은 특정 업스트림 필드에 대해 수행한 변경사항을 재설정하는 데 유용합니다.
- **모든 값 지우기** 단추를 사용하여 노드로 읽어온 **모든** 필드의 값을 재설정할 수 있습니다. 이 옵션은 모든 필드에 대해 효과적으로 **값** 열을 Read로 설정합니다. 이 옵션은 모든 필드의 값을 재설정하고 업스트림 조작에서 값 및 유형을 다시 읽는 데 유용합니다.
- 컨텍스트 메뉴를 사용하여 하나의 필드에서 다른 필드로 속성을 **복사**하도록 선택할 수 있습니다. 자세한 정보는 유형 속성 복사 주제를 참조하십시오.
- **사용하지 않는 필드 설정 보기** 옵션을 사용하여 데이터에 더 이상 존재하지 않거나 이 유형 노드에 연결된 적이 있는 필드의 유형 설정을 볼 수 있습니다. 이 옵션은 변경된 데이터 세트에 유형 노드를 재사용할 때 유용합니다.

① 유형에 대한 정보 보기 및 설정

다양한 소스 노드 및 유형 노드에서 IBM® SPSS® Modeler의 모델링 및 기타 작업에 대해 유용한 필드 메타데이터 및 특성을 지정할 수 있습니다. 이러한 특성에는 다음이 포함됩니다.

- 데이터 세트 내의 각 필드에 대해 범위, 변수군, 정렬된 변수군 또는 플래그 등의 사용 유형을 지정합니다.
- 결측값 및 시스템 널값을 처리하기 위한 옵션을 설정합니다.
- 모델링 목적으로 필드 역할을 설정합니다.
- 데이터 세트에서 자동으로 값을 읽기 위해 사용되는 옵션 및 필드에 대한 값을 지정합니다.
- 필드 및 값 레이블을 지정합니다.

아래 목록에서 사용자의 상황에 맞는 도움말을 선택하십시오.

② 측정 수준

측정 수준(이전의 "데이터 유형" 또는 "사용 유형")은 IBM® SPSS® Modeler에서 데이터 필드가 어떻게 사용되는지 설명합니다. 소스 또는 유형 노드의 유형 탭에서 측정 수준을 지정할 수 있습니다. 예를 들어, 값이 1 및 0인 정수 필드에 대한 측정 수준을 *플래그*로 설정할 수 있습니다. 이는 일반적으로 1 = *True* 및 0 = *False*임을 나타냅니다.

저장 공간 대 측정. 필드의 측정 수준은 해당 저장 유형과 다릅니다. 저장 유형은 데이터가 문자열, 정수, 실수, 날짜, 시간 또는 시간소인으로 저장되는지 여부를 표시합니다. 데이터 유형은 유형 노드를 사용하여 스트림의 어느 위치에서든 수정할 수 있는 반면, 저장 공간은 IBM SPSS Modeler로 데이터를 읽을 때 소스에서 결정해야 합니다(단, 이후에 변환 함수를 사용하여 변경할 수 있음). 자세한 정보는 필드 저장 공간 및 형식화 설정의 내용을 참조하십시오.

일부 모델링 노드는 해당 필드 탭에서 아이콘으로 해당 입력 및 대상 필드에 허용되는 측정 수준 유형을 표시합니다.

측정 수준 아이콘

표 1. 측정 수준 아이콘

아이콘	측정수준
	기본값
	연속
	범주형
	플래그

아이콘	측정수준
	명목
	순서
	유형없음
	요약도표
	지리공간적

사용 가능한 측정 수준은 다음과 같습니다.

- **기본값** 저장 유형 및 값을 알 수 없는 데이터(예를 들어, 아직 읽지 않았기 때문)는 **<기본값>**으로 표시됩니다.
- **연속형** 0-100 또는 0.75-1.25 범위 등의 숫자 값을 설명하는 데 사용됩니다. 연속형 값은 정수, 실수 또는 날짜/시간일 수 있습니다.
- **범주형** 고유 값의 정확한 숫자를 알 수 없는 경우 문자열 값에 사용됩니다. 이는 **인스턴스화 되지 않은** 데이터 유형입니다(데이터의 사용 및 저장에 대해 가능한 모든 정보가 아직 알려져 있지 않음을 의미함). 데이터를 읽고 나면 스트림 특성 대화 상자에서 지정된 명목 필드의 최대 멤버 수에 따라 측정 수준은 *플래그*, *명목* 또는 *유형 없음*이 됩니다.
- **플래그** 특성의 존재 여부를 표시하는 두 개의 고유 값(예: true 및 false, Yes 및 No 또는 0 및 1)을 가진 데이터에 사용됩니다. 사용된 값은 다를 수 있지만 항상 하나의 값은 "참" 값으로 지정하고 다른 하나의 값은 "거짓" 값으로 지정해야 합니다. 데이터는 텍스트, 정수, 실수, 날짜, 시간 또는 시간소인으로 표시될 수 있습니다.
- **명목** 각각 세트의 멤버로 처리되는 다중 고유 값을 가진 데이터(예: small/medium/large)를 설명하는 데 사용됩니다. 명목 데이터는 모든 저장 공간(숫자, 문자열 또는 날짜/시간)을 가질 수 있습니다. 측정 수준을 *명목*으로 설정해도 값이 문자열 저장 공간으로 자동으로 변경되지는 않습니다.
- **순서** 내재된 순서가 있는 다중 고유 값을 가진 데이터를 설명하는 데 사용됩니다. 예를 들어, 급여 범주 또는 만족도 순위를 순서 데이터로 유형화할 수 있습니다. 순서는 데이터 요소의 자연 정렬 순서에 의해 정의됩니다. 예를 들어, 1, 3, 5는 정수 세트에 대한 기본 정렬 순서이고 HIGH, LOW, NORMAL(알파벳 오름차순)은 문자열 세트에 대한 순서입니다. 순서 측정 수준을 사용하면 시각화, 모델 작성 및 순서 데이터를 고유 유형으로 인식하는 다른 애플리케이션(예: IBM SPSS Statistics)에 내보내기를 위해 범주형 데이터 세트를 순서 데이터로 정의할 수 있습니다. 명목 필드를 사용할 수 있는 모든 위치에서 순서 필드를 사용할 수 있습니다. 또한 모든 유형(실수, 정수, 문자열, 날짜, 시간 등)의 필드를 순서로 정의할 수 있습니다.
- **유형 없음** 위 유형을 준수하지 않는 데이터, 단일 값을 가진 필드 또는 세트에 정의된 최대값보다 많은 수의 멤버가 포함된 명목 데이터에 사용됩니다. 이는 그렇지 않으면 측정 수준이 다수의 멤버를 가진 세트(예: 계정 번호)가 되는 경우에도 유용합니다. 필드에 대해 **유형 없음**을 선택하면 역할이 **없음**으로 자동으로 설정되고 **레코드 ID**가 유일한 대안입니다. 세트의 기본 최대 크기는 250개의 고유 값입니다. 이 숫자는 도구 메뉴에서 액세스할 수 있는 스트림 특성 대화 상자의 옵션 탭에서 조정되거나 사용 안함으로 설정될 수 있습니다.

- **콜렉션** 목록에서 기록되는 비지리 공간적 데이터를 식별하는 데 사용됩니다. 콜렉션은 깊이가 0(영)인 목록 필드이며 이 목록에 있는 요소는 다른 측정 수준 중 하나를 가집니다.
- 목록에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.
- **지리 공간적** 지리 공간적 데이터를 식별하기 위해 목록 저장 유형과 함께 사용됩니다. 목록은 깊이가 0과 2 사이(경계값 포함)인 목록을 가진 정수 목록 또는 실수 목록 필드가 될 수 있습니다. 추가 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.

측정 수준을 수동으로 지정하거나, 소프트웨어가 데이터를 읽고 읽은 값을 기반으로 측정 수준을 결정하도록 할 수 있습니다.

선택적으로, 범주형 데이터로 처리해야 하는 연속형 데이터 필드가 있는 경우 이를 변환하는 옵션을 선택할 수 있습니다. 자세한 정보는 연속형 데이터 변환의 내용을 참조하십시오.

자동 입력 사용

1. 유형 노드 또는 소스 노드의 유형 탭에서 원하는 필드의 **값** 열을 **<Read>**로 설정하십시오. 그러면 모든 다운스트림 노드에서 메타데이터를 사용할 수 있습니다. 대화 상자의 선글라스 단추를 사용하여 모든 필드를 신속하게 **<Read>** 또는 **<Pass>**로 설정할 수 있습니다.
2. 데이터 소스에서 바로 값을 읽으려면 **값 읽기**를 클릭하십시오.

필드의 측정 수준을 수동으로 설정

1. 테이블에서 필드를 선택하십시오.
2. **측정** 열의 드롭 다운 목록에서 필드의 측정 수준을 선택하십시오.
3. 또는, Ctrl-A 또는 Ctrl-클릭을 사용하여 여러 필드를 선택한 후 드롭 다운 목록을 사용하여 측정 수준을 선택할 수 있습니다.

가. 지리 공간적 측정 수준

목록 저장 유형과 함께 사용되는 지리 공간적 측정 수준에는 다양한 유형의 지리 공간적 데이터를 식별하는 데 사용되는 6개의 하위 수준이 있습니다.

- **점** - 특정 지점(예: 도시의 중심)을 식별합니다.
- **다각형** - 지역의 단일 경계와 해당 위치를 식별하는 일련의 점입니다(예: County).
- **LineString** - 폴리라인 또는 라인이라고도 하는 LineString은 선의 경로를 식별하는 일련의 점입니다. 예를 들어, LineString은 도로, 강, 철도 등의 고정 항목이거나 비행기의 비행 경로 또는 배의 항로와 같은 움직이는 물체의 이동 경로일 수 있습니다.

- **다중 점** - 데이터의 각 행에 지역당 여러 지점이 포함될 때 사용됩니다. 예를 들어, 각 행이 도시 거리를 나타내는 경우, 각 거리의 여러 지점이 모든 가로등을 식별하는 데 사용될 수 있습니다.
- **복수 다각형** - 데이터의 각 행에 여러 다각형이 포함될 때 사용됩니다. 예를 들어, 각 행이 국가의 윤곽선을 나타내는 경우, 미국은 본토, 알래스카, 하와이 등의 다양한 영역을 식별하는 여러 다각형으로 기록될 수 있습니다.
- **다중 LineString** - 데이터의 각 행에 여러 선이 포함될 때 사용됩니다. 선은 갈라질 수 없기 때문에 다중 LineString을 사용하여 선 그룹을 식별할 수 있습니다. 예를 들어, 각 국가의 철도망 또는 가항 수로와 같은 데이터입니다.

이러한 측정 하위 수준은 목록 저장 유형과 함께 사용됩니다. 추가 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

제한사항

지리 공간적 데이터를 사용할 때는 몇 가지 제한사항에 주의해야 합니다.

- 좌표계는 데이터의 형식에 영향을 줄 수 있습니다. 예를 들어, 평면직각 좌표계는 좌표값 x, y 및 z(필요한 경우)를 사용하는 반면, 지리 좌표계는 좌표값 경도 및 위도와 (필요한 경우) 고도 또는 깊이 값을 사용합니다.
  좌표계에 대한 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오.
- LineString은 자체를 가로지를 수 없습니다.
- 다각형은 자동으로 닫히지 않습니다. 다각형마다 첫 번째 및 마지막 점을 동일한 점으로 정의해야 합니다.
- 복수 다각형의 데이터 방향이 중요합니다. 시계방향은 단단한 구조를 표시하고 반시계 방향은 비어 있는 구조를 표시합니다. 예를 들어, 호수를 포함하는 국가의 영역을 기록하는 경우, 본토 영역 경계는 시계방향으로 기록하고 각 호수의 형태는 반시계 방향으로 기록할 수 있습니다.
- 다각형은 그 자체와 교차할 수 없습니다. 이러한 교차의 예는 다각형의 경계를 숫자 8 양식의 연속 선으로 구성하려는 경우입니다.
- 복수 다각형은 서로 겹칠 수 없습니다.
- 지리 공간적 필드의 경우, 관련된 유일한 저장 유형은 **실수** 및 **정수**입니다(기본 설정은 **실수**임).

지리 공간적 측정 하위 수준 아이콘

표 1. 지리 공간적 측정 하위 수준 아이콘

아이콘	측정수준
	점
	다각형

아이콘	측정수준
	선문자열
	다중점
	복수다각형
	다중선문자열

③ 연속형 데이터 변환

범주형 데이터를 연속형으로 처리하면 모델(특히, 목표 필드인 경우)의 품질에 심각한 영향을 미칠 수 있습니다(예를 들어, 2진 모형이 아닌 회귀 모형을 생성함). 이를 방지하기 위해 정수 범위를 범주형 유형(예: *순서* 또는 *플래그*)으로 변환할 수 있습니다.

1. 조작 및 메뉴 생성 단추(도구 기호가 있음)에서 **연속형 정수를 순서로 변환**을 선택하십시오. 변환 값 대화 상자가 표시됩니다.
2. 자동으로 변환되는 범위 크기를 지정하십시오. 이는 입력하는 크기까지의 모든 범위에 적용됩니다.
3. **확인**을 클릭하십시오. 적용되는 범위가 *플래그* 또는 *순서*로 변환되고 유형 노드의 유형 탭에 표시됩니다.

변환 결과

- 정수 저장 공간을 포함하는 연속형 필드가 순서로 변경되는 경우, 최소값과 최대값이 펼쳐져 최소값에서 최대값까지의 모든 정수 값이 포함됩니다. 예를 들어, 범위가 1, 5이면 값 세트는 1, 2, 3, 4, 5입니다.
- *연속형* 필드가 *플래그*로 변경되는 경우, 최소값 및 최대값은 플래그 필드의 false 및 true 값이 됩니다.

④ 인스턴스화란?

인스턴스화는 데이터 필드의 저장 유형 및 값과 같은 정보를 읽거나 지정하는 프로세스입니다. 인스턴스화는 시스템 자원을 최적화하기 위한 사용자 지시 프로세스입니다. 소스 노드의 유형 탭에서 옵션을 지정하거나 유형 노드를 통해 데이터를 실행하여 소프트웨어에 값을 읽도록 지시하십시오.

- 유형을 알 수 없는 데이터도 인스턴스화되지 않음으로 참조됩니다. 해당 저장 유형 및 값을 알 수 없는 데이터는 유형 탭의 *측정* 열에 **<기본값>**으로 표시됩니다.
- 필드의 저장 공간(예: 문자열 또는 숫자)에 대한 일부 정보가 있는 경우 해당 데이터는 *부분적으로 인스턴스화됨*이라고 합니다. **범주형** 또는 **연속형**은 부분적으로 인스턴스화된 측정 수준입

니다. 예를 들어, **범주형**은 필드가 기호 필드임을 지정하지만 명목, 순서 및 플래그 필드 중 어느 것인지 알 수 없습니다.
- 값을 포함하여 유형에 대한 모든 세부사항을 알고 있는 경우, *완전히 인스턴스화됨* 측정 수준(명목, 순서, 플래그 또는 연속형)이 이 열에 표시됩니다. *연속형* 유형은 부분적으로 인스턴스화된 데이터 필드 및 완전히 인스턴스화된 데이터 필드 모두에 사용됩니다. 연속형 데이터는 정수 또는 실수일 수 있습니다.

유형 노드가 있는 데이터 스트림의 실행 동안, 인스턴스화되지 않은 유형은 초기 데이터 값을 기반으로 즉시 부분적으로 인스턴스화됩니다. 모든 데이터가 노드를 통과하면, 값을 **<Pass>**로 설정하지 않은 한, 모든 데이터가 완전히 인스턴스화됩니다. 실행이 중단되면 데이터는 부분적으로 인스턴스화된 상태로 유지됩니다. 유형 탭이 인스턴스화되면 필드의 값은 스트림의 해당 위치에서 정적입니다. 이는 스트림을 다시 실행해도 업스트림 변경사항은 특정 필드의 값에 영향을 주지 않음을 의미합니다. 새 데이터 또는 추가된 조작을 기반으로 값을 변경하거나 업데이트하려면 유형 탭 자체에서 값을 편집하거나 필드의 값을 **<Read>** 또는 **<Read +>**로 설정해야 합니다.

인스턴스화 시기

일반적으로, 데이터 세트가 매우 크지 않고 나중에 스트림에 필드를 추가할 계획이 아닌 경우에는 소스 노드에서 인스턴스화하는 것이 가장 편리한 방법입니다. 그러나 다음 경우에는 별도의 유형 노드에서 인스턴스화하는 것이 유용합니다.

- 데이터 세트가 크고 스트림이 해당 유형 노드 이전에 서브세트를 필터링합니다.
- 스트림에서 데이터가 필터링되었습니다.
- 스트림에서 데이터가 병합되었거나 추가되었습니다.
- 처리 중에 새 데이터 필드가 파생됩니다.

> ⓘ **참고**: 데이터베이스 내보내기 노드에서 데이터를 내보내는 경우 데이터를 완전히 인스턴스화해야 합니다.

⑤ 데이터 값

유형 탭의 **값** 열을 사용하여 데이터에서 자동으로 값을 읽거나 별도의 대화 상자에서 측정 수준 및 값을 지정할 수 있습니다.

값 드롭 다운 목록에서 사용 가능한 옵션은 다음 표에 표시된 바와 같이 자동 입력을 위한 명령어를 제공합니다.

표 1. 자동 입력을 위한 명령어

옵션	기능
<Read>	노드가 실행될 때 데이터를 읽습니다.
<Read+>	데이터를 읽어 현재 데이터(있는 경우)에 추가합니다.
<Pass>	데이터를 읽지 않습니다.
<Current>	현재 데이터 값을 유지합니다.
지정...	값과 및 측정 수준 옵션을 지정하는 별도의 대화 상자가 열립니다.

유형 노드를 실행하거나 **값 읽기**를 클릭하면 자동 입력이 수행되고 선택항목을 기반으로 데이터 소스에서 값을 읽습니다. 이러한 값은 지정 옵션을 사용하거나 필드 열의 셀을 두 번 클릭하여 수동으로 지정할 수도 있습니다.

유형 노드에서 필드에 대한 변경을 수행한 후에는 대화 상자 도구 모음의 다음 단추를 사용하여 값 정보를 재설정할 수 있습니다.

- **값 지우기** 단추를 사용하여 이 노드에서 작성된 필드 값 변경사항(비상속 값)을 지우고 업스트림 조작에서 값을 다시 읽을 수 있습니다. 이 옵션은 특정 업스트림 필드에 대해 수행한 변경사항을 재설정하는 데 유용합니다.
- **모든 값 지우기** 단추를 사용하여 노드로 읽어온 모든 필드의 값을 재설정할 수 있습니다. 이 옵션은 모든 필드에 대해 효과적으로 *값* 열을 **Read**로 설정합니다. 이 옵션은 모든 필드의 값을 재설정하고 업스트림 조작에서 값 및 측정 수준을 다시 읽는 데 유용합니다.

값 열의 회색 텍스트

유형 노드 또는 소스 노드 내에서 **값** 열의 데이터가 검은색 텍스트로 표시된다면 필드 값을 읽어 해당 노드에 저장된 것입니다. 이 필드에 검은색 텍스트가 표시되지 않는다면 해당 필드의 값을 읽지 않아 이후 업스트림으로 판별된 것입니다.
데이터가 회색 텍스트로 표시되는 경우가 있습니다. SPSS® Modeler가 실제로 데이터를 읽고 저장하지 않아도 필드의 유효한 값을 식별하거나 유추할 수 있는 경우에 이와 같이 표시됩니다. 다음 노드 중 하나를 사용하는 경우에 이러한 상황이 발생할 수 있습니다.

- 사용자 입력 노드. 데이터는 노드 내에 정의되므로 값이 노드에 저장되지 않았더라도 필드의 값 범위는 항상 알려져 있습니다.
- 통계 파일 소스 노드. 데이터 유형에 대한 메타데이터가 있을 경우 데이터를 읽거나 저장하지 않아도 SPSS Modeler가 가능한 값 범위를 유추할 수 있습니다.

이 노드에서는 **값 읽기**를 클릭할 때까지 값이 회색 텍스트로 표시됩니다.

> **경고**: 스트림의 데이터를 인스턴스화하지 않았는데 데이터 값이 회색으로 표시될 경우 **검사** 열에서 설정한 유형 값 검사가 적용되지 않습니다.

가. 값 대화 상자 사용

유형 탭의 **값** 또는 **결측** 열을 클릭하면 사전 정의된 값의 드롭 다운 목록이 표시됩니다. 이 목록에서 **지정...** 옵션을 선택하면 선택된 필드에 대한 값을 읽고 지정하고 레이블 지정하고 처리하기 위한 옵션을 설정할 수 있는 별도의 대화 상자가 열립니다.

다수의 제어는 모든 유형의 데이터에 대해 공통입니다. 이 공통 제어에 대해 여기서 설명합니다.

측정 현재 선택된 측정 수준을 표시합니다. 데이터를 사용할 방식을 반영하도록 설정을 변경할 수 있습니다. 예를 들어, day_of_week라는 필드에 개별 날짜를 나타내는 숫자가 포함되어 있는 경우에는 각각의 범주를 개별적으로 조사하는 분포 노드를 작성하기 위해 이를 명목 데이터로 변경할 수 있습니다.

저장 공간 저장 유형을 표시합니다(알려진 경우). 저장 유형은 선택하는 측정 수준의 영향을 받지 않습니다. 저장 유형을 변경하기 위해 고정 파일 및 가변파일 소스 노드에서 데이터 탭을 사용하거나 채움 노드에서 변환 함수를 사용할 수 있습니다.

모델 필드 모델 너깃 스코어링의 결과로 생성되는 필드의 경우 모델 필드 세부사항도 볼 수 있습니다. 이 세부사항에는 대상 필드의 이름과 모델링에서 필드의 역할(예측값, 확률, 성향 등)이 포함됩니다.

값 선택된 필드에 대한 값을 판별할 방법을 선택하십시오. 여기서 작성하는 선택사항은 유형 노드 대화 상자의 **값** 열에서 이전에 작성한 선택사항을 대체합니다. 값을 읽기 위한 선택사항은 다음과 같습니다.

- 데이터에서 읽기 노드가 실행될 때 값을 읽으려면 선택하십시오. 이 옵션은 <읽기>와 동일합니다.
- 통과 현재 필드에 대한 데이터를 읽지 않으려면 선택하십시오. 이 옵션은 <통과>와 동일합니다.
- 값 및 레이블 지정 여기의 옵션은 선택된 필드에 대한 값 및 레이블을 지정하는 데 사용됩니다. 값 확인과 함께 사용하여 이 옵션을 통해 현재 필드에 대한 사용자의 지식을 기반으로 하는 값을 지정하십시오. 이 옵션은 각 유형의 필드에 대한 고유 제어를 활성화합니다. 값 및 레이블에 대한 옵션은 후속 주제에서 개별적으로 다룹니다.

> **참고**: 측정 수준이 유형 없음 또는 <기본값>인 필드에 대해서는 값 또는 레이블을 지정할 수 없습니다.

- **데이터의 값 확장** 여기서 입력하는 값을 사용하여 현재 데이터를 추가하려면 선택하십시오. 예를 들어, field_1의 범위가 (0,10)에서 시작하는 경우 (8,16)에서 시작하는 값 범위를 입력하면 원래 최소값은 제거하지 않고 16을 추가하여 범위가 확장됩니다. 새 범위는 (0,16)입니다. 이 옵션을 선택하면 자동 입력 옵션이 **<읽기+>**로 자동으로 설정됩니다.

최대 목록 길이 측정 수준이 지리 공간 또는 콜렉션인 데이터에만 사용할 수 있습니다. 목록이 포함할 수 있는 요소 수를 지정하여 목록의 최대 길이를 설정하십시오.

최대 문자열 길이 유형이 없는 데이터에만 사용할 수 있습니다. 테이블을 작성하기 위해 SQL을 생성할 때 이 필드를 사용하십시오. 데이터에서 가장 큰 문자열 값을 입력하십시오. 그러면 문자열을 포함하기에 충분한 열이 테이블에 생성됩니다. 문자열 길이 값을 사용할 수 없으면 데이터에 적합하지 않은 기본 문자열 값이 사용됩니다. 예를 들어, 값이 너무 작으면 테이블에 데이터를 쓸 때 오류가 발생할 수 있으며, 값이 너무 크면 성능에 불리한 영향을 미칠 수 있습니다.

값 확인 지정된 연속형, 플래그 또는 명목 값을 준수하도록 값을 강제하는 방법을 선택하십시오. 이 옵션은 유형 노드 대화 상자의 **확인** 열에 해당하며 여기서 작성되는 설정은 해당 대화 상자의 설정을 대체합니다. **값 및 레이블 지정** 옵션과 함께 사용되면 값 확인을 통해 예측값을 가진 데이터에서 값을 준수할 수 있습니다. 예를 들어, 값을 1, 0으로 지정한 후 **삭제** 옵션을 사용하는 경우 값이 1 또는 0이 아닌 모든 레코드를 삭제할 수 있습니다.

공백 정의 데이터에서 결측값 또는 공백을 선언하는 데 사용하는 다음과 같은 제어를 활성화하려면 선택하십시오.

- **결측값** 특정 값(예: 99 또는 0)을 공백으로 정의하려면 이 테이블을 사용하십시오. 값은 필드의 저장 유형에 대해 적합해야 합니다.
- **범위** 결측값의 범위(예: 연령 1-17 또는 65 초과)를 지정하는 데 사용됩니다. 경계값이 비어 있으면 범위는 한정되지 않습니다. 예를 들어, 상한 없이 하한으로 100이 지정되면 100 이상의 값은 모두 결측값으로 정의됩니다. 경계값도 포함됩니다. 예를 들어, 하한이 5이고 상한이 10인 범위의 범위 정의에는 5와 10이 포함됩니다. 결측값 범위는 날짜/시간 및 문자열(이 경우 값이 범위 내에 있는지 판별하기 위해 알파벳 정렬 순서가 사용됨)을 포함한 모든 저장 유형에 대해 정의할 수 있습니다.
- **널/공백** 시스템 널(데이터에서 $null$로 표시됨) 및 공백(표시되는 문자가 없는 문자열 값)을 공백으로 지정할 수도 있습니다.

> ⓘ **참고**: 유형 노드에서는 분석을 위해 빈 문자열도 공백으로 처리하지만 빈 문자열은 내부적으로 다르게 저장되며 특정 케이스에서 다르게 처리될 수 있습니다.

> ⓘ **참고**: 공백을 정의되지 않음 또는 $null$로 코딩하려면 채움 노드를 사용하십시오.

설명 필드 레이블을 지정하려면 이 텍스트 상자를 사용하십시오. 이 레이블은 스트림 특성 대화 상자에서 작성하는 선택사항에 따라 그래프, 테이블, 출력, 모델 브라우저 등의 다양한 위치에 표시됩니다.

나. 연속형 데이터의 값 및 레이블 지정

숫자 필드에는 *연속형* 측정 수준이 사용됩니다. 연속형 데이터의 경우 세 가지 저장 유형이 있습니다.

- 실수
- 정수
- 날짜/시간

모든 연속형 필드를 편집하는 데 동일한 대화 상자가 사용됩니다. 저장 유형은 참조용으로만 표시됩니다.

값 지정

다음 제어는 연속형 필드에 고유하고 값 범위를 지정하는 데 사용됩니다.

하한. 값 범위의 하한을 지정하십시오.

상한. 값 범위의 상한을 지정하십시오.

레이블 지정

범위 필드의 값에 레이블을 지정할 수 있습니다. **레이블** 단추를 클릭하여 값 레이블을 지정하기 위한 별도의 대화 상자를 여십시오.

ㄱ. 값 및 레이블 하위 대화 상자

범위 필드에 대한 값 대화 상자에서 **레이블**을 클릭하면 범위의 값에 대한 레이블을 지정할 수 있는 새 대화 상자가 열립니다.

이 테이블의 *값* 및 *레이블* 열을 사용하여 값 및 레이블 쌍을 정의할 수 있습니다. 현재 정의된 쌍이 여기에 표시됩니다. 비어 있는 셀을 클릭한 후 값 및 해당 레이블을 입력하여 새 레이블 쌍을 추가할 수 있습니다. *참고*: 이 테이블에 값/값-레이블 쌍을 추가해도 새 값이 필드에 추가되지는 않습니다. 대신 필드 값에 대한 메타데이터만 작성됩니다.

유형 노드에서 지정하는 레이블은 스트림 특성 대화 상자에서 작성하는 선택사항에 따라 많은 위치에(도구 팁, 출력 레이블 등으로) 표시됩니다.

다. 명목 및 순서 데이터의 값 및 레이블 지정

명목(세트) 및 순서(순서 지정된 세트) 측정 수준은 데이터 값이 세트의 멤버로서 따로따로 사용됨을 표시합니다. 세트의 저장 유형은 문자열, 정수, 실수 또는 날짜/시간입니다.

다음 제어는 명목 및 순서 필드에 고유하고 값 및 레이블을 지정하는 데 사용됩니다.

값. 테이블의 *값* 열을 사용하여 현재 필드에 대한 지식을 기반으로 값을 지정할 수 있습니다. 이 테이블을 사용하여 필드에 예상되는 값을 입력하고 값 검사 드롭 다운 목록을 사용하여 데이터 세트가 이러한 값과 일치하는지 검사할 수 있습니다. 화살표와 삭제 단추를 사용하면 기존 값을 수정할 뿐만 아니라 값을 다시 정렬하고 삭제할 수 있습니다.

레이블. *레이블* 열을 사용하여 세트에 있는 각 값에 대해 레이블을 지정할 수 있습니다. 이러한 레이블은 스트림 특성 대화 상자에서 선택하는 항목에 따라 그래프, 테이블, 출력 및 모델 브라우저 등의 다양한 위치에 표시됩니다.

라. 플래그의 값 지정

플래그 필드는 두 개의 고유 값을 갖는 데이터를 표시하는 데 사용됩니다. 플래그의 저장 유형은 문자열, 정수, 실수 또는 날짜/시간입니다.

참. 조건이 충족될 때 필드의 플래그 값을 지정합니다.

거짓. 조건이 충족되지 않을 때 필드의 플래그 값을 지정합니다.

레이블. 플래그 필드의 값마다 레이블을 지정합니다. 이러한 레이블은 스트림 특성 대화 상자에서 선택하는 항목에 따라 그래프, 테이블, 출력 및 모델 브라우저 등의 다양한 위치에 표시됩니다.

마. 콜렉션 데이터의 값 지정

콜렉션 필드는 목록에 있는 비지리 공간적 데이터를 표시하는 데 사용됩니다.
콜렉션 **측정** 수준에 대해 설정할 수 있는 유일한 항목은 **목록 측도**입니다. 기본적으로 이 측도는 유형 없음으로 설정되지만, 다른 값을 선택하여 목록 내에 있는 요소의 측정 수준을 설정할 수 있습니다. 다음 옵션 중 하나를 선택할 수 있습니다.

- 유형 없음
- 연속
- 명목
- 순서
- 플래그

바. 지리 공간 데이터의 값 지정

지리 공간적 필드는 목록에 있는 지리 공간적 데이터를 표시하는 데 사용됩니다.

지리 공간적 **측정** 수준의 경우, 다음 옵션을 설정하여 목록에 있는 요소의 측정 수준을 설정할 수 있습니다.

유형 지리 공간적 필드의 측정 하위 수준을 선택하십시오. 사용 가능한 하위 수준은 목록 필드의 깊이로 결정됩니다. 기본값은 Point(깊이 0), LineString(깊이 1) 및 Polygon(깊이 1)입니다.

하위 수준에 대한 자세한 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.

목록 깊이에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

좌표계 이 옵션은 측정 수준을 비지리 공간적 수준에서 지리 공간적 수준으로 변경한 경우에만 사용할 수 있습니다. 지리 공간적 데이터에 좌표계를 적용하려면 이 선택란을 선택하십시오. 기본적으로, **도구 > 스트림 특성 > 옵션 > 지리 공간적** 분할창에서 설정된 좌표계가 표시됩니다. 다른 좌표계를 사용하려면, **변경** 단추를 클릭하여 좌표계 선택 대화 상자를 표시하고 필요한 좌표계를 선택하십시오.

좌표계에 대한 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오.

⑥ 결측값 정의

유형 탭의 **결측** 열은 필드에 대해 결측값 처리가 정의되었는지 여부를 표시합니다. 가능한 설정은 다음과 같습니다.

On(*). 이 필드에 대해 결측값 처리가 정의되어 있음을 표시합니다. 지정 옵션(아래 참조)을 사용하여 명시적으로 지정하거나 다운스트림 채움 노드를 통해 결측값 처리를 정의할 수 있습니다.
Off. 필드에 대해 결측값 처리가 정의되어 있지 않습니다.

지정. 이 필드에 대해 결측값으로 간주할 명시적 값을 선언할 수 있는 대화 상자를 표시하려면 이 옵션을 선택하십시오.

⑦ 유형 값 검사

각 필드에 대해 검사 옵션을 사용하면 해당 필드의 모든 값을 조사하여 값이 현재 유형 설정을 따르는지 또는 값 지정 대화 상자에 지정한 값을 따르는지 판별합니다. 이 옵션은 단일 조작 내에서 데이터 세트를 정리하고 데이터 세트의 크기를 줄이는 데 유용합니다.

유형 노드 대화 상자에 있는 *검사* 열의 설정은 유형 한계를 벗어난 값이 발견될 때 발생하는 상황을 결정합니다. 필드에 대한 검사 설정을 변경하려면 *검사* 열에서 해당 필드에 대한 드롭 다운 목록을 사용하십시오. 모든 필드에 대한 검사 설정을 설정하려면 *필드* 열에서 클릭하고 Ctrl-A를 누르십시오. 그런 다음 검사 열에서 임의 필드에 대한 드롭 다운 목록을 사용하십시오.

다음 검사 설정을 사용할 수 있습니다.

없음. 값이 검사 없이 전달됩니다. 기본 설정입니다.

무효화. 한계를 벗어난 값을 시스템 널($null$)로 변경합니다.

강제 적용. 해당 측정 수준이 완전히 인스턴스화된 필드에 지정된 범위를 벗어난 값이 있는지 조사합니다. 지정되지 않은 값은 다음 규칙에 따라 해당 측정 수준에 유효한 값으로 변환됩니다.

- 플래그의 경우, true 및 false 값 이외의 값은 false 값으로 변환됩니다.
- 세트(명목 또는 순서)의 경우, 알 수 없는 값은 세트 값의 첫 번째 멤버로 변환됩니다.
- 범위의 상한보다 큰 숫자는 상한으로 대체됩니다.
- 범위의 하한보다 작은 숫자는 하한으로 대체됩니다.
- 범위 내의 널값에는 해당 범위의 중간 값이 제공됩니다.

삭제. 유효하지 않은 값이 있으면 전체 레코드가 삭제됩니다.

경고. 모든 데이터를 읽은 후에는 스트림 특성 대화 상자에서 유효하지 않은 항목 수가 계산되고 보고됩니다.

중단. 유효하지 않은 값이 처음 발견될 때 스트림 실행이 종료됩니다. 스트림 특성 대화 상자에서 오류가 보고됩니다.

⑧ 필드 역할 설정

필드의 역할은 모델 작성에 필드가 사용되는 방법을 지정합니다. 예를 들어, 필드가 입력인지 또는 목표(예측 목표)인지 여부입니다.

ⓘ **참고**: 파티션, 빈도 및 레코드 ID 역할은 각각 단일 필드에만 적용할 수 있습니다.

다음 역할을 사용할 수 있습니다.

입력. 필드가 시스템 학습에 대한 입력으로 사용됩니다(예측자 필드).

목표: 필드가 시스템 학습의 출력 또는 목표로 사용됩니다(모델이 예측하려는 필드 중 하나).

둘 다. 필드가 Apriori 노드에서 입력 및 출력 둘 다로 사용됩니다. 기타 모든 모델링 노드에서는 해당 필드를 무시합니다.

없음. 필드가 시스템 학습에서 무시됩니다. 해당 측정 수준이 **유형 없음**으로 설정된 필드는 **역할** 열에서 자동으로 **없음**으로 설정됩니다.

파티션. 훈련, 테스트 및 검증을 위해 데이터를 개별 표본으로 파티셔닝하는 데 사용되는 필드를 표시합니다. 필드는 두 개 또는 세 개의 가능한 값(필드 값 대화 상자에 정의된)을 갖는 인스턴스화된 세트 유형이어야 합니다. 첫 번째 값은 훈련 표본을 나타내고 두 번째 값은 테스트 표본을 나타내며 세 번째 값(있는 경우)은 검증 표본을 나타냅니다. 추가 값은 무시되며 플래그 필드를 사용할 수 없습니다. 분석에서 파티션을 사용하려면, 해당 모델 작성 또는 분석 노드의 모델 옵션 탭에서 파티셔닝을 사용 가능하게 설정해야 합니다. 파티셔닝이 사용될 때, 파티션 필드에 대해 널값을 갖는 레코드는 분석에서 제외됩니다. 스트림에 여러 파티션 필드를 정의한 경우, 적용 가능한 각 모델링 노드의 필드 탭에서 단일 파티션 필드를 지정해야 합니다. 데이터에 아직 적합한 필드가 없으면 파티션 노드 또는 파생 노드를 사용하여 필드를 작성할 수 있습니다. 자세한 정보는 파티션 노드 주제를 참조하십시오.

분할. (명목, 순서 및 플래그 필드만 해당) 필드의 가능한 값마다 모델이 작성되도록 지정합니다.

빈도. (숫자 필드만 해당) 이 역할을 설정하면 필드 값을 레코드의 빈도 가중치로 사용할 수 있습니다. 이 기능은 C&R 트리, CHAID, QUEST 및 선형 모델에서만 지원됩니다. 기타 모든 노드에서는 이 역할을 무시합니다. 빈도 가중치는 이 기능을 지원하는 모델링 노드의 필드 탭에서 **빈도 가중치 사용** 옵션을 사용하여 사용합니다.

레코드 ID. 필드가 고유 레코드 식별자로 사용됩니다. 이 기능은 대부분의 노드에서 무시되지만 선형 모형에서 지원되고 IBM Netezza In-Database 마이닝 노드에 필요합니다.

⑨ 유형 속성 복사

하나의 필드에서 다른 필드로 유형의 속성(예: 값, 검사 옵션 및 결측값)을 쉽게 복사할 수 있습니다.

1. 해당 속성을 복사할 필드에서 마우스 오른쪽 단추를 클릭하십시오.
2. 컨텍스트 메뉴에서 **복사**를 선택하십시오.
3. 해당 속성을 변경할 필드에서 마우스 오른쪽 단추를 클릭하십시오.
4. 컨텍스트 메뉴에서 **특수 속성 붙여넣기**를 선택하십시오. *참고*: Ctrl-클릭 방법을 사용하거나 컨텍스트 메뉴에서 **필드 선택** 옵션을 사용하여 여러 필드를 선택할 수 있습니다.

새 대화 상자가 열리고 여기서 붙여넣을 특정 속성을 선택할 수 있습니다. 여러 필드에 붙여넣는 경우, 여기서 선택하는 옵션은 모든 목표 필드에 적용됩니다.

다음 속성을 붙여넣으십시오. 아래의 목록에서 선택하여 하나의 필드에서 다른 필드로 속성을 붙여넣으십시오.

- **유형**. 측정 수준을 붙여넣으려면 선택하십시오.
- **값**. 필드 값을 붙여넣으려면 선택하십시오.
- **결측**. 결측값 설정을 붙여넣으려면 선택하십시오.
- **검사**. 값 검사 옵션을 붙여넣으려면 선택하십시오.
- **역할**. 필드의 역할을 붙여넣으려면 선택하십시오.

⑩ 필드 형식 설정 탭

테이블 및 유형 노드의 형식 탭에는 현재 또는 사용하지 않은 필드와 각 필드에 대한 형식 옵션의 목록이 표시됩니다. 필드 형식 필드의 각 열에 대한 설명은 다음과 같습니다.

필드. 선택된 필드의 이름을 표시합니다.

형식. 이 열의 셀을 두 번 클릭하면 열리는 대화 상자를 사용하여 필드에 대한 형식을 개별적으로 지정할 수 있습니다. 자세한 정보는 필드 형식 옵션 설정의 내용을 참조하십시오. 여기서 지정된 형식은 전체 스트림 특성에서 지정된 형식을 대체합니다.

참고: Statistics 내보내기 및 Statistics 출력 노드는 필드별 형식을 해당 메타데이터에 포함하는 .sav 파일을 내보냅니다. IBM® SPSS® Statistics .sav 파일 형식에서 지원하지 않는 필드별 형식이 지정된 경우 노드는 IBM SPSS Statistics 기본 형식을 사용합니다.

맞춤. 테이블 열 내에서 값을 맞추는 방식을 지정하려면 이 열을 사용하십시오. 기본 설정은 **자동**이며 이는 기호 값은 왼쪽으로 맞추고 숫자 값은 오른쪽으로 맞춥니다. **왼쪽**, **오른쪽** 또는 **가운데**를 선택하여 기본값을 대체할 수 있습니다.
열 너비. 기본적으로 열 너비는 필드의 값을 기반으로 자동으로 계산됩니다. 자동 너비 계산을 대체하려면 테이블 셀을 클릭한 후 드롭 다운 목록을 사용하여 새 너비를 선택하십시오. 여기에 나열되지 않는 사용자 정의 너비를 입력하려면 필드 또는 형식 열에서 테이블 셀을 두 번 클릭하여 필드 형식 하위 대화 상자를 여십시오. 또는 셀을 마우스 오른쪽 단추로 클릭한 후 **형식 설정**을 선택할 수 있습니다.

현재 필드 보기. 기본적으로 대화 상자에는 현재 활성 필드의 목록이 표시됩니다. 사용하지 않은 필드의 목록을 보려면 **사용하지 않은 필드 설정 보기**를 선택하십시오.

컨텍스트 메뉴. 이 탭에 대한 컨텍스트 메뉴는 다양한 선택사항 및 설정 업데이터 옵션을 제공합니다. 열을 마우스 오른쪽 단추로 클릭하여 이 메뉴를 표시하십시오.

- **모두 선택**. 모든 필드를 선택합니다.

- **선택 안함**. 선택사항을 선택 취소합니다.
- **필드 선택**. 유형 또는 저장 공간 특성을 기반으로 필드를 선택합니다. 옵션은 **범주형 선택**, 연속형 선택(숫자), **유형 없음 선택**, **문자열 선택**, **숫자 선택** 또는 **날짜/시간 선택**입니다. 자세한 정보는 측정 수준의 내용을 참조하십시오.
- **형식 설정**. 필드별로 날짜. 시간 및 소수점 옵션을 지정하는 데 필요한 하위 대화 상자를 엽니다.
- **맞춤 설정**. 선택된 필드에 대한 맞춤을 설정합니다. 옵션은 **자동**, **가운데**, **왼쪽** 또는 **오른쪽**입니다.
- **열 너비 설정**. 선택된 필드의 필드 너비를 설정합니다. 데이터에서 너비를 읽어오려면 **자동**을 지정하십시오. 또는 필드 너비를 5, 10, 20, 30, 50, 100 또는 200으로 설정할 수 있습니다.

가. 필드 형식 옵션 설정

필드 형식은 유형 및 테이블 노드의 형식 탭에서 사용할 수 있는 하위 대화 상자에서 지정됩니다. 이 대화 상자를 열기 전에 둘 이상의 필드를 선택한 경우에는 선택사항 첫 번째 필드의 설정이 모두에 대해 사용됩니다. 여기서 지정한 후 **확인**을 클릭하면 이 설정이 형식 탭에서 선택된 모든 필드에 적용됩니다.

필드별로 다음과 같은 옵션을 사용할 수 있습니다. 이들 옵션 중 다수는 스트림 특성 대화 상자에서도 지정할 수 있습니다. 필드 수준에서 작성된 설정은 스트림에 대해 지정된 기본값을 대체합니다.

날짜 형식. 날짜 저장 공간 필드에 사용하거나 문자열이 CLEM 날짜 함수에 의해 날짜로 해석될 때 사용할 날짜 형식을 선택합니다.

시간 형식. 시간 저장 공간 필드에 사용하거나 문자열이 CLEM 시간 함수에 의해 시간으로 해석될 때 사용할 시간 형식을 선택합니다.

숫자 표시 형식. 표준(**####.###**), 지수표기(**#.###E+##**) 또는 통화 표시 형식(**$###.##**)에서 선택할 수 있습니다.
소수점 기호. 쉼표(,) 또는 마침표(.)를 소수점 구분 문자로 선택하십시오.

그룹 기호. 숫자 표시 형식에 대해 값을 그룹화하는 데 사용된 기호를 선택하십시오(예: *3,000.00*의 쉼표). 옵션에는 없음, 마침표, 쉼표, 공백 및 정의된 로케일이 포함됩니다(이 경우 현재 로케일의 기본값을 사용함).

소수점 이하 자릿수(표준, 지수, 통화, 내보내기). 숫자 표시 형식에 대해 실수를 표시할 때 사용할 소수점 이하 자릿수를 지정합니다. 이 옵션은 각 표시 형식마다 별도로 지정됩니다. **소수점 이하 자릿수 내보내기** 설정은 플랫 파일 내보내기에만 적용되며 스트림 특성을 대체합니다. 플랫 파일 내보내기의 스트림 기본값은 스트림 특성의 **표준 소수점 이하 자릿수** 설저에 지정된 값입니다. XML 내보내기 노드에서 내보낸 소수점 이하 자릿수는 항상 6입니다.

맞춤. 열에서 값을 맞추는 방식을 지정합니다. 기본 설정은 자동이며 이는 기호 값은 왼쪽으로 맞추고 숫자 값은 오른쪽으로 맞춥니다. 왼쪽, 오른쪽 또는 가운데를 선택하여 기본값을 대체할 수 있습니다.

열 너비. 기본적으로 열 너비는 필드의 값을 기반으로 자동으로 계산됩니다. 목록 상자 오른쪽의 화살표를 사용하여 5구간의 사용자 너비를 지정할 수 있습니다.

(4) 필드 필터링 또는 이름 바꾸기

스트림의 어느 지점에서나 필드의 이름을 바꾸고 필드를 제외할 수 있습니다. 예를 들어, 의학 연구자로서 환자(레코드 수준 데이터)의 칼륨 수준(필드 수준 데이터)에 대해 관심이 없을 수 있으므로 K(칼륨) 필드를 필터링할 수 있습니다. 이는 소스 또는 출력 노드의 필터 탭을 사용하거나 별도의 필터 노드를 사용하여 수행할 수 있습니다. 액세스되는 노드에 관계없이 기능은 동일합니다.

- 가변파일, 고정 파일, Statistics 파일, XML 또는 확장 가져오기 등의 소스 노드에서 데이터를 IBM® SPSS® Modeler로 읽어올 때 필드의 이름을 바꾸거나 필드를 필터링할 수 있습니다.
- 필터 노드를 사용하면 스트림의 어느 지점에서나 필드의 이름을 바꾸거나 필드를 필터링할 수 있습니다.
- Statistics 내보내기, Statistics 변환, Statistics 모델 및 Statistics 출력 노드에서 IBM SPSS Statistics 이름 지정 표준을 준수하도록 필드를 필터링하거나 필드의 이름을 바꿀 수 있습니다. 자세한 정보는 IBM SPSS Statistics에 대한 필드 이름 변경 또는 필터링 주제를 참조하십시오.
- 위 노드의 필터 탭을 사용하여 다중 응답 세트를 정의하거나 편집할 수 있습니다. 자세한 정보는 다중 응답 세트 편집 주제를 참조하십시오.
- 최종적으로 필터 노드를 사용하여 한 소스 노드에서 다른 소스 노드로 필드를 맵핑할 수 있습니다.

① 필터링 옵션 설정

필터 탭에서 사용되는 테이블은 노드에 들어갈 때와 노드에서 나갈 때 각 필드의 이름을 표시합니다. 이 테이블의 옵션을 사용하여 중복이거나 다운스트림 조작에 불필요한 필드의 이름을 바꾸거나 해당 필드를 필터링할 수 있습니다.

- **필드**현재 연결된 데이터 소스의 입력 필드를 표시합니다.
- **필터**. 모든 입력 필드의 필터 상태를 표시합니다. 필터링된 필드에는 이 필드가 다운스트림으로 전달되지 않음을 나타내는 빨간색 X가 이 열에 포함되어 있습니다. 선택된 필드에 대해 *필터* 열을 클릭하여 필터링을 켜고 끄십시오. Shift+클릭 선택 방법을 사용하여 동시에 여러 필드에 대한 옵션을 선택할 수도 있습니다.
- **필드**. 필드가 필터 노드를 나갈 때 필드를 표시합니다. 중복 이름은 빨간색으로 표시됩니다. 이 열을 클릭한 후 새 이름을 입력하여 필드 이름을 편집할 수 있습니다. 또는 *필터* 열을 클릭하여 중복 필드를 사용 안함으로 설정하여 필드를 제거하십시오.

테이블의 모든 열은 열 헤더를 클릭하여 정렬할 수 있습니다.

현재 필드 보기. 필터 노드에 활성 상태로 연결된 데이터 세트에 대한 필드를 보려면 선택하십시오. 이 옵션은 기본적으로 선택되며 필터 노드를 사용하는 가장 일반적인 방법입니다.

사용하지 않은 필드 설정 보기. 필터 노드에 한 번 연결되었지만 더 이상 연결되지 않는 데이터 세트에 대한 필드를 보려면 선택하십시오. 이 옵션은 한 스트림에서 다른 스트림으로 필터 노드를 복사하거나 필터 노드를 저장하고 재로드할 때 유용합니다.

필터 단추 메뉴

대화 상자의 왼쪽 상단에 있는 필터 단추를 클릭하여 다수의 단축키 및 기타 옵션을 제공하는 메뉴에 액세스하십시오.

다음과 같은 작업을 수행하도록 선택할 수 있습니다.

- 모든 필드 제거
- 모든 필드 포함
- 모든 필드 토글
- 중복 제거. 이 옵션을 선택하면 중복 이름의 모든 발생(첫 번째 발생 포함)이 제거됩니다.
- 다른 애플리케이션과 부합하도록 필드 및 다중 응답 세트의 이름 바꾸기. 자세한 정보는 IBM SPSS Statistics에 대한 필드 이름 변경 또는 필터링의 내용을 참조하십시오.
- 필드 이름 자르기
- 필드 및 다중 응답 세트 이름 값 익명화
- 입력 필드 이름 사용
- 다중 응답 세트 편집. 자세한 정보는 다중 응답 세트 편집 주제를 참조하십시오.
- 기본 필터 상태 설정

대화 상자의 맨 위에 있는 화살표 토글 단추를 사용하여 기본적으로 필드를 포함할지 아니면 삭제할지 지정할 수도 있습니다. 이는 몇몇 필드만 다운스트림으로 포함되는 큰 데이터 세트의 경우 유용합니다. 예를 들어, 삭제할 모든 필드를 개별적으로 선택하는 대신 보존할 필드만 선택한 후 다른 모든 필드는 삭제하도록 지정할 수 있습니다.

가. 필드 이름 자르기

필터 단추 메뉴(필터 탭의 상단 왼쪽 구석)에서 필드 이름을 자르도록 선택할 수 있습니다.

최대 길이. 필드 이름 길이를 제한하기 위한 문자 수를 지정하십시오.

숫자 수. 필드 이름을 잘랐을 때 이름이 더 이상 고유하지 않는 경우, 필드 이름을 더 자르고 이름에 숫자를 추가하여 차별화할 수 있습니다. 사용되는 숫자 수를 지정할 수 있습니다. 화살표 단추를 사용하여 숫자를 조정하십시오.

예를 들어, 다음 표에서는 기본 설정(최대 길이=8 및 숫자 수=2)을 사용하여 의학 데이터 세트의 필드 이름을 자르는 방법을 보여줍니다.

표 1. 필드 이름 자르기

필드 이름	잘린 필드 이름
Patient Input 1	Patien01
Patient Input 2	Patien02
Heart Rate	HeartRat
BP	BP

나. 필드 이름 익명화

왼쪽 맨 위에서 필터 단추 메뉴를 클릭하고 필드 이름 익명화를 선택하여 필터 탭이 포함된 노드에서 필드 이름을 익명화할 수 있습니다. 익명화된 필드 이름은 문자열 접두부와 고유 숫자 기반 값으로 구성되어 있습니다.

해당 이름 익명화. 필터 탭에서 이미 선택된 필드의 이름만 익명화하려면 **선택된 필드만**을 선택하십시오. 기본값은 **모든 필드**이며 모든 필드 이름을 익명화합니다.

필드 이름 접두부. 익명화된 필드 이름의 기본 접두부는 anon_입니다. 다른 접두부를 원하면 **사용자 정의**를 선택하고 직접 접두부를 입력하십시오.

다중 응답 세트 익명화. 여러 응답 세트의 이름을 필드와 동일한 방식으로 익명화합니다. 자세한 정보는 다중 응답 세트 편집 주제를 참조하십시오.

원래의 필드 이름을 복원하려면 필터 단추 메뉴에서 **입력 필드 이름 사용**을 선택하십시오.

다. 다중 응답 세트 편집

왼쪽 상단의 필터 단추 메뉴를 클릭한 후 **다중 응답 세트 편집**을 선택하여 필터 탭을 포함하는 노드에서 다중 응답 세트를 추가하거나 편집할 수 있습니다.

다중 응답 세트는 각 케이스(예: 설문조사 응답자에게 방문한 박물관 또는 읽은 잡지를 묻는 경우)에 대해 둘 이상의 값을 가질 수 있는 데이터를 기록하는 데 사용됩니다. 다중 응답 세트는 Data Collection 소스 노드 또는 통계 파일 소스 노드를 사용하여 IBM® SPSS® Modeler로 가져오거나 필터 노드를 사용하여 IBM SPSS Modeler에서 정의할 수 있습니다.

새로 작성을 클릭하여 새 다중 응답 세트를 작성하거나 **편집**을 클릭하여 기존 세트를 수정하십시오.

이름 및 레이블. 세트에 대한 이름 및 설명을 지정합니다.

유형. 다중 응답 질문은 두 방법 중 하나로 처리할 수 있습니다.

- **다중 이분형 세트**. 각각의 가능한 응답에 대해 별도의 플래그 필드가 작성되므로 10개의 잡지가 있으면 10개의 플래그 필드가 있으며 각각의 플래그 필드는 *참* 또는 *거짓*에 대해 0 또는 1과 같은 값을 가질 수 있습니다. 계수된 값을 사용하면 참으로 계수되는 값을 지정할 수 있습니다. 이 방법은 응답자가 적용되는 모든 옵션을 선택할 수 있게 하려는 경우에 유용합니다.
- **다중 범주 세트**. 지정된 응답자의 최대 응답 수까지 각 응답에 대해 명목 필드가 작성됩니다. 각각의 명목 필드는 *Time*에 대한 *1*, *Newsweek*에 대한 *2*, *PC Week*에 대한 *3*등의 가능한 응답을 나타내는 값을 가지고 있습니다. 이 방법은 응답 수를 제한하려는 경우(예: 응답자에게 가장 자주 읽은 세 개의 잡지를 선택하도록 요구하는 경우)에 가장 유용합니다.

세트의 필드. 오른쪽의 아이콘을 사용하여 필드를 추가하거나 제거하십시오.

설명

- 다중 응답 세트에 포함된 모든 필드는 동일한 저장 공간을 가지고 있어야 합니다.
- 세트는 포함하는 필드와 구별됩니다. 예를 들어, 세트를 삭제해도 세트에 포함된 필드는 삭제되지 않고 해당 필드 간 링크만 삭제됩니다. 세트는 여전히 삭제 지점에서 업스트림으로 표시되지만 다운스트림으로는 표시되지 않습니다.
- 필터 메뉴의 IBM SPSS Statistics에 대해 이름 바꾸기, **자르기** 또는 **값 익명화** 옵션을 선택하거나 탭에서 직접 필터 노드를 사용하여 필드의 이름을 바꾸는 경우 다중 응답 세트에서 사용되는 이 필드에 대한 참조도 업데이트됩니다. 하지만 필터 노드에 의해 삭제되는 다중 응답 세트의 필드는 다중 응답 세트에서 제거되지 않습니다. 이러한 필드는 더 이상 스트림에 표시되지 않지만 여전이 다중 응답 세트에 의해 참조되므로 예를 들어, 내보내기 수행 시 이를 고려할 수 있습니다.

(5) 파생 노드

IBM® SPSS® Modeler에서 가장 강력한 기능 중 하나는 데이터 값을 수정하고 기존 데이터에서 새 필드를 파생시키는 기능입니다. 긴 데이터 마이닝 프로젝트 동안에는 웹 로그 데이터의 문자열에서 고객 ID를 추출하거나 트랜잭션 및 인구 통계 데이터를 기반으로 고객 생애 가치를 작성하는 등의 여러 파생 작업을 수행하는 것이 일반적입니다. 이 변환은 모두 다양한 필드 작업 노드를 사용하여 수행할 수 있습니다.

몇몇 노드는 새 필드를 파생시키는 기능을 제공합니다.

	파생 노드는 데이터 값을 수정하거나 하나 이상의 기존 필드로부터 새 필드를 작성합니다. 수식, 플래그, 명목형, 상태, 개수, 조건부 유형의 필드를 작성합니다.
	재분류 노드는 한 세트의 범주형 값을 다른 값으로 변환합니다. 재분류는 분석을 위해 범주를 접거나 데이터를 재그룹화하는 데 유용합니다.
	구간화 노드는 하나 이상의 기존 연속형(숫자 범위) 필드의 값을 기반으로 새 명목형(세트) 필드를 자동으로 작성합니다. 예를 들어, 연속형 수입 필드를 평균값에서의 편차로서 수입 그룹을 포함하는 새 범주형 필드로 변환할 수 있습니다. 새 필드에 대한 구간을 작성한 후에는 절단점을 기반으로 파생 노드를 생성할 수 있습니다
	플래그로 설정 노드는 하나 이상의 명목 필드에 대해 정의된 범주형 값을 바탕으로 다중 플래그 필드를 파생시킵니다.
	구조변환 노드는 명목 또는 플래그 필드를 아직 또 다른 필드의 값으로 채워질 수 있는 필드 그룹으로 변환합니다. 예를 들어, payment type이라는 이름의 필드와 credit, cash, debit의 값이 주어진 경우, 각각이 실제 이루어진 지불의 값을 포함할 수 있는 세 개의 새 필드(credit, cash, debit)가 작성됩니다.
	히스토리 노드는 이전 레코드의 필드에 있는 데이터를 포함하는 새 필드를 작성합니다. 히스토리 노드는 시계열 데이터 같은 순차 데이터에 가장 자주 사용합니다. 히스토리 노드를 사용하기 전에 정렬 노드를 사용하여 데이터를 정렬할 수 있습니다.

파생 노드 사용

파생 노드를 사용하면 하나 이상의 기존 필드에서 여섯 가지 유형의 새 필드를 작성할 수 있습니다.

- **수식** 새 필드는 임의의 CLEM 표현식의 결과입니다.
- **플래그** 새 필드는 지정된 조건을 나타내는 플래그입니다.
- **명목** 새 필드는 명목이며 이는 해당 멤버가 지정된 값의 그룹임을 의미합니다.
- **상태** 새 필드는 두 상태 중 하나입니다. 이 상태 사이에서의 전환은 지정된 조건에 의해 트리거됩니다.
- **개수** 새 필드는 조건이 참인 횟수를 기반으로 합니다.
- **조건부** 새 필드는 조건의 값에 따라 두 표현식 중 하나의 값입니다.

이 노드 각각의 파생 노드 대화 상자에는 특별한 옵션 세트가 포함되어 있습니다. 이 옵션은 후속 주제에서 다룹니다.

다음을 사용하면 행 순서가 변경될 수 있습니다.
- SQL 푸시백을 통해 데이터베이스에서 실행
- 원격 IBM SPSS Analytic Server를 통해 실행
- 임베드된 IBM SPSS Analytic Server에서 실행되는 함수 사용
- 목록 파생(예: 목록 또는 지리 공간적 필드 파생 참조)
- 공간 함수 호출

① 파생 노드에 대한 기본 옵션 설정

파생 노드에 대한 대화 상자의 맨 위에는 필요한 파생 노드의 유형을 선택할 수 있는 다수의 옵션이 있습니다.

모드. 다중 필드를 파생할지 여부에 따라 **단일** 또는 **다중**을 선택하십시오. **다중**을 선택하면 다중 파생 필드에 대한 옵션을 포함하도록 대화 상자가 변경됩니다.

파생 필드. 단순 파생 노드의 경우 파생하여 각 레코드에 추가할 필드의 이름을 지정하십시오. 기본 이름은 Derive*N*입니다. 여기서 *N*은 현재 세션 동안 지금까지 작성한 파생 노드의 수입니다.

파생 유형. 드롭 다운 목록에서 파생 노드의 유형(예: 수식 또는 명목)을 선택하십시오. 각각의 유형에 대해 유형별 대화 상자에서 사용자가 지정하는 조건을 기반으로 새 필드가 작성됩니다.

드롭 다운 목록에서 옵션을 선택하면 각 파생 노드 유형의 특성에 따라 새 제어 세트가 기본 대화 상자에 추가됩니다.

필드 유형. 새로 파생된 노드에 대해 측정 수준(예: 연속형, 범주형 또는 플래그)을 선택하십시오. 이 옵션은 파생 노드의 모든 양식에 공통입니다.

참고: 새 필드를 파생시키려면 특별한 함수 또는 수학 표현식을 사용해야 할 수 있습니다. 이 표현식 작성을 지원하기 위해 모든 유형의 파생 노드에 대해 대화 상자에서 표현식 작성기를 사용할 수 있으며 이 표현식 작성기는 CLEM 표현식의 전체 목록과 규칙 검사를 제공합니다.

② 다중 필드 파생

파생 노드 내에서 모드를 **다중**으로 설정하면 동일한 노드 내 동일한 조건을 기반으로 다중 필드를 파생시킬 수 있습니다. 이 기능은 데이터 세트의 여러 필드에서 동일한 변환을 작성하길 원할 때 시간을 절약합니다. 예를 들어, 최초 급여와 이전의 경험을 기반으로 현재 급여를 예측하는 회귀 모형을 작성하려는 경우 3개의 비대칭 변수 모두에 로그 변환을 적용하면 도움이 될 수 있습니다. 각각의 변환에 대해 새 파생 노드를 추가하는 대신 한 번에 모든 필드에 동일한 함수를 적용할 수 있습니다. 단순히 새 필드를 파생시킬 모든 필드를 선택한 후 필드 소괄호 내 @FIELD 함수를 사용하여 파생 표현식을 입력하십시오.

참고: @FIELD 함수는 동시에 여러 필드를 파생시키는 데 필요한 중요한 도구입니다. 이 함수를 사용하면 정확한 필드 이름을 지정하지 않고 현재 필드의 컨텐츠를 참조할 수 있습니다. 예를 들어, 다중 필드에 로그 변환을 적용하는 데 사용되는 CLEM 표현식은 log(@FIELD)입니다.

다중 모드를 선택하면 다음과 같은 옵션이 대화 상자에 추가됩니다.

파생 위치. 필드 선택기를 사용하여 새 필드를 파생시킬 필드를 선택하십시오. 각각의 선택된 필드에 대해 하나의 출력 필드가 생성됩니다. *참고*: 선택된 필드의 저장 유형은 동일하지 않아도 됩니다. 하지만 *모든* 필드에 대해 조건이 유효하지 않으면 파생 조작이 실패합니다.

필드 이름 확장. 새 필드 이름에 추가할 확장자를 입력하십시오. 예를 들어, *Current Salary*의 로그가 포함된 새 필드에 대해 *log_* 확장자를 필드 이름에 추가하여 *log_Current Salary*를 생성할 수 있습니다. 단일 선택 단추를 사용하여 확장자를 필드 이름의 접두부(시작 부분에)와 접미부(끝 부분에) 중 어느 것으로 추가할지 선택하십시오. 기본 이름은 Derive*N*입니다. 여기서 *N*은 현재 세션 동안 지금까지 작성한 파생 노드의 수입니다.

단일 모드 파생 노드에서와 같이 이제 새 필드 파생에 사용할 표현식을 작성해야 합니다. 선택된 파생 조작의 유형에 따라 조건을 작성하는 다수의 옵션이 있습니다. 이 옵션은 후속 주제에서 다룹니다. 표현식을 작성하기 위해 간단하게 수식 필드를 입력하거나 계산기 단추를 클릭하여 표현식 작성기를 사용할 수 있습니다. 다중 필드에 대한 조작을 참조할 때 @FIELD 함수를 사용하는 것을 기억하십시오.

가. 다중 필드 선택

다중 입력 필드에 대해 조작을 수행하는 모든 노드(예: 파생(다중 모드), 통합, 정렬, 다중 도표, 시간 구성)의 경우 필드 선택 대화 상자를 사용하여 쉽게 다중 필드를 선택할 수 있습니다.

정렬 기준. 다음 옵션 중 하나를 선택하여 볼 수 있는 필드를 정렬할 수 있습니다.

- **기본**. 필드가 데이터 스트림에서 현재 노드에 전달된 대로 필드의 순서를 보십시오.
- **이름**. 알파벳순으로 볼 필드를 정렬하십시오.
- **유형**. 측정 수준별로 정렬된 필드를 보십시오. 이 옵션은 특정 측정 수준을 갖는 필드를 선택할 때 유용합니다.

한 번에 하나씩 목록에서 필드를 선택하거나 Shift-클릭 및 Ctrl-클릭 방법을 사용하여 다중 필드를 선택하십시오. 또한 목록 아래의 단추를 사용하여 측정 수준을 기반으로 필드 그룹을 선택하거나 테이블의 모든 필드를 선택 또는 선택 취소할 수 있습니다.

③ 수식 파생 옵션 설정

수식 파생 노드는 CLEM 표현식의 결과를 기반으로 데이터 세트에서 각 레코드에 대한 새 필드를 작성합니다. 이 표현식은 조건부 표현식이 되어서는 안 됩니다. 조건부 표현식을 기반으로 하는 값을 파생시키려면 파생 노드의 플래그 또는 조건부 유형을 사용하십시오.

수식 CLEM 언어를 사용하여 새 필드에 대한 값을 파생시켜 수식을 지정하십시오.

> ⓘ **참고**: SPSS® Modeler는 파생된 목록 필드에 사용될 하위 측정 수준을 알 수 없으므로 콜렉션 및 지리 공간적 측정 수준에 대해 **지정...**을 클릭하여 값 대화 상자를 열고 필요한 하위 측정 수준을 설정할 수 있습니다. 추가 정보는 파생된 목록 값 설정의 내용을 참조하십시오.
> 지리 공간적 필드의 경우, 관련된 유일한 저장 유형은 **실수** 및 **정수**입니다(기본 설정은 **실수**임).

가. 파생된 목록 값 설정

파생 노드 수식 **필드 유형** 드롭 다운 목록에서 **지정...**을 선택하면 값 대화 상자가 표시됩니다. 이 대화 상자에서는 수식 **필드 유형** 측정 수준 콜렉션 또는 지리 공간적에 사용될 하위 측정 수준 값을 설정합니다.
측정 콜렉션 또는 **지리 공간적**을 선택하십시오. 기타 측정 수준을 선택하면 편집 가능한 값이 없다는 메시지가 대화 상자에 표시됩니다.

요약도표

콜렉션 **측정** 수준에 대해 설정할 수 있는 유일한 항목은 **목록 측도**입니다. 기본적으로 이 측도는 유형 없음으로 설정되지만 다른 값을 선택하여 목록 내 요소의 측정 수준을 설정할 수 있습니다. 다음 옵션 중 하나를 선택할 수 있습니다.

- 유형 없음
- 범주형
- 연속
- 명목
- 순서
- 플래그

지리 공간적

지리 공간적 **측정** 수준의 경우 다음과 같은 옵션을 선택하여 목록 내 요소의 측정 수준을 설정할 수 있습니다.

유형 지리 공간적 필드의 측정 하위 수준을 선택하십시오. 사용 가능한 하위 수준은 목록 필드의 깊이에 의해 결정되며 기본값은 다음과 같습니다.

- 점(깊이 0)
- 선 스트링(깊이 1)
- 다각형(깊이 1)
- 다중 점(깊이 1)
- 다중 선 스트링(깊이 2)
- 다중 다각형(깊이 2)

하위 수준에 대한 자세한 정보는 지리 공간적 측정 수준의 내용을 참조하십시오.
목록 깊이에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

좌표계 이 옵션은 측정 수준을 비지리 공간적 수준에서 지리 공간적 수준으로 변경한 경우에만 사용할 수 있습니다. 지리 공간적 데이터에 좌표계를 적용하려면 이 선택란을 선택하십시오. 기본적으로, **도구 > 스트림 특성 > 옵션 > 지리 공간적** 분할창에서 설정된 좌표계가 표시됩니다. 다른 좌표계를 사용하려면 **변경** 단추를 클릭하여 좌표계 선택 대화 상자를 표시하고 데이터와 일치하는 좌표계를 선택하십시오.

좌표계에 대한 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오.

나. 목록 또는 지리 공간적 필드 파생

목록 항목으로 기록되어야 하는 데이터를 잘못된 속성을 가진 SPSS® Modeler로 가져오는 경우가 있습니다. 예를 들어, 별도의 지리 공간적 필드(예: x 좌표 및 y 좌표 또는 위도 및 경도)로 또는 .csv 파일의 개별 행으로 가져오는 경우가 있습니다. 이 상황에서는 개별 필드를 단일 목록 필드로 결합해야 합니다. 이를 수행하는 한 가지 방법은 파생 노드를 사용하는 것입니다.

> ⓘ **참고**: 지리 공간적 데이터를 결합할 때는 x(또는 경도) 필드와 y(또는 위도) 필드를 알고 있어야 합니다. 결과 목록 필드가 지리 공간적 좌표의 표준 형식인 [x, y] 또는 [경도, 위도] 요소 순서를 가지도록 데이터를 결합해야 합니다.

다음의 단계에서는 목록 필드를 파생시키는 단순한 예를 보여줍니다.

1. 스트림에서 파생 노드를 소스 노드에 연결하십시오.
2. 파생 노드의 설정 탭에 있는 **파생 유형** 목록에서 **수식**을 선택하십시오.
3. **필드 유형**에서 **콜렉션**(지리 공간적 목록이 아닌 경우) 또는 **지리 공간적**을 선택하십시오. 기본적으로 SPSS Modeler는 "최선의 추측" 접근 방식을 사용하여 올바른 목록 세부사항을 설정합니다. **지정...**을 선택하여 값 대화 상자를 열 수 있습니다. 이 대화 상자는 목록의 데이터에 대한 자세한 정보를 입력하는 요약도표에 사용할 수 있으며, 데이터의 유형을 설정하고 데이터의 좌표계를 지정하는 데 사용할 수 있는 지리 공간에 사용할 수 있습니다.

   > ⓘ **참고**: 지리 공간의 경우 사용자가 지정하는 좌표계는 데이터의 좌표계와 정확하게 일치해야 합니다. 그렇지 않으면 지리 공간 기능이 잘못된 결과를 생성합니다.

4. **수식** 분할창에서 데이터를 올바른 목록 형식으로 결합하는 수식을 입력하십시오. 또는 계산기 단추를 클릭하여 표현식 작성기를 여십시오.
   목록을 파생시키는 수식의 단순한 예는 [x, y]입니다. 여기서 x 및 y는 데이터 소스에 있는 별도의 필드입니다. 작성되는 새 파생된 필드는 각 레코드의 값이 해당 레코드에 대해 연결된 x 및 y 값인 목록입니다.

   > ⓘ **참고**: 이 방식으로 목록에 결합되는 필드는 동일한 저장 유형을 가지고 있어야 합니다.

목록과 목록의 깊이에 대한 자세한 정보는 목록 저장 공간 및 연관된 측정 수준의 내용을 참조하십시오.

④ 파생 플래그 옵션 설정

파생 플래그 노드는 고혈압 또는 고객 계정 비활성화 등의 특정 조건을 표시하는 데 사용됩니다. 각각의 레코드에 대해 플래그 필드가 작성되며 참 조건이 충족되면 참에 대한 플래그 값이 필드에서 추가됩니다.

참 값. 아래에 지정된 조건과 일치하는 레코드에 대해 플래그 필드에 포함할 값을 지정하십시오. 기본값은 T입니다.

거짓 값. 아래에 지정된 조건과 일치하지 *않는* 레코드에 대해 플래그 필드에 포함할 값을 지정하십시오. 기본값은 F입니다.

참일 조건. CLEM 조건을 지정하여 각 레코드의 특정 값을 평가하고 해당 레코드에 위에서 정의된 참 값 또는 거짓 값을 제공하십시오. 거짓이 아닌 숫자 값인 경우 참 값이 레코드에 제공됩니다.

참고: 빈 문자열을 리턴하려면 ""와 같이 사이에 아무것도 없는 여는 따옴표 및 닫는 따옴표를 입력해야 합니다. 빈 문자열은 예를 들어, 테이블에서 참 값을 더 명확하게 나타내기 위해 거짓 값으로 사용될 수 있습니다. 마찬가지로 따옴표로 묶지 않으면 숫자로 처리되는 문자열 값을 원하는 경우 따옴표를 사용해야 합니다.

예제

12.0 이전의 IBM® SPSS® Modeler 릴리스에서는 값이 쉼표로 구분된 다중 응답을 단일 필드로 가져왔습니다. 예:

```
museum_of_design,institute_of_textiles_and_fashion
museum_of_design
archeological_museum
$null$national_art_gallery,national_museum_of_science,other
```

이 데이터의 분석을 준비하기 위해 hassubstring 함수를 사용하여 다음과 같은 표현식으로 각 반응에 대해 별도의 플래그 필드를 생성할 수 있습니다.

```
hassubstring(museums,"museum_of_design")
```

⑤ 파생 명목 옵션 설정

파생 명목 노드는 각각의 레코드가 충족하는 조건을 판별하기 위해 CLEM 조건 세트를 실행하는 데 사용됩니다. 각 레코드에 대해 조건이 충족되면 충족된 조건 세트를 표시하는 값이 파생된 새 필드에 추가됩니다.

기본값. 충족되는 조건이 없는 경우 새 필드에서 사용할 값을 지정하십시오.

필드 설정. 특정 조건이 충족되는 경우 새 필드에서 입력할 값을 지정하십시오. 목록에 있는 각각의 값은 인접 열에서 지정되는 연관된 조건을 가지고 있습니다.

이 조건이 참인 경우. 나열할 세트 필드에서 각 멤버에 대한 조건을 지정하십시오. 표현식 작성기를 사용하여 사용 가능한 함수 및 필드를 선택하십시오. 화살표 및 삭제 단추를 사용하여 조건을 다시 정렬하거나 제거할 수 있습니다.
조건은 데이터 세트에서 특정 필드의 값을 검정하여 작동합니다. 각각의 조건이 검정되면 위에서 지정된 값이 충족된 조건(있는 경우)을 표시하는 새 필드에 할당됩니다. 충족되는 조건이 없으면 기본값이 사용됩니다.

⑥ 파생 상태 옵션 설정

파생 상태 노드는 파생 플래그 노드와 비슷합니다. 플래그 노드는 현재 레코드에 대한 *단일* 조건의 이행에 따라 값을 설정하지만 파생 상태 노드는 *두 개의 독립* 조건을 이행하는 방식에 따라 필드의 값을 변경할 수 있습니다. 이는 각각의 조건이 충족되면 값이 변경됨(켜짐 또는 꺼짐)을 의미합니다.

초기 상태. 초기에 새 필드의 각 레코드에 **On** 값과 **Off** 값 중 어느 값을 지정할지 선택하십시오. 이 값은 각각의 조건이 충족되면 변경될 수 있습니다.

"On" 값. On 조건이 충족될 때 새 필드의 값을 지정하십시오.

"On"으로 전환 조건. 조건이 참일 때 상태를 On으로 변경할 CLEM 조건을 지정하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

"Off" 값. Off 조건이 충족될 때 새 필드에 대한 값을 지정하십시오.

"Off"로 전환 조건. 조건이 거짓일 때 상태를 Off로 변경할 CLEM 조건을 지정하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

참고: 빈 문자열을 지정하려면 ""와 같이 사이에 아무것도 없는 여는 따옴표 및 닫는 따옴표를 입력해야 합니다. 마찬가지로 따옴표로 묶지 않으면 숫자로 처리되는 문자열 값을 원하는 경우 따옴표를 사용해야 합니다.

⑦ 파생 개수 옵션 설정

파생 개수 노드는 일련의 조건을 데이터 세트의 숫자 필드 값에 적용하는 데 사용됩니다. 각각의 조건이 충족되면 파생된 개수 필드의 값이 설정된 증분만큼 증가합니다. 이 유형의 파생 노드는 시계열 데이터의 경우 유용합니다.

초기값. 새 필드의 실행에 사용되는 값을 설정합니다. 초기값은 숫자 상수여야 합니다. 화살표 단추를 사용하여 값을 늘리거나 줄이십시오.

증분 조건. 충족되면 증분량에서 지정된 수를 기반으로 파생된 값을 변경할 CLEM 조건을 지정하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

증분량. 개수를 증분시키는 데 사용되는 값을 설정하십시오. CLEM 표현식의 결과 또는 숫자 상수를 사용할 수 있습니다.

재설정 조건. 충족되면 파생된 값을 초기값으로 재설정할 조건을 지정하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

⑧ 파생 조건부 옵션 설정

파생 조건부 노드에서는 일련의 If-Then-Else문을 사용하여 새 필드의 값을 파생시킵니다.

If. 실행 시 각 레코드에 대해 평가될 CLEM 조건을 지정하십시오. 조건이 true인 경우(숫자의 경우 false가 아닌 경우) Then 표현식에 의해 아래에서 지정된 값이 새 필드에 제공됩니다. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

Then. 위의 If문이 true인 경우(또는 false가 아닌 경우) 새 필드에 대한 CLEM 표현식 또는 값을 지정하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

Else. 위의 If문이 false인 경우 새 필드에 대한 CLEM 표현식 또는 값을 지정하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

⑨ 파생 노드를 사용하여 값 코딩변경

파생 노드는 예를 들어, 범주형 값을 가진 문자열 필드를 숫자 명목(세트) 필드로 변환하여 값의 코딩을 변경하는 데도 사용할 수 있습니다.

1. 파생 유형에 대해 필드의 유형(명목, 플래그 등)을 적절하게 선택하십시오.
2. 값의 코딩변경에 대한 조건을 지정하십시오. 예를 들어, Drug='drugA'인 경우 값을 1로 설정하고 Drug='drugB'인 경우 값을 2로 설정하는 방식으로 값을 설정할 수 있습니다.

(6) 채움 노드

채움 노드는 필드 값을 바꾸고 저장 공간을 변경하는 데 사용됩니다. 지정된 CLEM 조건(예: @BLANK(FIELD))을 기반으로 값을 바꾸도록 선택할 수 있습니다. 또는 모든 공백 또는 널값을 특정 값으로 대체할 것을 선택할 수 있습니다. 채움 노드는 종종 유형 노드와 함께 사용되어 결측값을 바꿉니다. 예를 들어, @GLOBAL_MEAN과 같은 표현식을 지정하여 필드의 평균값으로

공백을 채울 수 있습니다. 이 표현식은 전역값 설정 노드에 의해 계산된 대로 평균값으로 모든 공백을 채웁니다.

필드 채우기. 텍스트 필드의 오른쪽에 있는 필드 선택기 단추를 사용하여 값을 검토하여 바꿀 필드를 데이터 세트에서 선택하십시오. 기본 작동은 아래에 지정된 조건 및 바꾸기 표현식에 따라 값을 바꾸는 것입니다. 아래의 바꾸기 옵션을 사용하여 대체 바꾸기 방법을 선택할 수도 있습니다.

> **참고**: 사용자 정의 값으로 바꿀 필드를 여럿 선택하는 경우에는 필드 유형이 비슷한 것이 중요합니다(모두 숫자 또는 모두 기호).

바꾸기. 다음 방법 중 하나를 사용하여 선택된 필드의 값을 바꾸려면 선택하십시오.

- **조건 기반**. 이 옵션은 지정된 값으로 바꾸기 위한 조건으로 사용되는 표현식을 작성할 수 있도록 조건 필드 및 표현식 작성기를 활성화합니다.
- **항상**. 선택된 필드의 모든 값을 바꿉니다. 예를 들어, 이 옵션을 통해 CLEM 표현식 (to_string(income))을 사용하여 수입의 저장 공간을 문자열로 변환할 수 있습니다.
- **공백값**. 선택된 필드에서 사용자가 지정한 공백값을 모두 바꿉니다. 표준 조건 @BLANK(@FIELD)를 사용하여 공백을 선택합니다. *참고*: 유형 노드 또는 소스 노드의 유형 탭을 사용하여 공백을 정의할 수 있습니다.
- **널값**. 선택된 필드에서 모든 시스템 널값을 바꿉니다. 표준 조건 @NULL(@FIELD)을 사용하여 널을 선택합니다.
- **공백값 및 널값**. 선택된 필드에서 공백값과 시스템 널을 모두 바꿉니다. 이 옵션은 널이 결측값으로 정의되었는지 확신할 수 없는 경우에 유용합니다.

조건. 이 옵션은 **조건 기반** 옵션을 선택한 경우에만 사용할 수 있습니다. 선택된 필드를 평가하기 위해 CLEM 표현식을 지정하려면 이 텍스트 상자를 사용하십시오. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

바꾸기. 선택된 필드에 새 값을 제공하려면 CLEM 표현식을 지정하십시오. 텍스트 상자에 undef를 입력하여 값을 널값으로 바꿀 수도 있습니다. 계산기 단추를 클릭하여 표현식 작성기를 여십시오.

> **참고**: 선택된 필드가 문자열인 경우에는 해당 필드를 문자열 값으로 바꿔야 합니다. 기본값인 0 또는 다른 숫자 값을 문자열 필드의 대체값으로 사용하면 오류가 발생합니다.

다음을 사용하면 행 순서가 변경될 수 있습니다.
- SQL 푸시백을 통해 데이터베이스에서 실행
- 원격 IBM® SPSS® Analytic Server를 통해 실행
- 임베드된 IBM SPSS Analytic Server에서 실행되는 함수 사용
- 목록 파생(예: 목록 또는 지리 공간적 필드 파생 참조)
- 공간 함수 호출

① 채움 노드를 사용한 저장 공간 변환

채움 노드의 바꾸기 조건을 사용하면 단일 또는 다중 필드에 대한 필드 저장 공간을 쉽게 변환할 수 있습니다. 예를 들어, 변환 함수 to_integer를 사용하면 CLEM 표현식 to_integer(income)를 사용하여 *income*을 문자열에서 정수로 변환할 수 있습니다.

사용 가능한 변환 함수를 보고 표현식 작성기를 사용하여 CLEM 표현식을 자동으로 작성할 수 있습니다. 함수 드롭 다운 목록에서 **변환**을 선택하여 저장 공간 변환 함수의 목록을 보십시오. 사용 가능한 변환 함수는 다음과 같습니다.

- to_integer(ITEM)
- to_real(ITEM)
- to_number(ITEM)
- to_string(ITEM)
- to_time(ITEM)
- to_timestamp(ITEM)
- to_date(ITEM)
- to_datetime(ITEM)

날짜 및 시간 값 변환. 변환 함수(및 날짜 또는 시간 값 등의 특정 입력 유형이 필요한 기타 함수)는 스트림 옵션 대화 상자에서 지정된 현재 형식에 따라 결정됩니다. 예를 들어, 값이 *Jan 2003, Feb 2003* 등의 값을 가진 문자열 필드를 날짜 저장 공간으로 변환하려면 **MON YYYY**를 스트림의 기본 날짜 형식으로 선택하십시오.

파생 계산 중 임시 변환의 경우 파생 노드에서도 변환 함수를 사용할 수 있습니다. 파생 노드를 사용하여 범주형 값을 가진 문자열 필드의 코딩 변경 등의 기타 조작을 수행할 수도 있습니다. 자세한 정보는 파생 노드를 사용하여 값 코딩변경의 내용을 참조하십시오.

(7) 재분류 노드

재분류 노드는 한 세트의 범주형 값을 다른 값으로 변환합니다. 재분류는 분석을 위해 범주를 접거나 데이터를 재그룹화하는 데 유용합니다. 예를 들어, *제품*에 대한 값을 *부엌 제품*, *욕실 및 침구*, *가전 제품*이라는 세 그룹으로 다시 분류할 수 있습니다. 종종 이 작업은 값을 분류하고 재분류 노드를 생성하여 분포 노드에서 직접 수행됩니다. 자세한 정보는 분포 노드 사용 주제를 참조하십시오.

재분류는 하나 이상의 기호 필드에 대해 수행될 수 있습니다. 또한 기존 필드를 새로운 값으로 대체하거나 새 필드를 생성할 수 있습니다.

재분류 노드를 사용하는 시기

재분류 노드를 사용하기 전에 또 다른 필드 작업 노드가 현재 작업에 더 적절하지 않은지 고려하십시오.

- 자동 방법을 사용하여 순자 범위를 변수군(순위 또는 백분위수 등)으로 변환하려면 구간화 노드를 사용해야 합니다. 자세한 정보는 구간화 노드 주제를 참조하십시오.
- 수치 범위를 수동으로 변수군으로 분류하려면 파생 노드를 사용해야 합니다. 예를 들어, 월급 값을 특정 월급 범위 범주로 축소하려면 파생 노드를 사용하여 각 범주를 수동으로 정의해야 합니다.
- 범주형 필드의 값을 기반으로 하여 *Mortgage_type*과 같은 하나 이상의 플래그 필드를 생성하려면 플래그로 설정 노드를 사용해야 합니다.
- 범주형 필드를 수치 저장 공간으로 변환하기 위해 파생 노드를 사용할 수 있습니다. 예를 들어, *아니오* 및 *예* 값을 각각 0 및 1로 변환할 수 있습니다. 자세한 정보는 파생 노드를 사용하여 값 코딩변경 주제를 참조하십시오.

① 재분류 노드에 대한 옵션 설정

재분류 노드를 사용하기 위한 세 단계가 있습니다.

1. 먼저 다중 필드 또는 단일 필드 재분류를 선택해야 합니다.
2. 기존 필드로 코딩변경을 수행할 것인지 새 필드를 작성할 것인지 선택하십시오.
3. 그런 다음 재분류 노드 대화 상자의 동적 옵션을 사용하여 변수군을 원하는 대로 맵핑하십시오.

   모드. 한 필드에 대한 범주를 재분류하려면 **단일**을 선택하십시오. 한 번에 둘 이상의 필드의 변환을 가능하게 하는 옵션을 활성화하려면 **다중**을 선택하십시오.

   재분류. 원래 명목 필드를 유지하고 재분류된 값을 포함하는 추가 필드를 파생시키려면 **새 필드**를 선택하십시오. 원래 필드의 값을 새 분류로 덮어쓰려면 **기존 필드**를 선택하십시오. 이는 본질적으로 "채우기" 작업입니다.
   일단 모드 및 대체 옵션을 지정했으면 변환 필드를 선택하고 대화 상자의 아래쪽에 있는 동적 옵션을 사용하여 새 분류 값을 지정해야 합니다. 이러한 옵션은 사용자가 위에서 선택한 모드에 따라 다릅니다.

   필드 분류. 오른쪽의 필드 선택기 단추를 사용하여 하나(단일 모드) 또는 다중(다중 모드) 범주형 필드를 선택하십시오.

   새 필드 이름. 코딩변경된 값을 포함하는 새 명목 필드의 이름을 지정하십시오. 이 옵션은 위에서 **새 필드**가 선택된 경우에 단일 모드에서만 사용 가능합니다. **기존 필드**가 선택된 경우에는 원래 필드 이름이 유지됩니다. 다중 모드에서 작업하는 경우, 각 새 필드에 추가된 확장자를 지정하는 제어가 이 옵션을 대체합니다. 자세한 정보는 다중 필드 재분류 주제를 참조하십시오.

값 재분류. 이 테이블을 사용하면 이전 변수군 값에서 사용자가 여기서 지정하는 변수군 값으로 맵핑을 지울 수 있습니다.

- **원래 값**. 이 열에는 선택 필드의 기존 값이 나열됩니다.
- **새로운 값**. 이 열을 사용하면 새 범주 값을 입력하거나 드롭 다운 목록에서 하나를 선택할 수 있습니다. 분포 차트의 값을 사용하여 재분류 노드를 자동으로 생성하는 경우, 해당 값이 드롭 다운 목록에 포함됩니다. 이로 인해 기존 값을 알려진 값 변수군에 신속하게 맵핑할 수 있습니다. 예를 들어, 의료 조직은 네트워크 또는 로케일에 따라 진단을 다르게 그룹화하는 경우가 있습니다. 합병 후에 새 데이터를 재분류해야 하며 심지어 기존 데이터를 일관성 있게 재분류해야 합니다. 기존 데이터에서 수동으로 각 목표 값을 입력하지 않고 여러 마스터 값을 IBM® SPSS® Modeler로 읽어온 다음 *진단* 필드에 대한 분포 차트를 실행하여 차트에서 직접 이 필드에 대한 재분류(값) 노드를 생성할 수 있습니다. 이 프로세스를 통해 새 값 드롭 다운 목록에서 모든 목표 진단 값을 사용할 수 있습니다.

4. **가져오기**를 클릭하여 위에서 선택한 하나 이상의 필드에 대한 원래 값을 읽을 수 있습니다.
5. **복사**를 클릭하여 아직 맵핑되지 않은 필드에 대한 *새 값* 열로 원래 값을 붙여넣을 수 있습니다. 맵핑되지 않은 원래 값은 드롭 다운 목록에 추가됩니다.
6. **새로 지우기**를 클릭하여 *새로운 값* 열에서 모든 지정 사항을 지우십시오. *참고*: 이 옵션은 드롭 다운 목록의 값은 지우지 않습니다.
7. 각 원래 값에 대한 연속 정수를 자동으로 생성하려면 **자동**을 클릭하십시오. 1.5, 2.5 등의 실수 값이 아니라 정수만 생성될 수 있습니다.

예를 들어, 제품 이름에 대해 자동으로 연속 제품 ID 번호를 생성하거나 대학 강좌에 대한 강좌 번호를 생성할 수 있습니다. 이 기능은 IBM SPSS Statistics의 집합에 대한 자동 코딩변경 변환에 해당됩니다.

지정되지 않은 값 사용. 이 옵션은 새 필드에서 지정되지 않은 값을 채우는 데 유용합니다. **원래 값**을 선택하여 원래 값을 유지하거나 기본값을 지정하십시오.

② 다중 필드 재분류

한 번에 둘 이상의 필드에 대한 범주 값을 맵핑하려면 모드를 **다중**으로 설정하십시오. 그러면 아래에서 버려지는 재분류 대화 상자 내의 새 설정을 사용할 수 있습니다.

필드 재분류. 오른쪽의 필드 선택기 단추를 사용하여 변환할 필드를 선택하십시오. 필드 선택기를 사용하여 한 번에 모든 필드 또는 유사한 유형의 필드(명목 또는 플래그 등)를 선택할 수 있습니다.

필드 이름 확장. 다중 필드의 코딩을 동시에 변경하는 경우, 개별 필드 이름이 아니라 모든 새 필드에 추가된 공통 확장자를 지정하는 것이 더 효율적입니다. _recode 등의 확장자를 지정하고 이 확장자를 원래 필드 이름에 첨부할 것인지 여부를 지정하십시오.

③ 재분류 필드에 대한 저장 공간 및 측정 수준

재분류 노드는 항상 코딩변경 작업에서 명목 필드를 생성합니다. 일부 경우에는 이로 인해 **기존 필드** 재분류 모드를 사용할 때 필드의 측정 수준이 변경될 수 있습니다.

새 필드의 저장 공간(데이터가 *사용되는* 방법이 아니라 *저장되는* 방법)은 다음과 같은 설정 탭 옵션을 기준으로 하여 계산됩니다.

- 지정되지 않은 값은 기본값을 사용하도록 설정된 경우, 기본값 및 새로운 값을 둘 다 검사하고 적절한 저장 공간을 판별하여 저장 유형이 결정됩니다. 예를 들어, 모든 값이 정수로 구문 분석되는 경우에는 필드가 정수 저장 유형을 사용하게 됩니다.
- 지정되지 않은 값은 원래 값을 사용하도록 사용하도록 설정된 경우, 저장 유형은 원래 필드의 저장 공간을 기반으로 합니다. 모든 값이 원래 필드의 저장 공간으로 구문 분석될 수 있는 경우, 해당 저장 공간이 유지됩니다. 그렇지 않으면 기존값 및 새로운 값을 둘 다 포함하여 가장 적절한 저장 유형을 발견하여 저장 공간이 판별됩니다. 예를 들어, 정수 집합 { 1, 2, 3, 4, 5 }를 재분류 4 => 0, 5 => 0을 사용하여 재분류하면 새 정수 집합 { 1, 2, 3, 0 }이 생성되며 재분류 4 => "Over 3", 5 => "Over 3"을 사용하여 재분류하면 { "1", "2", "3", "Over 3" } 문자열 집합이 생성됩니다.

참고: 원래 유형이 인스턴스화되지 않으면 새 유형도 인스턴스화되지 않습니다.

(8) 값 익명화 노드

익명화 노드를 사용하면 노드의 모델 다운스트림에 포함되는 데이터에 대해 작업할 때 필드 이름, 필드 값 또는 둘 다를 위장할 수 있습니다. 이러한 방식으로, 권한없는 사용자가 직원 기록 또는 환자의 의료 기록과 같은 기밀 데이터를 볼 수 있는 위험없이 생성된 모델을 예를 들어, 기술 지원에 자유롭게 분배할 수 있습니다.
스트림에서 익명화 노드를 배치하는 위치에 따라 다른 노드를 변경해야 할 수도 있습니다. 예를 들어, 선택 노드에서 업스트림으로 익명화 노드를 삽입하는 경우 선택 노드의 선택 기준이 현재 익명화된 값에 적용되면 이 기준을 변경해야 합니다.

익명화하는 데 사용하는 방법은 다양한 요인에 따라 다릅니다. 필드 이름 및 연속형 측정 수준을 제외한 모든 필드 값의 경우 데이터는 다음과 같은 형식의 문자열로 대체됩니다.

```
prefix_Sn
```

여기서 *prefix_ is*는 사용자 지정 문자열 또는 기본 문자열 anon_이고 *n*은 0부터 시작하는 정수 값이며 각 고유 값에 대해 증분됩니다(예: anon_S0, anon_S1 등).

숫자 범위가 문자열이 아닌 정수 또는 실수 값을 처리하므로 유형이 연속형인 필드 값을 변환해야 합니다. 이러한 방식으로 범위를 다른 범위로 변환하므로 원래 데이터를 위장해야만 이를 익명화할 수 있습니다. 범위에서 값 *x*의 변환은 다음과 같은 방식으로 수행됩니다.

```
A*(x + B)
```

여기서,

*A*는 0보다 커야 하는 환산 계수입니다.

*B*는 값에 추가할 변환 오프셋입니다.

예제
환산 계수 *A*가 7로 설정되고 변환 오프셋 *B*가 3으로 설정되는 필드 *AGE*의 경우 *AGE*의 값이 다음으로 변환됩니다.

```
7*(AGE + 3)
```

① 익명화 노드의 옵션 설정

여기에서 다운스트림으로 값을 위장할 필드를 선택할 수 있습니다.

익명화 작업을 수행하기 전에 익명화 노드에서 업스트림으로 데이터 필드를 인스턴스화해야 한다는 점을 참고하십시오. 유형 노드 또는 소스 노드의 유형 탭에서 **값 읽기** 단추를 클릭하여 데이터를 인스턴스화할 수 있습니다.

필드. 현재 데이터 세트의 필드를 나열합니다. 필드 이름이 이미 익명화된 경우 익명화된 이름이 여기에 표시됩니다.

측정. 필드의 측정 수준입니다.

값 익명화. 하나 이상의 필드를 선택하고 이 열을 클릭한 후 **예**를 선택하여 기본 접두부 anon_으로 필드 값을 익명화하십시오. 사용자가 직접 접두부를 입력하거나 유형이 *연속형*인 필드 값의 경우 필드 값의 변환이 무작위 또는 사용자 지정 값을 사용할 것인지를 지정할 수 있는 대화 상자를 표시하려면 **지정**을 선택하십시오. *연속형* 및 비*연속형* 필드 유형은 동일한 작업에서 지정할 수 없다는 점을 참고하십시오. 필드의 각 유형에 대해 개별적으로 지정해야 합니다.

현재 필드 보기. 익명화 노드에 활동적으로 연결된 데이터 세트의 필드를 보려면 이를 선택하십시오. 이 옵션은 기본적으로 선택됩니다.

사용하지 않은 필드 설정 보기. 노드에 한 번 연결되었지만 더 이상 연결되어 있지 않은 데이터 세트의 필드를 보려면 이를 선택하십시오. 이 옵션은 한 스트림에서 다른 스트림으로 노드를 복사하거나 노드를 저장하고 다시 로드하는 경우에 유용합니다.

가. 필드 값이 익명화되는 방식 지정

값 대체 대화 상자를 사용하면 익명화된 필드 값의 기본 접두부를 사용하거나 사용자 정의 접두부를 사용할 것인지 선택할 수 있습니다. 이 대화 상자에서 **확인**을 클릭하면 설정 탭에서 값 익명화의 설정이 선택된 필드에 대해 **예**로 변경됩니다.

필드 값 접두부. 익명화된 필드 값의 기본 접두부는 anon_입니다. 다른 접두부를 원하면 **사용자 정의**를 선택하고 직접 접두부를 입력하십시오.

값 변환 대화 상자는 유형이 연속형인 필드의 경우에만 표시되며 필드 값의 변환이 무작위 또는 사용자 지정 값을 사용하기 위한 것인지를 지정할 수 있습니다.

변량. 변환에 무작위 값을 사용하려면 이 옵션을 선택하십시오. **난수 시드 설정**이 기본적으로 선택되어 있습니다. **시드** 필드에 값을 지정하거나 기본값을 사용하십시오.

고정됨. 변환에 사용자의 값을 지정하려면 이 옵션을 선택하십시오.

- **척도 기준**. 변환에서 필드 값에 곱하는 수입니다. 최소값은 1이고 최대값은 일반적으로 10이지만 오버플로우를 방지하기 위해 이 값을 낮출 수 있습니다.
- **변환 기준**. 변환에서 필드 값에 추가되는 수입니다. 최소값은 0이고 최대값은 일반적으로 1000이지만 오버플로우를 방지하기 위해 이 값을 낮출 수 있습니다.

② 필드 값 익명화

설정 탭에서 익명화하기 위해 선택된 필드에는 익명화된 값이 있습니다.

- 익명화 노드가 포함된 스트림을 실행할 때
- 값을 미리 볼 때

값을 미리 보려면 익명화된 값 탭에서 **값 익명화** 단추를 클릭하십시오. 그런 다음, 드롭 다운 목록에서 필드 이름을 선택하십시오.

측정 수준이 연속형인 경우 다음과 같이 표시됩니다.

- 원래 범위의 최소값 및 최대값
- 값을 변환하는 데 사용되는 등식

측정 수준이 연속형 외의 수준인 경우 화면에는 해당 필드의 원래 및 익명화된 값이 표시됩니다.

화면이 노란색 배경으로 표시되는 경우 마지막으로 값이 익명화된 이후에 선택된 필드의 설정이 변경되었거나 익명화된 값이 더 이상 올바르지 않은 경우와 같이 익명화 노드의 데이터 업스트림이 변경되었음을 표시합니다. 현재 값 세트가 표시됩니다. **값 익명화** 단추를 다시 클릭하여 현재 설정에 따라 새 값 세트를 생성하십시오.

값 익명화. 선택된 필드의 익명화된 값을 작성하고 테이블에 이 값을 표시합니다. 연속형 유형의 필드에 대해 난수 시드를 사용하는 경우 이 단추를 반복적으로 클릭하면 매번 다른 값 세트가 작성됩니다.

값 지우기. 테이블에서 원래 및 익명화된 값을 지웁니다.

(9) 구간화 노드

구간화 노드를 사용하면 하나 이상의 기존 연속(숫자 범위) 필드 값에 기반하여 새 명목 필드를 자동으로 작성할 수 있습니다. 예를 들어, 연속 수입 필드를 동일한 너비의 수입 그룹이 포함된 새 범주형 필드로 변환하거나 평균의 편차로 변환할 수 있습니다. 또는 두 필드 간 원래 연관의 강도를 유지하기 위해 범주형 "수퍼바이저" 필드를 선택할 수 있습니다.

다음과 같은 여러 가지 이유로 구간화가 유용할 수 있습니다.

- **알고리즘 요구사항**. Naive Bayes, 로지스틱 회귀분석과 같은 특정 알고리즘에는 범주형 입력이 필요합니다.
- **성능**. 입력 필드의 고유 값 수가 감소되면 다항 로지스틱과 같은 알고리즘의 성능이 향상될 수 있습니다. 예를 들어, 원래 값을 사용하지 않고 각 구간에 대해 중앙값 또는 평균 값을 사용하십시오.
- **데이터 개인정보 보호정책**. 급여와 같은 민감한 개인 정보는 개인정보를 보호하기 위해 실제 급여 수치가 아닌 범위로 보고될 수 있습니다.

여러 구간화 방법을 사용할 수 있습니다. 새 필드에 대한 구간을 작성한 후에는 절단점을 기반으로 파생 노드를 생성할 수 있습니다.

구간화 노드 사용 시기

구간화 노드를 사용하기 전에 다른 기술이 즉시 사용되는 작업에 보다 적절한지 여부를 고려하십시오.

- 사전정의된 특정 급여 범위와 같은 범주의 절단점을 수동으로 지정하려면 파생 노드를 사용하십시오. 자세한 정보는 파생 노드의 내용을 참조하십시오.
- 기존 세트의 새 범주를 작성하려면 재분류 노드를 사용하십시오. 자세한 정보는 재분류 노드의 내용을 참조하십시오.

결측값 처리

구간화 노드는 다음과 같은 방법으로 결측값을 처리합니다.

- **사용자 지정 공백**. 공백으로 지정된 결측값은 변환 중 포함됩니다. 예를 들어, 유형 노드를 사용하여 공백 값을 표시하기 위해 -99를 지정한 경우 이 값이 구간화 프로세스에 포함됩니다. 구간화 중 공백을 무시하려면 채움 노드를 사용하여 공백 값을 시스템 널값으로 대체해야 합니다.
- **시스템 결측값($null$)**. 널값은 구간화 변환 중 무시되며 변환 이후에 널을 유지합니다.

설정 탭은 사용 가능한 기술에 대한 옵션을 제공합니다. 보기 탭은 노드를 통해 이전에 실행된 데이터에 대해 설정된 절단점을 표시합니다.

① 구간화 노드의 옵션 설정

구간화 노드를 사용하면 다음 기술을 사용하여 구간(범주)을 자동으로 생성할 수 있습니다.

- 고정 너비 구간화
- 분위수(동일 개수 또는 합계)
- 평균 및 표준 편차
- 순위
- 범주형 "수퍼바이저" 필드와 관련하여 최적화됨

대화 상자의 하단 부분은 사용자가 선택하는 구간화 방법에 따라 동적으로 변경됩니다.

구간화 필드. 변환 보류 중인 연속(숫자 범위) 필드가 여기에 표시됩니다. 구간화 노드를 사용하면 여러 필드를 동시에 구간화할 수 있습니다. 오른쪽의 단추를 사용하여 필드를 추가하거나 제거하십시오.

구간화 방법. 새 필드 구간(범주)의 절단점을 판별하는 데 사용되는 방법을 선택하십시오. 이후 주제에서는 각 케이스에서 사용할 수 있는 옵션에 대해 설명합니다.

구간 임계값. 구간 임계값을 계산하는 방법을 지정하십시오.

- **항상 재계산**. 절단점 및 구간 할당은 항상 노드가 실행될 때 재계산됩니다.

- **사용 가능한 경우 구간 값 탭에서 읽기**. 절단점 및 구간 할당은 필요한 경우(예: 새 데이터가 추가된 경우)에만 계산됩니다.

다음 주제에서는 사용 가능한 구간화 방법의 옵션에 대해 설명합니다.

② 고정 너비 구간

구간화 방법으로 **고정 너비**를 선택하는 경우 새 옵션 세트가 대화 상자에 표시됩니다.

이름 확장자. 생성 필드에 사용할 확장자를 지정하십시오. *_BIN*이 기본 확장자입니다. 확장자가 필드 이름의 처음(**접두어**) 또는 끝(**접미어**)에 추가되는지 여부를 지정할 수도 있습니다. 예를 들어, *income_BIN*이라는 새 필드를 생성할 수 있습니다.

구간 너비. 구간의 "너비"를 계산하는 데 사용되는 값(정수 또는 실수)을 지정하십시오. 예를 들어, 필드 *나이*를 구간화하기 위해 기본값 10을 사용할 수 있습니다. *나이*에 18-65의 범위가 있으므로 생성되는 구간은 다음 표에 표시된 바와 같습니다.

표 1. 범위 18-65인 나이의 구간

구간1	구간2	구간3	구간4	구간5	구간6
>=13부터 <23까지	>=23부터 <33까지	>=33부터 <43까지	>=43부터 <53까지	>=53부터 <63까지	>=63부터 <73까지

구간 간격의 시작은 스캐닝된 최저 값에서 구간 너비(지정됨)의 절반을 뺀 값을 사용하여 계산됩니다. 예를 들어, 위에 표시된 구간에서 13이 다음 계산에 따라 간격을 시작하는 데 사용됩니다. 18 *[최저 데이터 값]* - 5 *[0.5 × (구간 너비 10)]* = 13.

구간 수. 새 필드의 고정 너비 구간(범주) 수를 판별하는 데 사용되는 정수를 지정하려면 이 옵션을 사용하십시오.

스트림에서 구간화 노드를 실행한 경우 구간화 노드 대화 상자에서 **미리보기** 탭을 클릭하여 생성되는 구간 임계값을 볼 수 있습니다. 자세한 정보는 생성된 구간 미리보기의 내용을 참조하십시오.

③ 분위수(동일 개수 또는 합계)

분위수 구간화 방법은 각 그룹에 동일한 수의 레코드가 있거나 각 그룹에 있는 값의 합계가 동일하도록 스캐닝된 레코드를 백분위수 그룹(또는 사분위수, 십분위수 등)으로 분할하는 데 사용할 수 있는 명목 필드를 작성합니다. 레코드는 지정된 구간 필드의 값에 기반하여 오름차순으로

순위가 지정되므로 선택된 구간 변수의 최저 값이 있는 레코드에 1순위가 지정되고 다음 레코드 세트에 2순위가 지정되는 등과 같습니다. 각 구간의 임계값은 사용된 데이터 및 분위수 지정 방법에 기반하여 자동으로 생성됩니다.

분위수 이름 확장자. 표준 p-분위수를 사용하여 생성되는 필드에 사용되는 확장자를 지정하십시오. 기본 확장자는 *_TILE* 더하기 *N*으로, *N*은 분위수 번호입니다. 확장자가 필드 이름의 처음(**접두어**) 또는 끝(**접미어**)에 추가되는지 여부를 지정할 수도 있습니다. 예를 들어, *income_BIN4*라는 새 필드를 생성할 수 있습니다.

사용자 정의 분위수 확장자. 사용자 정의 분위수 범위에 사용되는 확장자를 지정하십시오. 기본값은 *_TILEN*입니다. 이 케이스의 *N*은 사용자 정의 수로 대체되지 *않는다*는 점을 참고하십시오.

사용 가능한 p-tile은 다음과 같습니다.

- **사분위수**. 각각 케이스의 25%를 포함하는 네 개의 구간을 생성하십시오.
- **5분위수**. 각각 케이스의 20%를 포함하는 다섯 개의 구간을 생성하십시오.
- **십분위수**. 각각 케이스의 10%를 포함하는 10개의 구간을 생성하십시오.
- **20분위수**. 각각 케이스의 5%를 포함하는 20개의 구간을 생성하십시오.
- **백분위수**. 각각 케이스의 1%를 포함하는 100개의 구간을 생성하십시오.
- **사용자 정의 N**. 구간 수를 지정하려면 이를 선택하십시오. 예를 들어, 값 3은 각각 케이스의 33.3%를 포함하는 세 개의 연결된 범주(두 개의 절단점)를 생성합니다.

데이터에 지정된 분위수보다 적은 이산 값이 있는 경우 모든 분위수가 사용되지 않는다는 점을 참고하십시오. 이러한 경우 새 분포에 데이터의 원래 분포가 반영될 가능성이 높습니다.

분위수 지정 방법. 구간에 레코드를 지정하는 데 사용되는 방법을 지정합니다.

- **레코드 개수**. 각 구간에 동일한 수의 레코드를 지정합니다.
- **값 합계**. 각 구간의 값의 합계가 동일하도록 구간에 레코드를 지정합니다. 예를 들어, 영업 성과를 목표로 할 때 이 방법을 사용하면 최고 값 가능성이 맨 위 구간에 있을 때 레코드 당 값에 기반하여 십분위수 그룹에 가능성을 지정할 수 있습니다. 예를 들어, 제약 회사는 작성하는 처방전 수에 기반하여 십분위수 그룹으로 내과 의사의 순위를 지정할 수 있습니다. 각 십분위수에 대략적으로 동일한 수의 스크립트가 포함되지만 이러한 스크립트에 기여하는 개인 수는 동일하지 않으며 개인은 십분위수 10에 집중된 대부분의 스크립트를 작성합니다. 이러한 방법에서는 모든 값이 0보다 크다고 가정하고 이 경우에 해당하지 않는 경우 예상하지 않은 결과가 나타날 수 있다는 점을 참고하십시오.

등순위. 절단점의 한 쪽에 있는 값이 동일한 경우 경계값 조건 결과입니다. 예를 들어, 십분위수를 지정하며 레코드의 10% 이상에 구간 필드에 대해 동일한 값이 있는 경우 임계값을 한 방향 또는 다른 방향으로 강제 실행하지 않고 이러한 모든 레코드를 동일한 구간에 넣을 수 없습니다. 경계값이 위로 다음 구간으로 이동되거나 현재 구간에서 유지될 수 있지만 동일한 값의 모

든 레코드가 동일한 구간에 속하도록 분석해야 합니다(이로 인해 일부 구간에 예상보다 많은 레코드가 있는 경우에도). 따라서 이후 구간의 임계값도 조정할 수 있으므로 경계값을 분석하는 데 사용되는 방법에 기반하여 동일한 번호 세트에 대해 다르게 값이 지정됩니다.

- **다음에 추가**. 위로 다음 구간으로 경계값을 이동하려면 이를 선택하십시오.
- **현재구간에 유지**. 등순위 값을 현재(더 낮은) 구간에 유지합니다. 이 방법을 사용하면 더 적은 수의 총 구간이 작성될 수 있습니다.
- **무작위 지정**. 한 구간에 경계값을 무작위로 할당하려면 이를 선택하십시오. 각 구간에서 레코드 수를 동일한 양으로 유지하려고 합니다.

예: 레코드 수별 분위수 지정

다음 표는 레코드 수별로 분위수를 지정할 때 간단해진 필드 값이 분위수로 순위가 지정되는 방식을 나타냅니다. 결과는 선택된 경계값 옵션에 따라 다를 수 있다는 점을 참고하십시오.

표 1. 레코드 수별 분위수 지정 예

값	다음에 추가	현재에서 유지
10	1	1
13	2	1
15	3	2
15	3	2
20	4	3

구간 당 항목 수는 다음과 같이 계산됩니다.

총 값 수 / 분위수

위의 간단한 예에서 구간 당 원하는 항목 수는 1.25(5개의 값 / 4 사분위수)입니다. 값 13(값 번호 2)은 1.25 원하는 개수 임계값을 스트래들하므로 선택된 경계값 옵션에 따라 다르게 처리됩니다. **다음에 추가** 모드에서는 구간 2에 추가됩니다. **현재에서 유지** 모드에서는 구간 1에 계속 남아 있어서 구간 4에 대한 값의 범위를 기존 데이터 값의 범위 외부에 넣습니다. 따라서 세 개의 구간만 작성되며 각 구간의 임계값이 적절하게 조정됩니다(다음 표 참조).

표 2. 구간화 예 결과

구간	하한	상한
1	>=10	<15
2	>=15	<20
3	>=20	<=20

참고: 분위수별로 구간화하는 속도는 병렬 처리를 사용할 경우 빨라질 수 있습니다.

④ **케이스 순위 지정**

순위를 구간화 방법으로 선택하면 새 옵션 세트가 대화 상자에 표시됩니다.

순위화는 아래에 지정된 옵션에 따라 숫자 필드의 순위, 분수 순위, 백분위수 값이 포함된 새 필드를 작성합니다.

순위 순서. **오름차순**(최저값이 1로 표시됨) 또는 **내림차순**(최고값이 1로 표시됨)을 선택하십시오.

순위. 위에 지정된 바와 같이 오름차순 또는 내림차순으로 케이스의 순위를 지정하려면 이를 선택하십시오. 새 필드에 있는 값의 범위는 1-*N*이며 여기서 *N*은 원래 필드에 있는 이산 값의 수입니다. 경계값에는 해당 순위의 평균이 제공됩니다.

분수 순위. 새 필드의 값이 순위를 비결측 케이스의 가중치 합계로 나눈 값인 케이스의 순위를 지정하려면 이를 선택하십시오. 분수순위는 0 - 1의 범위에 들어갑니다.

백분율 분수 순위. 각 순위를 유효한 값을 갖는 레코드 수로 나누고 100을 곱합니다. 퍼센트 분수순위는 1 - 100의 범위에 들어갑니다.

확장자. 모든 순위 옵션에 대해 사용자 정의 확장자를 작성하고 필드 이름의 처음(**접두부**) 또는 끝(**접미부**)에 추가되는지 여부를 지정할 수 있습니다. 예를 들어, *income_P_RANK*라는 새 필드를 생성할 수 있습니다.

⑤ **평균/표준 편차**

구간화 방법으로 **평균/표준 편차**를 선택하는 경우 새 옵션 세트가 대화 상자에 표시됩니다.

이 방법은 지정된 필드의 분포에 대한 평균 및 표준 편차의 값에 기반하여 연결된 범주가 있는 하나 이상의 새 필드를 생성합니다. 아래에서 사용할 편차 수를 선택하십시오.

이름 확장자. 생성 필드에 사용할 확장자를 지정하십시오. *_SDBIN*이 기본 확장자입니다. 확장자가 필드 이름의 처음(**접두어**) 또는 끝(**접미어**)에 추가되는지 여부를 지정할 수도 있습니다. 예를 들어, *income_SDBIN*이라는 새 필드를 생성할 수 있습니다.

- **+/- 1 표준 편차**. 세 개의 구간을 생성하려면 이를 선택하십시오.
- **+/- 2 표준 편차**. 다섯 개의 구간을 생성하려면 이를 선택하십시오.
- **+/- 3 표준 편차**. 일곱 개의 구간을 생성하려면 이를 선택하십시오.

예를 들어, +/-1 표준 편차를 선택하면 세 개의 구간이 다음 표와 같이 계산되고 표시됩니다.

표 1. 표준 편차 구간 예

구간1	구간2	구간3
x < (Mean - Std. Dev)	(Mean - Std. Dev) <= x <= (Mean + Std. Dev)	x > (Mean + Std. Dev)

정규 분포에서 케이스의 68%는 평균의 표준 편차 1에 속하고 95%는 표준 편차 2에 속하고 99%는 표준 편차 3에 속합니다. 그러나 표준 편차에 기반하여 연결된 범주를 작성하면 실제 데이터 범위 외부 및 가능한 데이터 값의 범위 외부(예: 음수 급여 범위)에도 일부 구간이 정의될 수 있습니다.

⑥ 최적 구간화

구간화할 필드가 다른 범주형 필드와 강력하게 연관되어 있는 경우 두 필드 간 원래 연관의 강도를 유지하기 위한 방식으로 구간을 작성하기 위해 범주형 필드를 "수퍼바이저" 필드로 선택할 수 있습니다.

예를 들어, 군집 분석을 사용하여 주택 자금 융자의 연체율에 기반하여 상태를 그룹화했다고 가정하십시오(최고 비율이 첫 번째 군집에 있음). 이 경우 구간 필드로 *기일 경과 후 퍼센트* 및 *압류 퍼센트*를 선택하고 수퍼바이저 필드로 모델에서 생성한 소속군집 필드를 선택할 수 있습니다.

이름 확장자 생성 필드에 사용할 확장자와 필드의 처음(**접두부**) 또는 끝(**접미부**)에 이를 추가할 것인지 여부를 지정하십시오. 예를 들어, pastdue_OPTIMAL이라는 새 필드와 inforeclosure_OPTIMAL이라는 다른 필드를 생성할 수 있습니다.

수퍼바이저 필드 구간을 작성하는 데 사용되는 범주형 필드입니다.

큰 데이터 세트로 성능을 향상시키는 사전 구간화 필드 사전 처리를 사용하여 최적 구간화를 간소화해야 하는지 여부를 표시합니다. 이는 간단한 무감독 구간화 방법을 사용하여 매우 많은 구간으로 척도 값을 그룹화하고 각 구간의 값을 평균으로 표시하고 감독 구간화를 계록하기 전에 케이스 가중치를 적절하게 조정합니다. 실제로 이 방법은 속도에 대한 정밀도를 교환하며 대형 데이터 세트의 경우에 권장됩니다. 이 옵션이 사용되는 경우 사전 처리한 후 변수가 사용하게 되는 최대 구간 수도 지정할 수 있습니다.

큰 이웃 항목이 있는 상대적으로 작은 케이스 개수가 있는 구간 병합. 사용되는 경우 구간의 크기(케이스 수)와 인접 구간의 크기 비율이 지정된 임계값보다 작은 경우 구간이 병합됨을 표시합니다. 임계값이 크면 추가 병합이 발생할 수 있다는 점을 참고하십시오.

가. 절단점 설정

절단점 설정 대화 상자를 사용하면 최적 구간화 알고리즘의 고급 옵션을 지정할 수 있습니다. 이러한 옵션은 목표 필드를 사용하여 구간을 계산하는 방법을 알고리즘에 지시합니다.

구간 끝점. 낮거나 높은 끝점을 포함하거나(lower <= *x*) 제외하는지(lower < *x*) 여부를 지정할 수 있습니다.

첫 번째 및 마지막 구간. 첫 번째 및 마지막 구간 모두의 경우 구간이 최저 또는 최고 데이터 점으로 무제한(양수 또는 음수 무한대로 확장) 또는 제한되어야 하는지 여부를 지정할 수 있습니다.

⑦ 생성된 구간 미리보기

구간화 노드의 구간 값 탭을 사용하면 생성된 구간의 임계값을 볼 수 있습니다. 생성 메뉴를 사용하면 하나의 데이터 세트에서 다른 데이터 세트로 이러한 임계값을 적용하는 데 사용할 수 있는 파생 노드를 생성할 수도 있습니다.

구간화된 필드. 드롭 다운 목록을 사용하여 보고자 하는 필드를 선택할 수 있습니다. 표시되는 필드 이름은 명확히 하기 위해 원래 필드 이름을 사용합니다.

분위수. 드롭 다운 목록을 사용하여 보고자 하는 분위수(예: 10 또는 100)를 선택할 수 있습니다. 이 옵션은 분위수 방법(동일한 개수 또는 합계)을 사용하여 구간이 생성된 경우에만 사용할 수 있습니다.

구간 임계값. 각 구간에 속하는 레코드 수와 함께 생성된 각 구간에 대한 임계값이 여기에 표시됩니다. 최적 구간화 방법의 경우에만 각 구간의 레코드 수가 전체의 백분율로 표시됩니다. 순위 구간화 방법이 사용되는 경우 임계값을 적용할 수 없다는 점을 참고하십시오.

값 읽기. 데이터 세트에서 구간화된 값을 읽습니다. 새 데이터가 스트림을 통해 실행되는 경우 임계값도 겹쳐쓴다는 점을 참고하십시오.

파생 노드 생성

생성 메뉴를 사용하면 현재 임계값에 기반하여 파생 노드를 작성할 수 있습니다. 설정된 구간 임계값을 하나의 데이터 세트에서 다른 데이터 세트로 적용하는 데 유용합니다. 또한 이러한 분할 지점이 알려진 경우 대형 데이터 세트에 대해 작업할 때 파생 작업이 구간화 작업보다 효율적입니다(속도가 빠름을 의미함).

(10) RFM 분석 노드

RFM(Recency, Frequency, Monetary) 분석 노드를 사용하면 얼마나 최근에 사용자로부터 구매했는지(최근성), 얼마나 자주 구매했는지(빈도) 및 모든 트랜잭션에서 얼마나 소비했는지(구매총액)를 조사하여 최고의 고객이 될 수 있는 고객을 정량적으로 판별할 수 있습니다.

RFM 분석의 추론은 제품 또는 서비스를 한 번 구매한 고객이 다시 구매하는 경향이 있다는 것입니다. 분류된 고객 데이터는 사용자가 요구하는 대로 조정된 구간화 기준이 있는 여러 개의 구간으로 분할됩니다. 각 구간에서 고객에게 점수가 지정되고 이러한 점수가 결합되어 전체 RFM 점수를 제공합니다. 이 점수는 각 RFM 모수에 대해 작성된 구간에 대한 고객의 소속을 나타냅니다. 이 구간화된 데이터는 사용자의 요구를 충분히 충족할 수 있습니다. 예를 들어, 가장 빈번한 높은 값의 고객을 식별하거나 추가 모델링 및 분석에 필요한 스트림에서 전달될 수 있습니다.

단, RFM 점수를 분석하고 순위를 매기는 기능이 유용한 도구이기는 하지만 이를 사용할 때 특정 요인을 인식하고 있어야 합니다. 가장 높은 순위의 고객을 대상으로 하려는 유혹이 있을 수 있으나 해당 고객에게 지나치게 구매를 요구하는 경우 분노를 표출하거나 실제로는 반복 비즈니스가 실패할 수 있습니다. 또한 낮은 점수의 고객을 무시하는 대신 더 나은 고객으로 만들 수 있도록 노력해야 합니다. 반대로 높은 점수 단독으로는 시장에 따라 반드시 판매 전망이 높음을 나타내지 않을 수 있습니다. 예를 들어, 최근 범주에서 매우 최근에 구매했음을 의미하는 구간 5에 속하는 고객은 자동차 또는 텔레비전과 같은 비싸고 오래 사용하는 제품을 판매하려는 사람에게는 최선의 목표 고객이 아닐 수 있습니다.

참고: 데이터가 저장되는 방법에 따라 데이터를 유용한 형식으로 변환하기 위해 RFM 통합 노드를 RFM 분석 노드에 앞서 사용해야 합니다. 예를 들어, 입력 데이터는 고객당 한 행인 고객 형식이어야 합니다. 고객의 데이터가 트랜잭션 양식인 경우, RFM 통합 노드 업스트림을 사용하여 최근, 빈도 및 구매총액 필드를 파생시키십시오. 자세한 정보는 RFM 통합 노드 주제를 참조하십시오.

IBM® SPSS® Modeler의 RFM 통합 및 RFM 분석 노드는 독립적 구간화를 사용하기 위해 설정됩니다. 즉, 해당 값 또는 다른 두 측도에 관계없이 RFM(Recency, Frequency, Monetary) 값의 각 측도에 대한 데이터를 순위화하고 구간화합니다.

① RFM 분석 노드 설정

최근. 필드 선택기(텍스트 상자의 오른쪽에 있는 단추)를 사용하여 최근 필드를 선택하십시오. 이것은 날짜, 시간소인 또는 단순 숫자일 수 있습니다. 날짜 또는 시간소인이 가장 최근 트랜잭션의 날짜를 나타내는 경우 최고값은 가장 최근으로 간주됩니다. 숫자가 지정되면 가장 최근 트랜잭션 이후의 경과 시간을 나타내며 최저값이 가장 최근으로 간주됩니다.

참고: RFM 분석 노드에 앞서 RFM 통합 노드의 스트림이 사용되면 RFM 통합 노드에 의해 생성된 최근, 빈도 및 구매총액 필드가 RFM 분석 노드에서 입력으로 선택되어야 합니다.

빈도. 필드 선택기를 사용하여 사용할 빈도 필드를 선택하십시오.

구매총액. 필드 선택기를 사용하여 사용할 구매총액 필드를 선택하십시오.

구간 수. 세 개의 출력 유형 각각에 대해 작성할 구간 수를 선택하십시오. 기본값은 5입니다.

참고: 최소 구간 수는 2이며 최대 구간 수는 9입니다.

가중치. 기본적으로 점수를 계산할 때 최고 중요도는 최근 데이터에 지정되며 빈도가 그 다음, 구매총액에 마지막 중요도가 지정됩니다. 필요한 경우 이중 하나 또는 여러 개에 대해 가중치 영향을 수정하여 최고 중요도 지정을 변경할 수 있습니다.

RFM 점수는 (최근성 점수 x 최근성 가중치) + (구매빈도 점수 x 빈도 가중치) + (구매총액 점수 x 구매총액 가중치)로 계산됩니다.

등순위. 동일한(등순위) 점수를 구간화하는 방법을 지정합니다. 옵션은 다음과 같습니다.

- **다음에 추가**. 위로 다음 구간으로 경계값을 이동하려면 이를 선택하십시오.
- **현재구간에 유지**. 등순위 값을 현재(더 낮은) 구간에 유지합니다. 이 방법을 사용하면 더 적은 수의 총 구간이 작성될 수 있습니다. 이 옵션이 기본값입니다.

구간 임계값. 노드가 실행될 때 RFM 점수 및 구간 할당이 항상 다시 계산되는지 또는 필요한 경우(새 데이터가 추가되는 경우 등)에만 계산되는지 지정하십시오. **사용 가능한 경우 구간 값 탭에서 읽기**를 선택한 경우에는 구간 값 탭에서 다른 구간에 대한 상한 및 하한 절단점을 편집할 수 있습니다.

실행될 때 RFM 분석 노드가 원래 최근, 빈도 및 구매총액 필드를 구간화하고 다음 새 필드를 데이터 세트에 추가합니다.

- 최근성 점수. 최근성에 대한 순위(구간 값)
- 구매빈도 점수. 빈도에 대한 순위(구간 값)
- 구매총액 점수. 구매총액에 대한 순위(구간 값)
- RFM 점수. 최근성, 빈도 및 구매총액 점수의 가중된 합계입니다.

이상치를 마지막 구간에 추가. 이 선택란을 선택하면 하한 구간 아래에 위치한 레코드는 하한 구간에 추가되고 최고 구간 위의 레코드는 최고 구간에 추가되며 널 값이 지정됩니다. 이 선택란은 **사용 가능한 경우 구간 값 탭에서 읽기**를 선택한 경우에만 사용 가능합니다.

② RFM 분석 노드 구간화

구간 값 탭을 사용하면 생성된 구간에 대한 임계값을 볼 수 있으며 특정 케이스에 이를 수정할 수 있습니다.

참고: 설정 탭에서 **사용 가능한 경우 구간 값 탭에서 읽기**를 선택한 경우에만 이 탭에서 값을 수정할 수 있습니다.

구간화된 필드. 드롭 다운 목록을 사용하여 구간으로 나눌 필드를 선택하십시오. 사용 가능한 값은 설정 탭에서 선택된 값입니다.

구간 값 테이블. 생성된 각 구간에 대한 임계값이 여기에 표시됩니다. 설정 탭에서 **사용 가능한 경우 구간 값 탭에서 읽기**를 선택한 경우에는 관련 셀을 두 번 클릭하여 각 구간에 대한 상한 및 하한 절단점을 수정할 수 있습니다.

값 읽기. 데이터 세트에서 구간화된 값을 읽고 구간 값 테이블을 채웁니다. 설정 탭에서 **항상 다시 계산**을 선택한 경우에는 새 데이터가 스트림을 통해 실행될 때 구간 임계값을 덮어씁니다.

(11) 앙상블 노드

앙상블 노드는 둘 이상의 모델 너깃을 결합하여 개별 모델에서 얻을 수 있는 것보다 정확한 예측을 얻습니다. 여러 모델의 예측을 결합하면 개별 모델의 제한을 피할 수 있어 전반적인 정확도가 향상됩니다. 이 방식으로 결합된 모델은 일반적으로 개별 모델 이상의 성능을 보여줄 수 있습니다.

이러한 노드의 결합은 자동 분류자, 자동 숫자 및 자동 군집 자동화된 모델링 노드에서 자동으로 발생합니다.

앙상블 노드를 사용한 후에는 분석 노드 또는 평가 노드를 사용하여 결합된 결과의 정확도를 각각의 입력 모델과 비교할 수 있습니다. 이를 수행하려면 앙상블 노드의 설정 탭에서 **앙상블 모델이 생성한 필드 필터링** 옵션이 선택되지 않았는지 확인하십시오.

출력 필드

각각의 앙상블 노드는 결합된 스코어가 포함된 필드를 생성합니다. 이름은 지정된 목표 필드를 기반으로 하며 필드 측정 수준(플래그, 명목(세트) 또는 연속형(범위))에 따라 각각 *$XF_*, *$XS_* 또는 *$XR_* 접두부가 지정됩니다. 예를 들어, 목표가 *response*라는 플래그 필드인 경우 출력 필드는 *$XF_response*입니다.

신뢰도 또는 성향 필드. 플래그 및 명목 필드의 경우 다음 표에 자세히 설명된 대로 앙상블 방법을 기반으로 추가적인 신뢰도 또는 성향 필드가 작성됩니다.

표 1. 앙상블 방법 필드 작성

앙상블 방법	필드 이름
투표 신뢰 가중 투표 원시 성향 가중 투표 수정된 성향 가중 투표 가장 높은 신뢰도 승리	$XFC_<field>
평균 원시 성향	$XFRP_<field>
평균 수정된 원시 성향	$XFAP_<field>

① **앙상블 노드 설정**

앙상블의 목표 필드. 둘 이상의 업스트림 모델에서 목표로 사용하는 단일 필드를 선택하십시오. 업스트림 모델은 플래그, 명목 또는 연속형 대상을 사용할 수 있지만 스코어를 결합하려면 둘 이상의 모델이 동일한 목표를 공유해야 합니다.

앙상블 모델이 생성한 필드 필터링. 앙상블 노드에 반영되는 개별 모델이 생성한 모든 추가 필드의 출력에서 제거합니다. 모든 입력 모델에서 결합된 스코어에만 관심이 있는 경우 이 선택란을 선택합니다. 예를 들어, 분석 노드 또는 평가 노드를 사용하여 각 개별 입력 모델의 정확도와 결합된 스코어의 정확도를 비교하려는 경우에는 이 옵션을 선택 취소해야 합니다.

사용 가능한 설정은 목표로 선택된 필드의 측정 수준에 따라 다릅니다.

연속형 목표

연속형 대상의 경우 스코어는 평균값을 구합니다. 이는 스코어 결합을 위해 사용할 수 있는 유일한 방법입니다.

스코어 또는 추정값의 평균을 구하는 경우 앙상블 노드는 표준 오차 계산을 사용하여 측정되거나 추정된 값과 참 값 사이의 차이를 계산하고 해당 추정값이 얼마나 일치하는지 표시합니다. 표준 오차 계산은 새 모델에 대해 기본적으로 생성되지만 기존 모델에 대해 선택란을 선택 취소할 수 있습니다(예: 재생성해야 하는 경우).

범주형 대상

범주형 대상의 경우 각각의 가능한 예측값이 선택되는 횟수를 기록하고 총계가 가장 큰 값을 선택하여 작동하는 **투표**를 포함한 다수의 방법이 지원됩니다. 예를 들어, 5개 모델 중 3개가 *yes*

를 예측하고 다른 두 모델은 *no*를 예측하는 경우 *yes*가 3 대 2 투표로 이깁니다. 또는 각 예측에 대한 신뢰도 및 성향 값을 기반으로 투표에 **가중치**가 부여될 수 있습니다. 그런 다음 가중치가 합계되고 총계가 가장 높은 값이 다시 선택됩니다. 최종 예측에 대한 신뢰도는 앙상블에 포함된 모델의 수로 나눈 승리 값에 대한 가중치의 합계입니다.

모든 범주형 필드. 플래그 필드와 명목 필드 모두에 대해 다음과 같은 방법이 지원됩니다.

- 투표
- 신뢰 가중 투표
- 가장 높은 신뢰도 승리

플래그 필드만. 플래그 필드만의 경우 성향을 기반으로 한 다수의 방법도 사용할 수 있습니다.

- 원시 성향 가중 투표
- 수정된 성향 가중 투표
- 평균 원시 성향
- 평균 수정된 성향

투표 동률. 투표 방법에 대해 동률을 해결하는 방법을 지정할 수 있습니다.

- **무작위 선택**. 동률 값 중 하나가 무작위로 선택됩니다.
- **가장 높은 신뢰도**. 가장 높은 신뢰도로 예측된 동률 값이 승리합니다. 이것이 반드시 모든 예측값의 가장 높은 신뢰도와 동일하지는 않습니다.
- **원시 또는 수정된 성향(플래그 필드만)**. 가장 큰 절대 성향으로 예측된 동률 값입니다. 여기서 절대 성향은 다음과 같이 계산됩니다.

```
abs(0.5 - propensity) * 2
```

수정된 성향의 경우 다음과 같이 계산됩니다.

```
abs(0.5 - adjusted propensity) * 2
```

(12) 파티션 노드

파티션 노드는 파티션 필드를 생성하는데 사용되며 이 필드는 모델 작성의 학습, 검정, 검증 단계를 위한 별개의 서브세트 또는 표본으로 데이터를 분할합니다. 일표본을 사용하여 모델을 생성하고, 별도의 표본으로 검정하면, 현재 데이터와 유사한 더 큰 데이터 세트에 대해 모델을 일반화할 때 효율성을 효과적으로 표시할 수 있습니다.

파티션 노드는 **파티션**으로 설정된 역할이 있는 명목 필드를 생성합니다. 또는 적절한 필드가 사용자의 데이터에 이미 있는 경우에는 유형 노드를 사용하여 파티션으로 지정될 수 있습니다. 이 케이스에서 별도의 파티션 노드가 필요하지 않습니다. 두 개 또는 세 개의 값이 있는 인스턴스화된 모든 명목 필드는 파티션으로 사용될 수 있으나 플래그 필드는 사용될 수 없습니다. 자세한 정보는 필드 역할 설정 주제를 참조하십시오.

다중 파티션 필드가 스트림에서 정의될 수 있으나 이 경우, 분할을 사용하는 각 모델링 노드의 필드 탭에서 단일 파티션 필드를 선택해야 합니다. (하나의 파티션만 존재하는 경우 파티션이 사용될 때마다 자동으로 사용됩니다.)

파티션 사용. 분석에서 파티션을 사용하려면 적절한 모델 작성 또는 분석 노드의 모형 옵션 탭에서 파티션을 사용할 수 있어야 합니다. 이 옵션을 선택 취소하면 필드를 제거하지 않은 채로 분할을 사용할 수 없게 됩니다.

날짜 범위 또는 위치와 같이 몇 가지 기타 기준을 기반으로 하여 파티션 필드를 작성하려면 파생 노드를 사용할 수 있습니다. 자세한 정보는 파생 노드의 내용을 참조하십시오.

예. 이전 마케팅 캠페인에 긍정적으로 반응한 최근 고객을 식별하기 위해 RFM 스트림을 작성하는 경우, 판매 회사의 마케팅 부서가 파티션 노드를 사용하여 데이터를 학습 및 검정 파티션으로 분할합니다.

① 파티션 노드 옵션

파티션 필드. 노드에 의해 생성된 필드의 이름을 지정합니다.

파티션. 데이터를 두 표본(학습 및 검정) 또는 세 표본(학습, 검정 및 검증)으로 파티셔닝할 수 있습니다.

- **학습 및 검정**. 데이터를 두 개의 표본으로 파티셔닝하여 한 표본으로 모델을 학습하고 다른 표본으로 모델을 검정할 수 있습니다.
- **학습, 검정 및 검증**. 데이터를 세 개의 표본으로 파티셔닝하여 한 표본으로 모델을 학습하고 두 번째 표본으로 모델을 검정하고 세분화하며 세 번째 표본으로 결과를 검증할 수 있습니다. 따라서 각 파티션의 크기는 축소되나 매우 큰 데이터 세트를 사용하여 작업할 때에 비해 더 적합할 수 있습니다.

파티션 크기. 각 파티션의 상대적인 크기를 지정합니다. 파티션 크기의 합계가 100% 미만이면 파티션에 포함되지 않은 레코드가 삭제됩니다. 예를 들어, 사용자에게 천만 레코드가 있으며 파티션 크기를 학습 5%, 검정 10%로 지정한 경우 노드를 실행한 후에 나머지는 삭제되고 대략 500,000개의 학습 레코드 및 백만 개의 검정 레코드가 있어야 합니다.

값. 데이터에서 각 파티션 표본을 나타내기 위해 사용하는 값을 지정합니다.

- **시스템 정의 값 사용("1," "2," 및 "3")**. 각 파티션을 나타내기 위해 정수를 사용합니다. 예를 들어, 학습 표본에 포함되는 모든 레코드는 파티션 필드의 값으로 1을 사용합니다. 그러면 데이터를 로케일 사이에서 이동할 수 있으며 파티션 필드가 다른 위치에서 다시 인스턴스화되는 경우(데이터베이스에서 다시 데이터를 읽는 경우 등), 정렬 순서가 유지됩니다. 따라서 1이 여전히 학습 파티션을 나타냅니다. 그러나 값에 몇 가지 해석이 필요합니다.
- **시스템 정의 값에 레이블 추가**. 정수와 레이블을 조합합니다. 예를 들어, 학습 파티션 레코드는 1_Training 값을 사용합니다. 그러면 정렬 순서를 유지하면서도 데이터를 보는 사용자가 값을 식별할 수 있습니다. 그러나 값은 지정된 로케일에 한정됩니다.
- **값으로 레이블 사용**. 정수 없이 레이블을 사용합니다. 예를 들어, **Training**과 같습니다. 이 경우, 레이블을 편집하여 값을 지정할 수 있습니다. 그러나 이로 인해 데이터가 로케일에 한정되며 파티션 열을 다시 인스턴스화하면 값이 기본 정렬 순서로 배열되어 "시맨틱" 순서와 일치하지 않을 수 있습니다.

시드. 반복 가능한 파티션 할당이 선택된 경우에만 사용 가능합니다. 임의 퍼센트를 기반으로 사용하여 표본추출 레코드 또는 파티셔닝을 수행하는 경우에 이 옵션을 사용하면 다른 세션에서 동일한 결과를 복제할 수 있습니다. 난수 생성기에서 사용하는 시작값을 지정하면 노드를 실행할 때마다 동일한 레코드를 지정하도록 보장할 수 있습니다. 원하는 시드 값을 입력하거나 **생성** 단추를 클릭하여 자동으로 임의 값을 생성하십시오. 이 옵션을 선택하지 않으면 노드를 실행할 때마다 다른 표본이 생성됩니다.

> ⓘ **참고**: 데이터베이스에서 읽은 레코드에서 **시드** 옵션을 사용하는 경우 노드를 실행할 때마다 동일한 결과를 보장하려면 표본 추출 전에 정렬 노드가 필요할 수도 있습니다. 난수 시드는 레코드 순서에 의존하여 관계형 데이터베이스에서는 동일하게 보장되지 않기 때문입니다. 자세한 정보는 정렬 노드의 내용을 참조하십시오.

고유 필드를 사용하여 파티션 지정. 반복 가능한 파티션 할당이 선택된 경우에만 사용 가능합니다. (계층 1 데이터베이스 전용) SQL 푸시백을 사용하여 레코드를 파티션에 지정하려면 이 선택란을 선택하십시오. 레코드가 무작위이지만 반복 가능한 방식으로 지정되도록 하려면 드롭 다운 목록에서 고유 값이 있는 필드(ID 필드 등)를 선택하십시오.

데이터베이스 계층에 대서는 데이터베이스 소스 노드에 대한 설명을 참조하십시오. 자세한 정보는 데이터베이스 소스 노드의 내용을 참조하십시오.

선택 노드 생성

파티션 노드에서 일반 메뉴를 사용하여 각 파티션에 대한 선택 노드를 자동으로 생성할 수 있습니다. 예를 들어, 학습 파티션에서 모든 레코드를 선택하여 이 파티션만 사용하여 추가 평가 또는 분석을 작성할 수 있습니다.

(13) 플래그로 설정 노드

플래그로 설정 노드는 하나 이상의 명목 필드에 대해 정의된 범주형 값을 바탕으로 다중 플래그 필드를 파생시키는 데 사용됩니다. 예를 들어, 사용자의 데이터 세트에 값이 *높음*, *보통* 및 *낮음*인 *BP*(혈압) 명목 필드가 포함될 수 있습니다. 데이터 조작을 더 쉽게 하려면 환자의 혈압이 높은지 여부를 표시하는 높은 혈압에 대한 플래그 필드를 작성할 수 있습니다.

① 플래그로 설정 노드에 대한 옵션 설정

세트 필드. 측정 수준이 *명목*(변수군)인 모든 데이터 필드를 나열합니다. 목록에서 하나를 선택하여 변수군 내의 값을 표시하십시오. 해당 값에서 선택하여 플래그 필드를 작성할 수 있습니다. 사용 가능한 명목 필드 및 해당 값을 보려면 데이터가 업스트림 소스 또는 유형 노드를 사용하여 완전히 인스턴스화되어야 합니다. 자세한 정보는 유형 노드 주제를 참조하십시오.

필드 이름 확장. 새 플래그 필드에 대한 접미문자 또는 접두문자로 추가될 확장자를 지정하기 위한 제어를 선택하십시오. 기본적으로 새 필드 이름은 원래 필드 이름과 필드 값을 레이블에 조합하여 자동으로 생성됩니다. 예를 들어, *Fieldname_fieldvalue*와 같습니다.

사용가능 변수군 값. 위에서 선택된 변수군 내의 값이 여기에 표시됩니다. 플래그를 생성할 값을 하나 이상 선택하십시오. 예를 들어, *blood_pressure* 필드의 값이 *높음*, *중간* 및 *낮음*이면 *높음*을 선택하고 이를 오른쪽 목록에 추가할 수 있습니다. 그러면 높은 혈압을 표시하는 값이 있는 레코드에 대한 플래그가 있는 필드가 생성됩니다.

플래그 필드 생성. 새로 생성된 플래그 필드가 여기에 나열됩니다. 필드 이름 확장자 제어를 사용하여 새 필드의 이름을 지정하기 위한 옵션을 지정할 수 있습니다.

참 값. 플래그를 설정할 때 노드가 사용하는 참 값을 지정합니다. 기본적으로 이 값은 **T**입니다.

거짓 값. 플래그를 설정할 때 노드가 사용하는 거짓 값을 지정합니다. 기본적으로 이 값은 **F**입니다.

통합 키. 아래에 지정된 키 필드를 기준으로 하여 레코드를 함께 그룹화하려면 선택하십시오. **통합 키**를 선택하면 *임의의* 레코드가 참으로 설정된 경우에 그룹 내의 모든 플래그 필드가 "켜집니다". 레코드를 통합하는 데 사용할 키 필드를 지정하려면 필드 선택기를 사용하십시오.

(14) 구조변환 노드

명목형 또는 플래그 필드의 값을 기준으로 하여 다중 필드를 생성하는 데 구조변환 노드가 사용될 수 있습니다. 새로 생성된 필드는 또 다른 필드 또는 숫자 플래그(0 및 1)의 값을 포함할 수 있습니다. 이 노드의 기능은 플래그로 설정 노드의 기능과 유사합니다. 단, 더 많은 융통성을 제공합니다. 이 노드를 사용하면 또 다른 필드의 값을 사용하여 숫자 플래그를 포함하여 모든 유형의 필드를 작성할 수 있습니다. 그런 다음 기타 노드 다운스트림을 사용하여 통합 또는 기타 조작을 수행할 수 있습니다. 플래그로 설정 노드를 사용하면 한 단계로 필드를 통합할 수 있으며 플래그 필드를 생성할 때 편리합니다.

예를 들어, 다음 데이터 세트에는 값이 *Savings* 및 *Draft*인 *Account*라는 명목 필드가 포함됩니다. 각 계정에 대해 시작 잔액 및 현재 잔액이 기록되고 일부 고객은 각 유형의 계정이 여러 개 있을 수 있습니다. 각 고객이 특정 계정 유형을 갖고 있는지 파악하고 해당되는 경우 각 계정 유형의 금액이 얼마인지 알고자 하는 경우를 가정해 보십시오. 구조변환 노드를 사용하면 각 *Account* 값에 대한 필드를 생성하고 값으로 *Current_Balance*를 선택할 수 있습니다. 각 새 필드는 지정된 레코드에 대한 현재 잔액으로 채워집니다.

표 1. 구조변환 전의 표본 데이터

CustID	Account	Open_Bal	Current_Bal
12701	Draft	1000	1005.3
12702	Savings	100	144.51
12703	Savings	300	321.2
12703	Savings	150	204.51
12703	Draft	1200	586.32

표 2. 구조변환 후의 표본 데이터

CustID	Account	Open_Bal	Current_Bal	Account_Draft_Current_Bal	Account_Savings_Current_Bal
12701	Draft	1000	1005.3	1005.3	$null$
12702	Savings	100	144.51	$null$	144.51
12703	Savings	300	321.2	$null$	321.2
12703	Savings	150	204.51	$null$	204.51
12703	Draft	1200	586.32	586.32	$null$

통합 노드와 함께 구조변환 노드 사용

구조변환 노드와 통합 노드를 한 쌍으로 만들고자 하는 경우가 많습니다. 이전 예에서 ID가 12703인 한 고객은 세 개의 계정을 갖고 있습니다. 통합 노드를 사용하여 각 계정 유형에 대한 총 잔액을 계산할 수 있습니다. 키 필드는 *CustID*이며 통합 필드는 새로 구조변환된 필드인 *Account_Draft_Current_Bal* 및 *Account_Savings_Current_Bal*입니다. 다음은 결과를 표시하는 표입니다.

표 3. 구조변환 및 통합 후의 표본 데이터

CustID	Record_Count	Account_Draft_Current_Bal_Sum	Account_Savings_Current_Bal_Sum
12701	1	1005.32	$null$
12702	1	$null$	144.51
12703	3	586.32	525.71

① 구조변환 노드에 대한 옵션 설정

사용 가능한 필드. 측정 수준이 *명목형*(변수군) 또는 *플래그*인 모든 데이터 필드를 나열합니다. 목록에서 하나를 선택하여 변수군 또는 플래그 내의 값을 표시한 다음 해당 값에서 선택하여 구조변환 필드를 작성하십시오. 사용 가능한 필드 및 해당 값을 보려면 데이터가 업스트림 소스 또는 유형 노드를 사용하여 완전히 인스턴스화되어야 합니다. 자세한 정보는 유형 노드 주제를 참조하십시오.

사용가능 값. 위에서 선택된 변수군 내의 값이 여기에 표시됩니다. 구조변환 필드를 생성할 값을 하나 이상 선택하십시오. 예를 들어, 혈압 필드의 값이 높음, 중간 및 낮음이면 높음을 선택하고 이를 오른쪽 목록에 추가할 수 있습니다. 그러면 값이 높음인 레코드에 대해 지정된 값(아래 참조)이 있는 필드가 생성됩니다.

구조변환 필드 생성. 새로 생성된 구조변환 필드가 여기에 나열됩니다. 기본적으로 새 필드 이름은 원래 필드 이름과 필드 값을 레이블에 조합하여 자동으로 생성됩니다. 예를 들어, *Fieldname_fieldvalue*와 같습니다.

필드 이름 포함. 새 필드 이름에서 접두문자로 원래 필드 이름을 제거하려면 선택 취소하십시오.

기타 필드의 값 사용. 구조변환 필드를 채우는 데 값을 사용할 필드를 한 개 이상 지정하십시오. 한 개 이상의 필드를 선택하려면 필드 선택기를 사용하십시오. 선택된 각 필드에 대해 하나의 새 필드가 생성됩니다. 값 필드 이름이 구조변환된 필드 이름에 추가됩니다. 예를 들어, *BP_High_Age* 또는 *BP_Low_Age*와 같습니다. 각 새 필드는 원래 값 필드의 유형을 상속합니다.

숫자 값 플래그 생성. 다른 필드의 값을 사용하지 않고 숫자 값 플래그(거짓의 경우 0, 참의 경우 1)로 새 필드를 채우려면 선택하십시오.

(15) 전치 노드

기본적으로 열은 필드이고 행은 레코드 또는 관측입니다. 필요한 경우, 필드가 레코드가 되고 레코드가 필드가 되도록 전치 노드를 사용하여 행과 열의 데이터를 스왑할 수 있습니다. 예를 들어, 각 계열이 열이 아니고 행인 시계열 데이터가 있는 경우, 분석 전에 데이터를 전치시킬 수 있습니다.

① 전치 노드의 옵션 설정

전치 방법 드롭 다운에서 전치 노드를 수행할 방법을 선택하십시오. **두 필드와 레코드 모두**, **레코드를 필드로** 또는 **필드를 레코드**로. 세 가지 방법의 각 설정은 다음 절에서 설명합니다.

> ⊘ **제한사항:** 레코드를 필드로 및 필드를 레코드로 방법은 Windows 64비트, Linux 64비트 및 Mac에서만 지원됩니다.

두 피드와 레코드 모두

새 필드 이름은 지정된 접두부를 기반으로 자동으로 생성하거나 데이터의 기존 필드에서 읽어올 수 있습니다.

접두부 사용. 이 옵션은 지정된 접두부(Field1, Field2 등)를 기반으로 자동으로 새 필드 이름을 생성합니다. 필요에 따라 접두부를 사용자 정의할 수 있습니다. 이 옵션을 사용하는 경우에는 원래 데이터의 행 수와 상관없이 작성할 필드 수를 지정해야 합니다. 예를 들어, **새 필드 수**를 100으로 설정하면 처음 100행 이후의 모든 데이터는 삭제됩니다. 원래 데이터에 100개 미만의 행이 있는 경우 일부 필드가 널이 됩니다. (필요에 따라 필드 수를 늘릴 수 있지만, 이 설정의 목적은 백만 개의 레코드를 백만 개의 필드로 전치하는 것을 피하기 위한 것입니다. 이러한 전치는 감당할 수 없는 결과를 초래할 수 있습니다.
예를 들어, 행에 계열이 있고 각 월에 대한 개별 필드(열)가 있는 데이터가 있다고 가정하십시오. 각 계열이 개별 필드에 있고 각 월이 하나의 행에 있도록 이 데이터를 전치할 수 있습니다.

필드에서 읽기. 기존 필드에서 필드 이름을 읽습니다. 이 옵션을 사용하는 경우, 새 필드 수는 데이터에 의해 지정된 최대값까지 결정됩니다. 선택된 필드의 각 값은 출력 데이터에서 새 필드가 됩니다. 선택된 필드의 저장 유형(정수, 문자열, 날짜 등)에는 제한이 없지만, 중복된 필드 이름을 피하려면 선택된 필드의 각 값이 고유해야 합니다(즉, 값 수가 행 수와 일치해야 함). 중복된 필드 이름이 발견되면 경고가 표시됩니다.

- **값 읽기**. 선택된 필드가 인스턴스화되지 않은 경우 이 옵션을 선택하여 새 필드 이름 목록을 채우십시오. 필드가 이미 인스턴스화된 경우에는 이 단계가 필요하지 않습니다.
- **읽을 값 최대 수**. 데이터에서 필드 이름을 읽을 때 지나치게 많은 수의 필드가 작성되지 않도록 하기 위해 상한을 지정합니다. (앞에서 언급한 바와 같이, 백만 개의 레코드를 백만 개의 필드로 전치하는 경우 감당할 수 없는 결과를 초래합니다.)

예를 들어, 데이터의 첫 번째 열이 각 계열의 이름을 지정하는 경우, 전치된 데이터에서 이러한 값을 필드 이름으로 사용할 수 있습니다.

전치. 기본적으로 연속형(숫자 범위) 필드만 전치됩니다(정수 또는 실제 저장 공간). 선택적으로 숫자 필드의 서브세트를 선택하거나 문자열 필드를 대신 전치할 수 있습니다. 그러나, 전치되는 모든 필드는 동일한 저장 유형(숫자 또는 문자열이어야 하고 둘 다가 아니어야 함)이어야 합니다. 그 이유는 입력 필드를 혼합할 경우 각 출력 열 내에 혼합된 값이 생성되며, 이는 필드의 모든 값이 동일한 저장 공간을 가져야 하는 규칙을 위반하기 때문입니다. 다른 저장 유형(날짜, 시간, 시간소인)은 전치할 수 없습니다.

- **모든 숫자**. 모든 숫자 필드(정수 또는 실제 저장 공간)를 전치합니다. 출력의 행 수가 원래 데이터의 숫자 필드 수와 일치합니다.
- **모든 문자열**. 모든 문자열 필드를 전치합니다.
- **사용자 정의** 숫자 필드의 서브세트를 선택할 수 있습니다. 출력의 행 수가 선택된 필드 수와 일치합니다. 이 옵션은 숫자 필드에만 사용할 수 있습니다.

행 ID 이름. 노드에 의해 작성되는 행 ID 필드의 이름을 지정합니다. 이 필드의 값은 원래 데이터에 있는 필드의 이름으로 결정됩니다.

> **팁**: 행의 시계열 데이터를 열로 전치할 때, 원래의 데이터에 각 측정 주기의 레이블을 지정하는 행(예: 날짜, 월, 연도 등)이 포함되는 경우, 데이터의 첫 번째 행에 해당 레이블을 포함시키기 보다는 이러한 레이블을 IBM® SPSS® Modeler에 필드 이름으로 읽어와야 합니다(위의 예에 설명된 바와 같이, 월 또는 날짜를 각각 원래 데이터의 필드 이름으로 표시함). 그러면 각 열에서 레이블과 값이 혼합되지 않습니다(열 내에서 저장 유형을 혼합할 수 없기 때문에 숫자가 문자열로 읽혀지도록 함).

레코드를 필드로

필드. 필드 목록에는 전치 노드에 입력되는 모든 필드를 포함합니다.

색인. 색인 섹션을 사용하여 색인 필드로 사용할 필드를 선택하십시오.

필드. 필드 섹션을 사용하여 필드로 사용할 필드를 선택하십시오.

값 값 섹션을 사용하여 값 필드로 사용할 필드를 선택하십시오.

집계 함수. 색인의 레코드가 두 개 이상이면 레코드를 하나로 통합해야 합니다. **집계 함수** 드롭다운을 사용하여 다음 함수 중 하나를 사용하여 레코드를 통합하는 방법을 지정하십시오. 통합은 모든 필드에 영향을 미칩니다.

- **평균**. 각 키 필드 조합의 평균값을 리턴합니다. 평균은 중심 경향의 측도이며 산술 평균(합계를 케이스 수로 나눈 값)입니다.
- **합계** 각 키 필드 조합의 합계 값을 리턴합니다. 합계는 결측값이 있는 모든 케이스에서 값의 총계입니다.
- **최소** 각 키 필드 조합의 최소값을 리턴합니다.
- **최대** 각 키 필드 조합의 최대값을 리턴합니다.
- **중앙값** 각 키 필드 조합의 중앙값을 리턴합니다.중앙값은 벗어난 값의 영향을 받지 않는 중심 경향 측도이며 평균과 달리 상한 극단값 또는 하한 극단값의 영향을 받을 수 있습니다. 50번째 백분위수 또는 제2사분위수라고도 합니다.
- **개수** 각 키 필드 조합의 널이 아닌 값의 개수를 리턴합니다.

필드에서 레코드로

필드. 필드 목록에는 전치 노드에 입력되는 모든 필드를 포함합니다.

색인. 색인 섹션을 사용하여 색인 필드로 사용할 필드를 선택하십시오.

값 값 섹션을 사용하여 값 필드로 사용할 필드를 선택하십시오. 값 필드를 선택하지 않으면 지정되지 않은 모든 숫자 필드를 값으로 사용합니다. 숫자 필드를 사용할 수 없으면 지정되지 않은 모든 문자열 필드가 사용됩니다.

(16) 히스토리 노드

히스토리 노드는 시계열 데이터 같은 순차 데이터에 가장 자주 사용합니다. 이전 레코드에 있는 필드의 데이터가 포함된 새 필드를 작성하는 데 사용됩니다. 히스토리 노드를 사용하는 경우 특정 필드별로 사전 정렬되는 데이터를 가지길 원할 수 있습니다. 정렬 노드를 사용하여 이를 수행할 수 있습니다.

① 히스토리 노드에 대한 옵션 설정

선택된 필드. 필드 선택기(텍스트 상자 오른쪽의 단추)를 사용하여 히스토리를 원하는 필드를 선택하십시오. 각각의 선택된 필드는 데이터 세트의 모든 레코드에 대한 새 필드를 작성하는 데 사용됩니다.

오프셋. 히스토리 필드 값을 추출할 현재 레코드 이전의 마지막 레코드를 지정하십시오. 예를 들어, 오프셋이 3으로 설정된 경우에는 각각의 레코드가 이 노드를 통과할 때 이전 세 번째 레코드에 대한 필드 값이 현재 레코드에 포함됩니다. 범위 설정을 사용하여 이전 몇 번째 레코드까지 추출할지 지정하십시오. 화살표를 사용하여 오프셋 값을 조정하십시오.

범위. 값을 추출할 이전 레코드의 수를 지정하십시오. 예를 들어, 오프셋이 3으로 설정되고 범위가 5로 설정되면 노드를 통과하는 각각의 레코드에는 선택된 필드 목록에서 지정된 각각의 필드에 대해 5개의 필드가 추가됩니다. 이는 노드가 레코드 10을 처리하면 레코드 7부터 레코드 3까지의 필드가 추가됨을 의미합니다. 화살표를 사용하여 범위 값을 조정하십시오.

히스토리가 사용 불가능한 경우. 히스토리 값이 없는 레코드를 처리하기 위해 다음 옵션 중 하나를 선택하십시오. 이는 일반적으로 히스토리로 사용할 이전 레코드가 없는 데이터 세트 맨 위의 처음 몇몇 레코드를 나타냅니다.

- **레코드 삭제**. 선택된 필드에 사용 가능한 히스토리 값이 없는 레코드를 삭제하려면 선택하십시오.
- **히스토리를 정의하지 않은 상태로 두기**. 히스토리 값을 사용할 수 없는 레코드를 보존하려면 선택하십시오. 히스토리 필드는 $null$로 표시된 정의되지 않은 값으로 채워집니다.
- **값 채우기**. 히스토리 값을 사용할 수 없는 레코드에 사용할 값 또는 문자열을 지정하십시오. 기본 대체값은 시스템 널인 *undef*입니다. 널값은 문자열 $null$를 사용하여 표시됩니다.

대체값을 선택할 때는 적절한 실행이 수행되도록 하기 위해 다음과 같은 규칙에 유의하십시오.

- 선택된 필드의 저장 유형은 동일해야 합니다.
- 선택된 필드가 모두 숫자 저장 공간을 가진 경우 대체값은 정수로 구문 분석되어야 합니다.
- 선택된 필드가 모두 실수 저장 공간을 가진 경우 대체값은 실수로 구문 분석되어야 합니다.
- 선택된 필드가 모두 기호 저장 공간을 가진 경우 대체값은 문자열로 구문 분석되어야 합니다.
- 선택된 필드가 모두 날짜/시간 저장 공간을 가진 경우 대체값은 날짜/시간 필드로 구문 분석되어야 합니다.

위 조건 중에 충족되는 조건이 없는 경우에는 히스토리 노드 실행 시 오류가 수신됩니다.

(17) 필드 다시 정렬 노드

필드 다시 정렬 노드를 사용하면 필드를 다운스트림으로 표시하는 데 사용하는 기본 순서를 정의할 수 있습니다. 이 순서는 테이블, 목록 및 필드 선택기 같은 다양한 장소에서 필드의 표시에 영향을 줍니다. 이 작업은 관심 있는 필드를 더 잘 보이게 만들기 위해 넓은 데이터 세트에 대해 작업하는 등의 경우에 유용합니다.

① **필드 재정렬 설정 옵션**

필드를 다시 정렬하는 방법은 사용자 정의 순서 및 자동 정렬 두 가지입니다.

사용자 정의 순서

사용자 정의 순서를 선택하면 모든 필드를 볼 수 있으며 화살표 단추를 사용할 수 있는 필드 이름 및 유형 테이블을 사용하여 사용자 정의 순서를 작성할 수 있습니다.

필드를 다시 정렬하려면 다음을 수행하십시오.

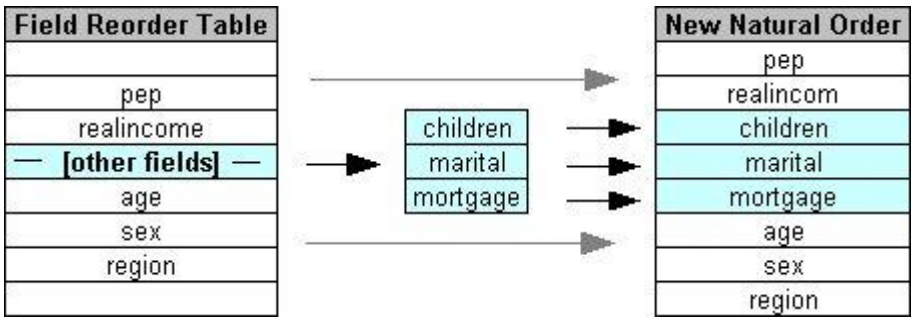
1. 테이블에서 필드를 선택하십시오. Ctrl-클릭 방법을 사용하여 다중 필드를 선택하십시오.
2. 단순 화살표 단추를 사용하여 필드를 한 행 위 또는 아래로 이동하십시오.
3. 선 화살표 단추를 사용하여 필드를 목록의 맨 아래나 맨 위로 이동하십시오.
4. [기타 필드]로 표시되는 구분선 행을 위 또는 아래로 이동하여 여기에 포함하지 않을 필드 정렬을 지정하십시오.

[기타 필드]에 대한 자세한 정보

기타 필드. **[기타 필드]** 구분선 행의 목적은 테이블을 두 개의 절반으로 나누는 것입니다.

- 구분선 행 위에 표시되는 필드는 테이블에 표시될 때 이 노드의 필드 다운스트림을 표시하는 데 사용된 모든 자연 순서의 맨 위에 정렬됩니다.
- 구분선 행 아래에 표시되는 필드는 테이블에 표시될 때 이 노드의 필드 다운스트림을 표시하는 데 사용된 모든 자연 순서의 맨 아래에 정렬됩니다.

그림 1. "기타 필드"가 새 필드 순서에 통합되는 방법을 설명하는 다이어그램



- 필드 다시 정렬 테이블에 표시되지 않는 모든 기타 필드는 구분선 행의 위치에 의해 표시되는 대로 "위쪽" 및 "아래쪽" 필드 사이에 표시됩니다.

추가적인 사용자 정의 정렬 옵션은 다음과 같습니다.

- 각 열 헤더(**유형**, **이름**, **저장 공간**) 위에 있는 화살표를 클릭하여 오름차순 또는 내림차순으로 필드를 정렬하십시오. 열 기준으로 정렬하는 경우, 여기서 지정되지 않은 필드([기타 필드] 행으로 표시됨)는 자연 순서 마지막에 정렬됩니다.
- **미사용 항목 지우기**를 클릭하여 필드 재정렬 노드에서 사용되지 않은 필드를 모두 삭제하십시오. 사용되지 않은 필드는 테이블에 빨강 글꼴로 표시됩니다. 이는 필드가 업스트림 조작에서 삭제되었음을 표시합니다.
- (새 필드 또는 지정되지 않은 필드를 표시하기 위한 빛나는 아이콘으로 표시되는) 새 필드에 대한 순서를 지정하십시오. **확인** 또는 **적용**을 클릭하면 아이콘이 없어집니다.

참고: 사용자 정의 순서가 적용된 후에 필드가 업스트림에 추가되면 새 필드가 사용자 정의 목록의 아래쪽에 추가됩니다.

자동 정렬

정렬 모수를 지정하려면 **자동 정렬**을 선택하십시오. 자동 정렬에 대한 옵션을 제공하기 위해 대화 상자 옵션이 동적으로 변경됩니다.

정렬기준. 필드 읽기를 재정렬 노드로 정렬하는 세 가지 방법 중 하나를 선택하십시오. 화살표 단추는 순서가 오름차순 또는 내림차순인지 표시합니다. 변경하려면 하나를 선택하십시오.

- 이름
- 유형
- 저장 공간

자동 정렬이 적용된 후에 필드 재정렬 노드의 업스트림에 추가된 필드는 선택된 정렬 유형을 기준으로 하여 자동으로 적절한 위치에 배치됩니다.

(18) 시간 구간 노드

SPSS® Modeler 버전 17.1 이하에서 사용 가능한 원래 시간 구간 노드는 AS(Analytic Server)와 호환되지 않으며 SPSS Modeler 릴리스 18.0에서는 더 이상 사용되지 않습니다.

대체 시간 구간 노드에는 원래 시간 구간 노드에서 변경된 사항의 수가 포함됩니다. 이 새 노드는 Analytic Server과 함께 사용되거나 SPSS Modeler에서 단독으로 사용될 수 있습니다.

시간 구간 노드를 사용하여 간격을 지정하고 추정하거나 예측하기 위한 새 시간 필드를 파생시키십시오. 초부터 년까지, 모든 범위의 시간 구간이 지원됩니다.

노드를 사용하여 새 시간 필드를 파생시키십시오. 새 필드에 사용자가 선택한 입력 시간 필드와 동일한 저장 유형이 있습니다. 노드는 다음 항목을 생성합니다.

- 선택한 접두부/접미부와 함께 필드 탭에 **시간 필드**로 지정된 필드. 기본적으로 접두부는 $TI_ 입니다.
- 필드 탭에 **차원 필드**로 지정된 필드.
- 필드 탭에 **통합할 필드**로 지정된 필드.

선택한 간격 또는 기간(예: 측정이 속하는 분 또는 초)에 따라 여러 추가 필드가 생성될 수도 있습니다.

① 시간 간격 - 필드 옵션

시간 구간 노드의 필드 탭을 사용하면 새 시간 구간이 파생되는 데이터를 선택할 수 있습니다.

필드 노드에 대한 모든 입력 필드와 해당 측정 유형 아이콘을 표시합니다. 모든 시간 필드에는 '연속형' 측정 유형이 있습니다. 입력으로 사용할 필드를 선택하십시오.

시간 필드 새 시간 간격이 파생되는 입력 필드를 표시합니다. 단일 연속형 필드만 허용됩니다. 이 필드는 시간 구간 노드에서 간격을 변환하기 위한 통합 키로 사용합니다. 새 필드에는 선택한 입력 시간 필드와 동일한 저장 유형이 있습니다. 정수 필드를 선택하는 경우 시간 지수로 간주됩니다.

차원 필드 선택적으로 여기에 필드를 추가하여 필드 값에 기반한 개별 시계열을 작성할 수 있습니다. 간단한 예로 특정 지역과 관련된 데이터를 사용하는 경우 차원으로 위치 필드를 사용할 수 있습니다. 이 예에서 시간 구간 노드의 데이터 출력이 위치 필드에 있는 각 위치 값의 시계열로 정렬됩니다.

차원은 TM1 노드에서 생성한 데이터와 비슷한 플랫 다차원 데이터를 사용하거나 지리공간과 같이 더 복잡한 데이터 유형을 지원하는 데 적합합니다. 근본적으로 **차원 필드**를 SQL 조회의 **Group By** 절과 동등한 기능 또는 통합 노드의 **키 필드**와 비슷하게 사용할 수 있습니다. 그러나 **차원 필드**는 기존 행 및 열 데이터보다 복잡한 데이터 구조를 처리하는 기능이 있으므로 더욱 정교한 특징을 갖습니다.

통합할 필드 시간 필드의 기간을 변경하기 위한 일부분으로 통합할 필드를 선택하십시오. 여기에서 선택하는 필드만 **지정된 필드의 사용자 정의 설정** 테이블의 작성 탭에서 사용할 수 있습니다. 포함되지 않은 필드는 노드에서 나가는 데이터에서 필터링되어 제거됩니다. 즉, **필드** 목록에 남아 있는 필드가 데이터에서 필터링되어 제거됩니다.

② 시간 구간 - 작성 옵션

작성 탭을 사용하면 시간 간격을 변경하기 위한 옵션과 해당 측정 유형에 기반하여 데이터의 필드를 통합하는 방식을 지정할 수 있습니다.

데이터를 통합할 때 기존 날짜, 시간 또는 시간소인 필드는 생성 필드로 대체되고 출력에서 삭제됩니다. 기타 필드는 사용자가 이 탭에 지정하는 옵션에 기반하여 통합됩니다.

시간 간격 계열을 작성하기 위한 간격 및 주기성을 선택하십시오.

추가 정보는 지원되는 구간의 내용을 참조하십시오.

기본 설정 다른 유형의 데이터에 적용할 기본 통합을 선택하십시오. 기본값은 측정 수준에 기반하여 적용됩니다. 예를 들어, 명목 필드는 모드를 사용하지만 연속형 필드는 합계로 통합됩니다. 세 가지 다른 측정 수준의 기본값을 설정할 수 있습니다.

- **연속형** 연속형 필드의 사용 가능한 함수로는 **Sum**, **Mean**, **Min**, **Max**, **Median**, **1st Quartile**, **3rd Quartile**이 있습니다.
- **명목** 옵션으로는 **Mode**, **Min**, **Max**가 있습니다.
- **플래그** 옵션은 **True if any true** 또는 **False if any false**입니다.

지정된 필드의 사용자 정의 설정 개별 필드의 기본 통합 설정에 대한 예외를 지정할 수 있습니다. 오른쪽의 아이콘을 사용하여 테이블에서 필드를 추가 또는 제거하거나 적절한 열의 셀을 클릭하여 해당 필드의 집계 함수를 변경하십시오. 유형이 없는 필드는 목록에서 제외되고 테이블에 추가될 수 없습니다.

새 필드 이름 확장자 노드에 의해 생성되는 모든 필드에 적용되는 **접두부** 또는 **접미부**를 지정하십시오.

(19) 재투영 노드

지리 공간적 데이터 또는 맵 데이터를 사용하는 경우, 좌표를 식별하기 위해 사용되는 가장 일반적인 두 가지 방법은 투영된 좌표 및 지리적 좌표계입니다. IBM® SPSS® Modeler에서 표현식 작성기 공간 기능, STP(Spatio-Temporal Prediction) 노드 및 맵 시각화 노드와 같은 항목은 투영된 좌표계를 사용하며, 따라서 지리적 좌표계로 기록된 가져온 데이터를 다시 투영해야 합니다. 가능하면 지리 공간적 필드(지리 공간적 측정 수준이 있는 모든 필드)가 가져올 때가 아니라 사용될 때 자동으로 재투영됩니다. 자동으로 재투영할 수 있는 필드가 없으면 재투영 노드를 사용하여 좌표계를 변경할 수 있습니다. 이 방법으로 재투영하면 잘못된 좌표계를 사용하여 오류가 발생하는 상황을 정정할 수 있습니다.

좌표계를 변경하기 위해 재투영해야 하는 상황의 예는 다음과 같습니다.

- *붙여쓰기* 지리 공간적 필드에 대한 좌표계가 다른 두 데이터 세트를 붙여쓰려고 시도하면 SPSS Modeler에서 다음과 같은 오류 메시지를 표시합니다. **Coordinate systems of *<Field1>* and *<Field2>* are not compatible. Reproject one or both fields to the same coordinate system.**

〈Field1〉 및 *〈Field2〉*는 오류를 발생시킨 지리 공간적 필드의 이름입니다.

- If/else 표현식 표현식의 두 파트 모두에 지리 공간적 필드 또는 리턴 유형이 있으나 좌표계가 다른 if/else 명령문을 포함하는 표현식을 사용하면 SPSS Modeler에서 다음과 같은 오류 메시지를 표시합니다. **The conditional expression contains incompatible coordinate systems: *〈arg1〉* and *〈arg2〉*.**

  〈arg1〉 및 *〈arg2〉*는 좌표계가 다른 지리 공간적 유형을 리턴하는 then 또는 else 인수입니다.

- 지리 공간적 필드 목록 생성 수많은 지리 공간적 필드로 구성된 목록 필드를 생성하려면 목록 표현식에 제공되는 모든 지리 공간적 필드 인수가 도일한 좌표계 내에 있어야 합니다. 그렇지 않으면 다음과 같은 오류 메시지가 표시됩니다. **Coordinate systems of 〈Field1〉 and 〈Field2〉 are not compatible. Reproject one or both fields to the same coordinate system.**

좌표계에 대한 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오.

① 재투영 노드에 대한 설정 옵션

필드

지리 필드

기본적으로 이 목록은 비어있습니다. **재투영할 필드** 목록에서 이 목록으로 지리 공간적 필드를 이동하여 해당 필드가 재투영되지 않도록 할 수 있습니다.

재투영할 필드

기본적으로 이 목록에는 이 노드에 대한 입력이 되는 모든 지리 공간적 필드가 포함됩니다. 이 목록 내의 모든 필드는 사용자가 **좌표계** 영역에서 설정한 좌표계로 재투영됩니다.

좌표계

스트림 기본값

기본 좌표계를 사용하려면 이 옵션을 선택하십시오.

지정

이 옵션을 선택하면 **변경** 단추를 사용하여 좌표계 선택 대화 상자를 표시하고 재투영에 사용할 좌표계를 선택할 수 있습니다.

좌표계에 대한 자세한 정보는 스트림의 지리 공간적 옵션 설정의 내용을 참조하십시오