

Lab4. Tableau를 이용한 Data Source 연동하기 I

1. Tableau 설치하기

- 1) Tableau를 사용하기 위해서는 먼저 PC나 노트북에 Tableau Desktop이라는 프로그램을 설치해야 한다.
- 2) <https://www.tableau.com/ko-kr> 사이트를 방문한다.
- 3) 페이지 오른쪽 상단의 [지금 체험] 버튼을 클릭한다.
- 4) 페이지 하단에서 14일 [무료 평가판 다운로드]를 클릭한다.
- 5) 학생이나 강사인 경우 <https://www.tableau.com/ko-kr/community/academic>을 방문하여 무료 버전을 이용할 수 있다.
- 6) [Tableau Desktop: 14일 무료 평가판 시작] 페이지에서 각각의 값을 입력한 후, [무료 평가판 다운로드] 버튼을 클릭한다.
- 7) 다운로드 받은 TableauDesktop-64bit-2023-1-2.exe를 더블클릭하여 실행한다.
- 8) 설치가 끝나면 자동으로 Tableau 프로그램이 시작되고, [Tableau 등록]이라는 팝업창이 나타난다.
- 9) 필요한 정보를 입력한 후, [지금 평가판 시작] 버튼을 클릭한다.

2. MySQL Driver 설치하기

- 1) 설치 및 등록 작업을 마치면 첫 화면으로 [연결]이 나온다.
- 2) 화면의 좌측 메뉴 중 [서버에 연결] > [자세히...]에서 MySQL을 검색 혹은 직접 찾아서 연결할 수 있고, 또는 MySQL은 자주 사용하기 때문에 바로 [서버에 연결] > [MySQL]로 연결해도 된다.
- 3) 첫번째 방법
 - [MySQL] "이 커넥터를 사용하려면 먼저 드라이버를 다운로드하여 설치해야 합니다." 팝업창에서 [드라이버 다운로드 ▶]를 클릭한다.
 - [드라이버 다운로드] 페이지에서 먼저 로그인을 수행하고, 로그인 후 다음과 같이 입력한다.
 - [데이터 원본] : MySQL
 - [운영 체제] : Windows or Mac
 - [비트 버전] : 64-bit
 - [MySQL] 섹션에서, [Tableau Desktop: 2019.3 - 2023.1.2]의 순서대로 진행한다.
- 4) 두번째 방법
 - * 첫번째 방법은 결국 MySQL 홈페이지에서 MySQL 드라이버를 다운로드하기 때문에 권장하지 않는다.
 - <https://dev.mysql.com/downloads/> 페이지로 이동한다.
 - 목록에서 [Connector/ODBC]를 클릭한다.
 - [MySQL Community Downloads Connector/ODBC] 페이지에서
 - [Select Operating System:] : Microsoft Windows or macOS
 - [Select OS Version:] : Windows (x86, 64-bit) or macOS 13 (x86, 64-bit) or macOS 13 (ARM, 64-bit)
 - Windows의 경우 MSI Installer의 [Download] 버튼을 클릭한다.
 - macOS의 경우 DMG Archive or Compressed TAR Archive [Download]를 클릭하여 다운로드한다.
 - [Login Now or Sign Up for a free account] 페이지에서, [No. thanks, just start my download.] 링크를 클릭한다.
 - 다운로드 받은 파일을 실행한다. (Windows의 경우 mysql-connector-odbc-8.0.33-winx64.msi)

3. MySQL Data Source 준비하기

- 1) Github의 4주차 실습예제 중 [Lab4. Running Spark Script on AWS Glue Studio.pdf]의 "5. AWS Glue Studio에서 spark-rds-demo3.py 실행하기"를 실행한다.
- 2) 권한문제 때문에 [Run]에 오류가 발생하면 바로 위의 "2. Parameter Store에 새로운 패스워드 설정하기"와 "3. AWS Glue Studio에서 spark-rds-demo1.py 실행하기"를 참조한다.
- 3) [Run]이 성공적으로 마치면, Amazon RDS의 {계정}-db의 newyork_taxi 데이터베이스에 2개의 테이블 즉, taxi_zone_lookup과 taxi_output 테이블이 생성된다.
- 4) Database 연결 프로그램(예: HeidiSQL or MySQL Workbench)를 이용하여 해당 데이터베이스의 taxi_output 테이블을 오픈한다.
- 5) 해당 테이블의 컬럼 중 do_location_id의 [데이터 유형]을 TEXT에서 INT로 변경한 후, 저장한다.

4. MySQL Data Source 연동하기

- 1) MySQL 드라이버 설치 후, 다시 [연결] > [서버에 연결] > [MySQL]을 선택한다.
- 2) [MySQL] 팝업창에서 다음과 같이 Amazon RDS와 연결한다.
 - [일반]
 - [서버] : Amazon RDS {계정}-db의 엔드포인트, ex)henry-db.cx1hah81fdfdf.ap-northeast-2.rds.amazonaws.com
 - [포트] : 3306
 - [데이터베이스] : newyork_taxi
 - [사용자 이름] : admin
 - [비밀 번호] : datalakemysql
 - [로그인] 버튼 클릭
- 3) 로그인 성공하면, 화면 좌측에 taxi_zone_lookup 테이블이 확인된다.
- 4) 혹은 화면 오른쪽에 테이블이 보이지 않으면 화면 왼쪽의 [테이블] > taxi_zone_lookup을 Drag & Drop 하면 된다.
- 5) 테이블 로딩 후, [지금 업데이트] 버튼을 클릭한다.
- 6) 데이터가 로딩 후, 화면 좌측의 테이블 중, taxi_output 테이블을 Drag & Drop 한다.
- 7) 두 개의 테이블이 서로 연결되었지만, 조인 조건을 넣지 않았기 때문에 주황색 느낌표가 있는 라인이다.
- 8) 주황색 라인에 마우스를 올리면, "카디널리티: 다대다(기본값), 이 관계에는 일치하는 필드가 없습니다. 관계를 편집하여 일치하는 필드를 선택하십시오." 메시지를 확인할 수 있다.
- 9) 화면의 아래에 "관계를 생성하는 것과 조인하는 것의 차이점은 무엇입니까?"의 섹션에서, taxi_zone_lookup 테이블에서는 #Location ID를 선택하고, taxi_output 테이블에서는 Do Location Id를 선택한다.
- 10) 그리고 [지금 업데이트]를 클릭한다.
- 11) 데이터를 확인한다.

5. sales_data_sample.csv 파일 연동하기

- 1) Tableau Desktop을 실행한다.
- 2) [연결] 페이지에서 [파일에 연결] > [자세히...]를 클릭한다.
- 3) [열기] 팝업창에서 Lab2에서 다운로드했던 sales_data_sample.csv 파일을 선택한다.

6. sales_data_sample.csv 파일로 시각화하기

- 1) sales_data_sample.csv이 로딩됐으면 화면 아래 [데이터 원본] 우측의 [시트1]을 클릭한다.
- 2) 화면 좌측의 측정값 중에 #Sales를 Drag하여 시트1의 중앙에 Drop한다.

79 3)Sales의 값인 10,032,629가 표시된다.
80 4)화면 중간에 있는 [시트 1]을 더블클릭하여 [제목 편집]창이 나오게 해서 Total of Sales라고 입력 후 [확인] 버튼을 클릭한다.
81 5)Sheet 하나에 한 값을 표시할 수 있다.
82
83 6)화면 하단에서 [시트 1] 오른쪽에 있는 [새 대시보드 1]을 클릭한다.
84 7)새로운 대시보드가 생성되고 화면 왼쪽에는 방금 작업한 [시트] > [시트 1]이 보인다.
85 8)시트1을 Drag해서 화면 중간에 있는 [여기에 시트 놓기]에 Drop한다.
86 9)즉, 대시보드는 미리 시트를 작업해서 Drag & Drop하면 된다.
87 10)화면 하단의 [대시 보드 1]을 더블클릭하여 이름을 "Sales Dashboard"로 변경한다.
88
89 11)화면 하단에서 [새 워크시트]를 클릭하여 [시트 2]를 생성한다.
90 12)화면 좌측의 측정값 중에 Orderdate가 있는데, 마우스를 올려서 아래 화살표를 클릭하면 현재 텍스트형 값인데, [데이터 유형 변경] > [날짜]로 변경한다.
91 13)방금 유형 변경한 Orderdate를 Drag하여 [시트 2]의 [열]에 Drop한다.
92 14)처음에는 년(Orderdate)가 표시되고 라벨 앞에 있는 +를 클릭하면 년에서 분기로, 분기에서 월, 월에서 일로 즉 작은 단위로 계속 라벨과 각 단위의 값들이 시트에 표시된다.
93 15)표시하고 싶지 않으면 해당 라벨을 Drag하여 화면 중앙으로 내리면 즉, [열]에서 제거하면 된다.
94 16)[열]이 아닌 [행]으로도 볼 수 있는데, 현재 [열]에 있는 년(Orderdate)를 Drag하여 [행]으로 Drop하면 된다.
95 17)역시 동일하게 라벨 앞의 +를 클릭하면 더 작은 단위로 시트에 표시된다.
96 18)년, 분기, 월까지 라벨로 표시하고, 분기를 제거하면 각 년의 각 월에 대한 값이 시트에 표시될 것이다.
97 19)현재 시트에는 년도와 월만 표시되고 값은 Abc라고 표시되어 있다.
98 20)이 Abc에 값을 넣으려면 화면 좌측의 측정값중에서 특정한 측정값을 Drag & Drop하면 되는데, 지금은 좌측 측정값 중에서 Quantityordered를 Drag & Drop하기로 한다.
99 21)이렇게 하면 현재 시트에는 주문년도와 주문월, 그리고 합계가 표시될 것이다.
100 22)시트 2 좌측에 보면 [마크] 섹션이 있다.
101 23)합계(Quantityordered)를 Drag하여 [마크] > [색상]으로 Drop한다.
102 24)값에 대해 색상으로 표시되는데, 기본적으로 글자색이 아닌 배경색이다보니 값들이 보이지 않게 된다.
103 25)이럴 때는 [마크] > [레이블]을 클릭해본다.
104 26)조그만 창이 팝업되는데, 여기서 [마크 레이블 표시]를 체크한다.
105 27)이 팝업창에서 폰트나 글자스타일을 지정할 수 있다.
106 28)화면 오른쪽 상단에 [표현 방식]이 있다.
107 29)[표현 방식]을 클릭해보면 다양한 차트가 있는데, Tableau에서 해당 값을 보고 권장차트가 자동 선택되어 있다.
108 30)현재의 권장 차트인 차트를 마우스로 선택해보자.
109 31)년도별로 구분된 라인차트를 확인할 수 있다.
110 32)다른 차트로 선택해서 확인해본다.
111 33)차트 중에 [누적 막대]를 선택한다.
112 34)각 막대에 마우스를 올려놓으면 각 값들을 확인할 수 있는데, [마크] > [레이블]을 클릭하여 팝업창을 띄운 뒤, [마크 레이블 표시]를 체크하면 해당 막대에 값들이 표시된다.
113 35)[누적 막대] 차트가 있는 화면 하단의 [시트 2]의 시트 이름을 더블클릭하여 QuantityOrdered라고 변경한다.
114 36)그리고 화면 하단의 Sales Dashboard로 이동하여, 방금 이름 변경한 QuantityOrdered를 화면 왼쪽 [시트]에서 찾아서 화면 중앙으로 Drag & Drop한다.
115 37)이미 있는 Total of Sales와 지금 추가한 [누적 막대]를 마우스로 적절히 이동하여 크기와 위치를 배치한다.
116
117 38)화면 하단에서 [새 워크시트]를 클릭하여 추가한다.
118 39)[시트 3]에는 좌측의 측정값 중 Country를 Drag & Drop해 본다.
119 40)그러면 지구 각 나라의 맵이 그려진다.
120 41)현재 [열]과 [행]은 각각 경도(생성됨)과 위도(생성됨)의 값으로 표시된 것이다.
121 42)지도에 Sales의 값을 표시하기 위해 좌측의 측정값 중 Sales를 Drag & Drop 한다.
122 43)그러면 Sales의 값을 기준으로 각 나라별로 원모양의 크기가 구별되는 것을 확인할 수 있다.
123 44)현재의 권장인 아닌 다른 지도 차트를 클릭해 보면 Sales 값 기준으로 각 나라의 색상이 다른 것을 확인할 수 있다.
124 45)역시 [마크] > [레이블]을 클릭하여 팝업창을 띄운 뒤, [마크 레이블 표시]를 체크하면 해당 나라에 값들이 표시된다.
125 46)화면 하단에서 [시트 3]을 더블클릭하여 Sales by Country로 변경한다.
126 47)화면 하단에서 Sales Dashboard를 선택한 후, 지금 생성한 Sales by Country를 Drag & Drop하여 적절한 크기와 위치에 배치한다.
127 48)지금까지 작업한 Dashboard를 저장하기 위해 저장(Ctrl + S)을 클릭한다.
128 49)[다른 이름으로 저장] 창이 나타나고, twb 포맷으로 저장할 수 있다.
129 50)상단의 메뉴바에 보면 [저장], [프레젠테이션 모드(F7)], [다른 사람과 통합 문서 공유]을 사용할 수 있다.
130