

1 Lab7. Superset에서 Data Source 연동하기

1. 데이터 준비하기

1) Database Connect Program(HeidiSQL or MySQL Workbench)로 Amazon RDS에 연결하여 확인

- RDS Database : {계정}-db
- 호스트명/IP : RDS Endpoint(ex:henry-db.cxlhah81ocl3.ap-northeast-2.rds.amazonaws.com)
- Port : 3306
- User : admin
- Password : datalakemysql
- Schema : newyork_taxi
- Table : taxi_output, taxi_zone_lookup

2) 새로운 Schema 생성

- Query Editor에 다음과 같이 입력하여 실행한다.
CREATE DATABASE `Northwind` /*!40100 COLLATE 'utf8mb4_general_ci' */;

3) CSV 파일로 Schema Table Import하기

- MySQL Workbench에서
- Northwind > Tables에서 마우스 right-click > [Table Data Import Wizard] 선택
- [Table Data Import] 창에서
 - [File Path] : [Browse...]를 클릭하여 ales_data_sample.csv 파일 선택
 - [Next] 버튼 클릭
- [Select Destination] > [Create new table: Northwind.sales_data_sample]
 - [Next] 버튼 클릭
- [Configure Import Settings] > [Encoding] : utf-8
 - [Next] 버튼 클릭
- [Import Data] > [Next] 버튼 클릭
- Data Import가 모두 끝나면 [Next] 버튼이 활성화된다. ※대략 10분 정도 소요됨.
- [Next] 버튼 클릭
- [Finish] 버튼 클릭

4) Table Data 확인하기

- 화면 좌측의 [SCHEMAS] 프레임에서 마우스 right-click > [Refresh All]
- [Northwind] > [Tables] > [sales_data_sample] > 마우스 right-click > [Select Rows - Limit 1000] 클릭
- 결과 확인

2. Superset에서 Data 연동하기

1) 화면 우상단의 [Settings] > [Database Connections]

2) 화면 우상단의 [+DATABASE] 버튼 클릭

3) [Connect a database] 창에서

- [Select a database to connect] : [MySQL] 클릭
- [Enter the required MySQL credentials]
 - [HOST] : Amazon RDS Endpoint(ex:henry-db.cxlhah81ocl3.ap-northeast-2.rds.amazonaws.com)
 - [PORT] : 3306
 - [DATABASE NAME] : Northwind
 - [USERNAME] : admin
 - [PASSWORD] : datalakemysql
 - [DISPLAY NAME] : RDS-Northwind
 - [CONNECT] 버튼 클릭
- [Database connected] > [FINISH] 버튼 클릭

4) [Database] 페이지에서 [RDS-Northwind] 확인

5) 페이지 상단의 메뉴 중 [SQL] > [SQL Lab] > 선택

6) [SQL Lab] 페이지에서

- [DATABASE] : (mysql) RDS-Northwind 선택
- [SCHEMA] : Northwind 선택
- [SEE TABLE SCHEMA] : sales_data_sample
- 이렇게 하면 페이지 중앙하단의 Datagrid에서 결과를 확인할 수 있다.
- 페이지 좌측 하단에는 해당 테이블의 각각의 컬럼의 이름과 데이터 타입을 보여준다.
- Query 입력 창에 다음과 같은 질의를 입력하고, [RUN SELECTION]을 클릭하여 결과를 확인한다.
SELECT PRODUCTLINE, sum(SALES) as sum_of_sales
FROM sales_data_sample
GROUP BY PRODUCTLINE

- 이 조회 결과를 가지고 차트를 만들려면 [CREATE CHART] 버튼을 클릭하면 브라우저의 새 탭이 열리면서 차트가 만들어진다.

7) 차트 페이지에서

- 페이지 좌측의 [Chart Source]의 "SELECT PRODUCTLINE, ..." 바로 오른쪽의 세로 점 3개 버튼 클릭 > [Query preview] 클릭
- [Query preview] 창이 열리며 방금 입력했던 Query 확인가능
- 이 창에서 바로 [SAVE AS DATASET] 버튼을 클릭해서 DataSet을 생성할 수 있다.
- [SAVE AS DATASET] 버튼을 클릭
- [Save or Overwrite Dataset] 창에서
 - [Save as new] : sum_by_productline
 - [SAVE] 버튼 클릭
- 페이지 상단의 [Datasets] 메뉴 클릭
- [Datasets] 목록에서 방금 생성한 [sum_by_productline] 확인할 수 있다.