

1 Lab2. Data Source 연동하기

1. Kaggle에서 Sample Sales Data Set 다운로드하기

- 1)Googling으로 "Kaggle sample sales data"를 검색한다.
- 2)검색결과에서 "Sample Sales Data" 링크를 클릭한다.
- 3)<https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>
- 4)페이지 오른쪽 상단의 [Download (79 KB)] 버튼 클릭
- 5)Download할 때 로그인 필요
- 6)다운로드 받은 archive.zip파일의 압축을 푼다. > sales_data_sample.csv

2. Kaggle CSV 데이터로 새 Datasets 생성하기

- 1)[Amazon S3] 페이지로 이동한다.
- 2)전에 이미 생성했던 {계정}-datalake-bucket을 클릭하여 해당 bucket으로 이동한다.
- 3)페이지에서 [폴더 만들기] 버튼 클릭
- 4)[폴더 만들기] 페이지에서 다음과 같이 설정한다.
 - [폴더 이름] : Sales
 - [폴더 만들기] 버튼 클릭
- 5)방금 생성한 "Sales" 폴더 링크 클릭
- 6)위 1번에서 다운로드 받고 압축을 풀었던 sales_data_sample.csv 파일을 업로드한다.

7)업로드 후 업로드한 파일인 sales_data_sample.csv 파일을 클릭한다.

8)해당 버킷의 업로드한 해당 파일 페이지로 이동한다.

9)다시 QuickSight 페이지로 이동하여 좌측 메뉴 중 [Datasets]를 클릭한다.

10)페이지 오른쪽 상단의 [New dataset] 버튼을 클릭한다.

11)[Create a Dataset] 페이지의 여러 목록에서 [S3] 버튼을 클릭한다.

12)[New S3 data source] 팝업창이 나타난다.

13)"manifest file" 링크를 클릭한다.

-<https://docs.aws.amazon.com/quicksight/latest/user/supported-manifest-file-format.html>

-해당 페이지의 "Manifest file format for Amazon QuickSight" 내용을 참조한다.

-VScode를 오픈하여 "Manifest file format for Amazon QuickSight"의 샘플 코드를 복사하여 붙여넣는다.

-다음과 같이 편집한다.

-아래 코드 중 URIs 항목의 값은 S3 버킷에 업로드한 csv 파일의 [S3 URI] 값을 복사해서 넣는다.

```
{
  "fileLocations": [
    {
      "URIs": [
        "s3://henry-datalake-bucket/Sales/sales_data_sample.csv"
      ]
    },
  ],
  "globalUploadSettings": {
    "format": "CSV",
    "delimiter": ",",
    "textqualifier": "\"",
    "containsHeader": "true"
  }
}
```

-파일을 sales_data_sample.json으로 저장한다.

14)다시 [New S3 data source] 팝업창으로 돌아와서,

-[Data source name] : sales_data_sample

-[Upload a manifest file] : "Upload" 선택 후 바로 위에서 생성한 파일인 sales_data_sample.json을 선택한다.

-[Connect] 버튼 클릭

15)[Finish dataset creation] 팝업창에서,

-[Edit/Preview data] 버튼을 클릭하여 미리 sales_data_sample 데이터를 확인할 수 있다.

-데이터 확인 및 미리 보기를 마친 후, 페이지 오른쪽 상단의 3개의 버튼 중 [PUBLISH & VISUALIZE] 버튼을 클릭한다.

16)[New sheet] 팝업창이 나타난다.

-[Interactive sheet]이 선택되어 있고, [Layout] : Titled, [Optimize for viewing on] : 1600px 그대로 수정하기 않고,

-[CREATE] 버튼 클릭

17)페이지에 방금 연동한 sales_data_sample을 확인할 수 있다.

18)페이지 좌측 상단의 [QuickSight] 옆에 방금 생성한 Dataset인 [sales_data_sample analysys]의 왼쪽의 별을 클릭하여 [Favorites]에 등록한다.

19)[QuickSight]를 클릭하면 [QuickSight]의 첫 페이지로 이동하게 된다.

20)여기서 좌측 메뉴의 [Favorites]페이지를 클릭하면 방금 [Favorites]에 추가한 [sales_data_sample analysis]를 확인할 수 있다.

3. RDS 데이터로 새 Datasets 생성하기

※ Refer to 4주차 실습 파일 [Lab3. Using Amazon RDS.pdf]

1)[QuickSight]의 첫 페이지에서 좌측 메뉴 중 [Datasets]를 클릭한다.

2)페이지 우측 상단의 [New dataset] 버튼을 클릭하여 새로운 Dataset을 생성한다.

3)[Create a Dataset] 페이지의 목록 중에서, [RDS]를 선택한다.

4)[New RDS data source] 팝업창에서

-[Instance ID] : 목록에서 {계정}-db를 선택한다.

-[Connection type] : Public network

-[Database name] : newyork_taxi

-[Data source name] : 동일하게 newyork_taxi

85 -[Username] : admin
86 -[Password] : datalakemysql
87 -[Validate connection] 버튼을 클릭하여 연결테스트를 수행한다.
88 -[Create data source] 버튼 클릭한다.
89
90 5)[Choose your table] 팝업 창에서,
91 -[Tables: contain the data you can visualize.] 목록에서 taxi_zone_lookup을 선택한다.
92 -[Select] 버튼을 클릭한다.
93
94 6)[Finish dataset creation] 창에서,
95 -[Import to SPICE for quicker analytics]가 선택되어 있음을 확인하고 마지막으로 [Visualize] 버튼을 클릭한다.
96
97 7)[New sheet] 팝업창에서,
98 -[Interactive sheet] 선택
99 -[Layout] : Titled
100 -[Optimize for viewing on] : 1600px
101 -[CREATE] 버튼을 클릭한다.
102
103 8)페이지에 방금 연동한 sales_data_sample을 확인할 수 있다.
104 9)페이지 좌측 상단의 [QuickSight] 옆에 방금 생성한 Dataset인 [taxi_zone_lookup analysys]의 왼쪽의 별을 클릭하여 [Favorites]에 등록한다.
105 10)[QuickSight]를 클릭하면 [QuickSight]의 첫 페이지로 이동하게 된다.
106 11)여기서 좌측 메뉴의 [Favorites]페이지를 클릭하면 방금 [Favorites]에 추가한 [taxi_zone_lookup analysys]를 확인할 수 있다.
107