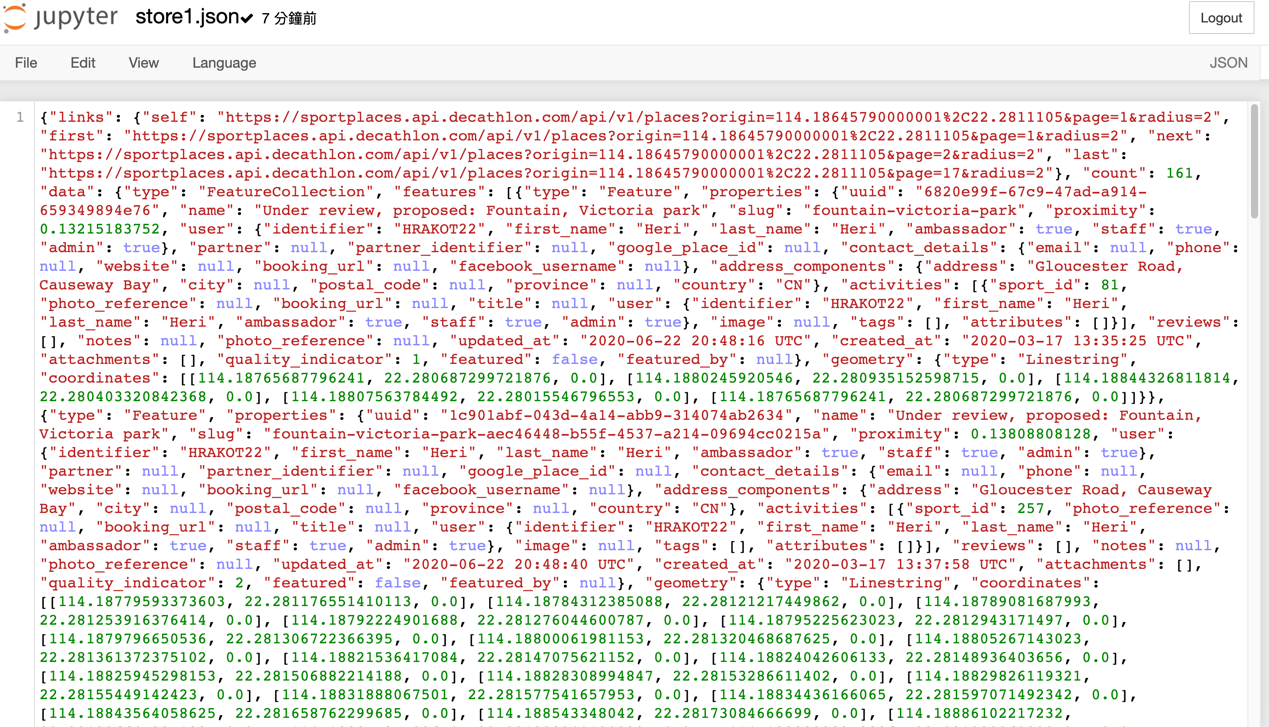
Dacathlon Data Engineer Case Study

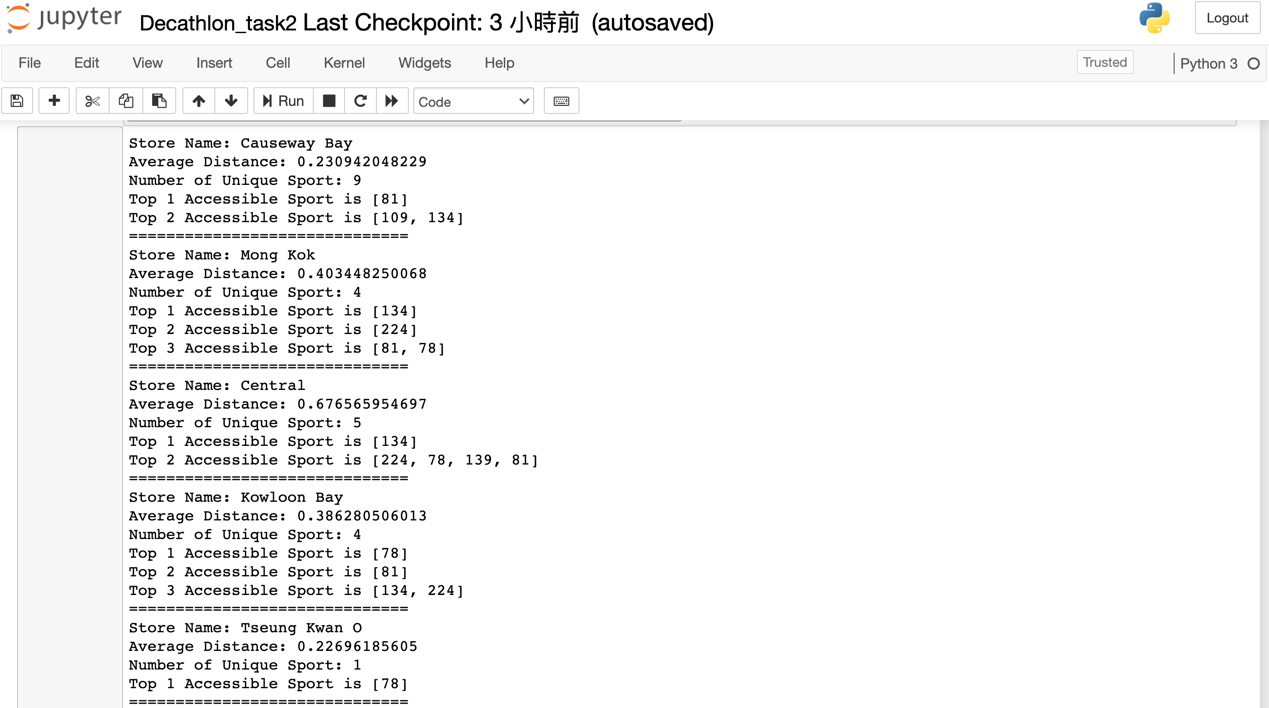
Task 1: Data extraction – Python

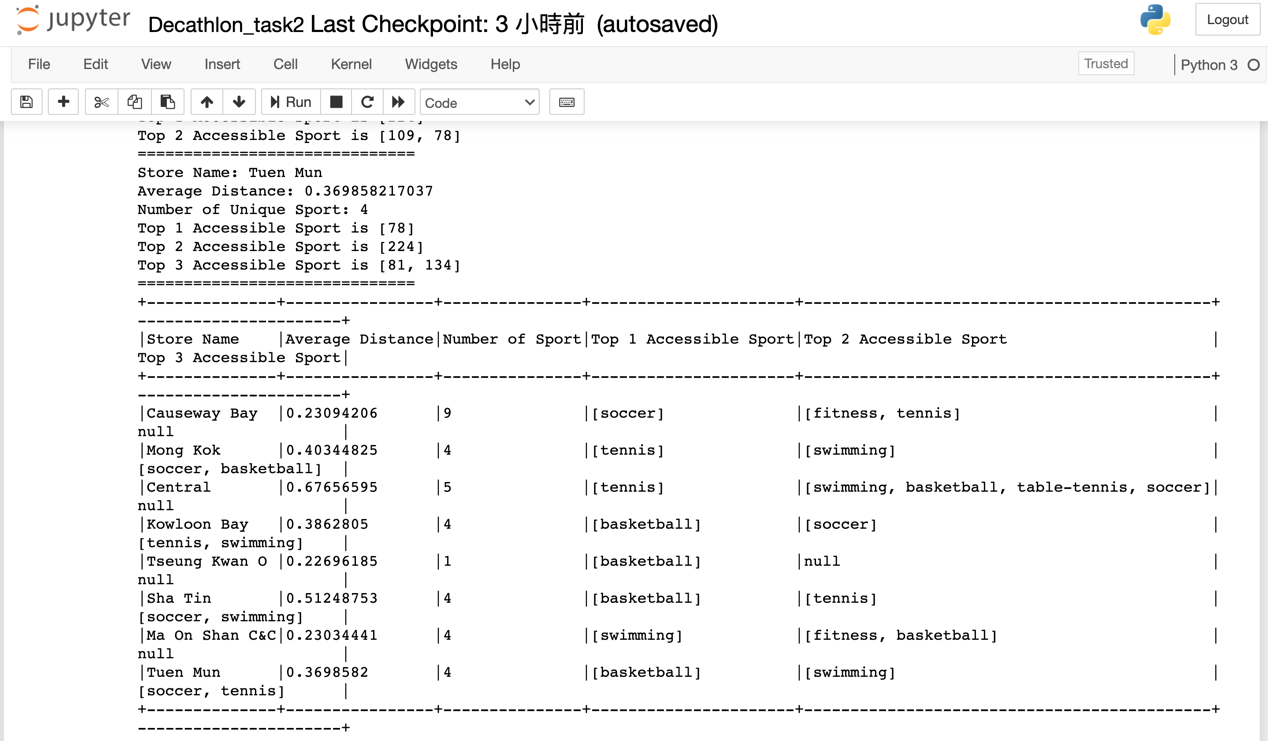
Example: store1 request output of sport places



Task 2: Data processing: A) Pyspark pipeline – PySpark

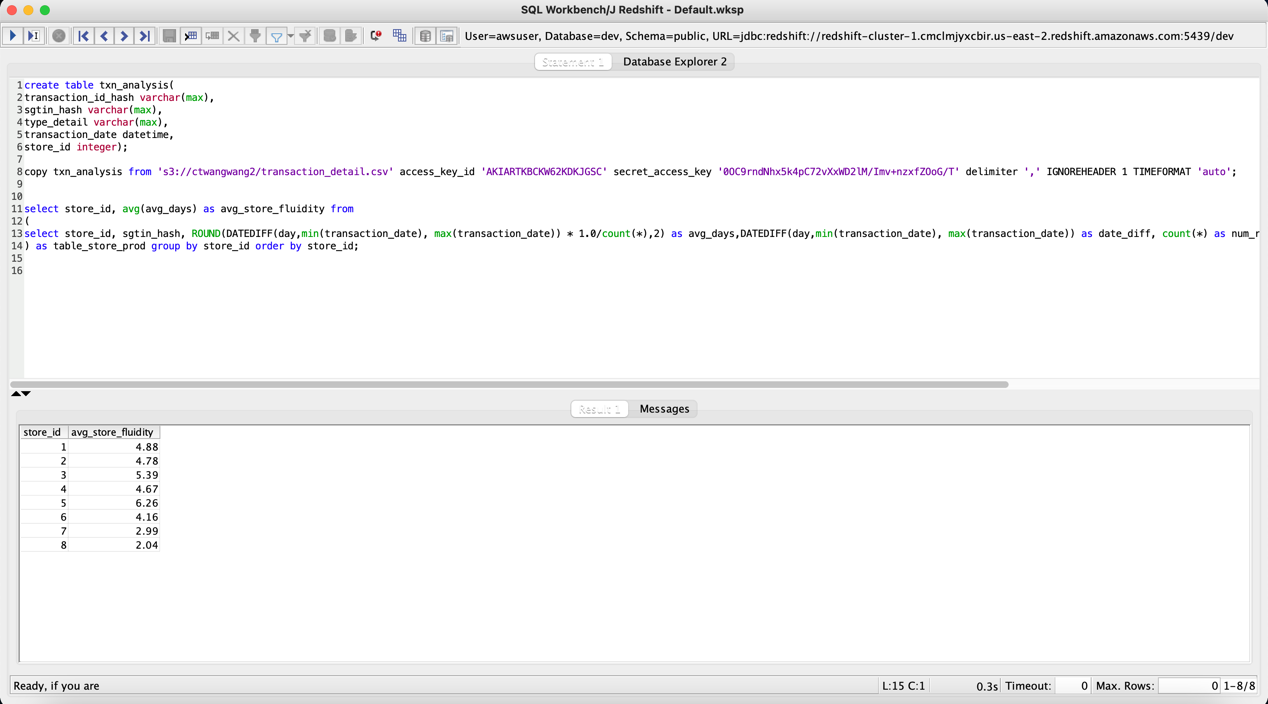
The final result output of Pyspark DataFrame





Task 2: Data processing: B) Redshift pipeline – SQL

The result of calculate the average days between sale and return of all items per store



3 [Bonus] Data Exposure – Python

By the UI of Streamlit, you can request the data based on the Table Info

