**PROJECT BRIEF –**

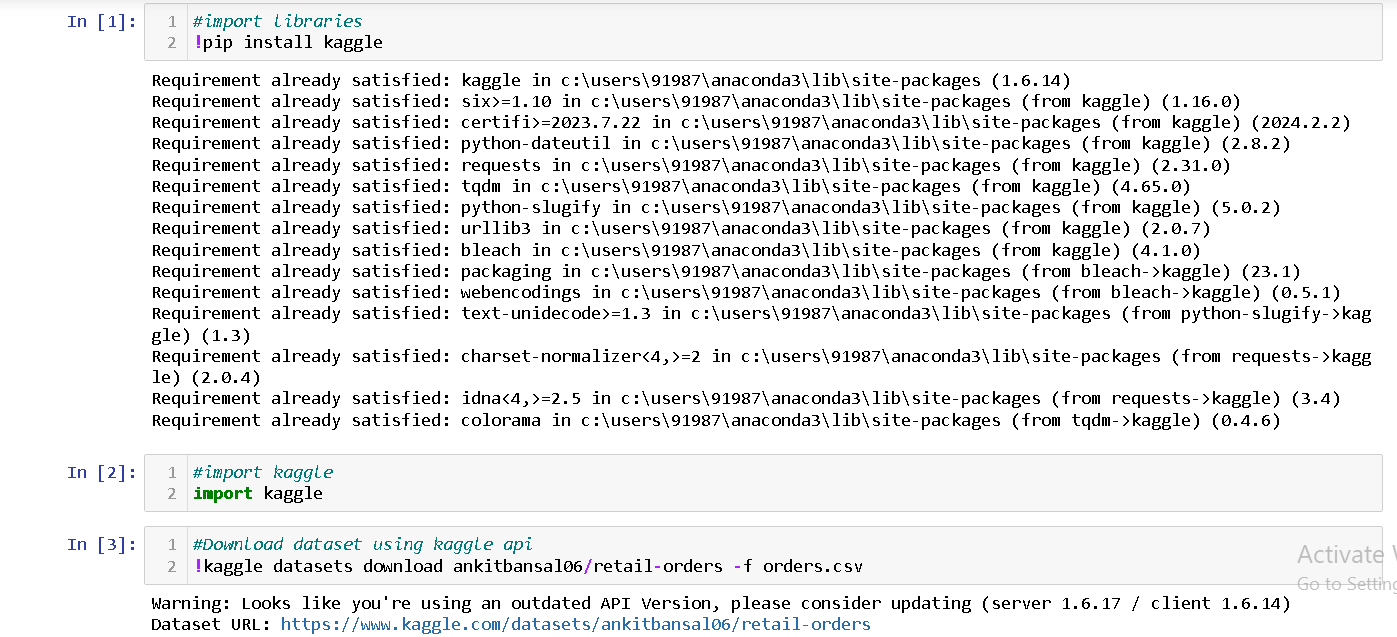
XYZ stores sales data need to be imported from kaggle api to pandas framework and do some data cleaning then load the data to MySQL and perform below data analysis.

Requirements

1. Find top 10 highest revenue generating products
2. Find top 5 highest selling products in each region
3. Find sales growth comparison MoM for 2022 and 2023 eg : jan 2022 vs jan 2023
4. For each category which month has the highest sales
5. Which sub category have the highest growth by profit in 2023 compare to 2022

**Solution Approach:-**

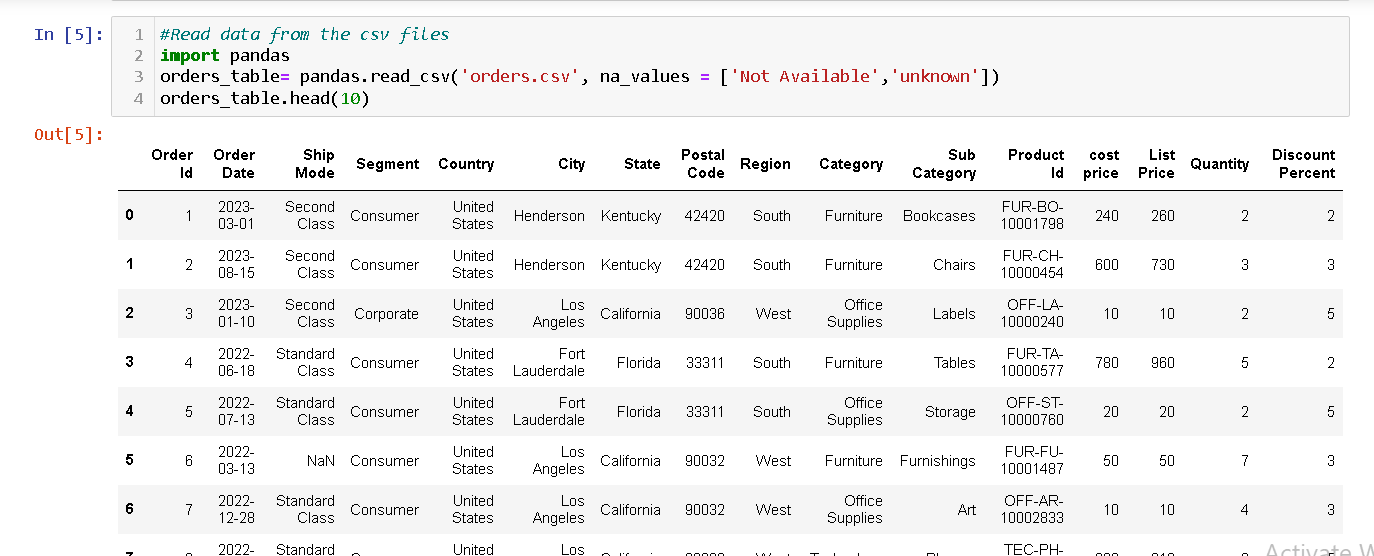
1. Generated new token and saved in the user file in C drive under .kaggle folder
2. Workings in Jupyter notebook where we use python libraries to fetch the data from kaggle.
   1. Install kaggle
   2. import kaggle
   3. Download dataset using kaggle api

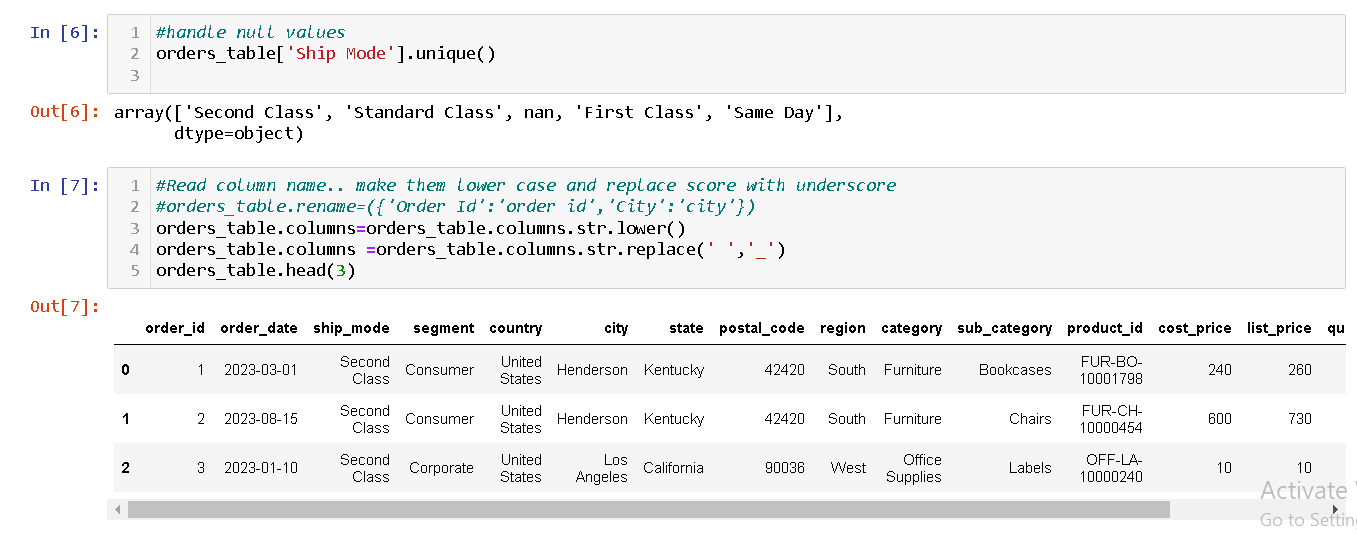


* 1. Extract file from zip file

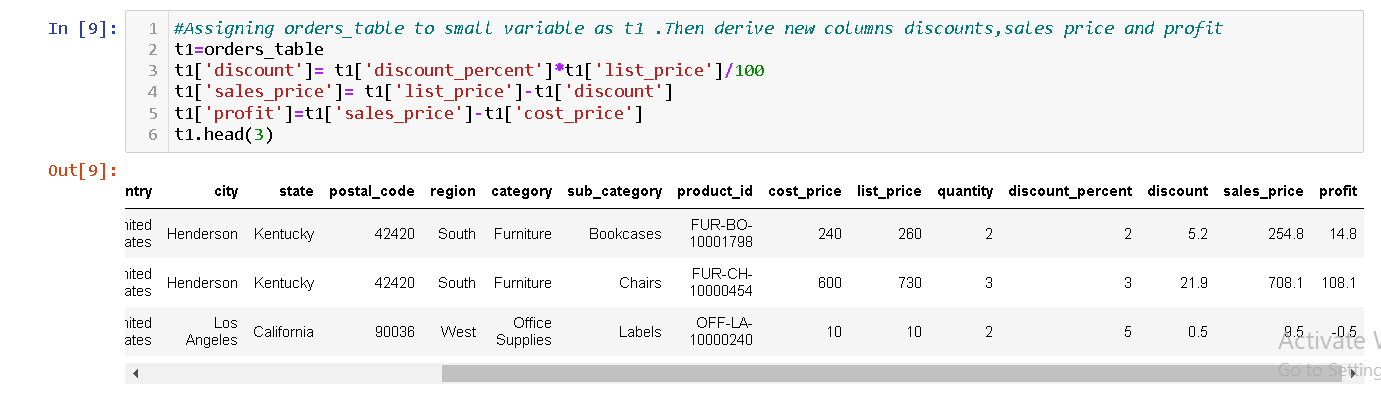


* 1. Import pandas data frame Read columns name..
  2. Replace values with null and make them lower case and replace space with underscore

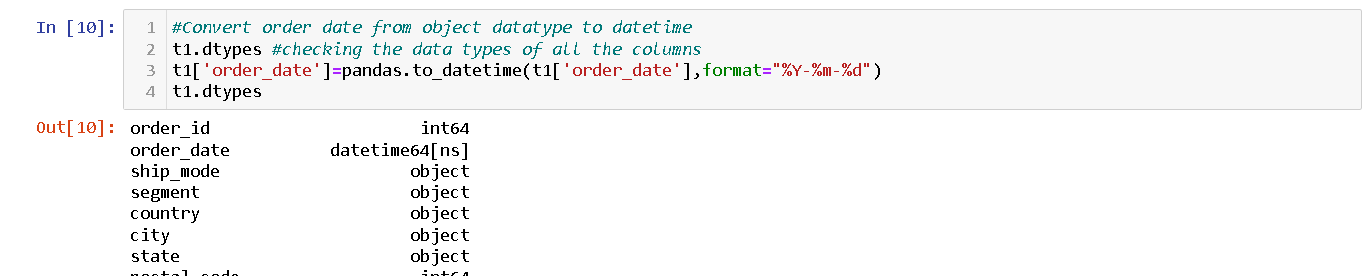




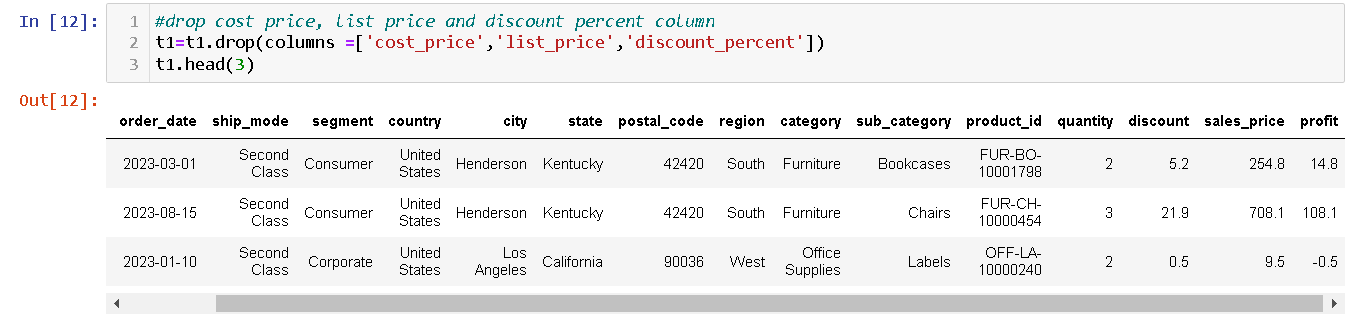
* 1. Derive new columns discount, sale price and profit



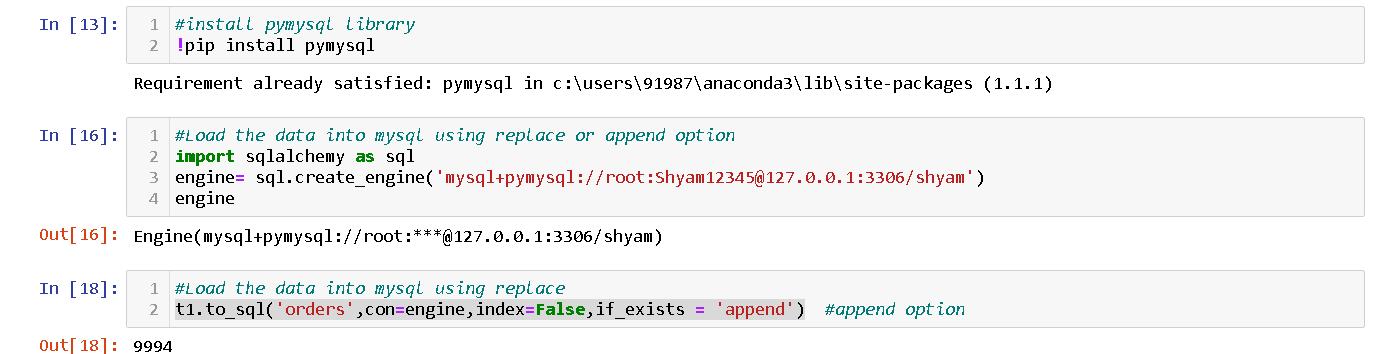
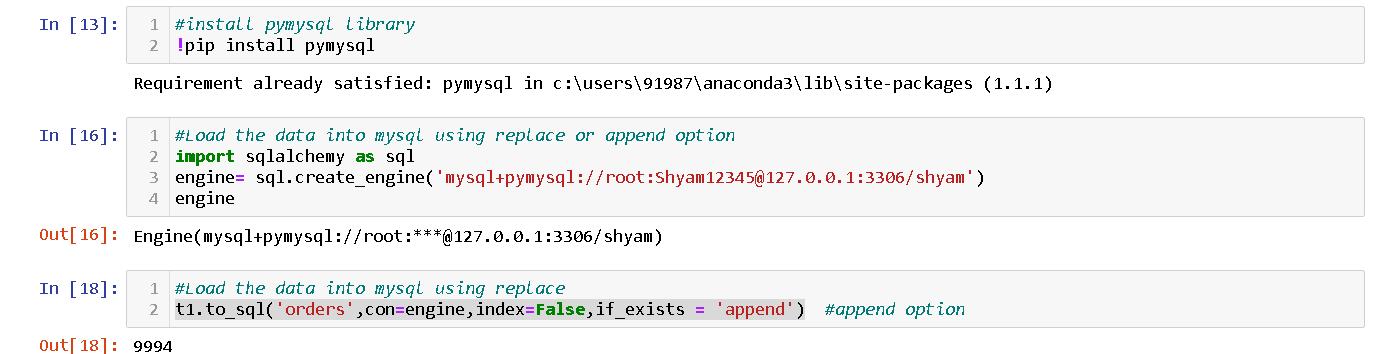
* 1. Convert order datae from object data type to datetime



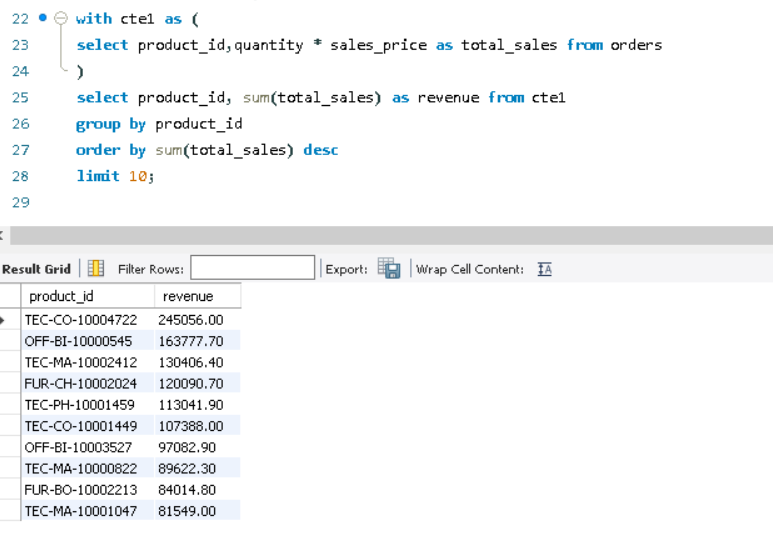
* 1. Drop cost price list price and discount percent columns

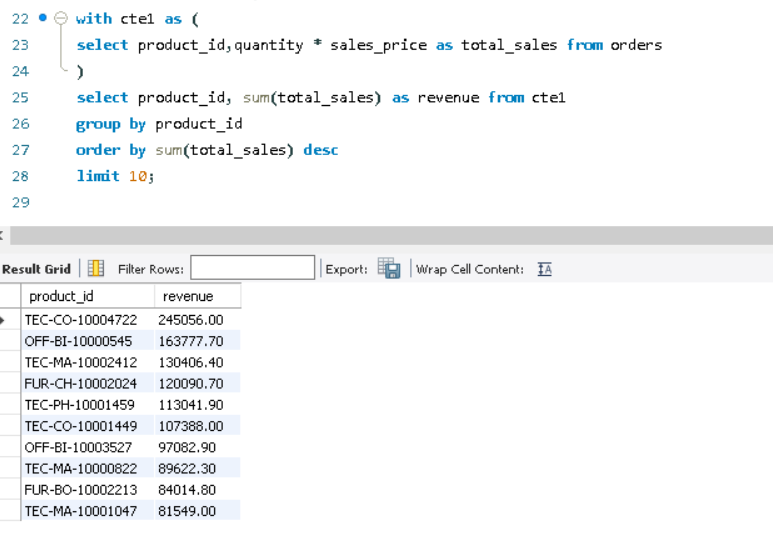


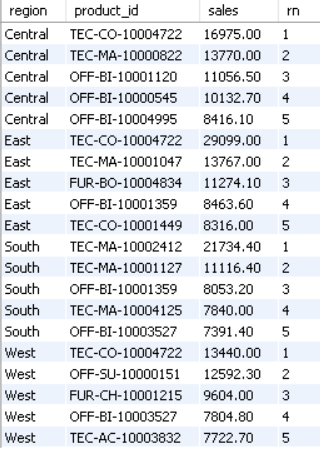
* 1. Load the data into sql server using append option with **memory optimization**

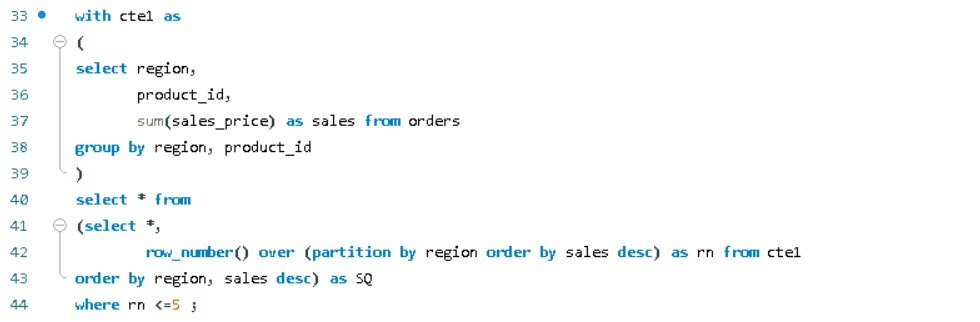


**Note** : We can also use replace option instead of append but by default it will use max character size for each object type taking unnecessary memory space. That’s why we took the approach to create a table with same name and columns with memory limit saving the memory space and then did to\_sql with append option to load the data to the db.

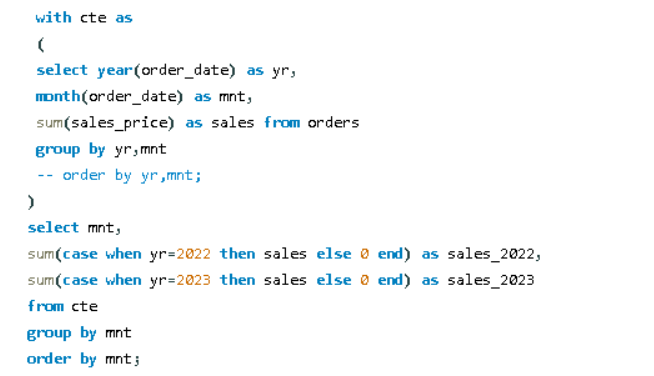
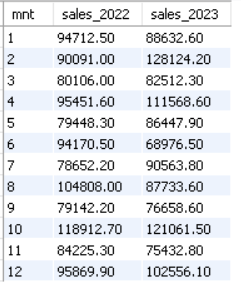
1. **MySQL**
   1. Find top 10 highest revenue generating products



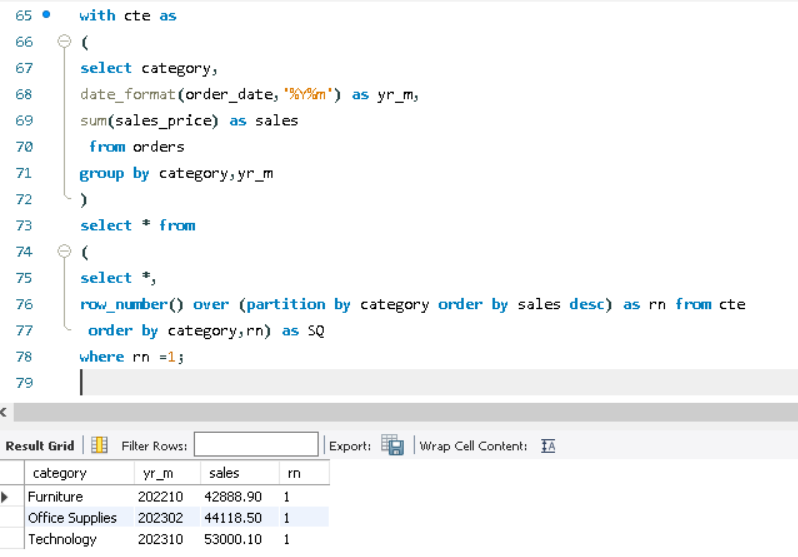
* 1. Find top 5 products with highest selling price in each region



* 1. Find sales growth comparison MoM for 2022 and 2023 eg : jan 2022 vs jan 2023



* 1. For each category which month has the highest sales

****

* 1. Which sub category have the highest growth by profit in 2023 compare to 2022

