

Project Development

Sprint 1

Team ID : PNT2022TMID35387

- Import library and load dataset in Python

The screenshot shows a Jupyter Notebook interface with the following components:

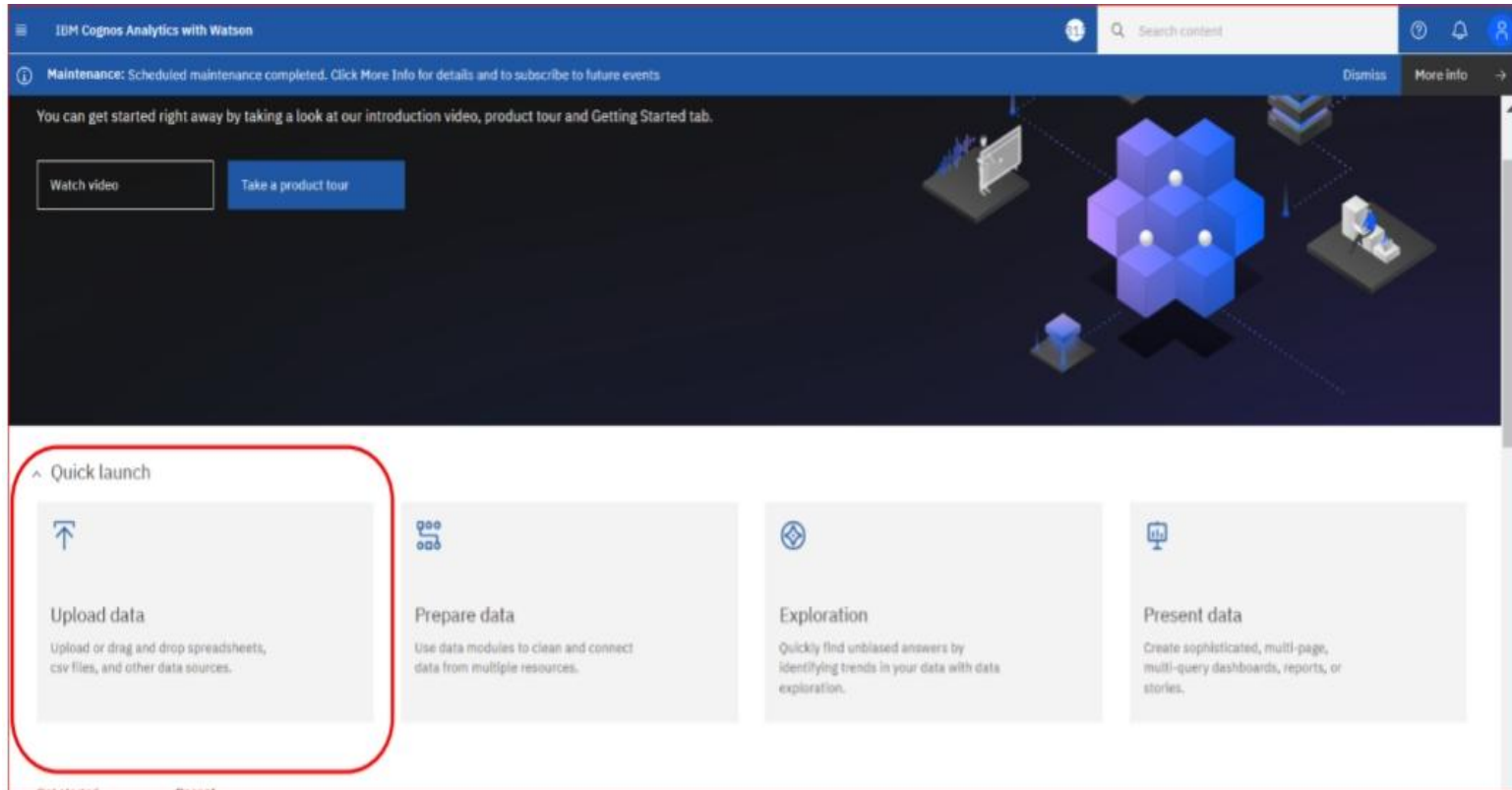
- Code Cell [2]:** Imports necessary libraries: `import numpy as np`, `import pandas as pd`, `import matplotlib.pyplot as plt`, `import seaborn as sns`, and `import sklearn as sk`.
- Code Cell [3]:** Loads the dataset: `data = pd.read_csv("Global_Superstore2.csv", encoding = 'unicode_escape', engine = 'python')` followed by `data`.
- Data Preview:** A table showing the first four rows of the dataset.

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Product ID	Category	Sub-Category	Product Name	Sales
0	32298	CA-2012-124891	31-07-2012	31-07-2012	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wir...	2309.650
1	26341	IN-2013-77878	05-02-2013	07-02-2013	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	New South Wales	FUR-CH-10003950	Furniture	Chairs	Novimex Executive Leather Armchair, Black	3709.395
2	25330	IN-2013-71249	17-10-2013	18-10-2013	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	TEC-PH-10004664	Technology	Phones	Nokia Smart Phone, with Caller ID	5175.171
3	13524	ES-2013-	28-01-	30-01-	First	KM-	Katherine	Home	Berlin	Berlin	TEC-PH-	Technology	Phones	Motorola Smart	2892.510

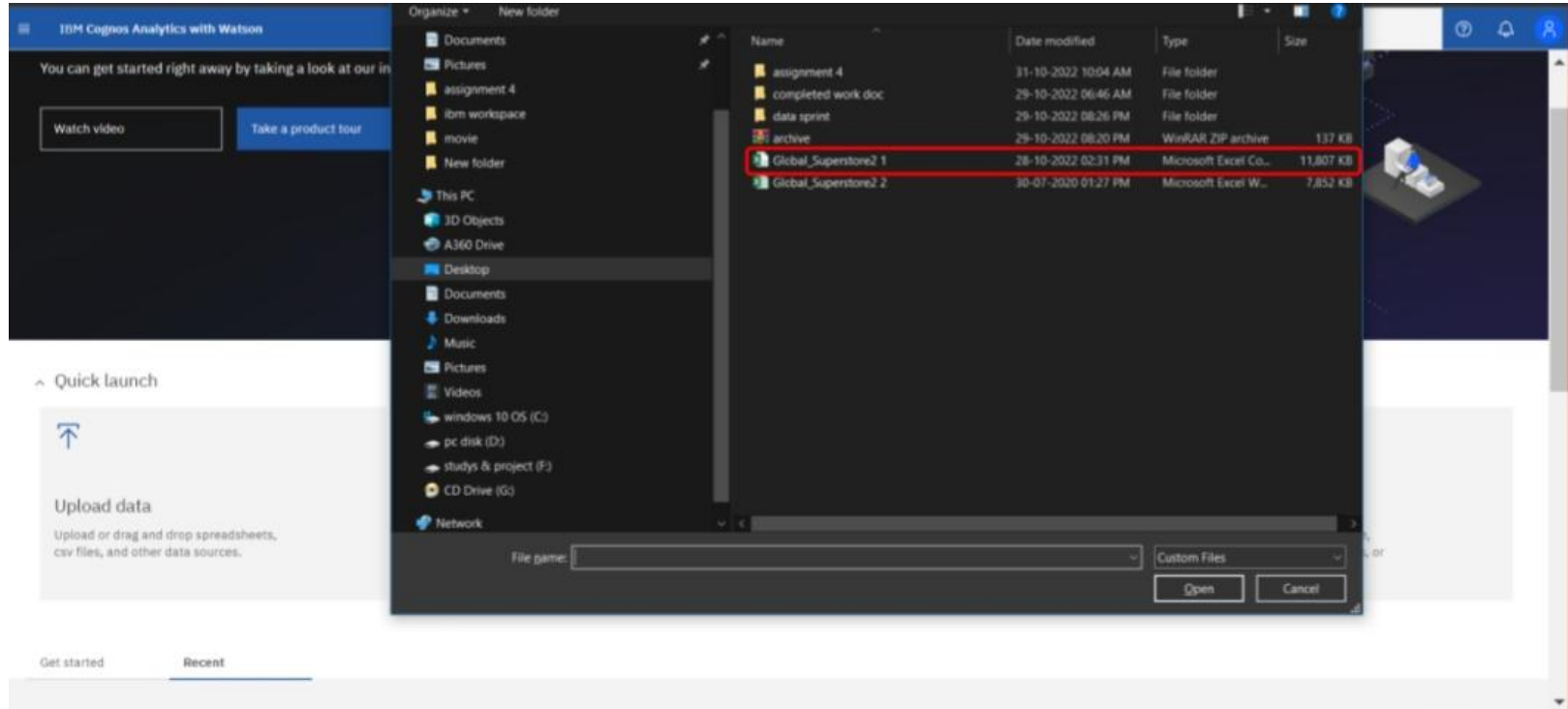
- Understanding and analyzing the dataset by using Python library's

```
+ Code + Markdown ▶ Run All ☒ Clear Outputs of All Cells ↺ Restart 📄 Variables 📄 Outline ... Python 3.10.4 64-bit
[+] Python Python
... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 51290 entries, 0 to 51289
Data columns (total 24 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Row ID                 51290 non-null  int64
1   Order ID               51290 non-null  object
2   Order Date             51290 non-null  object
3   Ship Date              51290 non-null  object
4   Ship Mode              51290 non-null  object
5   Customer ID            51290 non-null  object
6   Customer Name          51290 non-null  object
7   Segment               51290 non-null  object
8   City                   51290 non-null  object
9   State                  51290 non-null  object
10  Country                51290 non-null  object
11  Postal Code            9994 non-null   float64
12  Market                 51290 non-null  object
13  Region                 51290 non-null  object
14  Product ID             51290 non-null  object
15  Category               51290 non-null  object
16  Sub-Category           51290 non-null  object
17  Product Name           51290 non-null  object
18  Sales                  51290 non-null  float64
19  Quantity               51290 non-null  int64
20  Discount               51290 non-null  float64
21  Profit                 51290 non-null  float64
22  Shipping Cost          51290 non-null  float64
23  Order Priority          51290 non-null  object
dtypes: float64(5), int64(2), object(17)
memory usage: 9.4+ MB
```

● Open IBM COGNOS ANALYTICS



- Load dataset “**Global_Superstore2 1**” in IBM COGNOS ANALYTICS



- Data loading successfully in IBM COGNOS

