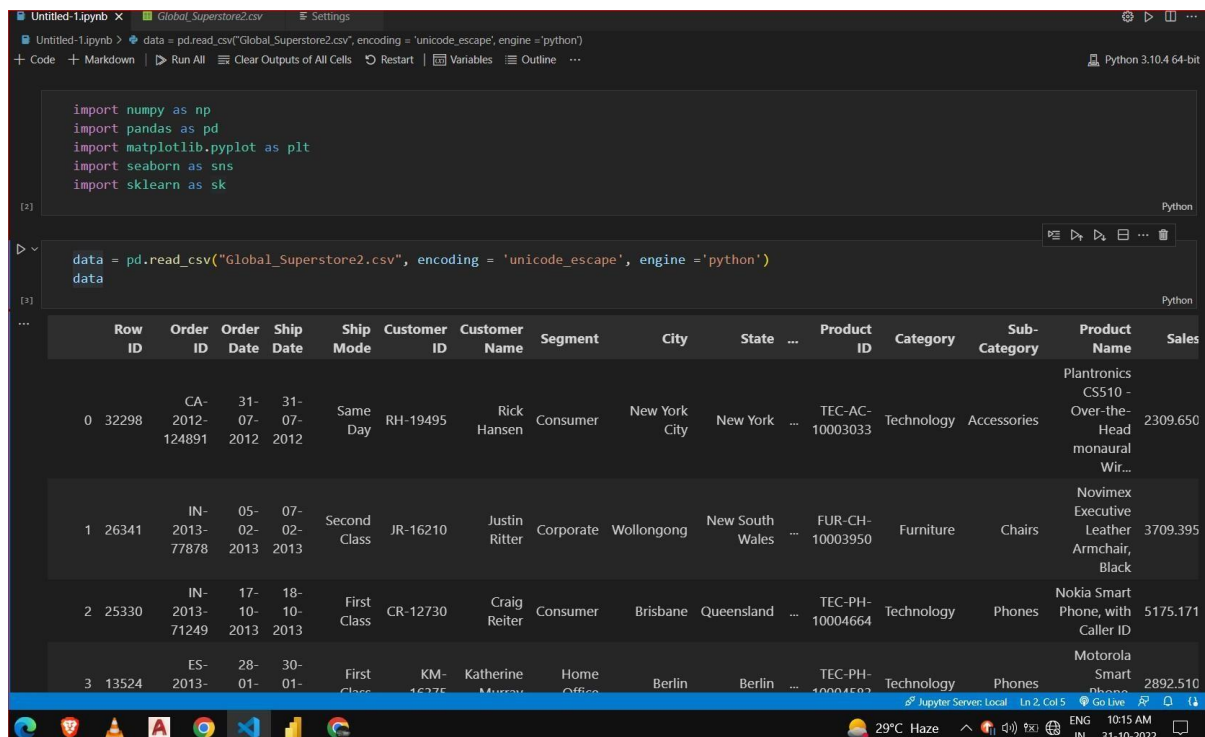


# PROJECT DEVELOPMENT PHASE

## DELIVERY OF SPRINT-1

DATE	12 NOVEMBER 2022
TEAM ID	PNT2022TMID01950
PROJECT NAME	Global Sales Data Analytics
MAXIMUM MARK	4 Marks

- Import library and load dataset in python



The screenshot shows a Jupyter Notebook interface with the following code and output:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn as sk
```

```
data = pd.read_csv("Global_Superstore2.csv", encoding = 'unicode_escape', engine = 'python')
data
```

The output displays a preview of the dataset with the following columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, City, State, Product ID, Category, Sub-Category, Product Name, and Sales. The first four rows of data are shown:

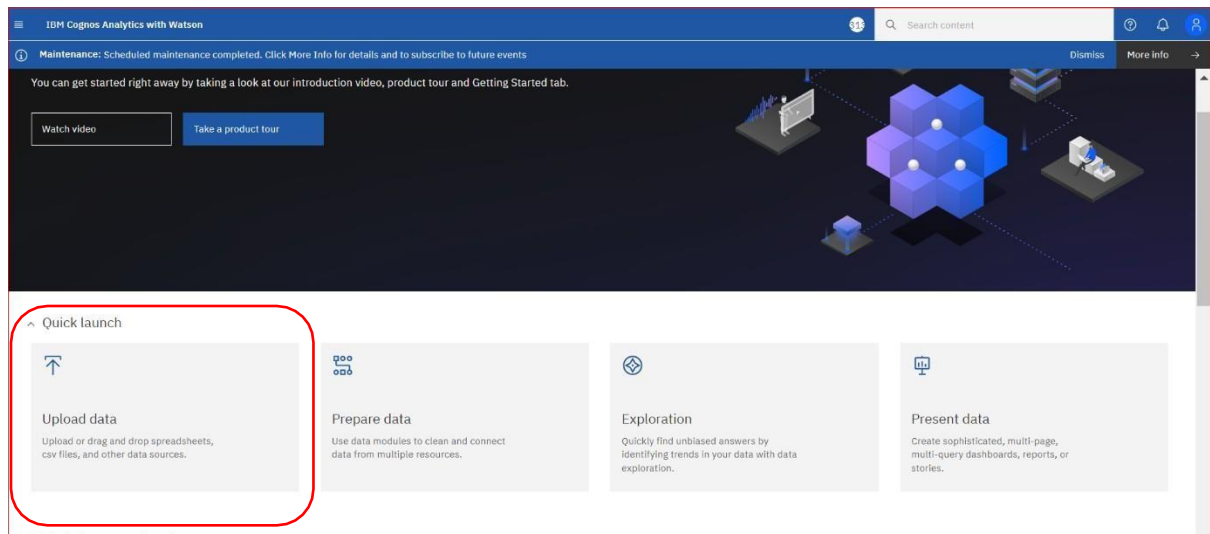
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Product ID	Category	Sub-Category	Product Name	Sales
0	32298	2012-07-12	2012-07-31	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	TEC-AC-10003033	Technology	Accessories	Plantronics CS510 - Over-the-Head monaural Wir...	2309.650
1	26341	2013-02-07	2013-05-02	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	2013-10-07	2013-10-17	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	TEC-PH-10004664	Technology	Phones	Nokia Smart Phone, with Caller ID	5175.171
3	13524	2013-01-28	2013-01-30	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	TEC-PH-10004592	Technology	Phones	Motorola Smart Phone	2892.510

- Understanding and Analyzing the dataset by using python library's.

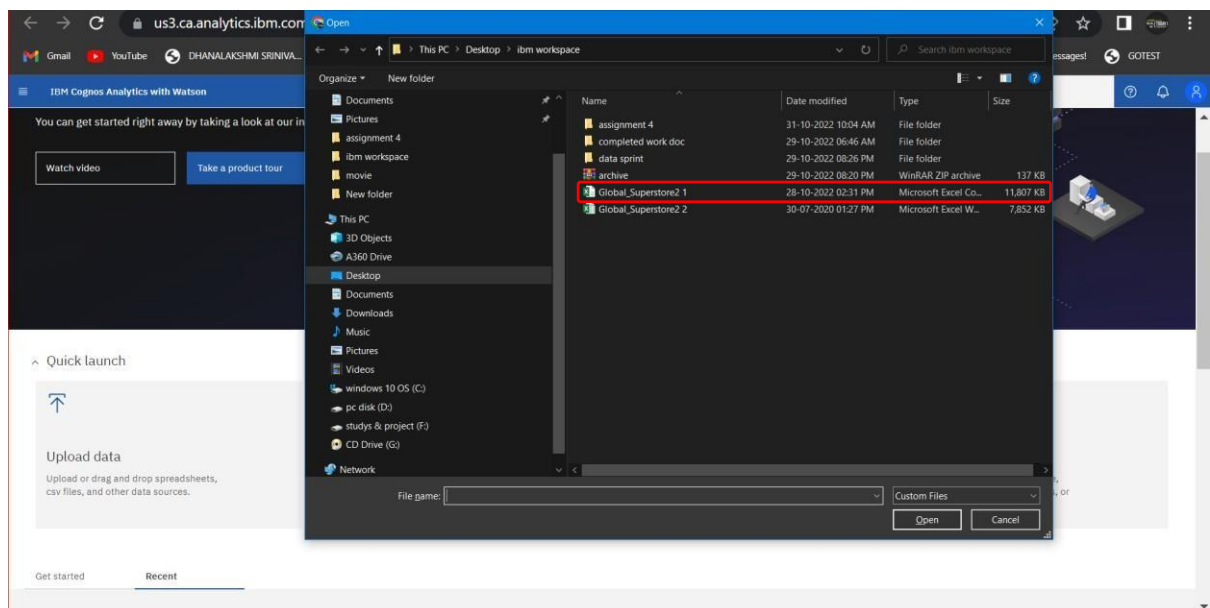
```
Untitled-1.ipynb x Global_Superstore2.csv Settings
Untitled-1.ipynb > data = pd.read_csv('Global_Superstore2.csv', encoding = 'unicode_escape', engine = 'python')
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Variables | Outline ... Python 3.10.4 64-bit

[4]:
...
<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

