

# PROJECT DEVELOPMENT PHASE-I

## SPRINT-I

Team ID	PNT2022TMID39710
Project Name	Global Sales Data Analytics

### Import library and load dataset in python

```
Untitled-1.ipynb x Global_Superstore2.csv Settings
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

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
```

	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-124891	2012-07-12	2012-07-12	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-77878	2013-02-07	2013-02-07	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-71249	2013-10-10	2013-10-10	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-16375	2013-01-16	2013-01-16	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	TEC-PH-10004664	Technology	Phones	Motorola Smart Phone	2892.510

### Understanding and Analysing the Dataset

```
Untitled-1.ipynb x Global_Superstore2.csv Settings
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

<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
```

## **Understanding The Datasets:**

- 1.Row ID** - It gives the idea of the row.
- 2.Order ID** - This ID is generated when the order is placed.
- 3.Order Date** - This gives the date on which the order is Placed.
- 4.Ship Date** - The Shipment date of the product.
- 5.Ship Mode** - The mode in which the shipment process is carried out.
- 6.Customer ID** - This ID is generated when the customer Places the first order.
- 7.Customer Name** - It gives the name of the customer.
- 8.Segment** - It gives the segment of the customer.
- 9.City** - The city in which the customer lives.
- 10.State** - The state in which the customer resides.
- 11.Country** - It gives the country of the customer.
- 12.Market** - Market where the order is placed.
- 13.Region** - It gives the region of the market.
- 14.Product ID** - This is unique ID generated for the each Product.
- 15.Category** - Category of the product.
- 16.Sub Category** - Sub Category to which the product Belongs.
- 17.Product Name** - The name of the product is mentioned.
- 18.Sales** - It gives sales details.
- 19.Quantity** - The number of product is ordered.
- 20.Discount** - Offer provided.
- 21.Profit** - Profit earned by the retailer.
- 22.Shipping cost** - The additional cost for shipping.
- 23.Order Priority** - Categorizes the order based on Priority level.