

# **AIM OF THE PROJECT**

## **CUSTOMER DATA MANAGEMENT**

### **CONTENT:**

- Purpose
- Process Involved in Our Analysis
- Data Source
- Tools used
- Aim of the Project
- Scope
- Data Collection
- Data Cleansing
- Data Profiling
- Loading Data into SQL Server
- Python
- Data loading into python
- Scatter plot
- Bar Graph
- Data Preparation for Model Building
- Model Building
- Model Performance
- Tableau
- Dashboard
- Conclusion

## **PURPOSE:**

To highlight the steps to acquire data and maintaining the database which would help in targeting potential customer and their needs.

## **PROCESS INVOLVED IN OUR ANALYSIS:**

Step1- Data Collection & Cleansing

Step2- Data Transform

Step3- Data Loading

Step4- Data Export through query

Step5- Data Visualization

## **DATA SOURCE:**

- Sales Database
- Internal Database
- Customer Database

## **TOOLS USED:**

- ETL
- MS-Excel
- Tableau
- SQL Server
- Python

## **AIM OF THE PROJECT:**

- To acquire data from various sources such as Customer data, Sales data, Internal data, CRM and collate it into single database.
- This can help in identifying potential customers for the company for its future business and marketing
- Thus, different marketing campaigns and advertising strategies can be directed towards the customer segments.

## **SCOPE:**

The scope of the project is to analyse the past data and visualize the Sales strategies, sales operations, sales analysis to ensure

- Sufficient sales volume
- Increased growth of the organization
- Current profit gained by the company

# DATA COLLECTION:

- Data pertaining to various attributes such as branch, city, customer type, gender, product line, unit price, quantity, tax, total, date, time, payment, cogs, gross income, rating in excel sheets from external sources such as Internal data, Customer data of the client.
- Initially a template in MS-Excel for the master data sheet was prepared and a few important attributes which need to be captured in it and subsequently into the database were finalized after thorough discussions with the team.
- Detailed description of the steps followed for the collection of data and collating it into the master database as follows:

The screenshot shows a Microsoft Excel spreadsheet titled "MasterData.xlsx". The "Master" tab is active, displaying a table with the following columns and data:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	concat	City	State	State_Code	Country	
2	40998	CA-2014-AH10015140-41954	11-11-2014	13-11-2014	First Class	AB-100151402	Aaron Bergman	Consumer	Consumer Aaron Bergman	Oklahoma City	Oklahoma	United States	
3	26341	IN-2014-JR162107-41675	05-02-2014	07-02-2014	Second Class	JR-162107	Justin Ritter	Corporate	Corporate Justin Ritter	Wollongong	New South Wales	Australia	
4	25330	IN-2014-CR127307-41929	17-10-2014	18-10-2014	First Class	CR-127307	Craig Reiter	Consumer	Consumer Craig Reiter	Brisbane	Queensland	Australia	
5	13924	ES-2014-KM1637548-41667	28-01-2014	30-01-2014	First Class	KM-1637548	Katherine Murray	Home Office	Home Office Katherine Murray	Berlin	Ber	Germany	
6	47375	IN-2014-MB10015140-41958	05-02-2014	07-02-2014	First Class	MB-100151402	James Mitchell	Corporate	Corporate James Mitchell	Dubai	Dub	Singapore	
7	22732	IN-2014-JM156557-41818	28-06-2014	01-07-2014	Second Class	JM-156557	Jim Mitchell	Corporate	Corporate Jim Mitchell	Sydney	New South Wales	Australia	
8	30579	IN-2012-TS2134099-41219	06-11-2012	08-11-2012	First Class	TS-21340992	Toby Swindell	Consumer	Consumer Toby Swindell	Porirua	Wellington	New Zealand	
9	31192	IN-2014-MB1008592-41379	18-04-2013	19-04-2013	Standard Class	MB-1008592	Mick Brown	Consumer	Consumer Mick Brown	Hamilton	Waikato	New Zealand	
10	40999	CA-2014-AH10015140-41954	11-11-2014	13-11-2014	First Class	AB-100151402	Aaron Bergman	Consumer	Consumer Aaron Bergman	Oklahoma City	Ok	United States	
11	36258	CA-2012-AB10015140-40974	06-03-2012	07-03-2012	First Class	AB-100151404	Aaron Bergman	Consumer	Consumer Aaron Bergman	Seattle	Wig	United States	
12	36259	CA-2012-AB10015140-40974	06-03-2012	07-03-2012	First Class	AB-100151404	Aaron Bergman	Consumer	Consumer Aaron Bergman	Seattle	Wig	United States	
13	22096	IN-2013-JS15217518-41253	13-09-2013	15-09-2013	First Class	JS-15217518	Corporate	Corporate Magdalene Morse	Kabul	Kab	Afghanistan		
14	45794	SA-2013-MM97360110-41269	26-12-2012	28-12-2012	Second Class	MM-97360110	Madelaine Morse	Consumer	Corporate Magdalene Morse	Izran	Jilan	Saudi Arabia	
15	41323	MW-2013-S2171518-41591	13-11-2013	13-11-2013	Same Day	WV-2171518	Vicky Freymann	Home Office	Home Office Vicky Freymann	Toledo	Parana	Pa	
16	27794	IN-2014-PF191202-41796	06-06-2014	08-06-2014	Second Class	PF-191202	Peter Fuller	Consumer	Consumer Peter Fuller	Mudanjiang	He	China	
17	13779	ES-2015-BP1118545-42216	31-07-2015	03-08-2015	Second Class	BP-1118545	Ben Peterman	Corporate	Corporate Ben Peterman	Paris	ile-de-France	Fr	
18	39519	CA-2012-AB10015140-40958	19-02-2012	20-02-2012	Standard Class	AB-100151402	Aaron Bergman	Consumer	Consumer Aaron Bergman	Arlington	Te	United States	
19	12669	ES-2015-P11883564-42255	08-09-2015	14-09-2015	Standard Class	PJ-11883564	Patrick Jones	Corporate	Corporate Patrick Jones	Prato	Tuscany	Tu	
20	22996	IN-2015-JS15217518-42020	31-01-2015	01-02-2015	First Class	JS-15217518	Corporate	Corporate Jim Sink	Townsville	Queensland	Qs	Australia	
21	45363	TZ-2013-AB00050512-41243	08-09-2013	10-09-2013	Second Class	TZ-00050512	Corporate	Corporate Jim Slink	Vientiane	Kip	Laos		
22	40539	IN-2013-AB800103-41494	08-08-2013	10-08-2013	First Class	AB-000103	Ann Blume	Corporate	Corporate Ann Blume	Bytom	Silesia	Pl	
23	36266	CA-2012-AB10015140-40974	06-03-2012	07-03-2012	First Class	AB-100151404	Aaron Bergman	Consumer	Consumer Aaron Bergman	Seattle	Wig	United States	
24	21856	IN-2012-JK152257-41930	01-05-2012	02-05-2012	First Class	JK-152257	Jason Klamczynski	Corporate	Corporate Jason Klamczynski	Suzhou	Anhui	China	
25	13928	ES-2014-LB16795139-41697	27-02-2014	01-03-2014	Second Class	LB-16795139	Lauren Beltran	Home Office	Home Office Lauren Beltran	Edinburgh	Scotland	Sc	United Kingdom
26	26570	US-2015-NP1832582-42216	31-07-2015	01-08-2015	First Class	NP-1832582	Naresj Patel	Consumer	Consumer Naresj Patel	Juarez	Chihuahua	Ch	Mexico
27	34019	IN-2012-PB19210102-41259	16-12-2012	19-12-2012	First Class	PB-19210102	Philip Breyer	Corporate	Corporate Phillip Breyer	Taipei	Taipei City	IN/A	Taiwan
28	11283	ES-2015-JS15217518-42020	13-01-2015	15-01-2015	Second Class	JS-15217518	Corporate	Corporate Jim Sink	Ulaanbaatar	Saxen	Germany	IN/A	
29	38469	CA-2012-AB100850140-41928	23-04-2012	25-04-2012	Second Class	AB-100850140	Augustine Hawkins	Corporate	Corporate Augustine Hawkins	Troy	New York	IN/A	United States
30	22999	IN-2013-BP1123095-41329	04-02-2013	04-02-2013	Same Day	BP-11230958	Benjamin Patterson	Consumer	Consumer Benjamin Patterson	Surat	Gujarat	IN/A	India
31	220	US-2012-RH1952536-41279	27-12-2012	29-12-2012	Second Class	RH-1952536	Rick Reed	Corporate	Corporate Rick Reed	Santo Domingo	Santo Domingo	IN/A	Dominican Republic
32	10048	ES-2013-BS1180545-41472	17-07-2013	19-07-2013	First Class	BS-1180545	Bill Shoney	Corporate	Corporate Bill Shoney	Saint-Brieuc	Brittany	IN/A	France
33	40977	CA-2013-AB10030340-41635	27-12-2013	31-12-2013	Standard Class	AB-100303404	Aaron Hawkins	Corporate	Corporate Aaron Hawkins	San Francisco	California	IN/A	United States
35	21286	IN-2012-DP131057-41215	02-11-2012	04-11-2012	Second Class	DP-131057	Dave Poirier	Corporate	Corporate Dave Poirier	Gold Coast	Queensland	Qs	Australia
36	40976	CA-2013-AB10030340-41635	27-12-2013	31-12-2013	Standard Class	AB-100303404	Aaron Hawkins	Corporate	Corporate Aaron Hawkins	San Francisco	California	IN/A	United States
37	40975	CA-2013-AB10030340-41635	18-05-2013	20-05-2013	Standard Class	AB-100303404	Aaron Hawkins	Corporate	Corporate Aaron Hawkins	Karlsruhe	Karlsruhe	IN/A	Germany
38	36651	CA-2012-AB10030340-41041	18-05-2012	18-05-2012	Standard Class	AB-100303404	Aaron Hawkins	Corporate	Corporate Aaron Hawkins	Los Angeles	California	IN/A	United States
39	40109	CA-2013-AB10030340-41273	30-12-2012	31-12-2012	First Class	AB-100303406	Aaron Hawkins	Corporate	Corporate Aaron Hawkins	New York City	New York	IN/A	United States
40	40978	CA-2013-AB10030340-41635	27-12-2013	31-12-2013	Standard Class	AB-100303404	Aaron Hawkins	Corporate	Corporate Aaron Hawkins	Cardiff	Cardiff	IN/A	United Kingdom

The "Sales\_data" tab is also visible, showing a table with columns: Sales\_ID, Sales Date, Sales Type, Sales Value, and Sales Status.

WPS Office MasterData.xlsx

Menu Home Insert Page Layout Formulas Data Review View Tools Smart Toolbox Click to find commands

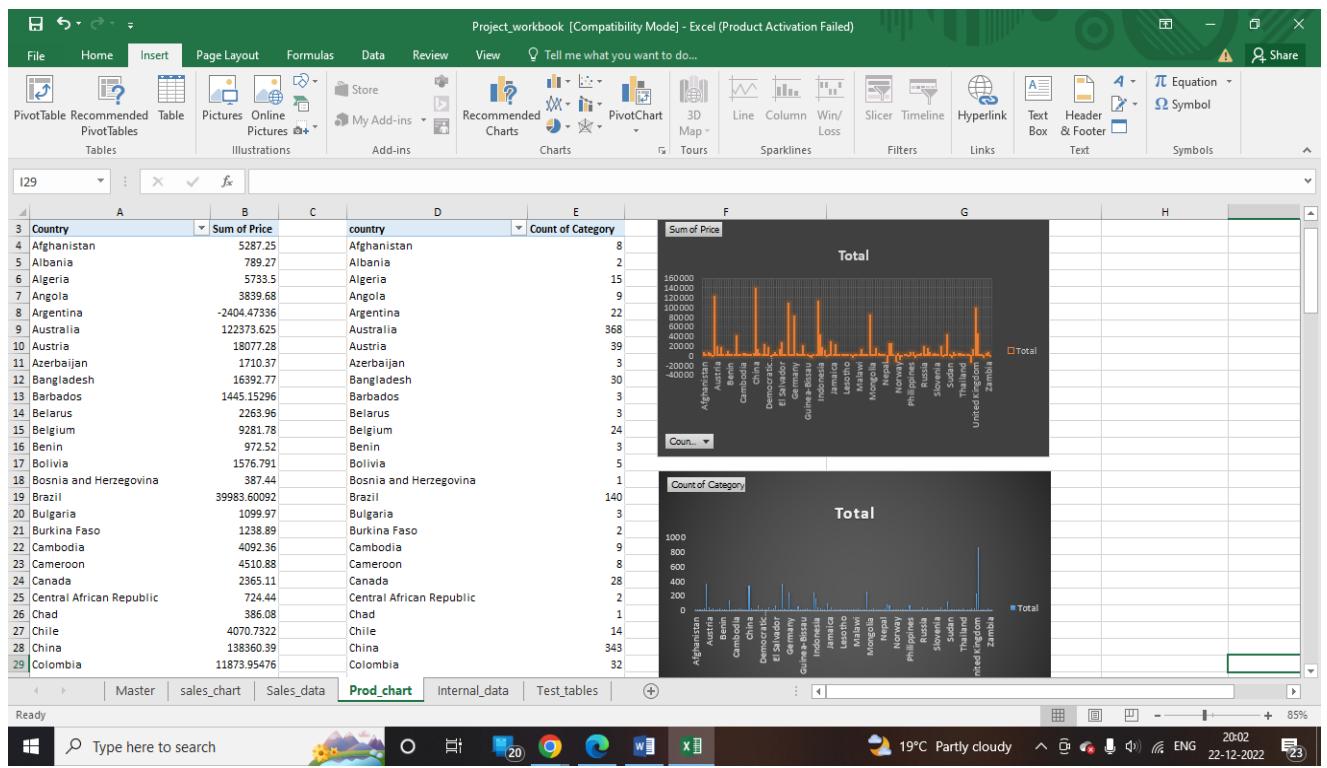
ProtTable AutoFilter Advanced Sort \* Highlight Duplicates \* Manage Duplicates \* Text to Columns \* Fill \* Validation \* Insert Drop-Down List Consolidate Form Group Ungroup Subtotal Hide Detail Split Sheet \* Merge Sheet \* Import Data \* Refresh All \* Settings

J12 fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	Country																											
2	United States US																											
3	Sydney	Syd																										
4	Berlin	Ber																										
5	Delhi	Del																										
6	London	Lon																										
7	Paris	Par																										
8	Aachen	Aac																										
9	Aalen	Aal																										
10	Abbotsford	Abbot																										
11	Australia	Aus																										
12	Germany	Ger																										
13	Senegal	Sen																										
14	New Zealand	Nz																										
15	Afghanistan	Afg																										
16	Saudi Arabia	Sa																										
17	Brazil	Brz																										
18	China	Ch																										
19	France	Fran																										
20	Italy	Ita																										
21	Tanzania	Tan																										
22	Poland	Pol																										
23	Chihuahua	Ch																										
24																												
25																												
26																												
27																												
28																												
29																												
30																												
31																												
32																												
33																												
34																												
35																												
36																												
37																												
38																												
39																												

Master sales\_chart Sales\_data Prod\_chart Internal\_data Test\_tables

27°C Haze ENG 22-12-2022 15:11



Project\_workbook [Compatibility Mode] - Excel (Product Activation Failed)

**Insert**

PivotTable Recommended Charts PivotChart 3D Map Line Column Win/Loss Slicer Timeline Hyperlink Text Box Header & Footer Text Symbols

**Sum of Profit**

	Total
FIRST CLASS	244132.2091
SAME DAY	276577.0681
SECOND CLASS	374036.2042
STANDARD CLASS	483132.9567
<b>Grand Total</b>	<b>727234.3452</b>

**Sum of sales**

Years	Sum of sales
2012	244132.2091
2013	276577.0681
2014	374036.2042
2015	483132.9567
<b>Grand Total</b>	<b>1377878.438</b>

**Shipping\_mode**

**Sum of Profit**

	Total
First Class	123971.4882
Same Day	44095.36832
Second Class	160460.6031
Standard Class	398706.8855
<b>Grand Total</b>	<b>727234.3452</b>

**Sales\_data**

**Prod\_chart**

**Internal\_data**

**Test\_tables**

Ready

19°C Partly cloudy 20:01 22-12-2022

Project\_workbook [Compatibility Mode] - Excel (Product Activation Failed)

**View**

Ruler Formula Bar Gridlines Headings Show Zoom 100% Zoom to Selection Window New Arrange Freeze Hide Synchronous Scrolling Window All Panes Unhide Reset Window Position Switch Windows Macros

**Sum of Profit**

	Total
First Class	65930.77384
Same Day	20505.34216
Second Class	74048.76896
Standard Class	128512.0262
<b>Grand Total</b>	<b>288996.9112</b>

**Sheet2**

**Master**

**sales\_chart**

**Sales\_data**

**Prod\_chart**

**Sheet1**

**Internal\_data**

**Test\_tables**

Ready

19°C Partly cloudy 20:39 22-12-2022

## **DATA CLEANSING:**

Before transferring data from these worksheets into a single worksheet containing master data we need to cleanse some data or modify some of the data in order to make the data uniform before it goes to master data sheet.

The attribute whose data needs modification are as follows:

Customer data : Customer ID, Customer name, Segment, Order ID, Row ID

Sales data: Order date, Ship date, Ship mode

Internal Data: City, State, State code, Country

## **DATA PROFILING:**

The process of examining the data available from existing information source file or Database. It will provide an important amount of useful insight about the quality of the data.

# LOADING DATA INTO SQL SERVER:

Why SQL server?

- In SQL database it is easier to extract data as per our requirement.
- In any organization there may be a large number of master data files and as a result maintaining a database can help.
- MS-Excel has a limited capacity to store up to 10 lakhs.
- Hence under those circumstances where we need to deal with much larger volumes of data importing into SQL is useful.

So basically we will now query the data according to our requirements, below are some examples:

## 1. Write a query to find payment mode-wise sales?

--1) Find the payment mode wise sales?

```
select Payment,sum(total) as Sales from data_manage group by Payment;
```

Results

	Product_line	Sales
1	Fashion accessories	54305.895
2	Electronic accessories	54337.5315
3	Food and beverages	56144.844
4	Sports and travel	55122.8265
5	Home and lifestyle	53861.913

## 2. Write a query to find the product line sales?

--2) Find the product line sales less than 50000?

```
select Product_line,sum(total) as Sales from data_manage group by Product_line having sum(total)>50000;
```

100 % <

Results Messages

	Product_line	Sales
1	Fashion accessories	54305.895
2	Electronic accessories	54337.5315
3	Food and beverages	56144.844
4	Sports and travel	55122.8265
5	Home and lifestyle	53861.913

## 3. Write a query Find the ratings <4.2 and payment mode is credit card?

--3) Find the ratings <4.2 in payment mode=credit card?

```
Select city,Customer_type,Product_line,Gender,Payment,Rating from data_manage where Customer_type='Member' and Rating<4.2 and Payment='credit card';
```

1% <

Results Messages

city	Customer_ty...	Product_line	Gender	Payment	Rating
Mandalay	Member	Health and beauty	Male	Credit card	4
Yangon	Member	Food and beverages	Male	Credit card	4
Naypyitaw	Member	Food and beverages	Female	Credit card	4
Yangon	Member	Fashion accessories	Female	Credit card	4.1
Yangon	Member	Home and lifestyle	Female	Credit card	4.1

#### 4. Write a query to find customer type is member & COGS>850?

--4) find customer type= member & COGS>850?

```
select city, customer_type, gender, Product_line, cogs from data_manage where Customer_type='member' and cogs>850;
```

Results

	city	customer_type	gen...	Product_line	cogs
1	Yangon	Member	Male	Electronic accessories	886.7
2	Mand...	Member	Male	Sports and travel	899.64
3	Nayp...	Member	Fem...	Sports and travel	898
4	Nayp...	Member	Male	Health and beauty	905
5	Mand...	Member	Male	Health and beauty	874.98
6	Nayp...	Member	Fem...	Fashion accessories	993
7	Nayp...	Member	Fem...	Fashion accessories	972.1
8	Yangon	Member	Fem...	Home and lifestyle	906.5
9	Nayp...	Member	Fem...	Food and beverages	985.2
10	Nayp...	Member	Fem...	Food and beverages	890.73
11	Nayp...	Member	Fem...	Food and beverages	871
12	Nayp...	Member	Male	Fashion accessories	898.38
13	Yangon	Member	Fem...	Food and beverages	887.94
14	Yangon	Member	Fem...	Sports and travel	877.32

#### 5. Write a query to find Count the on product line-wise rating?

--5) Count on product line?

```
select Product_line, count(Rating) as Rate from data_manage group by Product_line;
```

Results

	Product_line	Rate
1	Fashion accessories	178
2	Health and beauty	152
3	Electronic accessories	170
4	Food and beverages	174
5	Sports and travel	166
6	Home and lifestyle	160

## 6. Write a query to find the Top 10 where gross income is greater than 30 & city is Yangon?

```
--6)Write a query to find the Top 10 where gross income > 30 & city = Yangon?  
select top 10 percent city,Customer_type,Product_line,Gender,Payment from data_manage where gross_income>30 and city='yangon';
```

100 % <

Results Messages

	Product_line	Rate
1	Fashion accessories	178
2	Health and beauty	152
3	Electronic accessories	170
4	Food and beverages	174
5	Sports and travel	166
6	Home and lifestyle	160

## 7. Write a query to find the city's rank per total sales?

```
--7) Write a query to find the city's rank per total sales?  
select city,total,rank() over (order by rating desc) as myrank from data_manage;
```

100 % <

Results Messages

	city	total	myrank
1	Mandalay	520.4115	1
2	Naypyitaw	77.931	1
3	Mandalay	588.357	1
4	Yangon	339.36	1
5	Mandalay	554.295	1
6	Yangon	536.592	6
7	Mandalay	351.603	6
8	Naypyitaw	914.55	6
9	Yangon	561.078	6
10	Mandalay	335.895	6
11	Yangon	214.935	6
12	Mandalay	523.3725	6
13	Mandalay	193.011	6
14	Naypyitaw	103.824	6
15	Naypyitaw	273.42	6
16	Yangon	77.7735	6
17	Yangon	171.7275	6
18	Mandalay	614.943	6
19	Yangon	181.44	6
20	Naypyitaw	401.73	6
21	Yangon	367.038	6
22	Yangon	216.846	22
23	Naypyitaw	451.3635	22
24	Mandalay	333.207	22
25	Naypyitaw	14.679	22
26	Naypyitaw	264.7575	22

## 8. Write a query to find Gender wise sales?

```
--8) Write a query to find Gender wise sales?  
select Gender,sum(total) as Sales from data_manage group by Gender;
```

100 % < Results Messages

	Gender	Sales
1	Male	155083.824
2	Female	167882.925

## 9. Write a query to find the ratings greater than 9.8 payment mode and gender Female?

```
-- 9) Write a query to find the ratings > 9.8 payment mode and gender Female?  
select city,Customer_type,Product_line,Gender,Rating,Payment from data_manage where Payment='credit card' and Rating>=9.8 and Gender='Female';
```

100 % < Results Messages

	city	Customer_ty...	Product_line	Gender	Rati...	Payment
1	Naypyitaw	Member	Sports and travel	Female	10	Credit card
2	Yangon	Normal	Health and beauty	Female	10	Credit card
3	Naypyitaw	Member	Food and beverages	Female	9.9	Credit card
4	Naypyitaw	Normal	Fashion accessories	Female	9.8	Credit card

## 10. Write a query to find the product line sales greater than 50000?

```
--10) Write a query to find the product line sales greater than 50000?  
select Product_line,sum(total) as Sales from data_manage group by Product_line having sum(total)>50000;
```

100 % <

Results Messages

	Product_line	Sales
1	Fashion accessories	54305.895
2	Electronic accessories	54337.5315
3	Food and beverages	56144.844
4	Sports and travel	55122.8265
5	Home and lifestyle	53861.913

## 11. Write a query to find the product Line is 'health and beauty' and gross\_income greater than 10 and gender is 'female'?

```
--11) Write a query to find the product Line is 'health and beauty' and gross_income grater than 10 and gender is 'female'?  
select city,gender,product_line,payment,gross_income,Rating from data_manage  
where Product_line='Health and beauty' and gross_income>10 and rating>=9 and gender='female';
```

100 % <

Results Messages

	city	gender	product_line	payment	gross_inco...	Rating
1	Yangon	Female	Health and beauty	Ewallet	26.1415	9.1
2	Naypyitaw	Female	Health and beauty	Cash	34.3	9.1
3	Naypyitaw	Female	Health and beauty	Credit card	13.878	9.5
4	Yangon	Female	Health and beauty	Credit card	16.16	10
5	Mandalay	Female	Health and beauty	Cash	26.208	9.7
6	Mandalay	Female	Health and beauty	Cash	16.743	9.9

## **PYTHON:**

In order to visualize data and prediction to extract the important insights from it regarding potential business opportunities as well as future prediction, so that we can take important decisions such as:

- Identify revenues for individual products.
- Modify or increase the products and sales portfolio and future prediction.
- Identify important areas for investment and come up with conclusion for building marketing strategies, etc.

Python can be used to connect to any type of file. In this case we would be connecting to excel file and our analysis would be based on historical data and predicting future performance.

## **DATA LOADING INTO PYTHON:**

Import associated library such as pandas, matplotlib for data manipulation and visualization.

Read files using pd.read () function so that it will store into a dataframe.

Project - Jupyter Notebook

localhost:8888/notebooks/Data\_Management/Project.ipynb

jupyter Project Last Checkpoint: Yesterday at 10:27 AM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel) Logout

In [18]: `import pandas as pd  
import matplotlib as plt`

In [5]: `df=pd.read_excel('C:/Users/nehab/Desktop/Project_workbook.xls')`

In [6]: `df`

Out[6]:

Row_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	City	State	Product_ID	Category	
0	1	CA-2014-AB10015140-41954	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman	Consumer	Oklahoma City	Oklahoma	TEC-PH-5816	Technology
1	2	IN-2014-JR162107-41675	2014-02-05	2014-02-07	Second Class	JR-162107	Justin Ritter	Corporate	Wollongong	New South Wales	FUR-CH-5379	Furniture
2	3	IN-2014-CR127307-41929	2014-10-17	2014-10-18	First Class	CR-127307	Craig Reiter	Consumer	Brisbane	Queensland	TEC-PH-5356	Technology
3	4	ES-2014-KM1637548-41667	2014-01-28	2014-01-30	First Class	KM-1637548	Katherine Murray	Home Office	Berlin	Berlin	TEC-PH-5267	Technology
4	5	SG-2014-RH9495111-41948	2014-11-05	2014-11-06	Same Day	RH-9495111	Rick Hansen	Consumer	Dakar	Dakar	TEC-CO-6011	Technology
...	...	SY-2013-	...	...	...	...	...	...	...	...	...	...

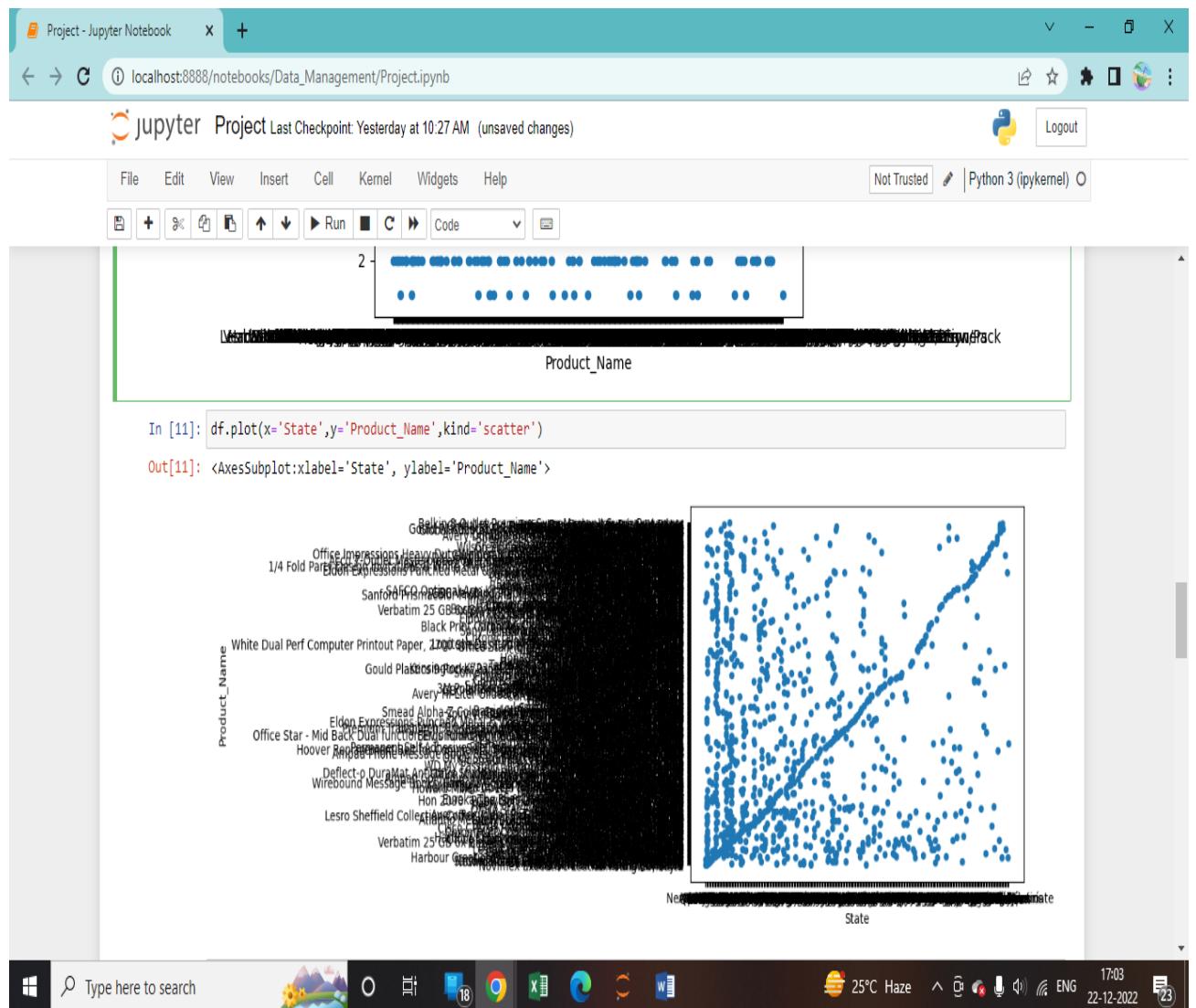
Type here to search

2018-09-01 16:58 25°C Haze ENG 22-12-2022 16:58 23

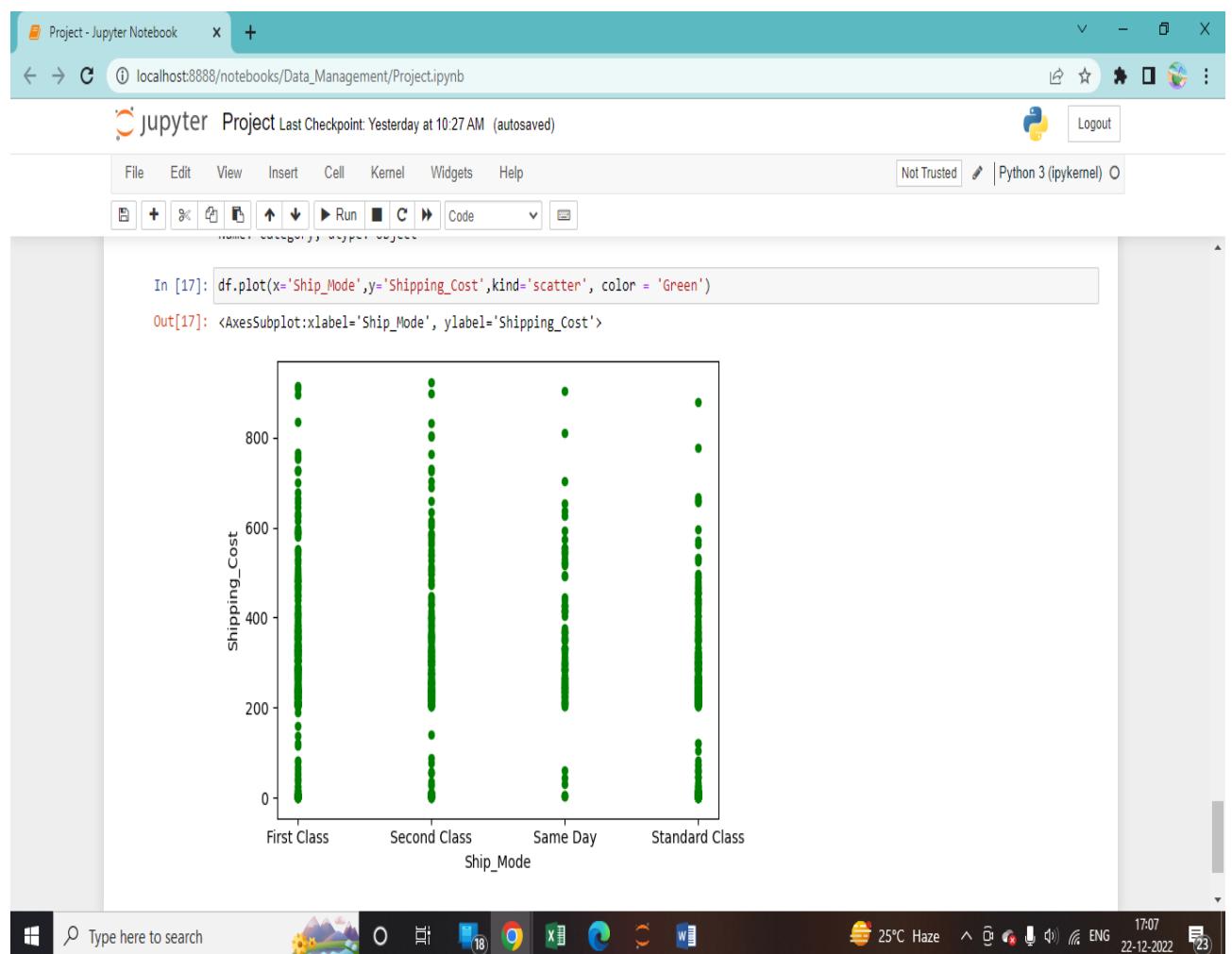
# PLOT SCATTER PRODUCT\_NAME VS QUANTITY FOR BETTER UNDERSTANDING OF DATA



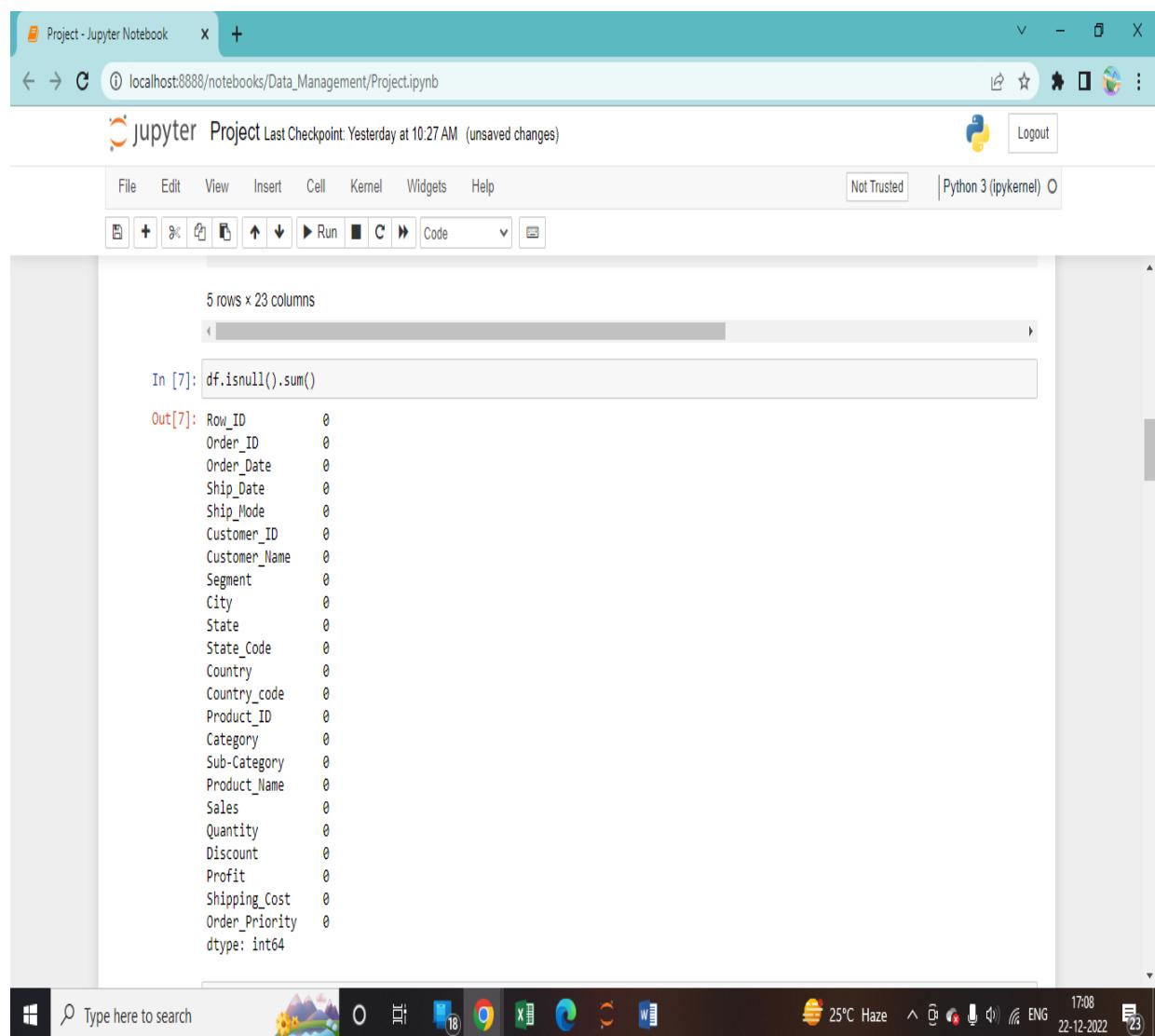
# PLOT SCATTER STATE VS PRODUCT\_NAME FOR BETTER UNDERSTANDING



# PLOT SCATTER SHIP\_MODE VS SHIPPING\_COST FOR BETTER UNDERSTANDING



**Using df.info we are checking detailed information about our data such as null values count, data type, etc.**



The screenshot shows a Jupyter Notebook interface running in a browser window titled "Project - Jupyter Notebook". The URL is "localhost:8888/notebooks/Data\_Management/Project.ipynb". The notebook header includes a logo, the title "Jupyter Project Last Checkpoint: Yesterday at 10:27 AM (unsaved changes)", and user information. The toolbar below has buttons for File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Not Trusted, Python 3 (ipykernel), and Logout. The main area displays the output of the command `df.isnull().sum()`. It shows a table with 23 columns and 5 rows, all of which have a value of 0. The columns are: Row\_ID, Order\_ID, Order\_Date, Ship\_Date, Ship\_Mode, Customer\_ID, Customer\_Name, Segment, City, State, State\_Code, Country, Country\_code, Product\_ID, Category, Sub-Category, Product\_Name, Sales, Quantity, Discount, Profit, Shipping\_Cost, and Order\_Priority. The data type for all columns is int64.

```
5 rows × 23 columns

In [7]: df.isnull().sum()

Out[7]:
Row_ID      0
Order_ID    0
Order_Date  0
Ship_Date   0
Ship_Mode   0
Customer_ID 0
Customer_Name 0
Segment     0
City        0
State       0
State_Code  0
Country     0
Country_code 0
Product_ID  0
Category    0
Sub-Category 0
Product_Name 0
Sales       0
Quantity    0
Discount    0
Profit      0
Shipping_Cost 0
Order_Priority 0
dtype: int64
```

Project - Jupyter Notebook

localhost:8888/notebooks/Data\_Management/Project.ipynb

Jupyter Project Last Checkpoint: Yesterday at 10:27 AM (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel) O

In [12]: df.info

```
Out[12]: <bound method DataFrame.info of
   Row_ID      Order_ID Order_Date Ship_Date     Ship_Mode \
0    CA-2014-AB10015140-41954 2014-11-11 2014-11-13 First Class
1    IN-2014-JR162107-41675 2014-02-05 2014-02-07 Second Class
2    IN-2014-CR127307-41929 2014-10-17 2014-10-18 First Class
3    ES-2014-KM1637548-41667 2014-01-28 2014-01-30 First Class
4    SG-2014-RH9495111-41948 2014-11-05 2014-11-06 Same Day
.. ...
994   SY-2013-SV10815126-41541 2013-09-24 2013-09-28 Standard Class
995   IN-2015-JF1541527-42095 2015-04-01 2015-04-05 Standard Class
996   CA-2015-AS10225140-42347 2015-12-09 2015-12-13 Second Class
997   IN-2014-CK1259592-41929 2014-10-17 2014-10-19 Second Class
998   MX-2014-SS2014018-41658 2014-01-19 2014-01-26 Standard Class

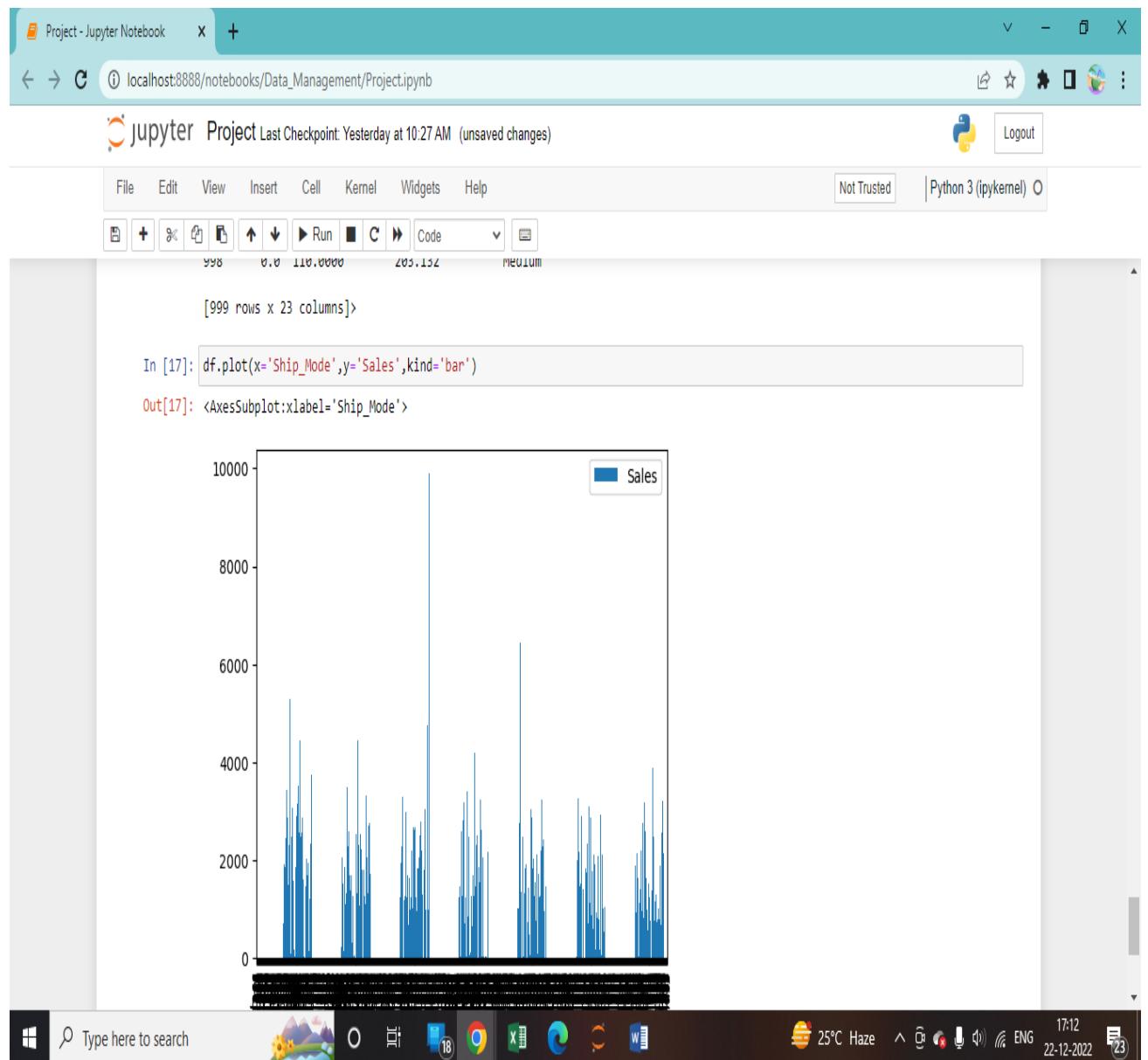
   Customer_ID Customer_Name Segment City \
0    AB-100151402    Aaron Bergman  Consumer Oklahoma City
1    JR-162107    Justin Ritter  Corporate Wollongong
2    CR-127307    Craig Reiter  Consumer Brisbane
3    KM-1637548  Katherine Murray  Home Office Berlin
4    RH-9495111    Rick Hansen  Consumer Dakar
.. ...
994   SV-10815126    Stuart Van  Corporate Homs
995   JF-1541527  Jennifer Ferguson  Consumer Liaoyang
996   AS-102251404    Alan Schoenberger  Corporate San Francisco
997   CK-1259592    Clytie Kelty  Consumer Taupo
998   SS-2014018    Saphhira Shifley  Corporate Araguaína

   State ... Product_ID Category Sub-Category \
0  Oklahoma ... TEC-PH-5816 Technology  Phones
1 New South Wales ... FUR-CH-5379 Furniture Chairs
```

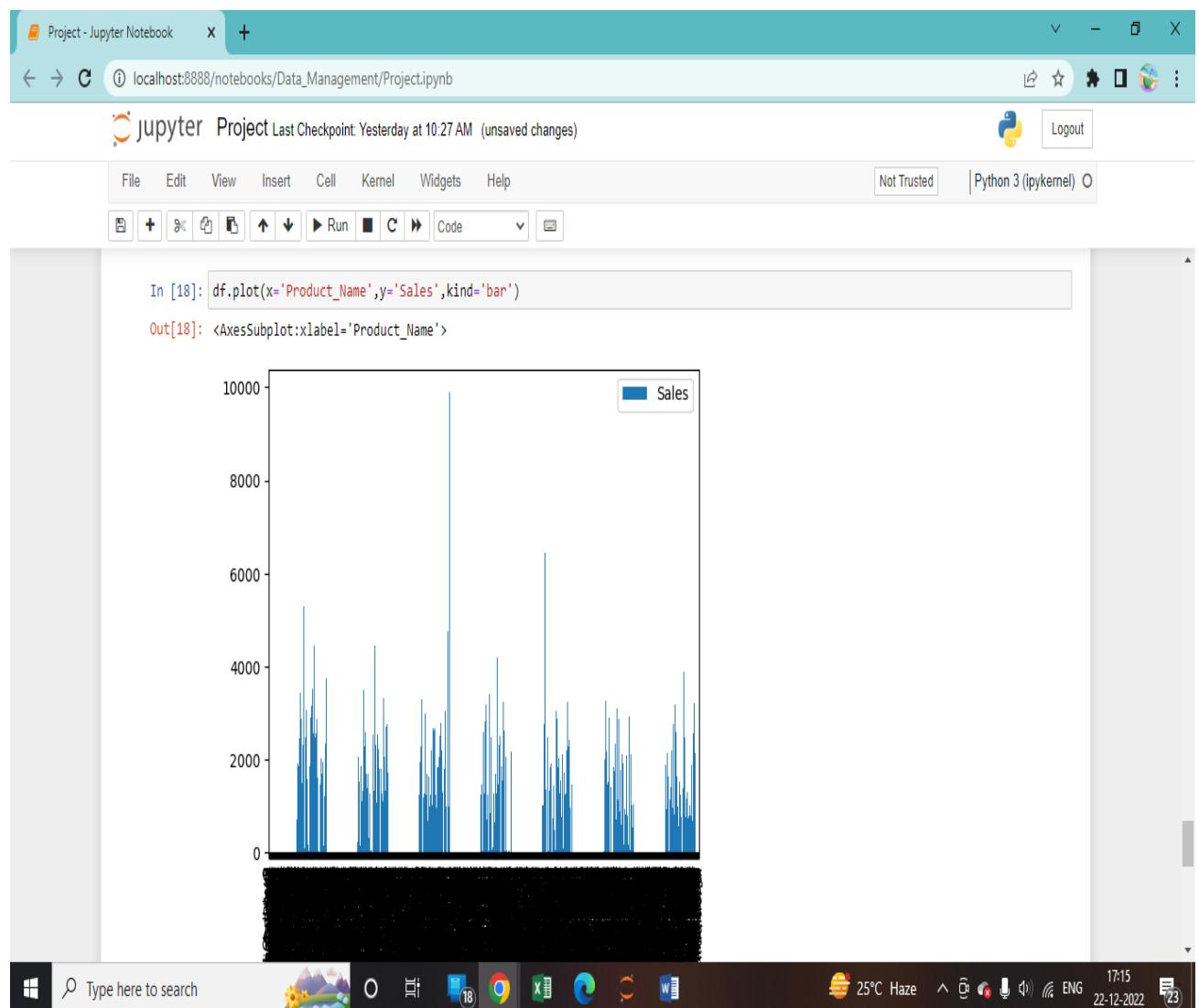
Type here to search

25°C Haze 17:08  
22-12-2022 ENG 23

# PLOT GRAPH SHIP\_MODE VS SALES FOR BETTER UNDERSTANDING

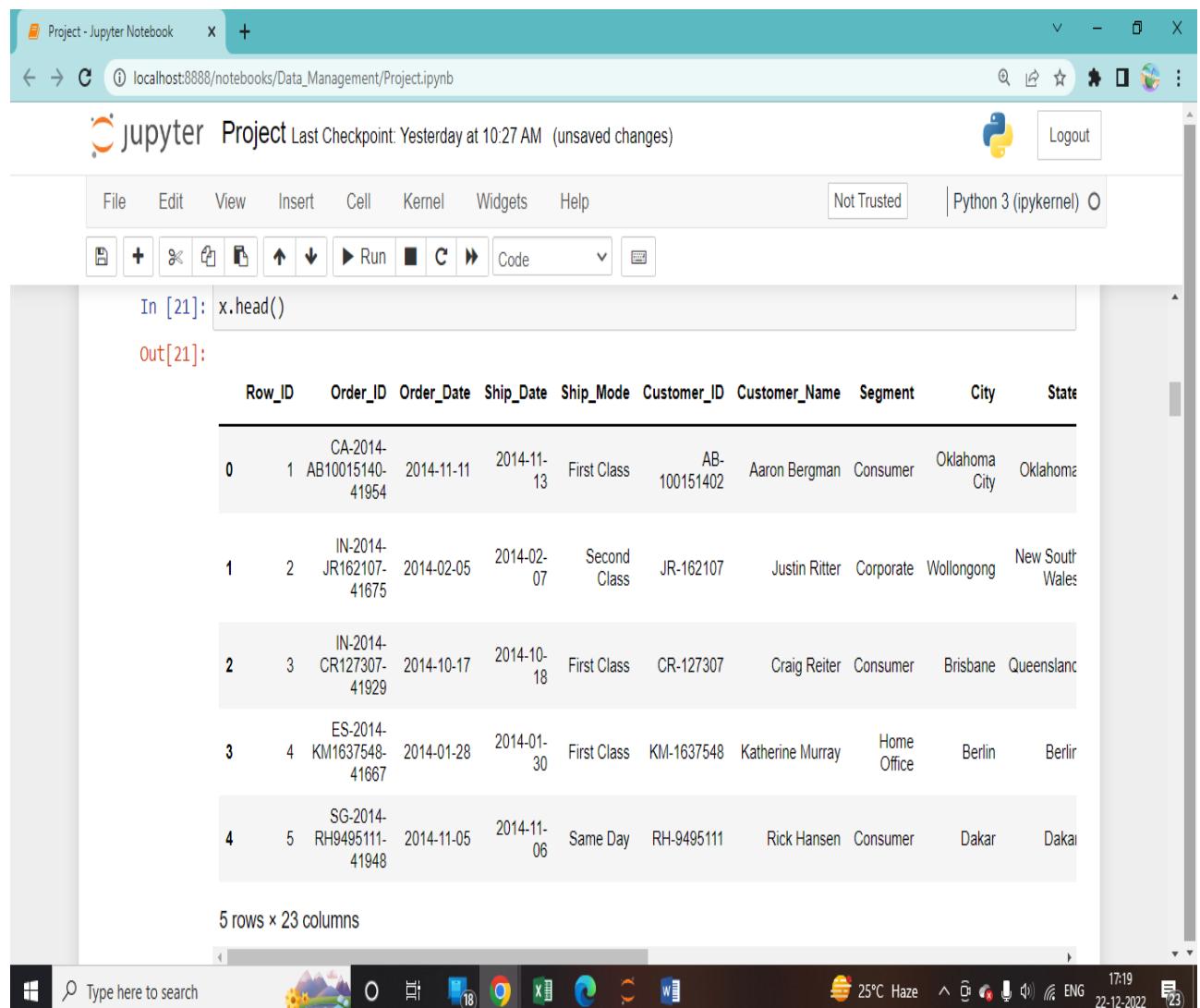


# PLOT GRAPH PRODUCT\_NAME VS SALES FOR BETTER UNDERSTANDING



# DATA PREPARATION FOR MODEL BUILDING

We have to divide the data into two parts as x(Independent Variable) and y(Dependent Variable) for model Building.



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

- Title Bar:** Project - Jupyter Notebook
- URL:** localhost:8888/notebooks/Data\_Management/Project.ipynb
- Header:** Jupyter Project Last Checkpoint: Yesterday at 10:27 AM (unsaved changes)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Not Trusted, Python 3 (ipykernel)
- Input Cell:** In [21]: `x.head()`
- Output Cell:** Out[21]: A Pandas DataFrame displaying 5 rows of data with 23 columns. The columns are: Row\_ID, Order\_ID, Order\_Date, Ship\_Date, Ship\_Mode, Customer\_ID, Customer\_Name, Segment, City, State.

	Row_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	City	State
0	1	CA-2014-AB10015140-41954	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman	Consumer	Oklahoma City	Oklahoma
1	2	IN-2014-JR162107-41675	2014-02-05	2014-02-07	Second Class	JR-162107	Justin Ritter	Corporate	Wollongong	New South Wales
2	3	IN-2014-CR127307-41929	2014-10-17	2014-10-18	First Class	CR-127307	Craig Reiter	Consumer	Brisbane	Queensland
3	4	ES-2014-KM1637548-41667	2014-01-28	2014-01-30	First Class	KM-1637548	Katherine Murray	Home Office	Berlin	Berlin
4	5	SG-2014-RH9495111-41948	2014-11-05	2014-11-06	Same Day	RH-9495111	Rick Hansen	Consumer	Dakar	Dakar

5 rows × 23 columns

- Taskbar:** Shows various application icons including File Explorer, Task View, Google Chrome, Microsoft Edge, and Microsoft Word.
- System Tray:** Displays weather (25°C Haze), battery level (17:19), date (22-12-2022), and system status (ENG).

Project - Jupyter Notebook

localhost:8888/notebooks/Data\_Management/Project.ipynb

jupyter Project Last Checkpoint: Last Wednesday at 10:27 AM (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel) 0

In [14]: `x=df.drop(["State","Sales"],axis=1)`

In [15]: `x`

Out[15]:

Row_ID	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Customer_Name	Segment	City	State_Code	Country_code	Product_I
0	AB10015140-41954	2014-11-11	2014-11-13	First Class	AB-100151402	Aaron Bergman	Consumer	Oklahoma City	Ok ...	Un	TEC-Pf 581
1	JR162107-41675	2014-02-05	2014-02-07	Second Class	JR-162107	Justin Ritter	Corporate	Wollongong	Ne ...	Au	FUR-Cf 537
2	CR127307-41929	2014-10-17	2014-10-18	First Class	CR-127307	Craig Reiter	Consumer	Brisbane	Qu ...	Au	TEC-Pf 535
3	KM1637548-41667	2014-01-28	2014-01-30	First Class	KM-1637548	Katherine Murray	Home Office	Berlin	Be ...	Ge	TEC-Pf 526
4	RH9495111-41948	2014-11-05	2014-11-06	Same Day	RH-9495111	Rick Hansen	Consumer	Dakar	Da ...	Se	TEC-CC 601
...	...	...	...	...	...	...	...	...	...	...	...
994	SV10815126-41514	2013-09-24	2013-09-28	Standard Class	SV-10815126	Stuart Van	Corporate	Homs	Hi ...	Sy	TEC-Pf 380

Type here to search

27 25°C Sunny 11:23 ENG 23-12-2022

Project - Jupyter Notebook

localhost:8888/notebooks/Data\_Management/Project.ipynb

# jupyter Project (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) O

Code

5 rows x 23 columns

In [7]: `x.tail()`

Out[7]:

	Unnamed: 0	Unnamed: 1
0	NaN	NaN
1	Sum of Profit	NaN
2	Ship_Mode	Total
3	First Class	65930.77384
4	Same Day	20505.34216

Type here to search

25°C Sunny 11:26 23-12-2022

Project - Jupyter Notebook

localhost:8888/notebooks/Data\_Management/Project.ipynb

# jupyter Project (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) O

In [22]: `y=df['Sales']`

In [23]: `y`

Out[23]:

0	221.980
1	3709.395
2	5175.171
3	2892.510
4	2832.960
...	
994	1570.176
995	2614.689
996	69.480
997	636.780
998	2751.200

Name: Sales, Length: 999, dtype: float64

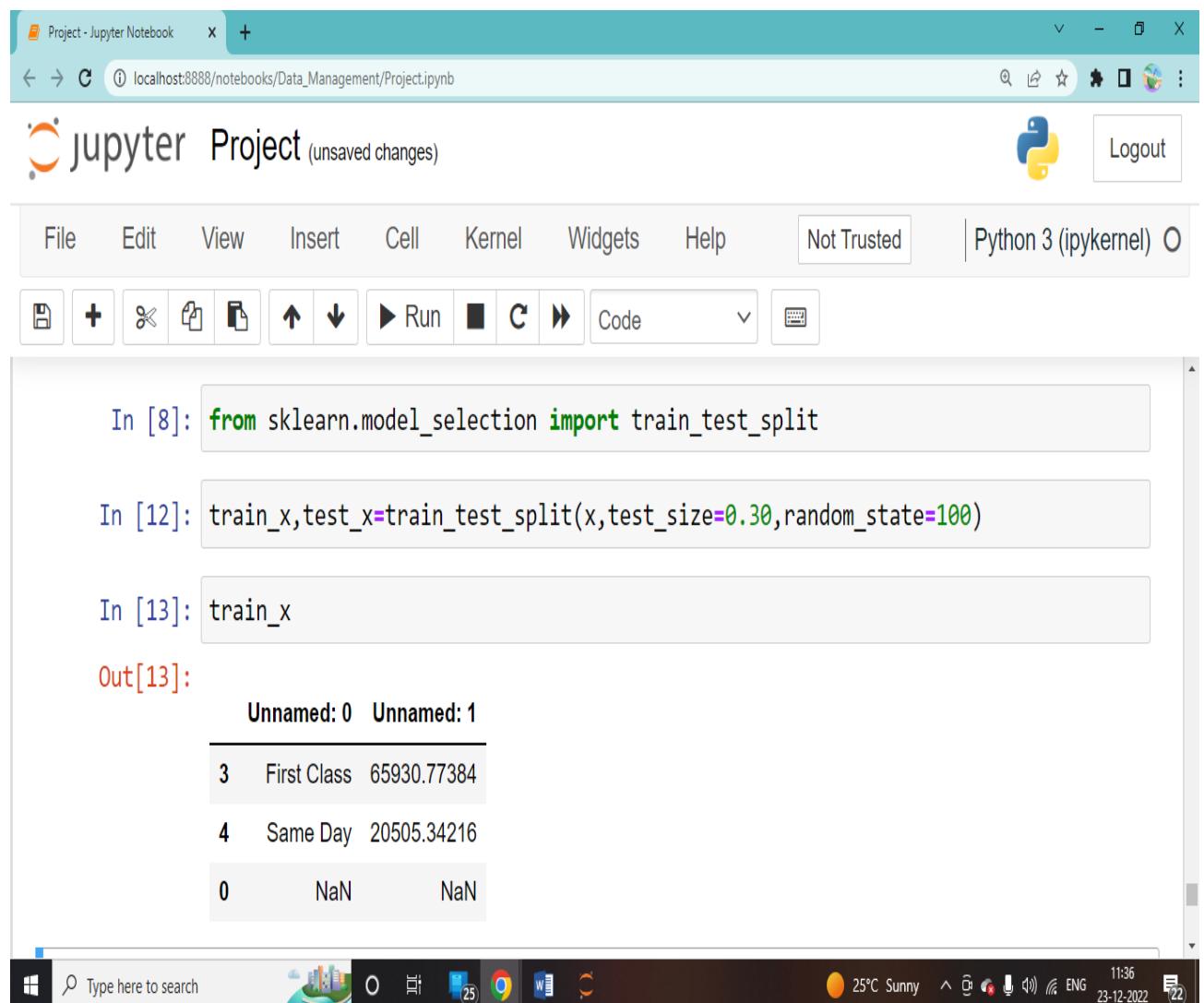
Type here to search

Market Brief 18 18 ENG 22-12-2022

# MODEL BUILDING

Using sklearn library we are splitting our data train and test inorder to build our model.

Then from sklearn package import linear regression with the help of linear model.



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

- Header:** Project - Jupyter Notebook, URL: localhost:8888/notebooks/Data\_Management/Project.ipynb, Logout button.
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Not Trusted, Python 3 (ipykernel).
- Buttons:** Save, New, Cell, Run, Stop, Kernel, Cell Type (Code), Cell Type (Text).
- In [8]:** `from sklearn.model_selection import train_test_split`
- In [12]:** `train_x,test_x=train_test_split(x,test_size=0.30,random_state=100)`
- In [13]:** `train_x`
- Out[13]:** Data frame output:

	Unnamed: 0	Unnamed: 1
3	First Class	65930.77384
4	Same Day	20505.34216
0	Nan	Nan
- System Tray:** Type here to search, Task View, Start button, Weather (25°C Sunny), Volume, Network, ENG, Date (23-12-2022), Battery (22%).

Project - Jupyter Notebook

localhost:8888/notebooks/Data\_Management/Project.ipynb

# jupyter Project (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) O

Out[13]:

	Unnamed: 0	Unnamed: 1
3	First Class	65930.77384
4	Same Day	20505.34216
0	NaN	NaN

In [14]: test\_x

Out[14]:

	Unnamed: 0	Unnamed: 1
1	Sum of Profit	NaN
2	Ship_Mode	Total

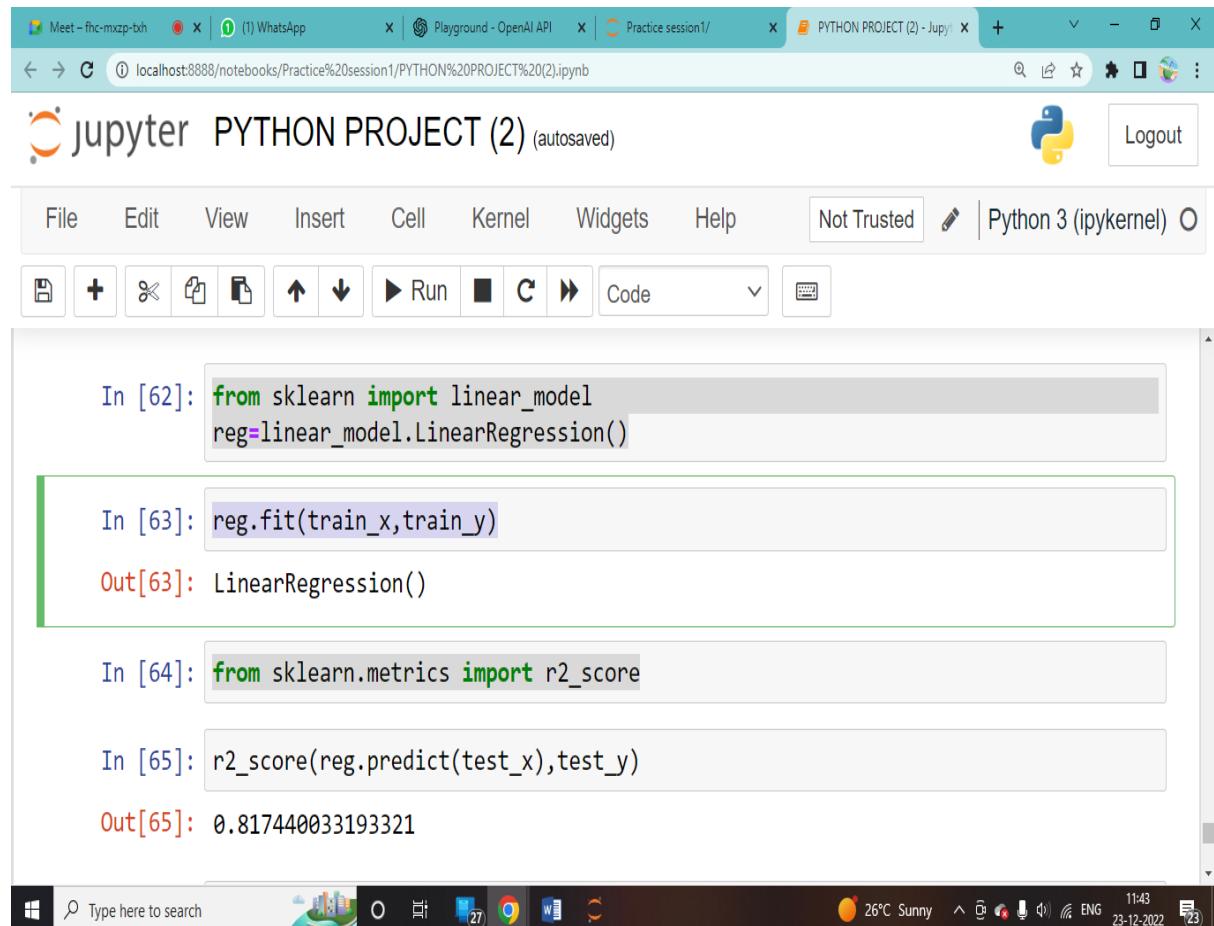
Type here to search

26°C Sunny 11:38  
ENG 23-12-2022

# MODEL PERFORMANCE

Using r2 metrices we checking model accuracy which is 82%.

Then we can predict by passing the values for future revenue prediction.



The screenshot shows a Jupyter Notebook interface running in a browser window titled "PYTHON PROJECT (2) - Jupyter". The notebook has several cells:

- In [62]: `from sklearn import linear_model  
reg=linear_model.LinearRegression()`
- In [63]: `reg.fit(train_x,train_y)`  
Out[63]: `LinearRegression()`
- In [64]: `from sklearn.metrics import r2_score`
- In [65]: `r2_score(reg.predict(test_x),test_y)`  
Out[65]: `0.817440033193321`

The browser tabs at the top include "Meet - fhc-mxzp-bh", "(1) WhatsApp", "Playground - OpenAI API", "Practice session1/", and "PYTHON PROJECT (2) - Jupyter". The system tray at the bottom shows a Windows taskbar with icons for search, file explorer, and various applications, along with system status like battery level (11:43), temperature (26°C), and language (ENG).

## TABLEAU

In order to visualize data to draw the important insights from it regarding potential business opportunities and target areas so that we can take important decisions such as:

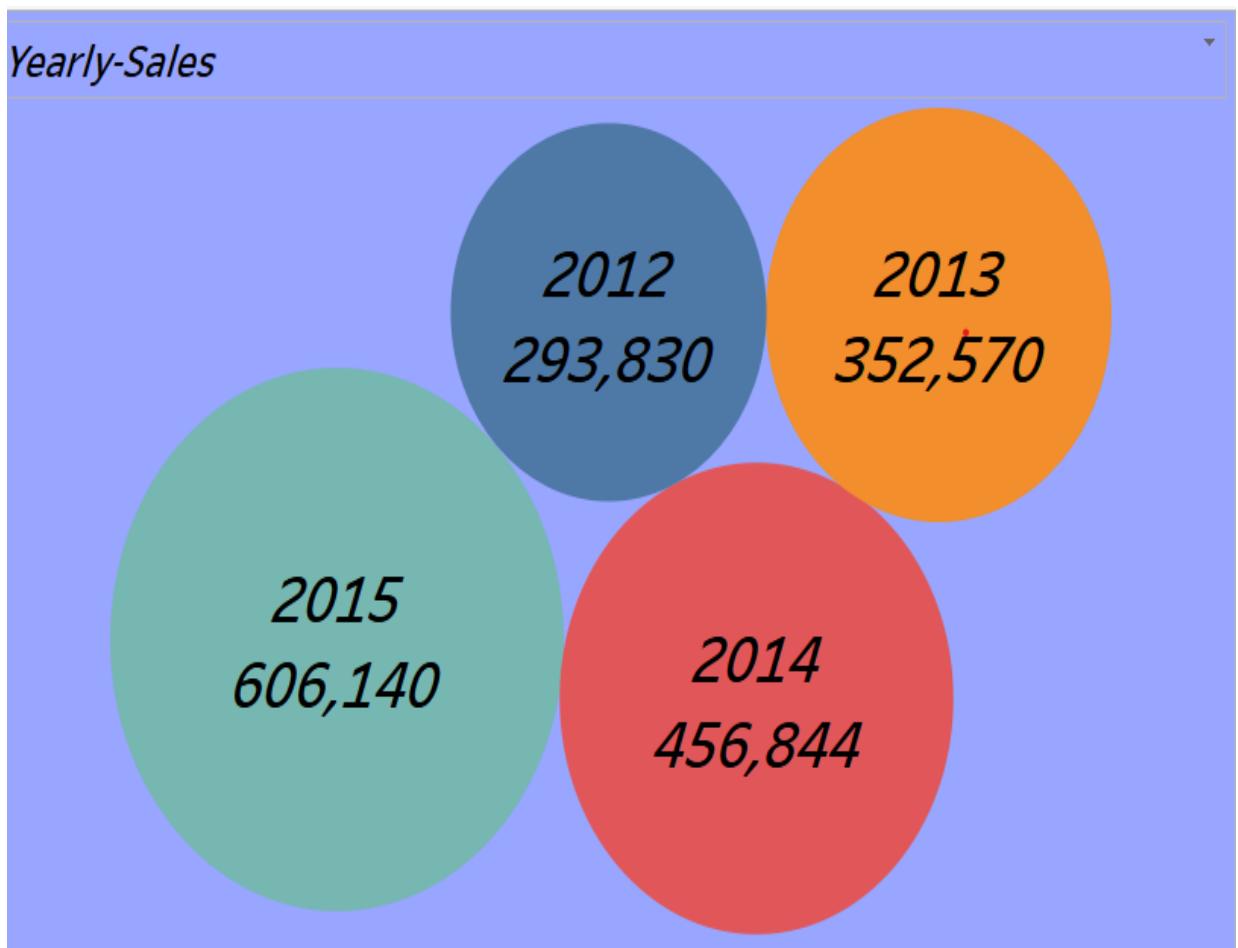
- Identify those places or areas where it can set up centers.
- Modify or increase the products and sales portfolio.
- Identify important areas for investment and come up with appropriate marketing strategies.

Tableau can be used to connect to any type of file. In this case we would be connecting to any excel file and our analysis would be based on historical data rather than live data.

Here we are highlighting KPI (Key Performance Indicator) i.e. Total Profit, Total Revenue and Total Sales for the entire Period.

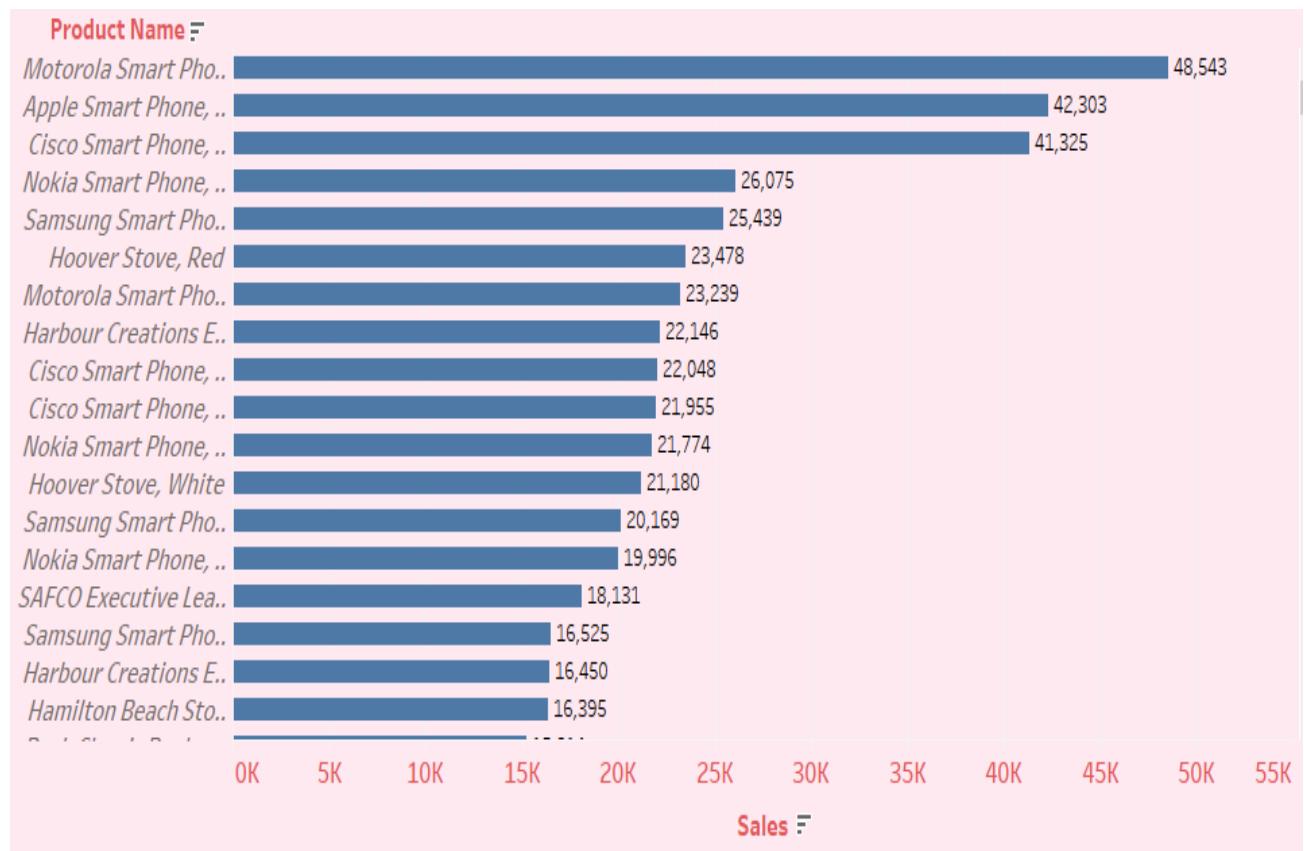
We are plotting Year vs Total Sales generated in each Year using Packed Bubbles.

The Lowest sales generated is in the year 2012:293,830 and the Highest sales generated is in the year 2015:606,140.

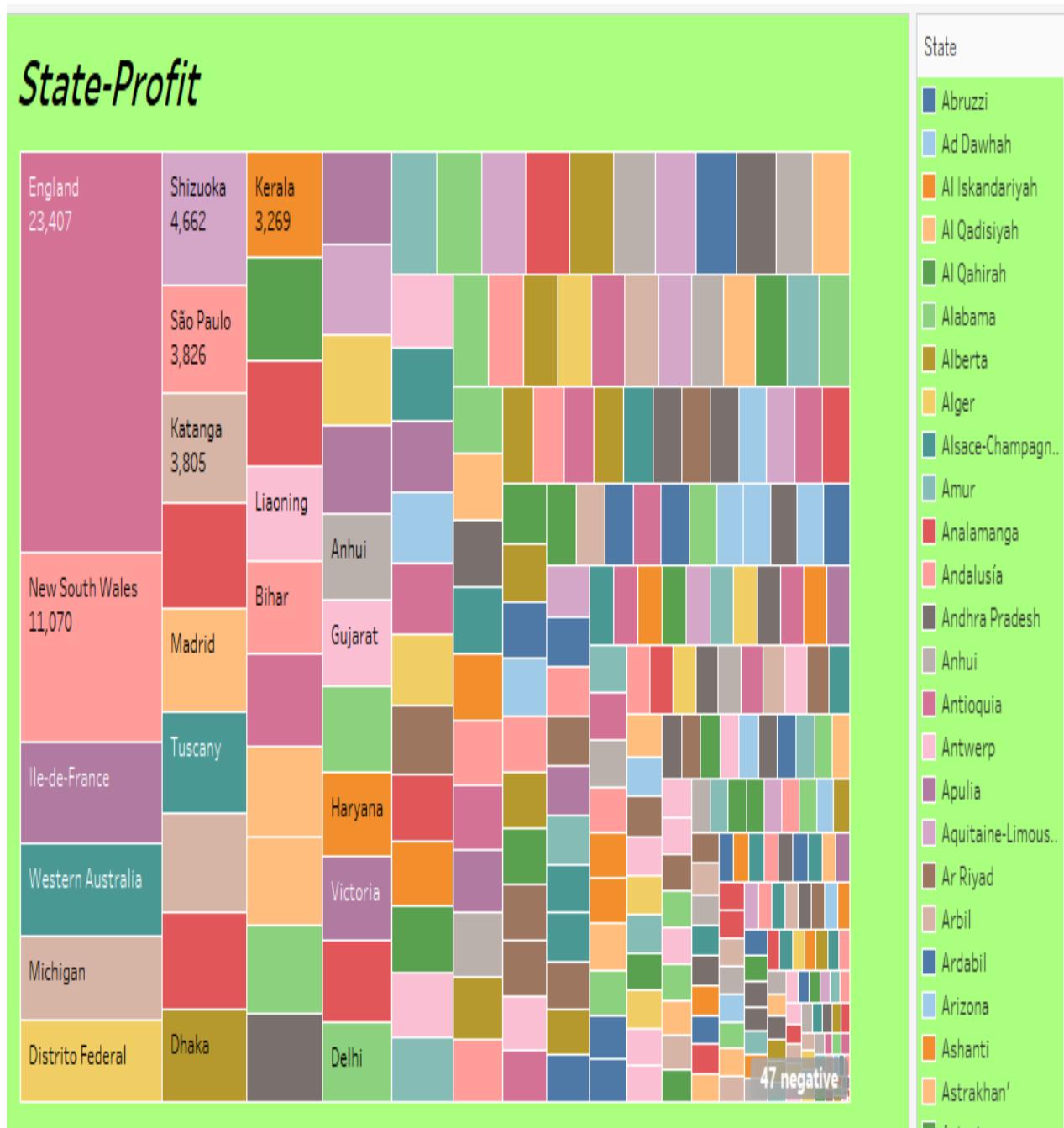


Here, we are showing Product vs Total Sales of each Product using Horizontal Bar.

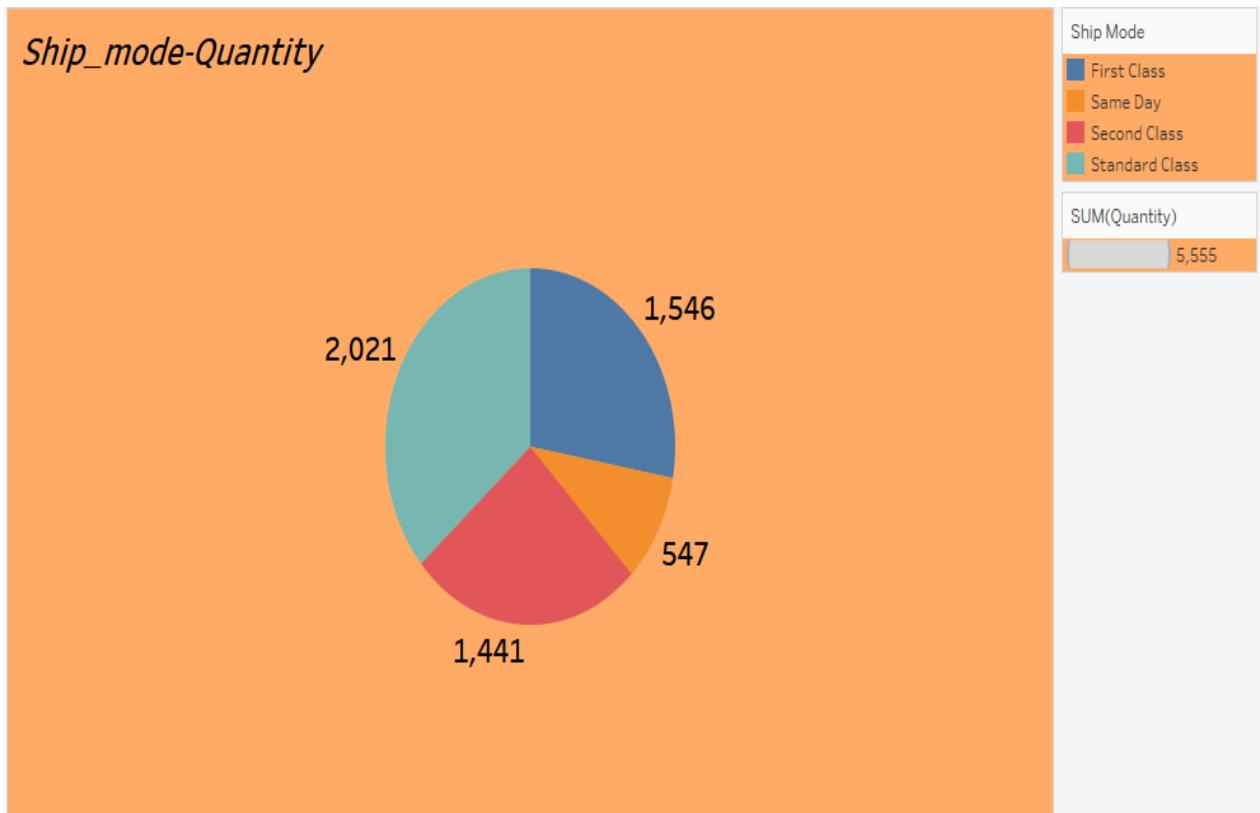
The Lowest Sales for Product Motorola Smart Phone is 48,543 and the Highest Sales for Product Insertable Tab Index is 2



Here, we are using tree map to show State vs Profit. We can easily identify the profit that is generated for each state.

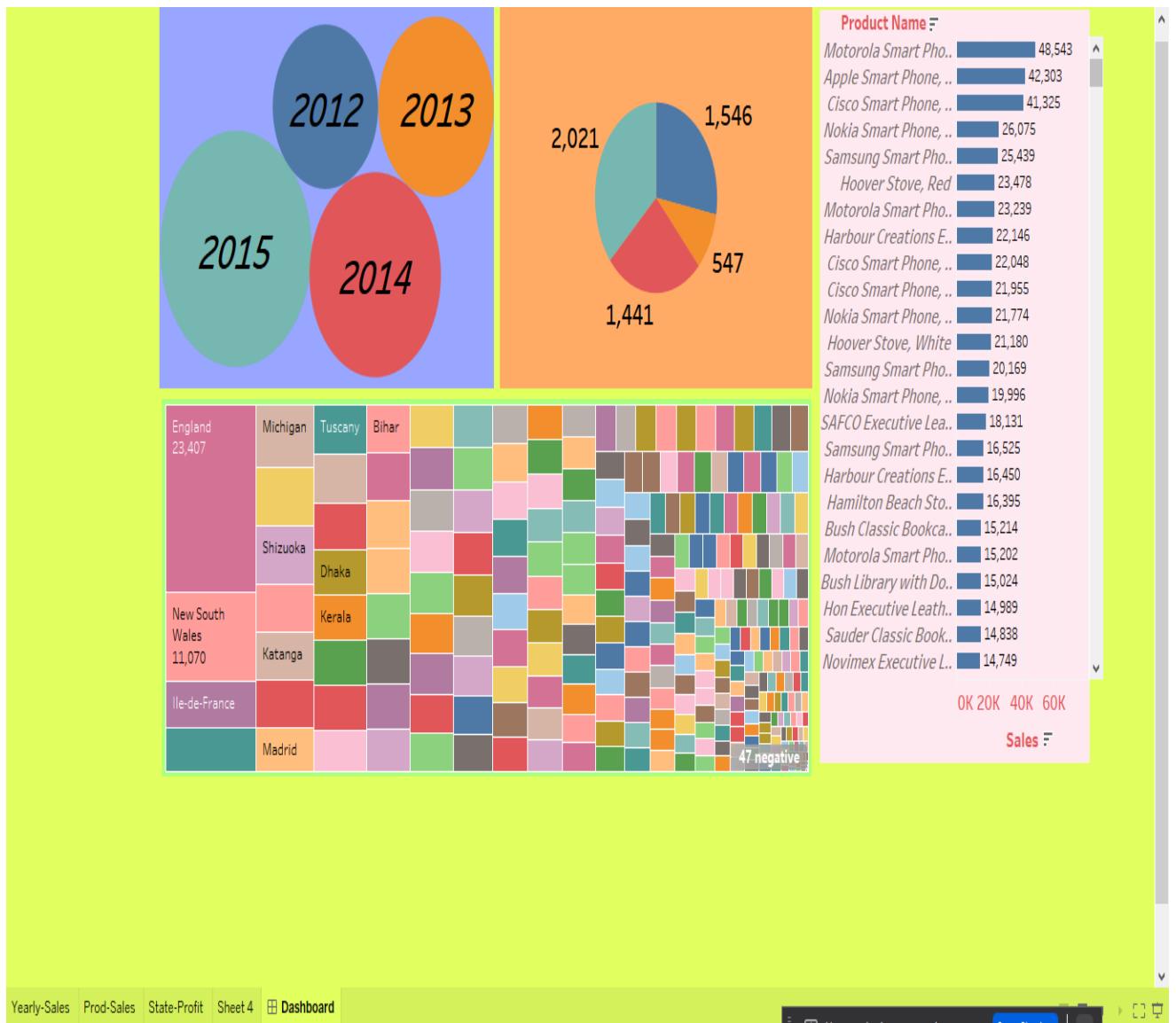


Here, in the Pie-Chart we are representing the Quantity of Product that is Shipped.

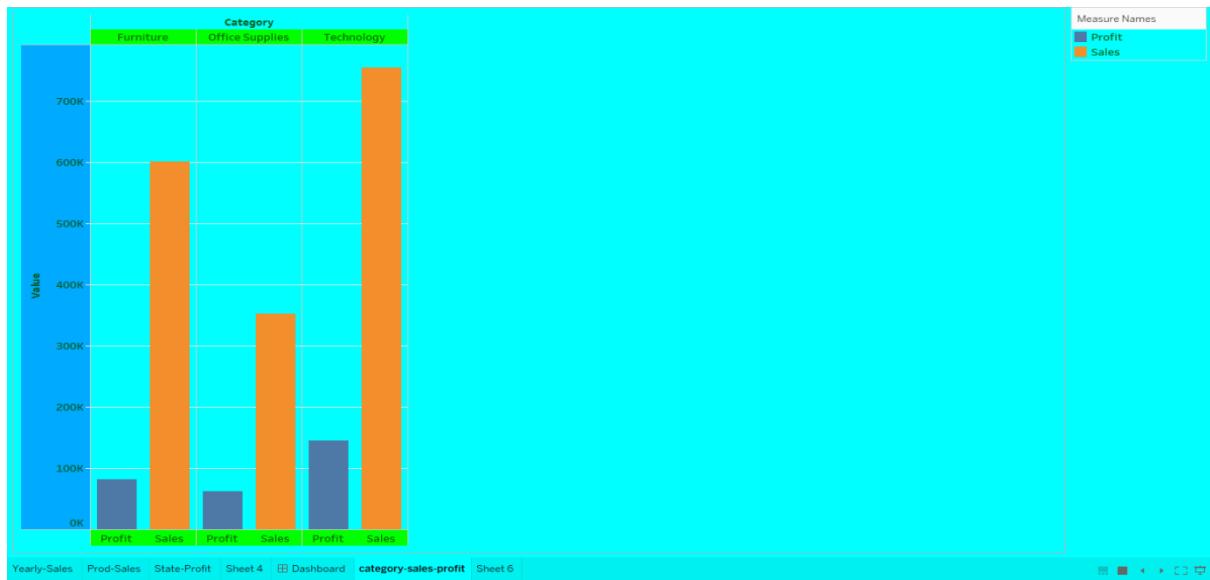


# DASHBOARD

- It is to show a comprehensive overview of datas from different sources that we have taken.
- It is to monitor, measure and analyze the relevant datas in key areas.



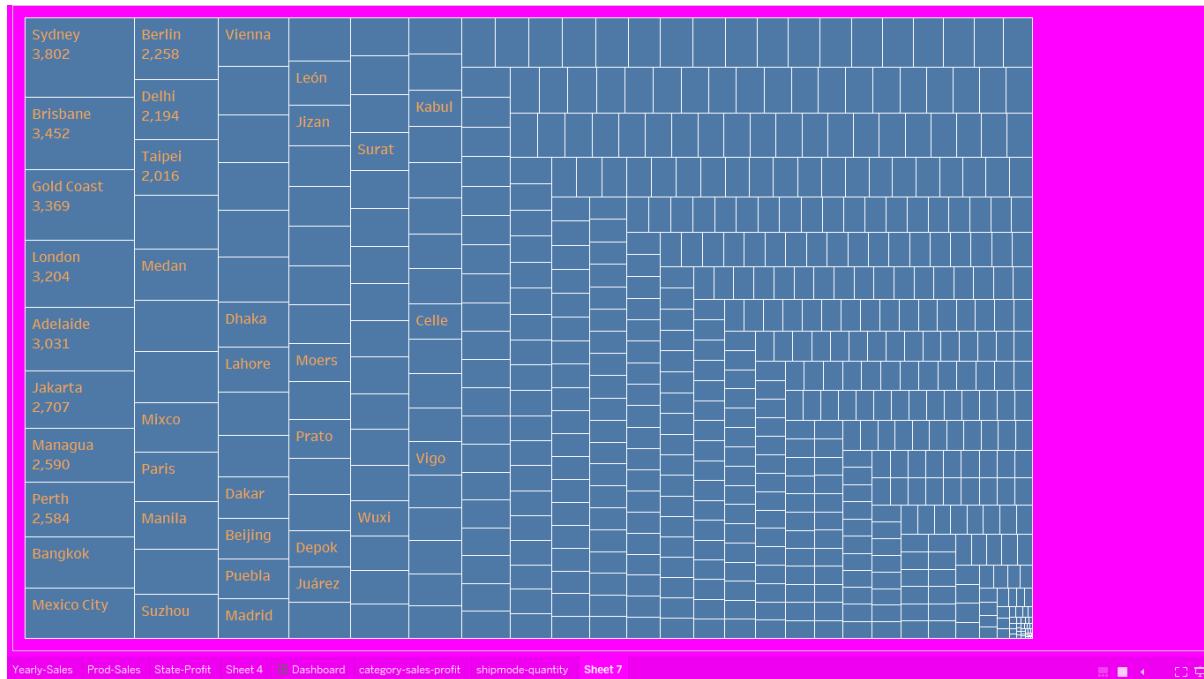
Here, the bar graph represents category wise sales and profit.



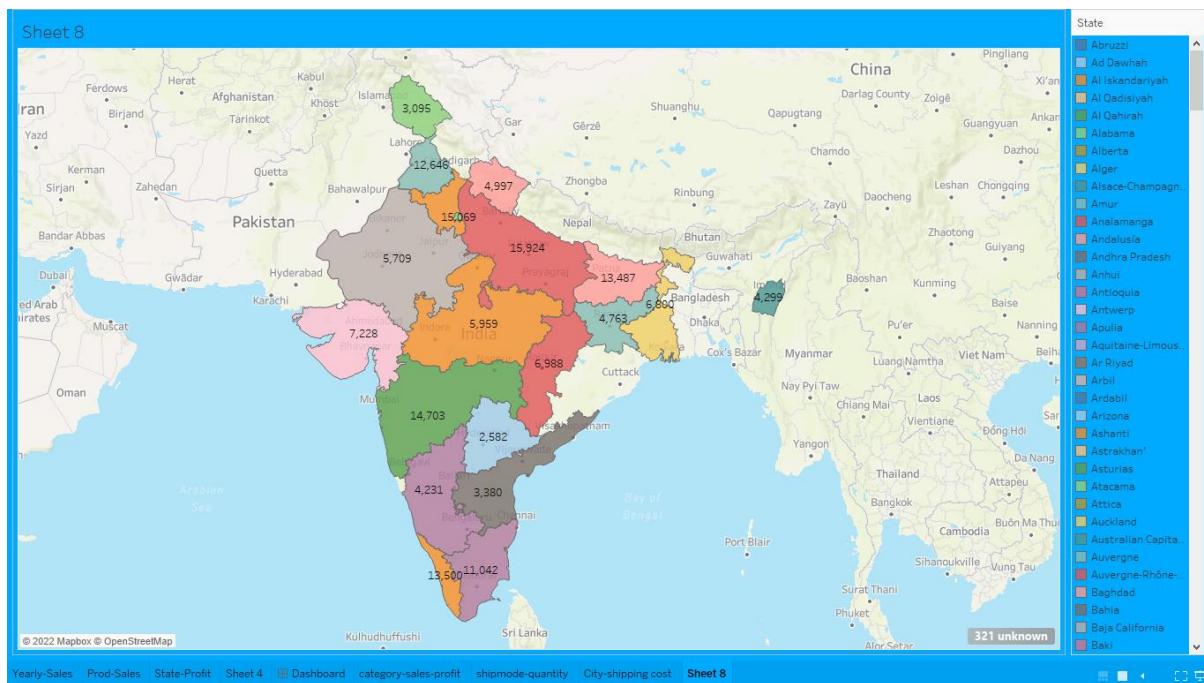
Here, the bar graph represents ship mode wise Quantity.



# Here, the tree map represents City wise Shipping cost.

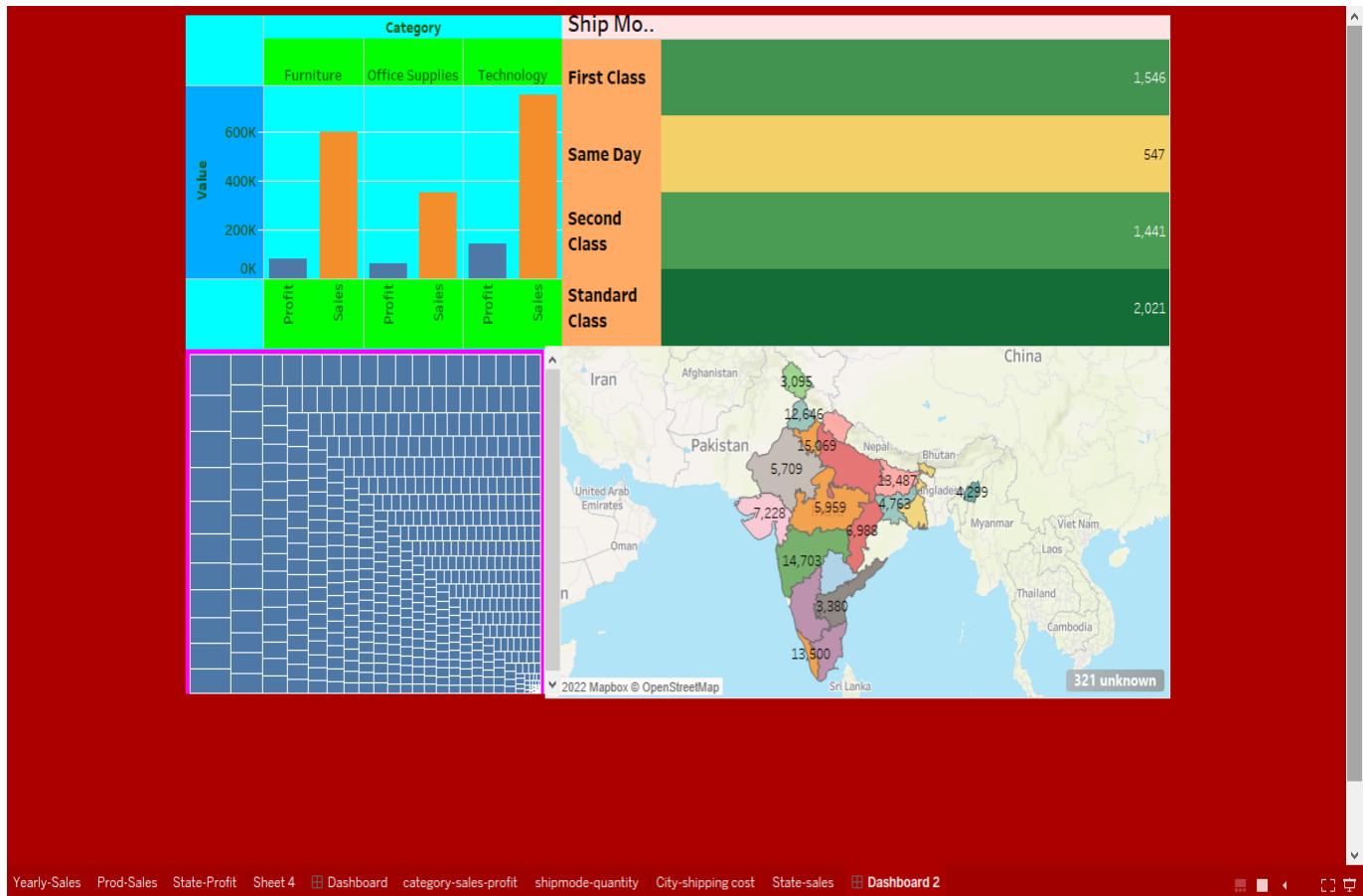


# Here, the bar graph represents state wise Sales.



# DASHBOARD

- It is to show a comprehensive overview of datas from different sources that we have taken.
- It is to monitor, measure and analyze the relevant datas in key areas.



## **CONCLUSION**

- Data Analytic tools are Powerful Tools because they can help a company to make better-informed decisions.
- The visualization tools helped to identify the target areas for the business.
- The company should focus on investment in the target areas and should also acquire the right resources for its business.
- As shown by the Data Visualization tools offering attractive discounts during the peak months will help to attract the potential customers which can further help to enhance revenue and reach in the market.