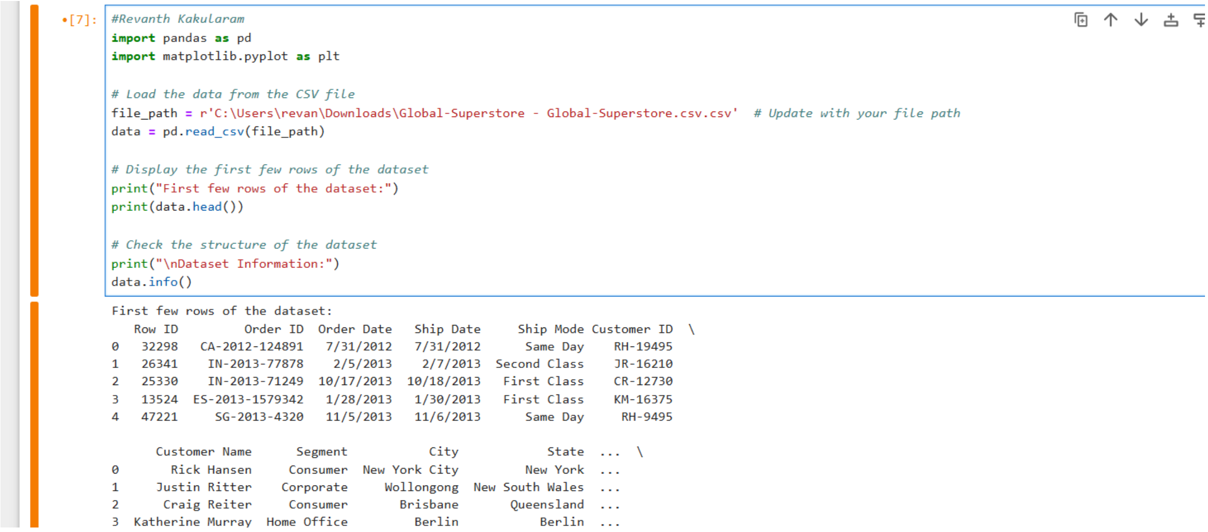
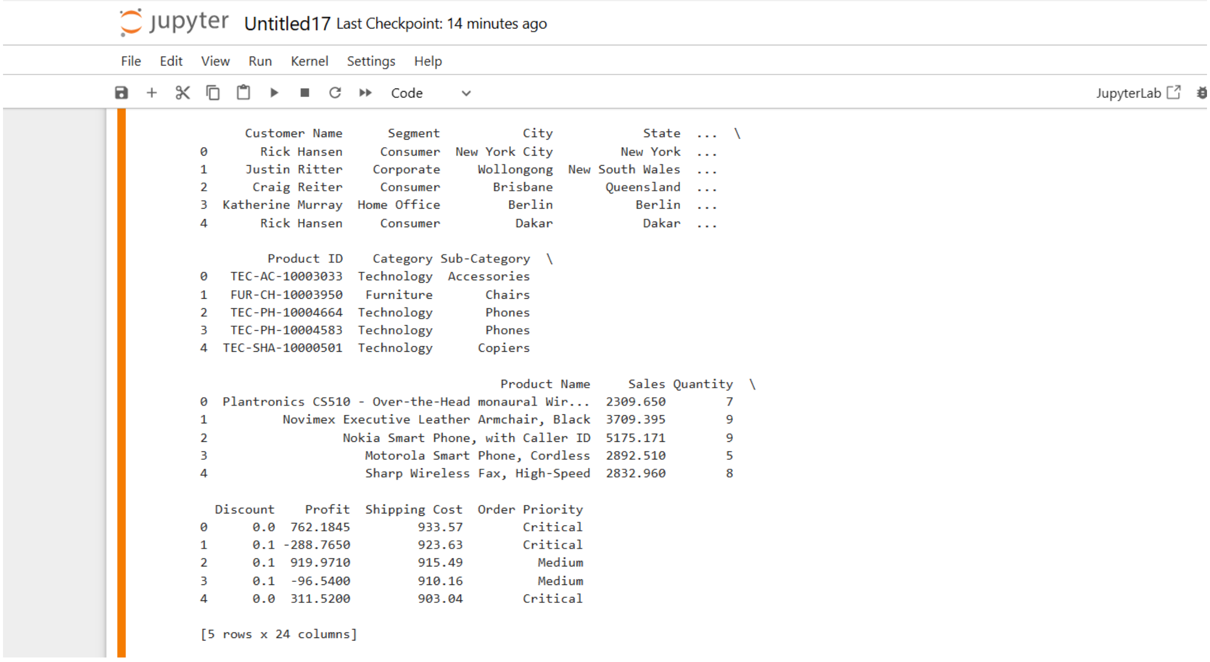
Name : Revanth Kakularam

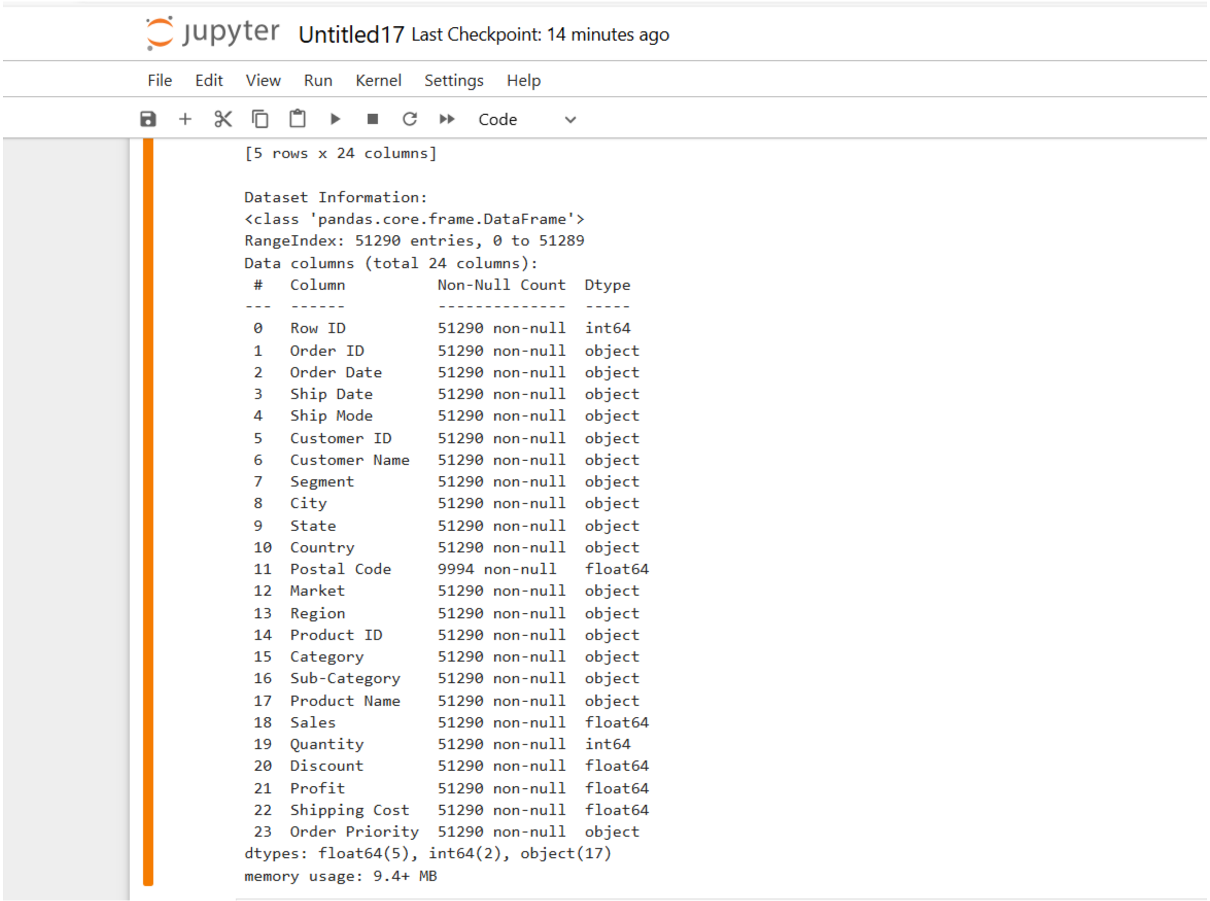


```python
#Revanth Kakularam
import pandas as pd
import matplotlib.pyplot as plt

# Load the data from the CSV file
file_path = r'C:\Users\revan\Downloads\Global-Superstore - Global-Superstore.csv.csv'  # Update with your file path
data = pd.read_csv(file_path)

# Display the first few rows of the dataset
print("First few rows of the dataset:")
print(data.head())

# Check the structure of the dataset
print("\nDataset Information:")
data.info()
```

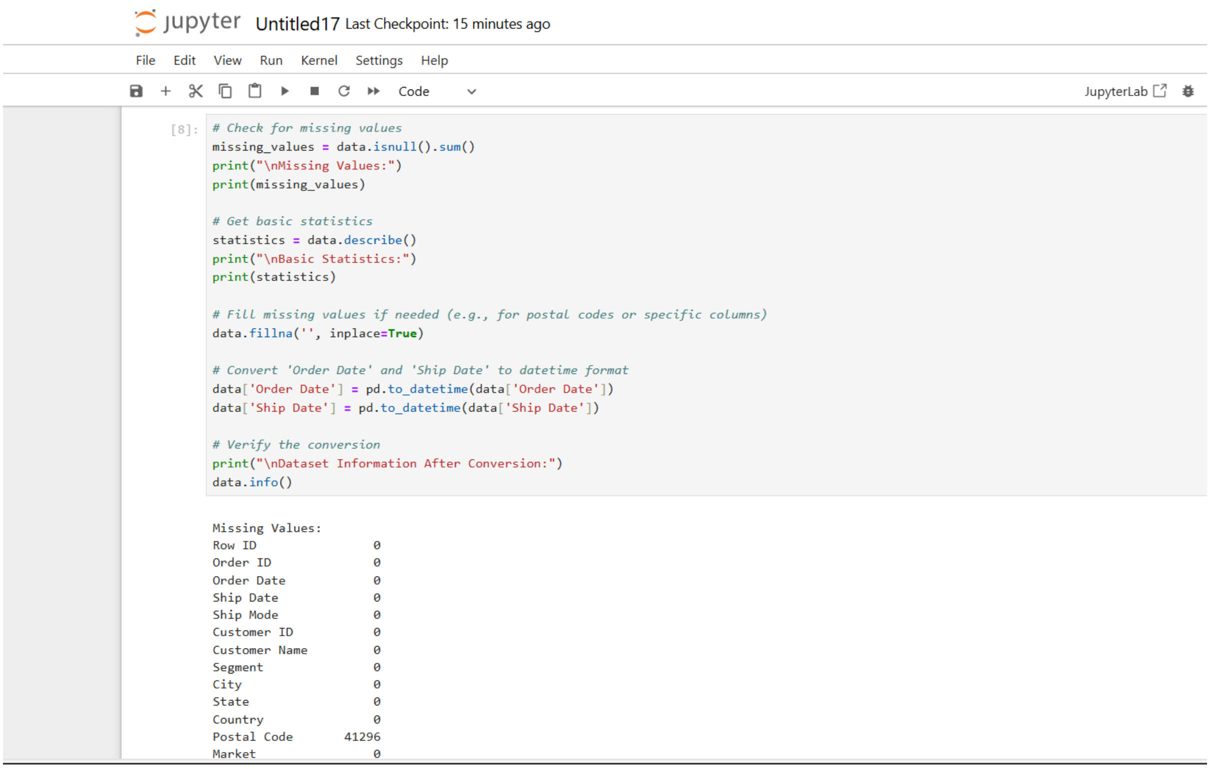
```
First few rows of the dataset:
   Row ID         Order ID  Order Date   Ship Date     Ship Mode Customer ID  \
0   32298   CA-2012-124891   7/31/2012   7/31/2012      Same Day    RH-19495   
1   26341    IN-2013-77878    2/5/2013    2/7/2013  Second Class    JR-16210   
2   25330    IN-2013-71249  10/17/2013  10/18/2013   First Class    CR-12730   
3   13524  ES-2013-1579342   1/28/2013   1/30/2013   First Class    KM-16375   
4   47221     SG-2013-4320    11/5/2013   11/6/2013      Same Day     RH-9495   

      Customer Name      Segment           City            State  ...  \
0       Rick Hansen     Consumer  New York City         New York  ...   
1     Justin Ritter    Corporate     Wollongong  New South Wales  ...   
2      Craig Reiter     Consumer       Brisbane       Queensland  ...   
3  Katherine Murray  Home Office         Berlin           Berlin  ...   
```



```
      Customer Name      Segment           City            State  ...  \
0       Rick Hansen     Consumer  New York City         New York  ...   
1     Justin Ritter    Corporate     Wollongong  New South Wales  ...   
2      Craig Reiter     Consumer       Brisbane       Queensland  ...   
3  Katherine Murray  Home Office         Berlin           Berlin  ...   
4       Rick Hansen     Consumer          Dakar            Dakar  ...   

        Product ID    Category Sub-Category  \
0   TEC-AC-10003033  Technology  Accessories   
1   FUR-CH-10003950   Furniture       Chairs   
2   TEC-PH-10004664  Technology       Phones   
3   TEC-PH-10004583  Technology       Phones   
4  TEC-SHA-10000501  Technology      Copiers   

                                      Product Name     Sales Quantity  \
0  Plantronics CS510 - Over-the-Head monaural Wir...  2309.650         7   
1          Novimex Executive Leather Armchair, Black  3709.395         9   
2              Nokia Smart Phone, with Caller ID  5175.171         9   
3               Motorola Smart Phone, Cordless  2892.510         5   
4               Sharp Wireless Fax, High-Speed  2832.960         8   

  Discount     Profit  Shipping Cost Order Priority  
0      0.0   762.1845         933.57       Critical  
1      0.1  -288.7650         923.63       Critical  
2      0.1   919.9710         915.49         Medium  
3      0.1   -96.5400         910.16         Medium  
4      0.0   311.5200         903.04       Critical  

[5 rows x 24 columns]
```

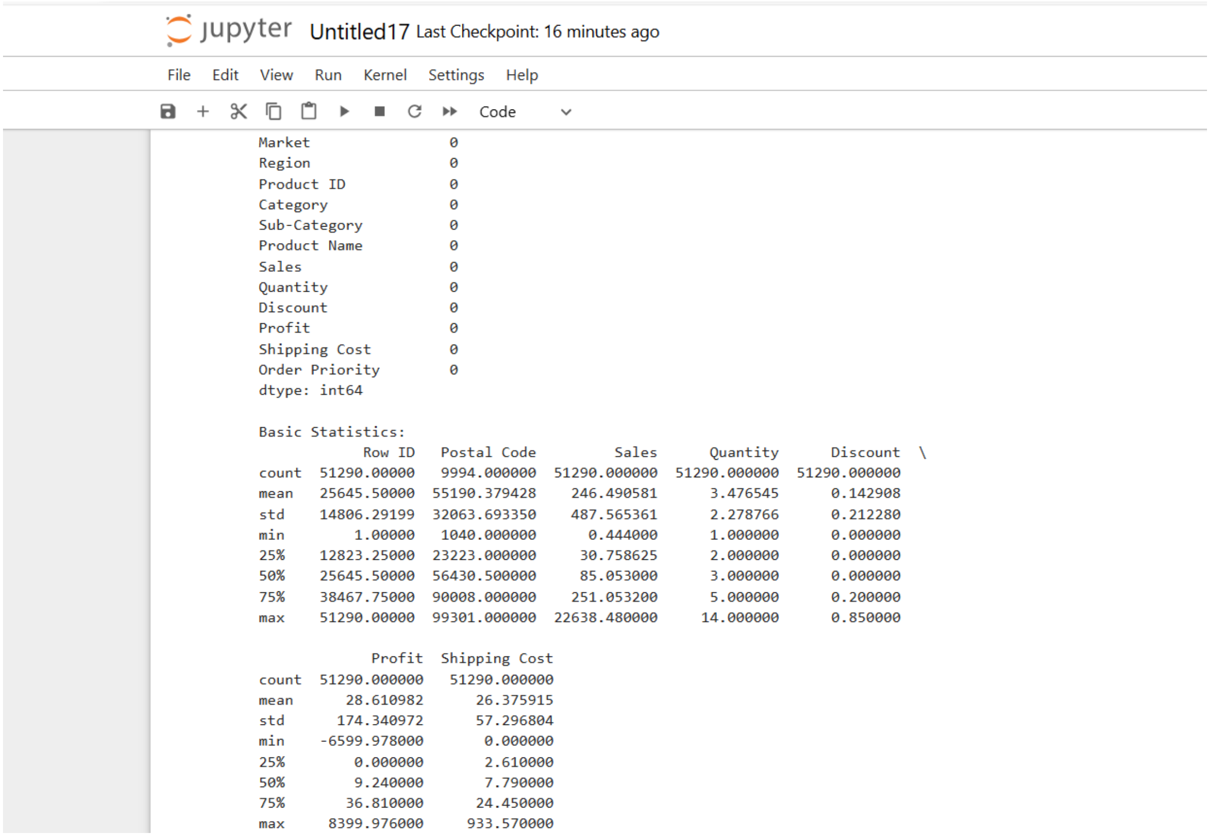

jupyter Untitled17 Last Checkpoint: 14 minutes ago
File Edit View Run Kernel Settings Help
Code

```
[5 rows x 24 columns]

Dataset Information:
<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
```

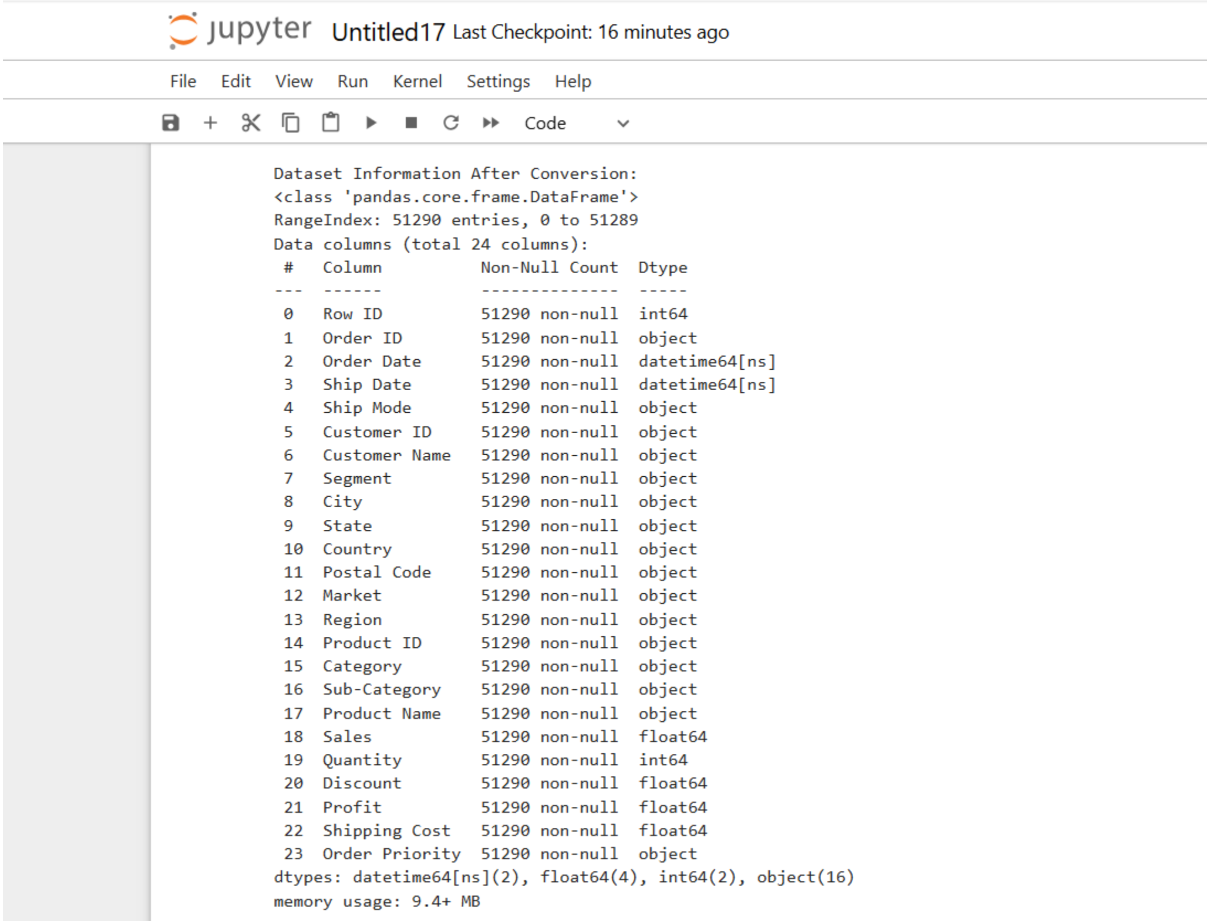

jupyter Untitled17 Last Checkpoint: 15 minutes ago
File Edit View Run Kernel Settings Help
Code
JupyterLab
[8]:

```python
# Check for missing values
missing_values = data.isnull().sum()
print("\nMissing Values:")
print(missing_values)

# Get basic statistics
statistics = data.describe()
print("\nBasic Statistics:")
print(statistics)

# Fill missing values if needed (e.g., for postal codes or specific columns)
data.fillna('', inplace=True)

# Convert 'Order Date' and 'Ship Date' to datetime format
data['Order Date'] = pd.to_datetime(data['Order Date'])
data['Ship Date'] = pd.to_datetime(data['Ship Date'])

# Verify the conversion
print("\nDataset Information After Conversion:")
data.info()
```

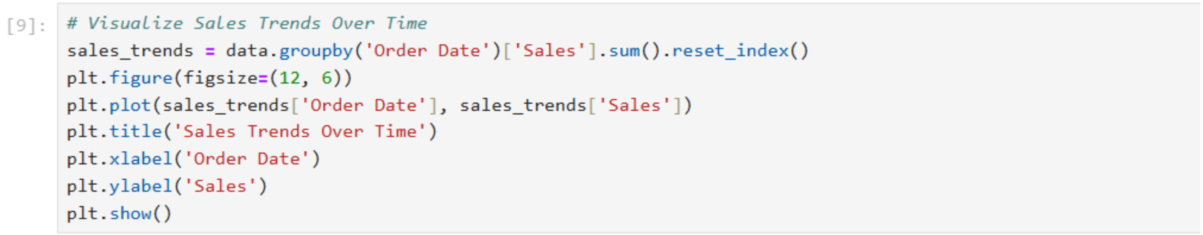
```
Missing Values:
Row ID                0
Order ID              0
Order Date            0
Ship Date             0
Ship Mode             0
Customer ID           0
Customer Name         0
Segment               0
City                  0
State                 0
Country               0
Postal Code       41296
Market                0
```


jupyter
Untitled17 Last Checkpoint: 16 minutes ago
File Edit View Run Kernel Settings Help
Code

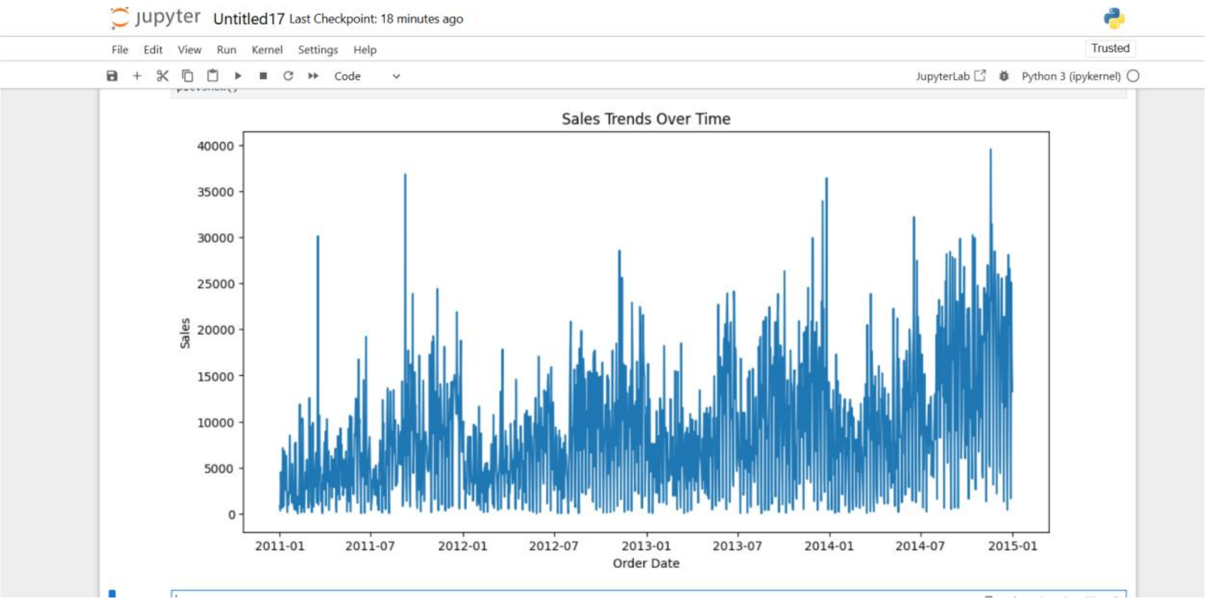
```
Market              0
Region              0
Product ID          0
Category            0
Sub-Category        0
Product Name        0
Sales               0
Quantity            0
Discount            0
Profit              0
Shipping Cost       0
Order Priority      0
dtype: int64

Basic Statistics:
            Row ID   Postal Code         Sales      Quantity      Discount  \
count  51290.00000   9994.000000  51290.000000  51290.000000  51290.000000
mean   25645.50000  55190.379428    246.490581      3.476545      0.142908
std    14806.29199  32063.693350    487.565361      2.278766      0.212280
min        1.00000   1040.000000      0.444000      1.000000      0.000000
25%    12823.25000  23223.000000     30.758625      2.000000      0.000000
50%    25645.50000  56430.500000     85.053000      3.000000      0.000000
75%    38467.75000  90008.000000    251.053200      5.000000      0.200000
max    51290.00000  99301.000000  22638.480000     14.000000      0.850000

             Profit  Shipping Cost
count  51290.000000   51290.000000
mean      28.610982      26.375915
std      174.340972      57.296804
min    -6599.978000       0.000000
25%        0.000000       2.610000
50%        9.240000       7.790000
75%       36.810000      24.450000
max     8399.976000     933.570000
```

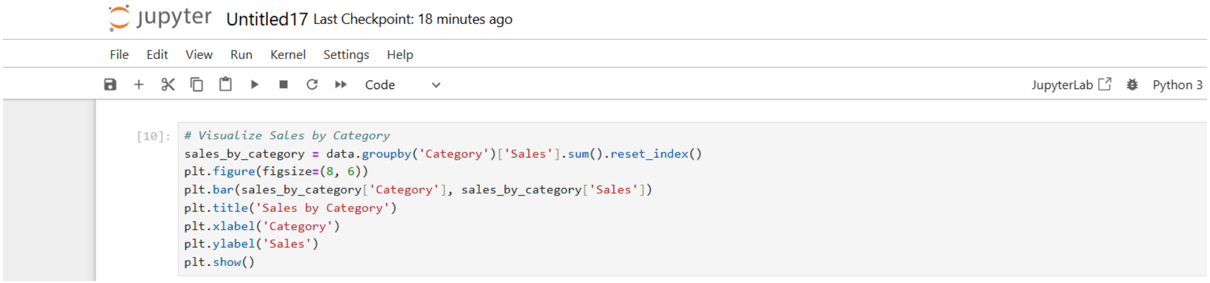

jupyter
Untitled17 Last Checkpoint: 16 minutes ago
File Edit View Run Kernel Settings Help
Code

```
Dataset Information After Conversion:
<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  datetime64[ns]
 3   Ship Date       51290 non-null  datetime64[ns]
 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     51290 non-null  object
 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: datetime64[ns](2), float64(4), int64(2), object(16)
memory usage: 9.4+ MB
```

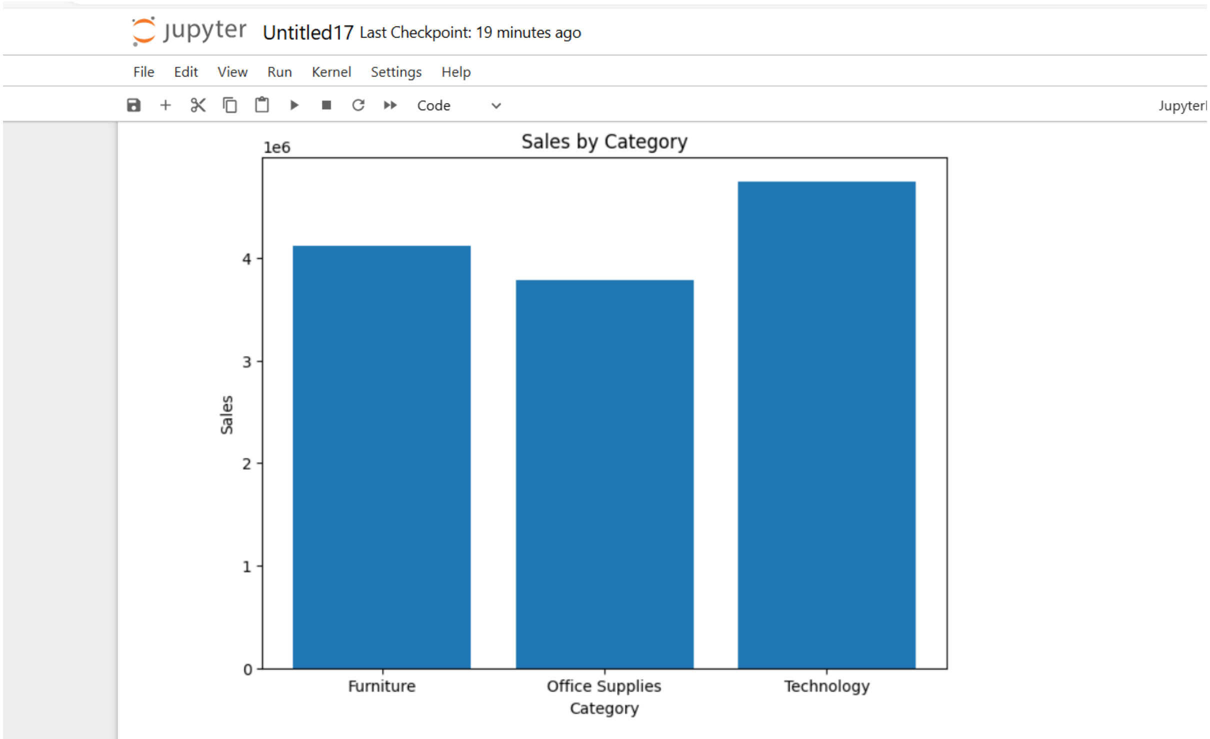


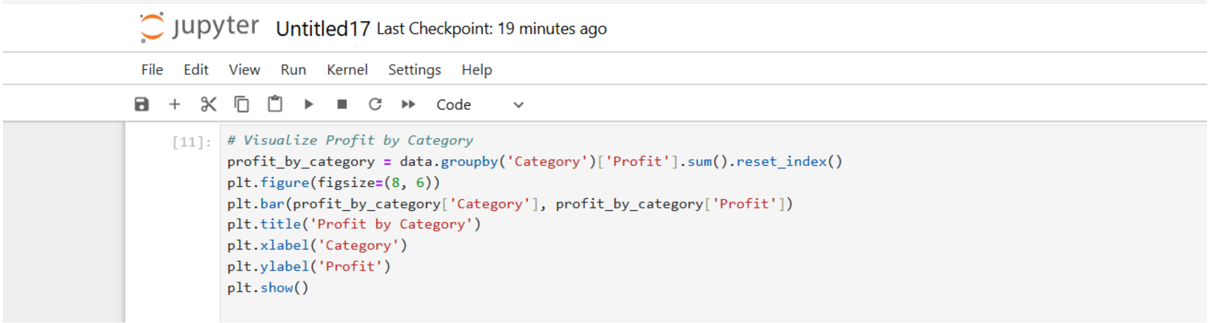
```
[9]: # Visualize Sales Trends Over Time
     sales_trends = data.groupby('Order Date')['Sales'].sum().reset_index()
     plt.figure(figsize=(12, 6))
     plt.plot(sales_trends['Order Date'], sales_trends['Sales'])
     plt.title('Sales Trends Over Time')
     plt.xlabel('Order Date')
     plt.ylabel('Sales')
     plt.show()
```



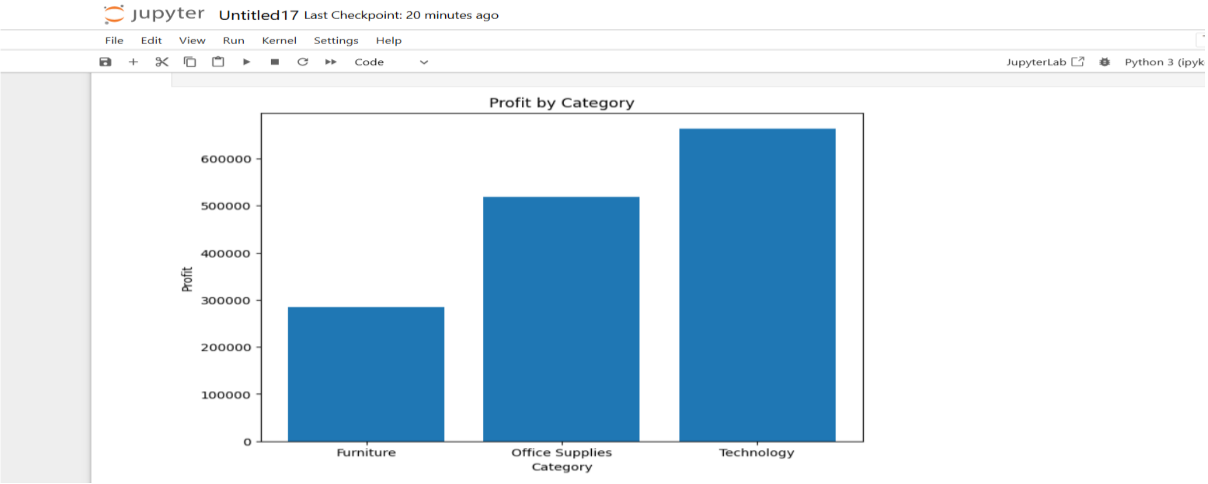


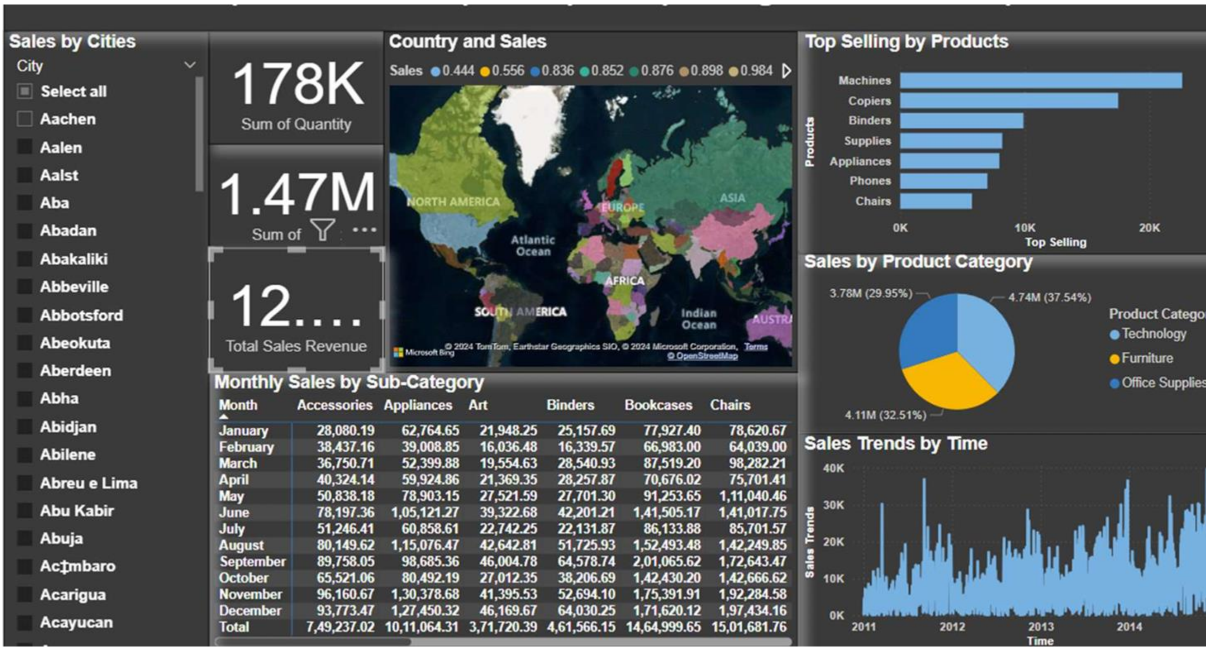
```
[10]: # Visualize Sales by Category
      sales_by_category = data.groupby('Category')['Sales'].sum().reset_index()
      plt.figure(figsize=(8, 6))
      plt.bar(sales_by_category['Category'], sales_by_category['Sales'])
      plt.title('Sales by Category')
      plt.xlabel('Category')
      plt.ylabel('Sales')
      plt.show()
```





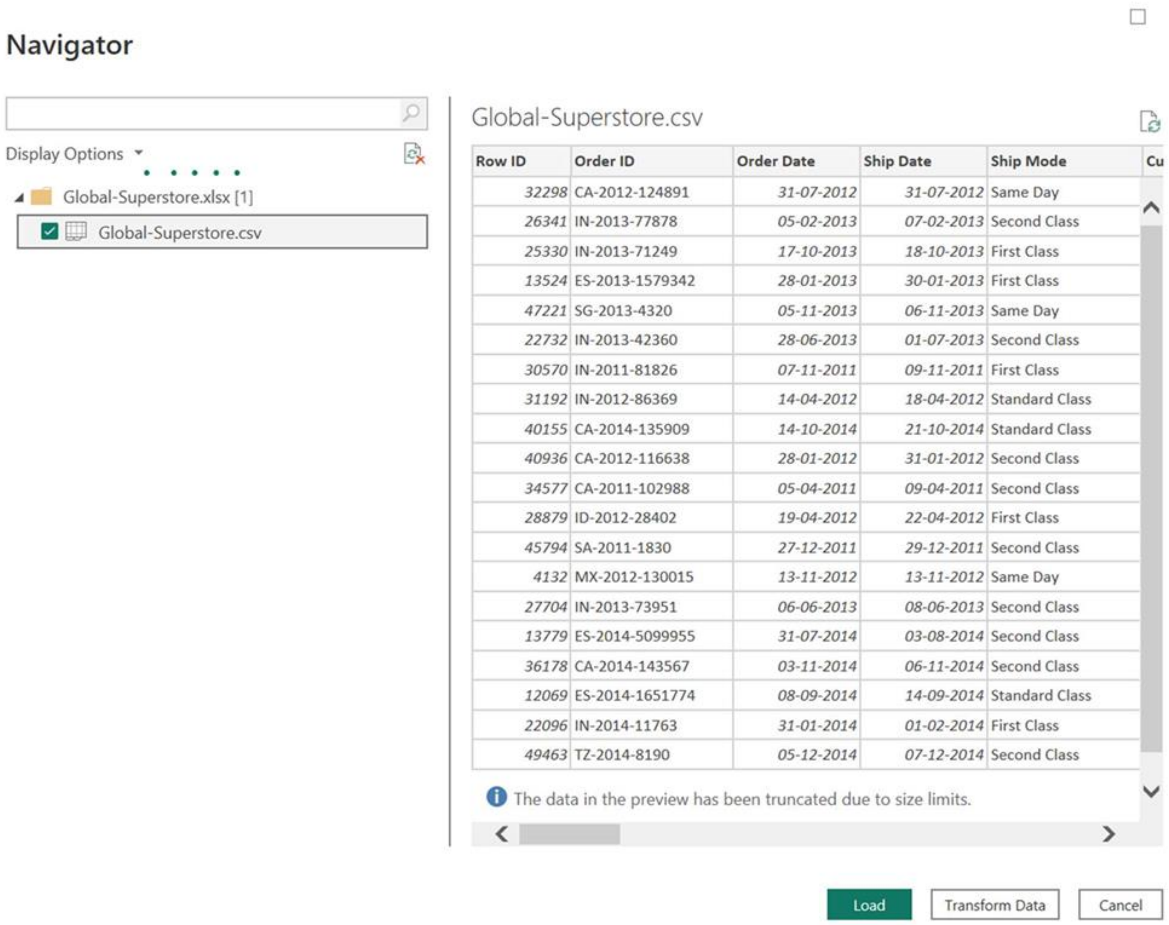
```
# Visualize Profit by Category
profit_by_category = data.groupby('Category')['Profit'].sum().reset_index()
plt.figure(figsize=(8, 6))
plt.bar(profit_by_category['Category'], profit_by_category['Profit'])
plt.title('Profit by Category')
plt.xlabel('Category')
plt.ylabel('Profit')
plt.show()
```





Monthly Sales by Sub-Category

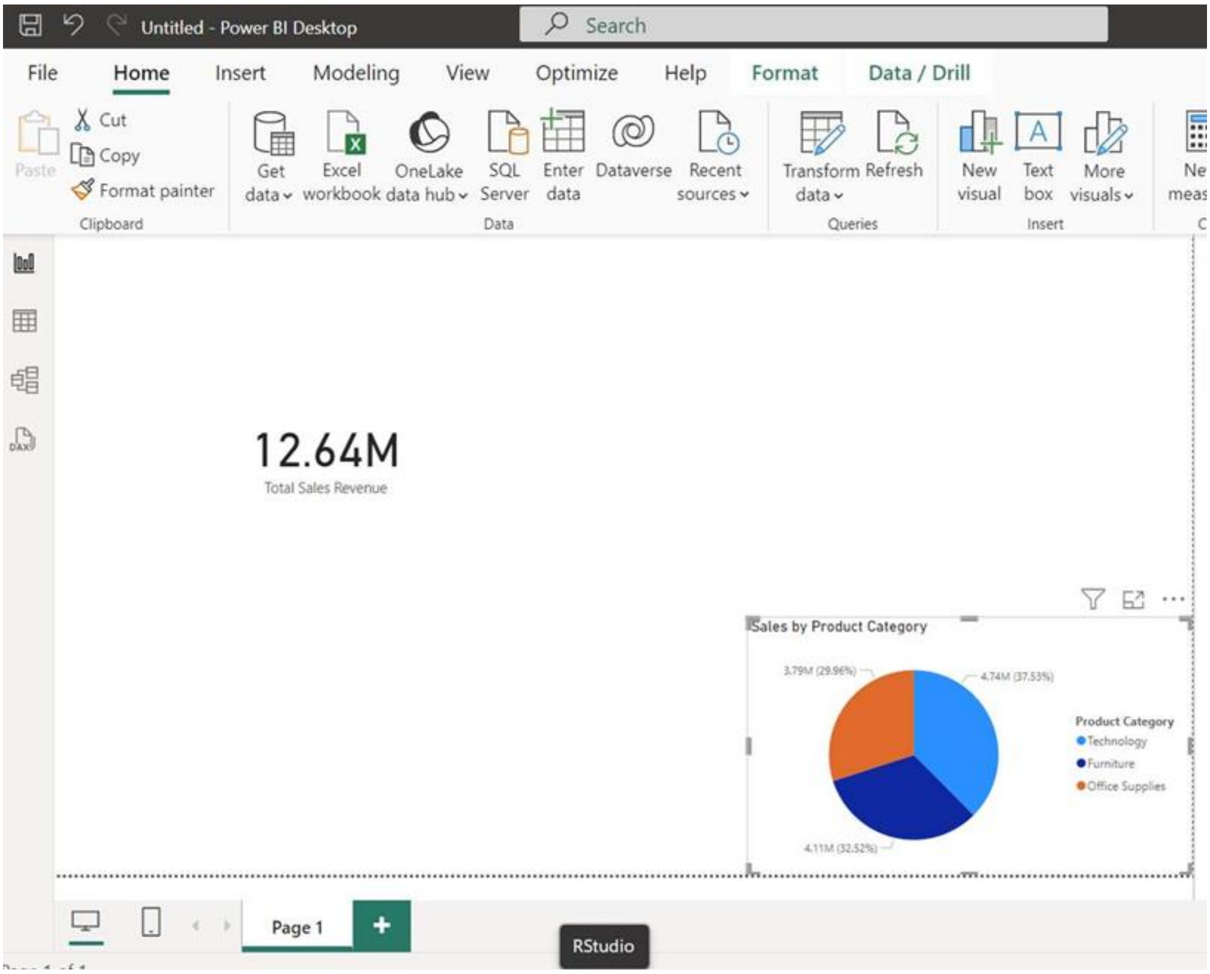
Month	Accessories	Appliances	Art	Binders	Bookcases	Chairs
January	28,080.19	62,764.65	21,948.25	25,157.69	77,927.40	78,620.67
February	38,437.16	39,008.85	16,036.48	16,339.57	66,983.00	64,039.00
March	36,750.71	52,399.88	19,554.63	28,540.93	87,519.20	98,282.21
April	40,324.14	59,924.86	21,369.35	28,257.87	70,676.02	75,701.41
May	50,838.18	78,903.15	27,521.59	27,701.30	91,253.65	1,11,040.46
June	78,197.36	1,05,121.27	39,322.68	42,201.21	1,41,505.17	1,41,017.75
July	51,246.41	60,858.61	22,742.25	22,131.87	86,133.88	85,701.57
August	80,149.62	1,15,076.47	42,642.81	51,725.93	1,52,493.48	1,42,249.85
September	89,758.05	98,685.36	46,004.78	64,578.74	2,01,065.62	1,72,643.47
October	65,521.06	80,492.19	27,012.35	38,206.69	1,42,430.20	1,42,666.62
November	96,160.67	1,30,378.68	41,395.53	52,694.10	1,75,391.91	1,92,284.58
December	93,773.47	1,27,450.32	46,169.67	64,030.25	1,71,620.12	1,97,434.16
Total	7,49,237.02	10,11,064.31	3,71,720.39	4,61,566.15	14,64,999.65	15,01,681.76

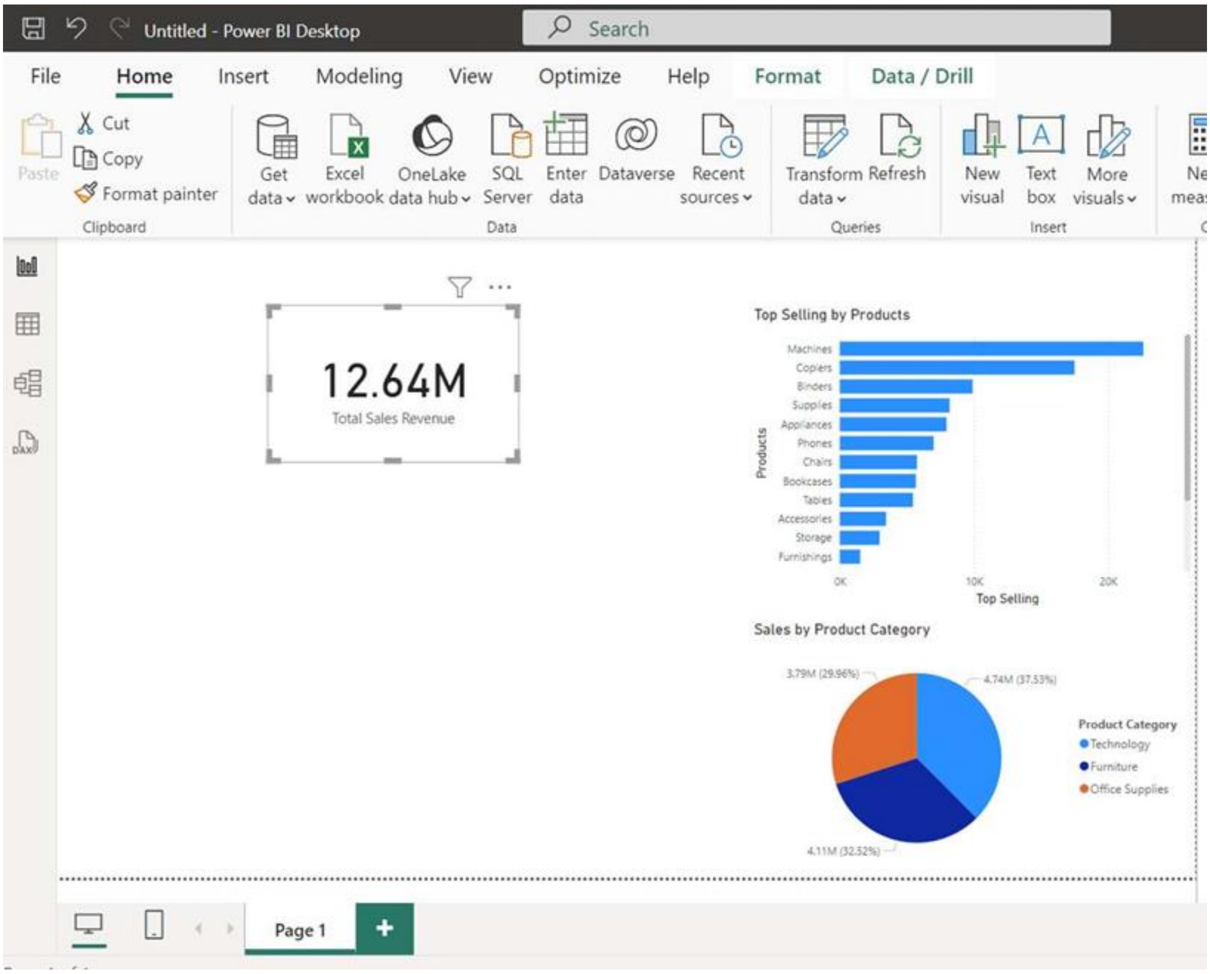


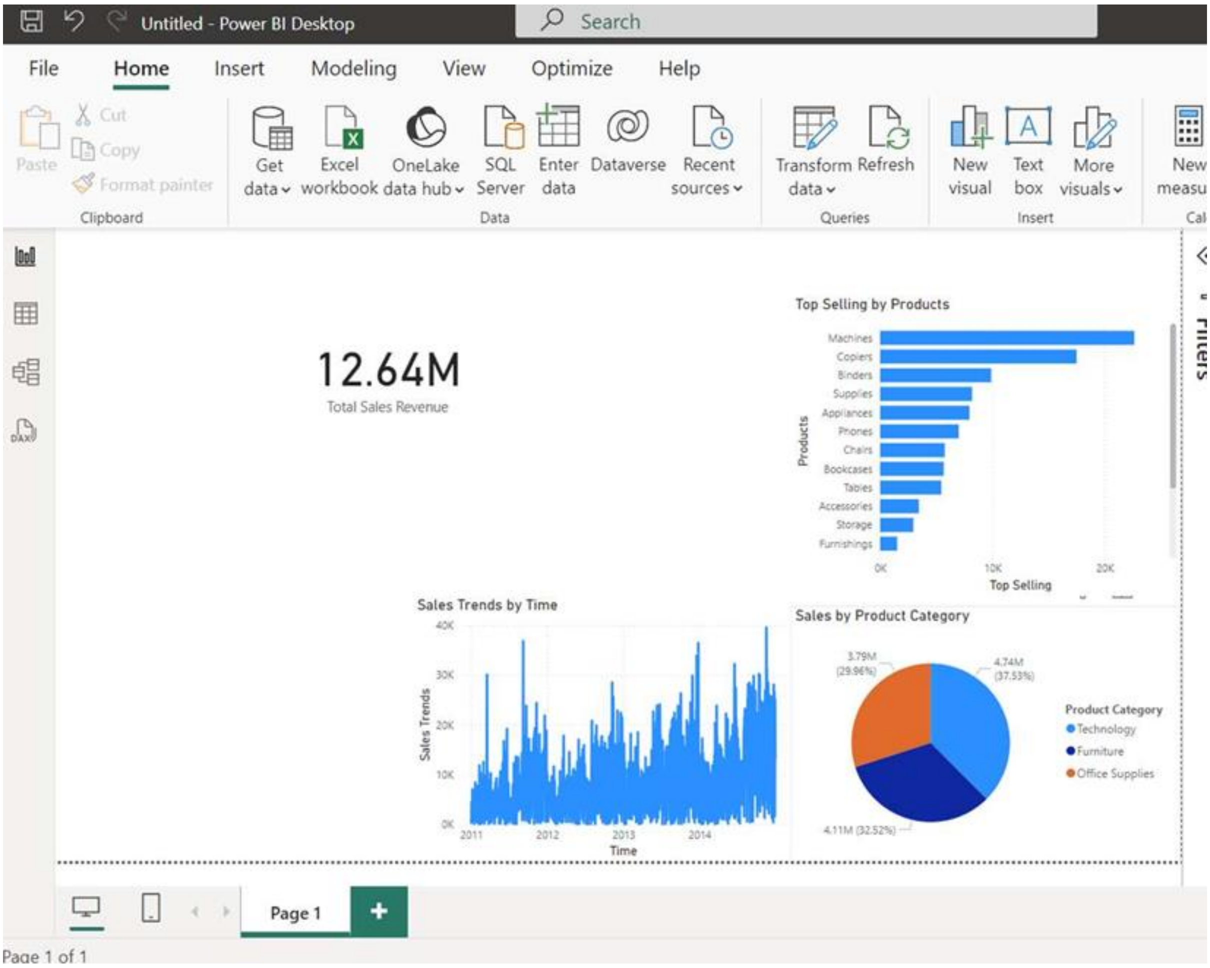
Global-Superstore.csv

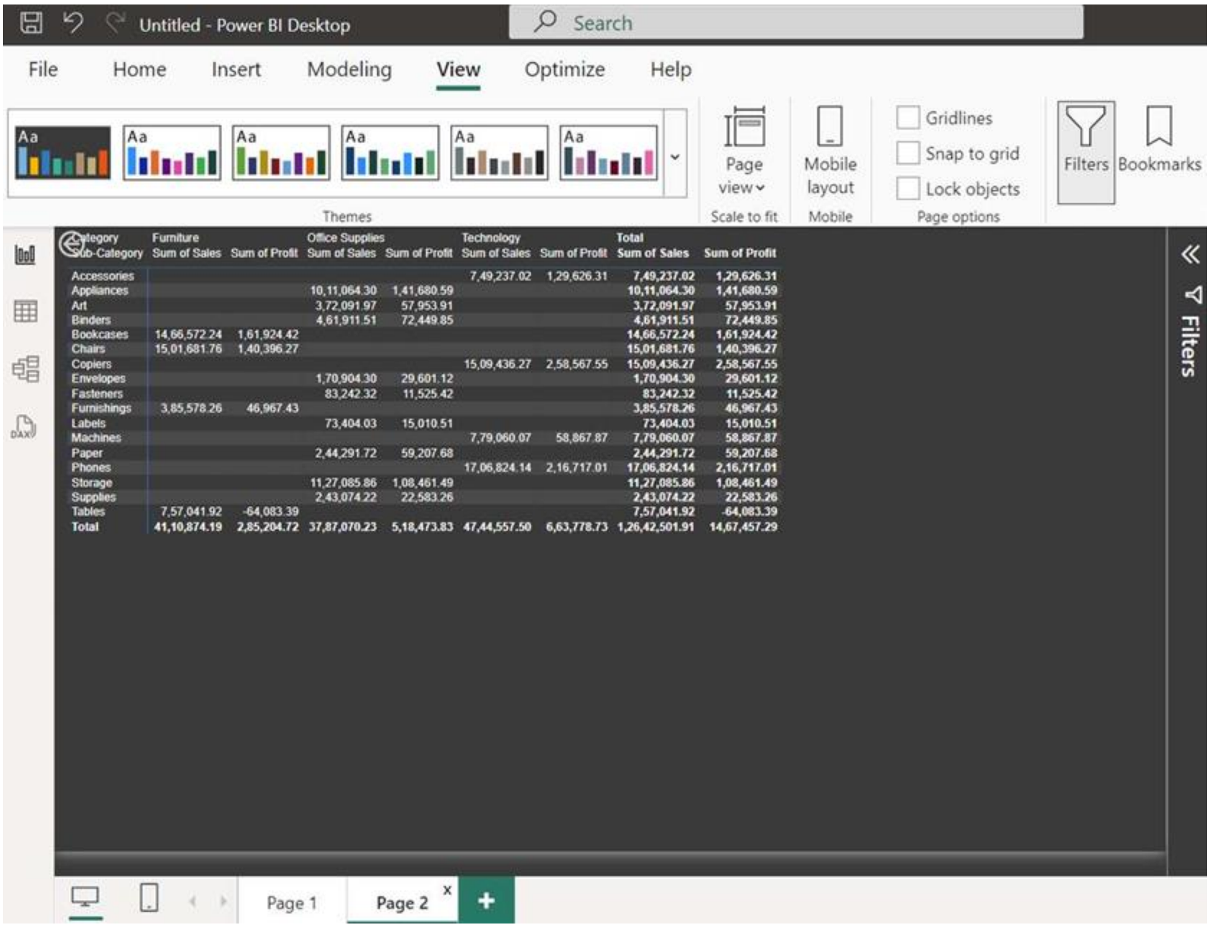
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Cu
32298	CA-2012-124891	31-07-2012	31-07-2012	Same Day	
26341	IN-2013-77878	05-02-2013	07-02-2013	Second Class	
25330	IN-2013-71249	17-10-2013	18-10-2013	First Class	
13524	ES-2013-1579342	28-01-2013	30-01-2013	First Class	
47221	SG-2013-4320	05-11-2013	06-11-2013	Same Day	
22732	IN-2013-42360	28-06-2013	01-07-2013	Second Class	
30570	IN-2011-81826	07-11-2011	09-11-2011	First Class	
31192	IN-2012-86369	14-04-2012	18-04-2012	Standard Class	
40155	CA-2014-135909	14-10-2014	21-10-2014	Standard Class	
40936	CA-2012-116638	28-01-2012	31-01-2012	Second Class	
34577	CA-2011-102988	05-04-2011	09-04-2011	Second Class	
28879	ID-2012-28402	19-04-2012	22-04-2012	First Class	
45794	SA-2011-1830	27-12-2011	29-12-2011	Second Class	
4132	MX-2012-130015	13-11-2012	13-11-2012	Same Day	
27704	IN-2013-73951	06-06-2013	08-06-2013	Second Class	
13779	ES-2014-5099955	31-07-2014	03-08-2014	Second Class	
36178	CA-2014-143567	03-11-2014	06-11-2014	Second Class	
12069	ES-2014-1651774	08-09-2014	14-09-2014	Standard Class	
22096	IN-2014-11763	31-01-2014	01-02-2014	First Class	
49463	TZ-2014-8190	05-12-2014	07-12-2014	Second Class	

The data in the preview has been truncated due to size limits.


Untitled - Power BI Desktop
Search
File
Home
Insert
Modeling
View
Optimize
Help
Format
Data / Drill
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
12.64M
Total Sales Revenue
Sales by Product Category
3.79M (29.96%)
4.74M (37.53%)
4.11M (32.52%)
Product Category
● Technology
● Furniture
● Office Supplies
Page 1
+
RStudio


Untitled - Power BI Desktop
Search
File
Home
Insert
Modeling
View
Optimize
Help
Format
Data / Drill
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
Ne mea
12.64M
Total Sales Revenue
Top Selling by Products
Machines
Copiers
Binders
Supplies
Appliances
Phones
Chairs
Bookcases
Tables
Accessories
Storage
Furnishings
Products
0K
10K
20K
Top Selling
Sales by Product Category
3.79M (29.96%)
4.74M (37.53%)
4.11M (32.52%)
Product Category
● Technology
● Furniture
● Office Supplies
Page 1
+


Untitled - Power BI Desktop
Search
File
Home
Insert
Modeling
View
Optimize
Help
Paste
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New measu
Cal
12.64M
Total Sales Revenue
Top Selling by Products
Products
Machines
Copiers
Binders
Supplies
Appliances
Phones
Chairs
Bookcases
Tables
Accessories
Storage
Furnishings
0K
10K
20K
Top Selling
Sales Trends by Time
Sales Trends
40K
30K
20K
10K
0K
2011
2012
2013
2014
Time
Sales by Product Category
3.79M (29.96%)
4.74M (37.53%)
4.11M (32.52%)
Product Category
● Technology
● Furniture
● Office Supplies
Filters
Page 1
+
Page 1 of 1



Category	Furniture		Office Supplies		Technology		Total	
Sub-Category	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit
Accessories					7,49,237.02	1,29,626.31	7,49,237.02	1,29,626.31
Appliances			10,11,064.30	1,41,680.59			10,11,064.30	1,41,680.59
Art			3,72,091.97	57,953.91			3,72,091.97	57,953.91
Binders			4,61,911.51	72,449.85			4,61,911.51	72,449.85
Bookcases	14,66,572.24	1,61,924.42					14,66,572.24	1,61,924.42
Chairs	15,01,681.76	1,40,396.27					15,01,681.76	1,40,396.27
Copiers					15,09,436.27	2,58,567.55	15,09,436.27	2,58,567.55
Envelopes			1,70,904.30	29,601.12			1,70,904.30	29,601.12
Fasteners			83,242.32	11,525.42			83,242.32	11,525.42
Furnishings	3,85,578.26	46,967.43					3,85,578.26	46,967.43
Labels			73,404.03	15,010.51			73,404.03	15,010.51
Machines					7,79,060.07	58,867.87	7,79,060.07	58,867.87
Paper			2,44,291.72	59,207.68			2,44,291.72	59,207.68
Phones					17,06,824.14	2,16,717.01	17,06,824.14	2,16,717.01
Storage			11,27,085.86	1,08,461.49			11,27,085.86	1,08,461.49
Supplies			2,43,074.22	22,583.26			2,43,074.22	22,583.26
Tables	7,57,041.92	-64,083.39					7,57,041.92	-64,083.39
Total	**41,10,874.19**	**2,85,204.72**	**37,87,070.23**	**5,18,473.83**	**47,44,557.50**	**6,63,778.73**	**1,26,42,501.91**	**14,67,457.29**