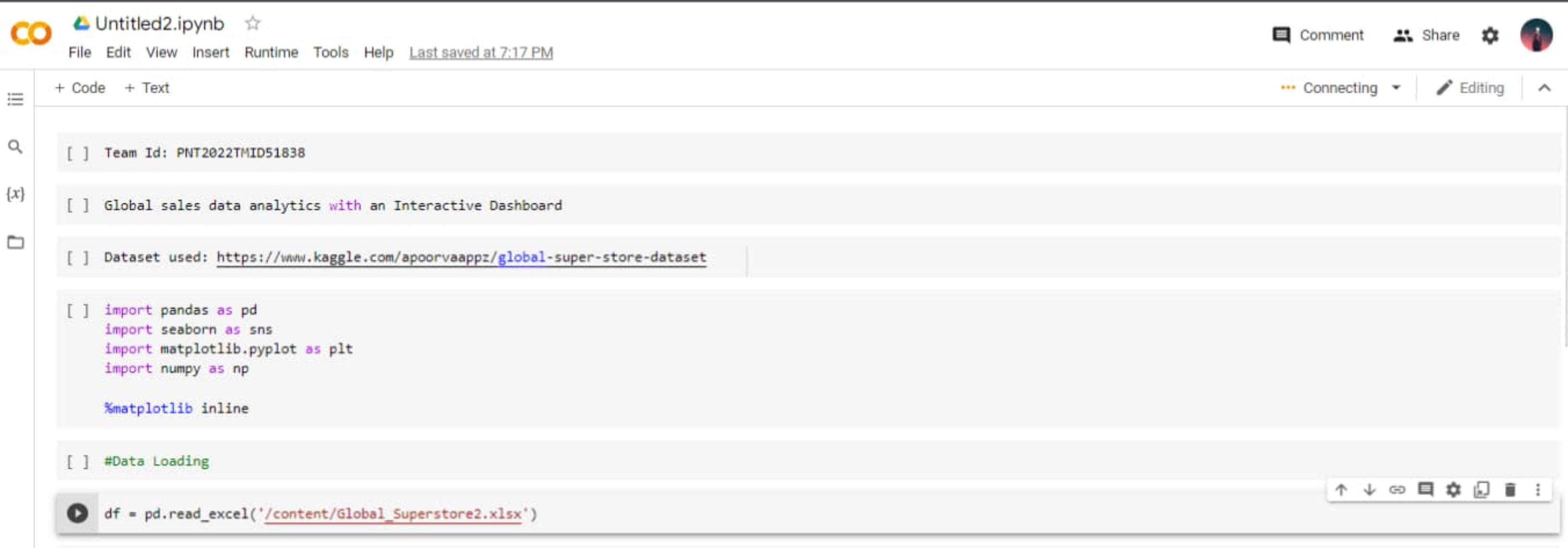

Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 7:17 PM
Comment
Share
+ Code + Text
Connecting
Editing

```
[ ] Team Id: PNT2022TMID51838
```

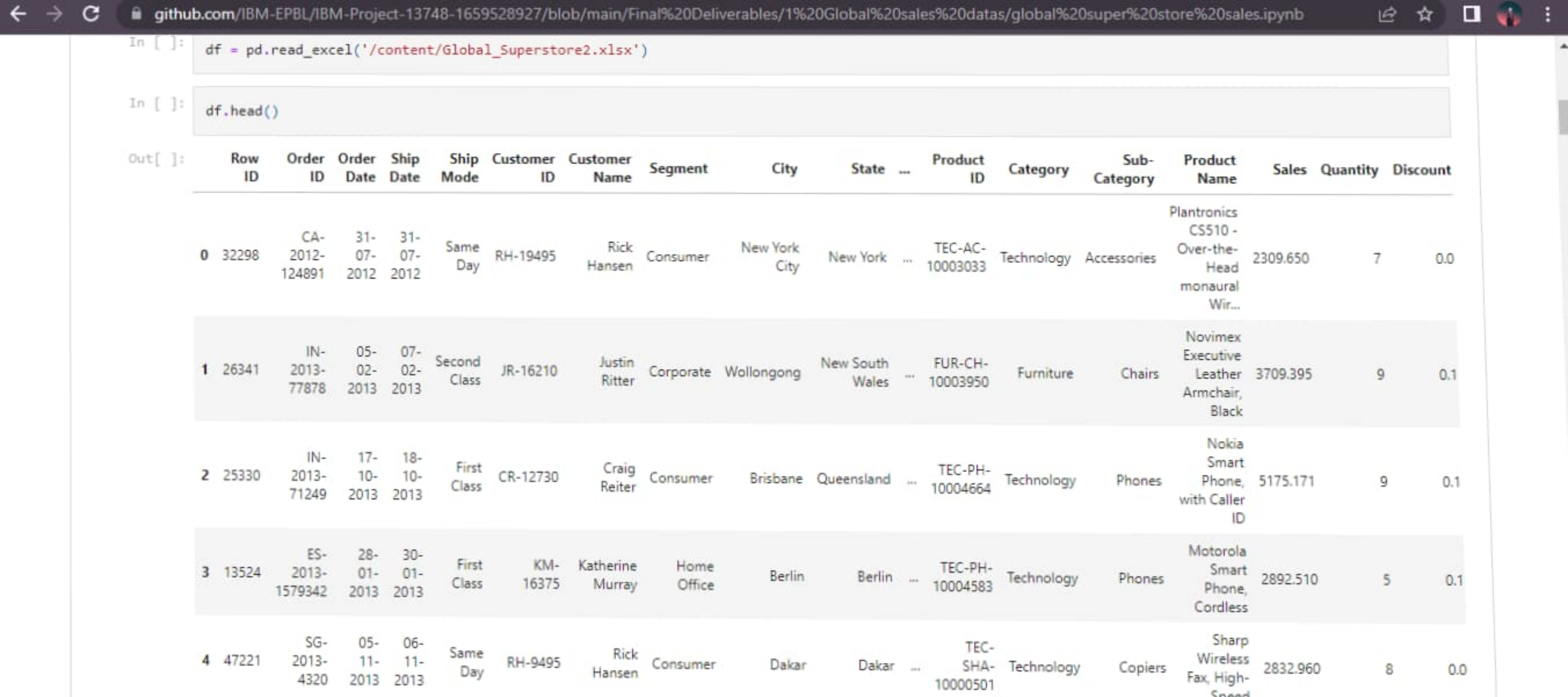
```
[ ] Global sales data analytics with an Interactive Dashboard
```

```
[ ] Dataset used: https://www.kaggle.com/apoorvaappz/global-super-store-dataset
```

```
[ ] import pandas as pd
    import seaborn as sns
    import matplotlib.pyplot as plt
    import numpy as np

    %matplotlib inline
```

```
[ ] #Data Loading
```

```
df = pd.read_excel('/content/Global_Superstore2.xlsx')
```



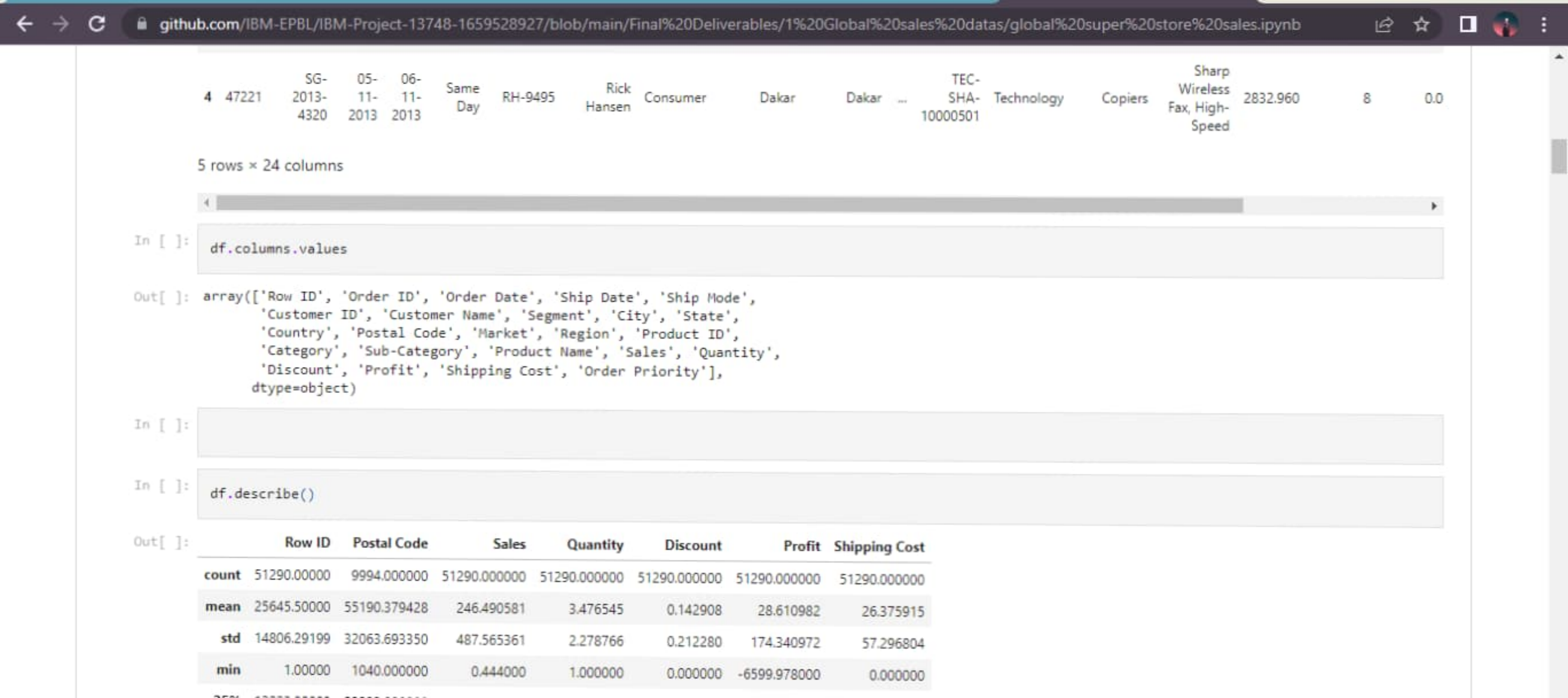
github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

In [ ]:
```
df = pd.read_excel('/content/Global_Superstore2.xlsx')
```

In [ ]:
```
df.head()
```

Out[ ]:

	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	Quantity	Discount
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	7	0.0
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	9	0.1
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	9	0.1
3	13524	ES-2013-1579342	28-01-2013	30-01-2013	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	...	TEC-PH-10004583	Technology	Phones	Motorola Smart Phone, Cordless	2892.510	5	0.1
4	47221	SG-2013-4320	05-11-2013	06-11-2013	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar	...	TEC-SHA-10000501	Technology	Copiers	Sharp Wireless Fax, High-Speed	2832.960	8	0.0


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

4	47221	SG-2013-4320	05-11-2013	06-11-2013	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar	...	TEC-SHA-10000501	Technology	Copiers	Sharp Wireless Fax, High-Speed	2832.960	8	0.0

5 rows × 24 columns

In [ ]:
```
df.columns.values
```

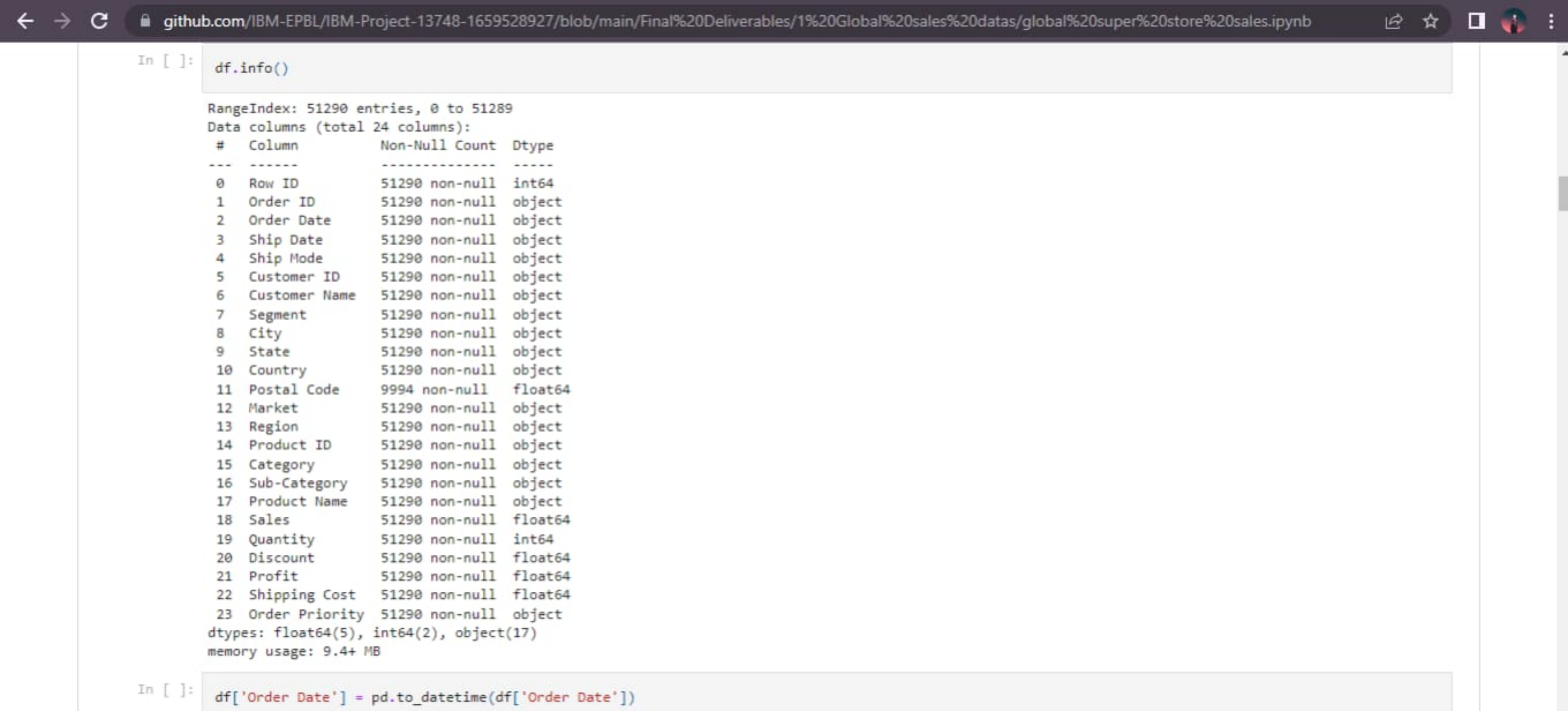
Out[ ]:
```
array(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'City', 'State',
       'Country', 'Postal Code', 'Market', 'Region', 'Product ID',
       'Category', 'Sub-Category', 'Product Name', 'Sales', 'Quantity',
       'Discount', 'Profit', 'Shipping Cost', 'Order Priority'],
      dtype=object)
```

In [ ]:

In [ ]:
```
df.describe()
```

Out[ ]:

	Row ID	Postal Code	Sales	Quantity	Discount	Profit	Shipping Cost
count	51290.00000	9994.000000	51290.000000	51290.000000	51290.000000	51290.000000	51290.000000
mean	25645.50000	55190.379428	246.490581	3.476545	0.142908	28.610982	26.375915
std	14806.29199	32063.693350	487.565361	2.278766	0.212280	174.340972	57.296804
min	1.00000	1040.000000	0.444000	1.000000	0.000000	-6599.978000	0.000000


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

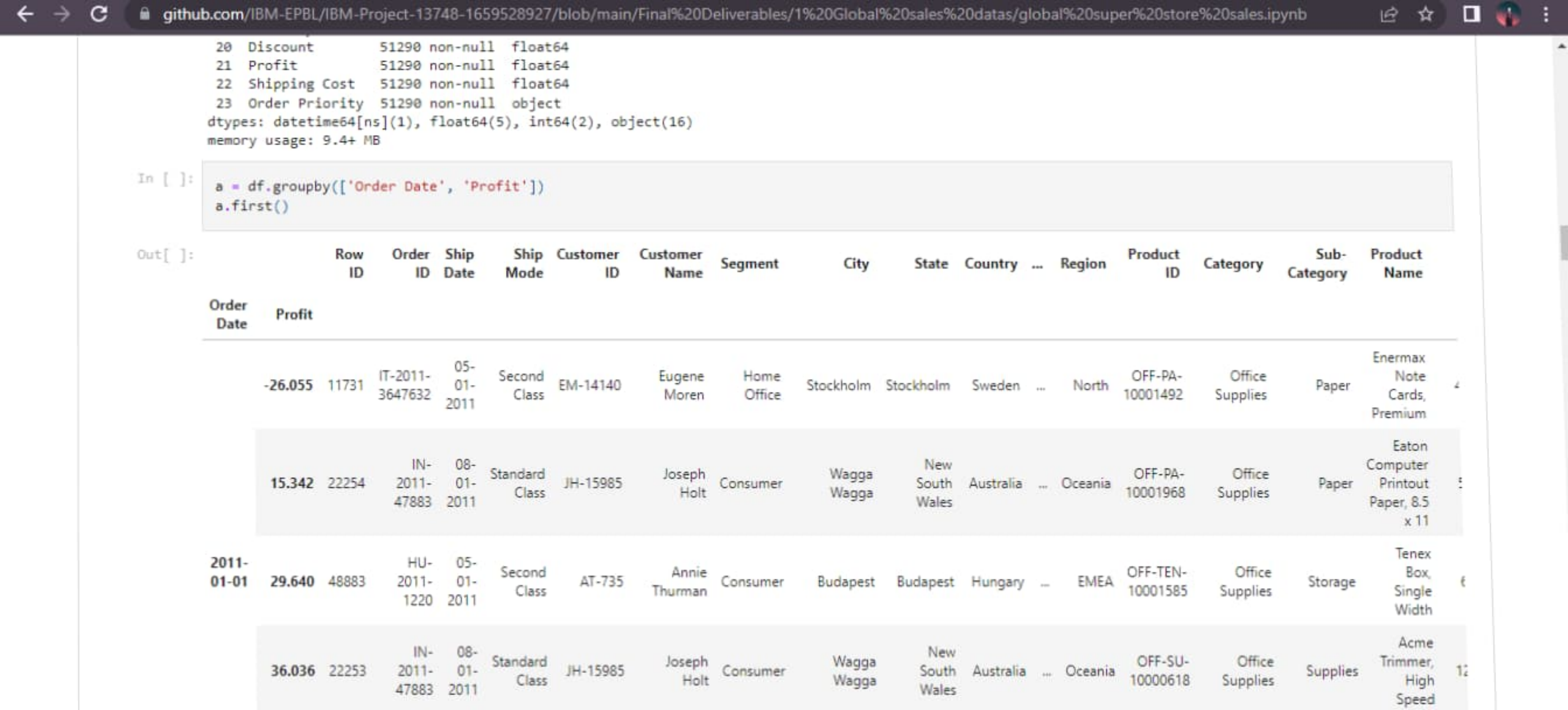
In [ ]:

```
df.info()
```

```
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
```

In [ ]:

```
df['Order Date'] = pd.to_datetime(df['Order Date'])
```



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

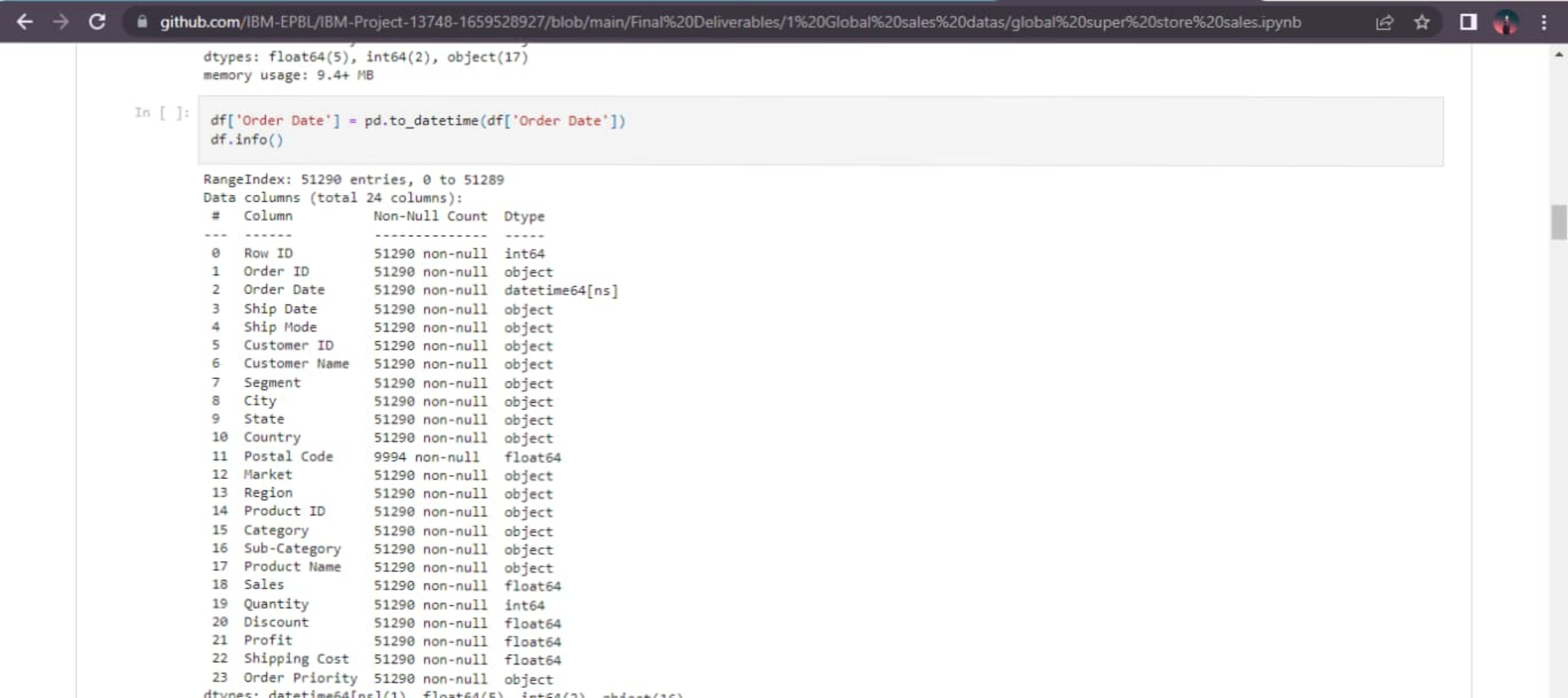
```
 20  Discount        51290 non-null  float64
 21  Profit          51290 non-null  float64
 22  Shipping Cost   51290 non-null  float64
 23  Order Priority  51290 non-null  object
dtypes: datetime64[ns](1), float64(5), int64(2), object(16)
memory usage: 9.4+ MB
```

In [ ]:

```
a = df.groupby(['Order Date', 'Profit'])
a.first()
```

Out[ ]:

		Row ID	Order ID	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	...	Region	Product ID	Category	Sub-Category	Product Name	
Order Date	Profit																	
2011-01-01	-26.055	11731	IT-2011-3647632	05-01-2011	Second Class	EM-14140	Eugene Moren	Home Office	Stockholm	Stockholm	Sweden	...	North	OFF-PA-10001492	Office Supplies	Paper	Enermax Note Cards, Premium	4
	15.342	22254	IN-2011-47883	08-01-2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	Australia	...	Oceania	OFF-PA-10001968	Office Supplies	Paper	Eaton Computer Printout Paper, 8.5 x 11	5
	29.640	48883	HU-2011-1220	05-01-2011	Second Class	AT-735	Annie Thurman	Consumer	Budapest	Budapest	Hungary	...	EMEA	OFF-TEN-10001585	Office Supplies	Storage	Tenex Box, Single Width	6
	36.036	22253	IN-2011-47883	08-01-2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	Australia	...	Oceania	OFF-SU-10000618	Office Supplies	Supplies	Acme Trimmer, High Speed	12



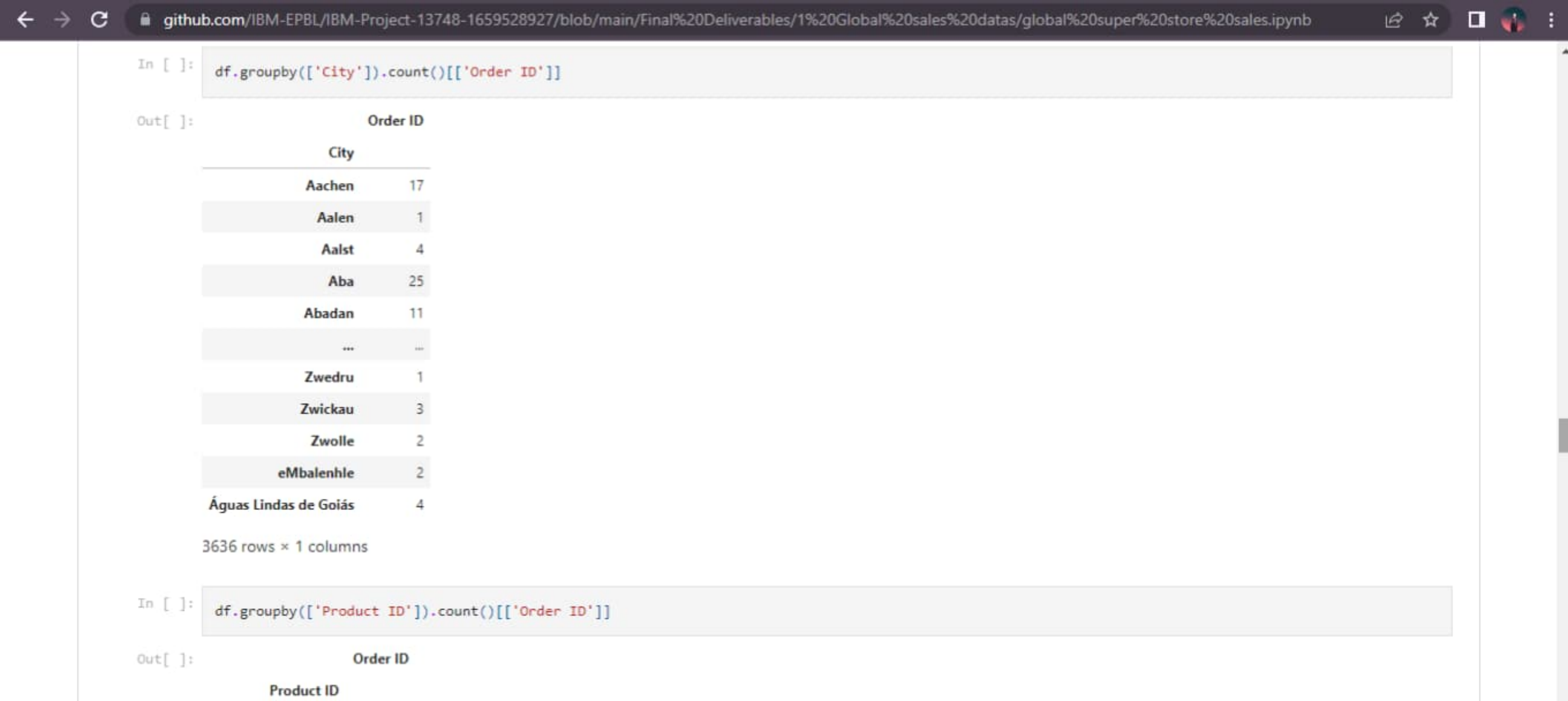
github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

```
dtypes: float64(5), int64(2), object(17)
memory usage: 9.4+ MB
```

In [ ]:

```
df['Order Date'] = pd.to_datetime(df['Order Date'])
df.info()
```

```
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  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: datetime64[ns](1), float64(5), int64(2), object(16)
```


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

In [ ]:

```
df.groupby(['City']).count()[['Order ID']]
```

Out[ ]:

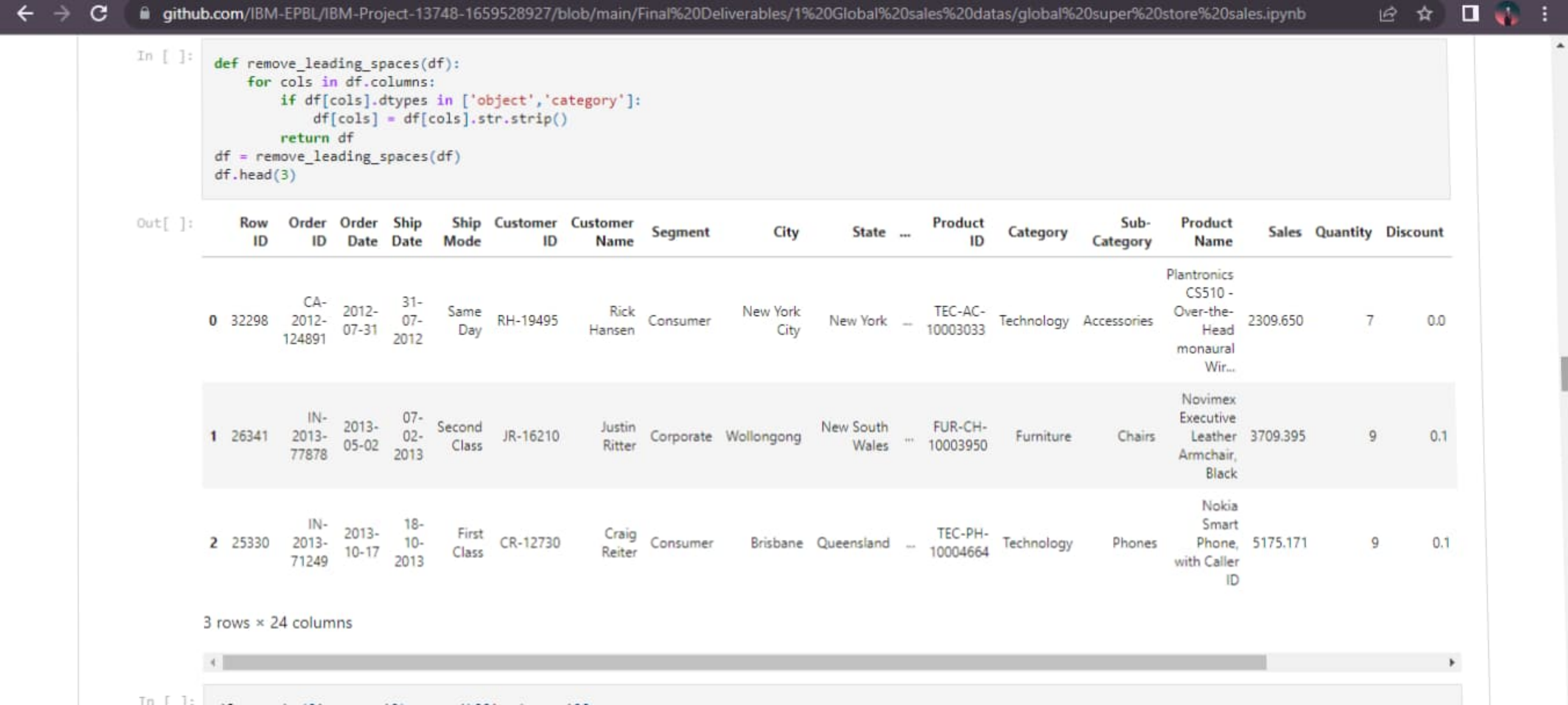
```
                      Order ID
                City
             Aachen         17
              Aalen          1
              Aalst          4
                Aba         25
             Abadan         11
                ...        ...
             Zwedru          1
            Zwickau          3
             Zwolle          2
          eMbalenhle         2
Águas Lindas de Goiás        4

3636 rows × 1 columns
```

In [ ]:

```
df.groupby(['Product ID']).count()[['Order ID']]
```

Out[ ]:

```
                 Order ID
      Product ID
```



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

In [ ]:

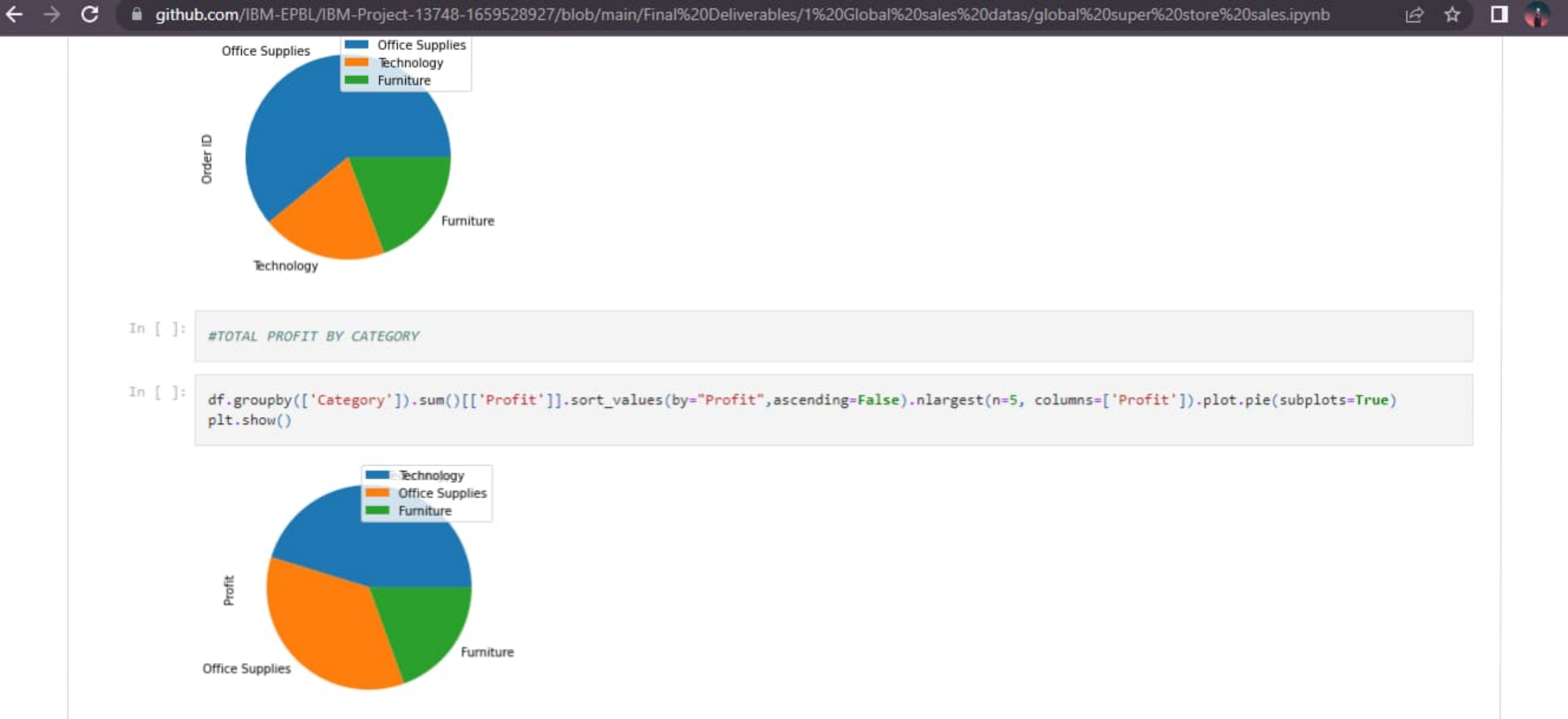
```
def remove_leading_spaces(df):
    for cols in df.columns:
        if df[cols].dtypes in ['object','category']:
            df[cols] = df[cols].str.strip()
        return df
df = remove_leading_spaces(df)
df.head(3)
```

Out[ ]:

	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	Quantity	Discount
0	32298	CA-2012-124891	2012-07-31	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	7	0.0
1	26341	IN-2013-77878	2013-05-02	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	9	0.1
2	25330	IN-2013-71249	2013-10-17	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	9	0.1

3 rows × 24 columns

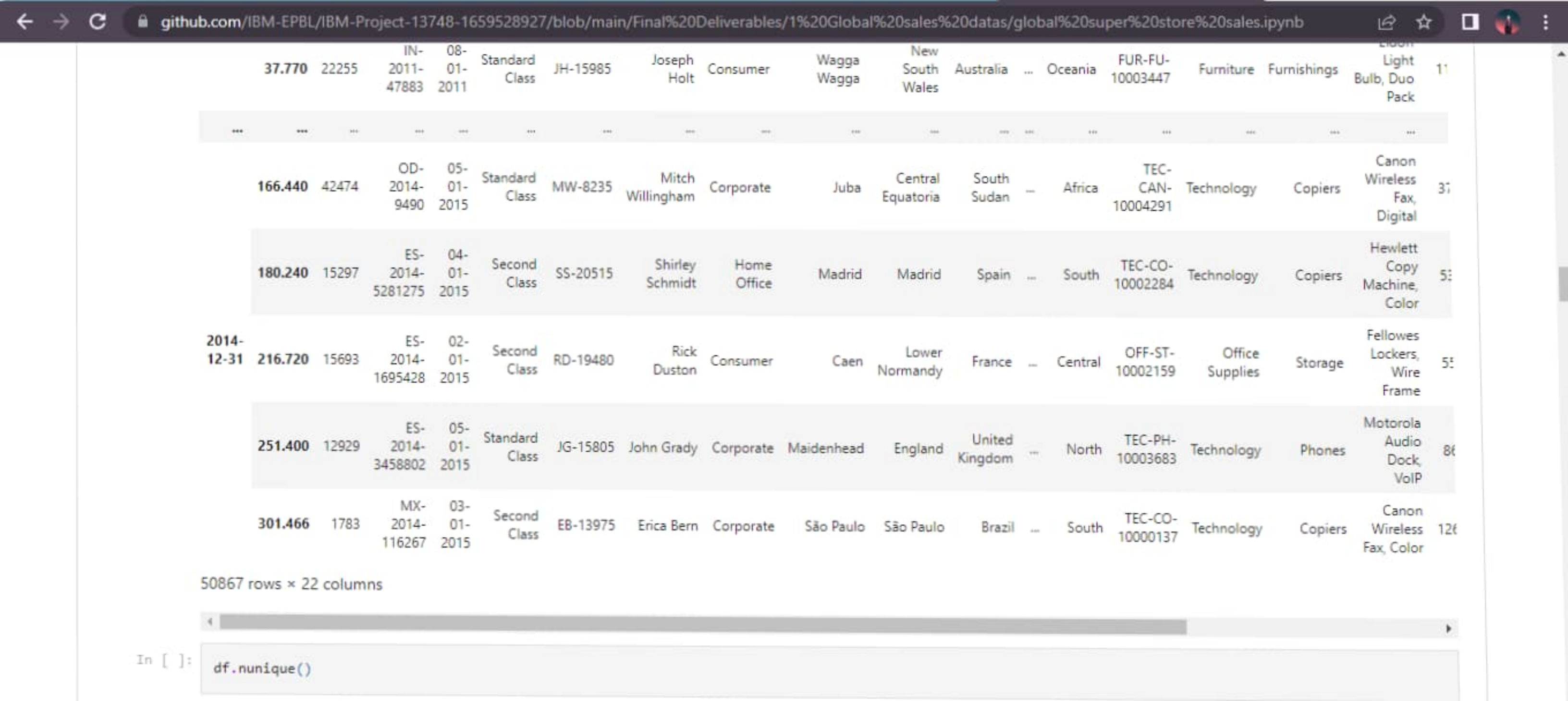
In [ ]:



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb
Office Supplies
🟦 Office Supplies
🟧 Technology
🟩 Furniture
Order ID
Furniture
Technology

```
In [ ]: #TOTAL PROFIT BY CATEGORY
```

```
In [ ]: df.groupby(['Category']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.pie(subplots=True)
        plt.show()
```

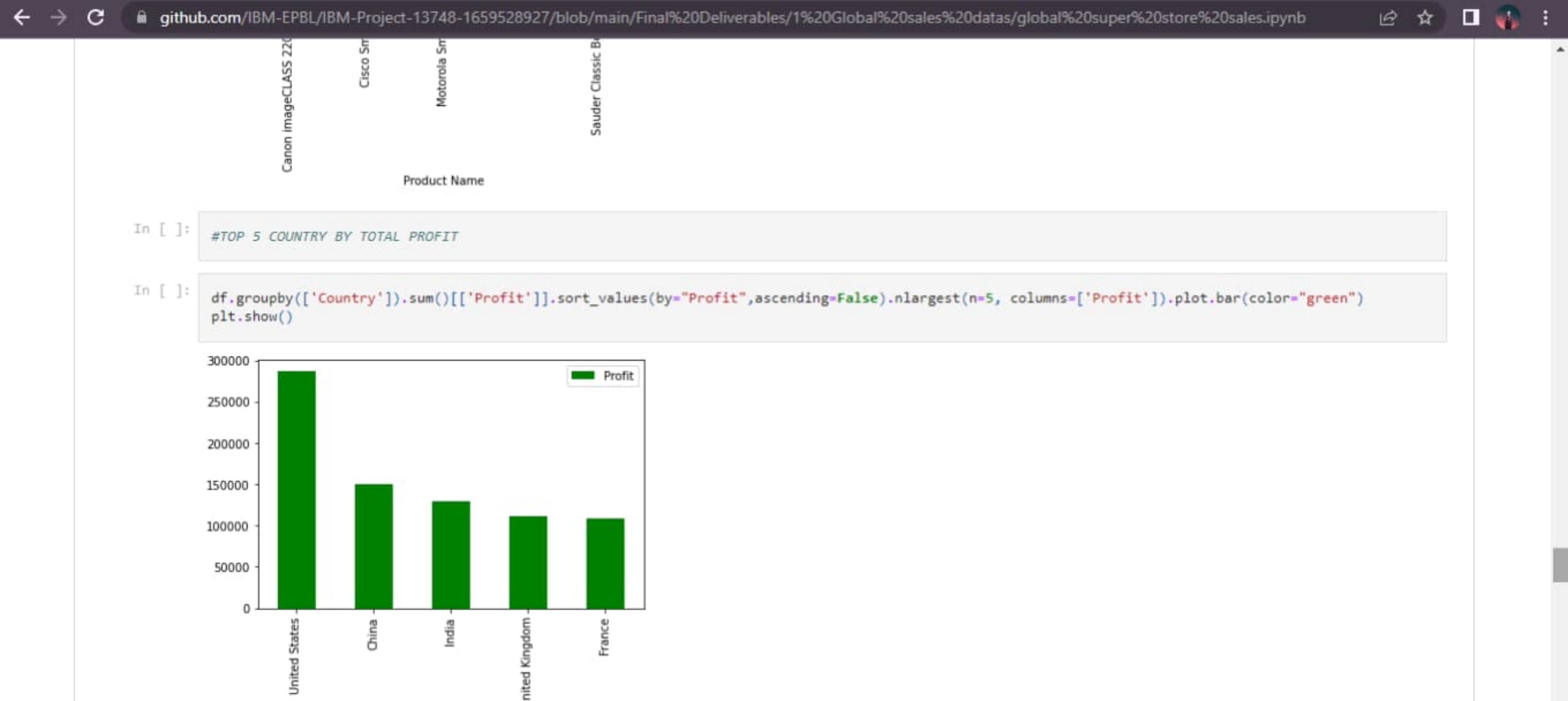


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

	37.770	22255	IN-2011-47883	08-01-2011	Standard Class	JH-15985	Joseph Holt	Consumer	Wagga Wagga	New South Wales	Australia	...	Oceania	FUR-FU-10003447	Furniture	Furnishings	Eldon Light Bulb, Duo Pack	11
...	**...**	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2014-12-31	**166.440**	42474	OD-2014-9490	05-01-2015	Standard Class	MW-8235	Mitch Willingham	Corporate	Juba	Central Equatoria	South Sudan	...	Africa	TEC-CAN-10004291	Technology	Copiers	Canon Wireless Fax, Digital	3
	180.240	15297	ES-2014-5281275	04-01-2015	Second Class	SS-20515	Shirley Schmidt	Home Office	Madrid	Madrid	Spain	...	South	TEC-CO-10002284	Technology	Copiers	Hewlett Copy Machine, Color	5
	216.720	15693	ES-2014-1695428	02-01-2015	Second Class	RD-19480	Rick Duston	Consumer	Caen	Lower Normandy	France	...	Central	OFF-ST-10002159	Office Supplies	Storage	Fellowes Lockers, Wire Frame	5
	251.400	12929	ES-2014-3458802	05-01-2015	Standard Class	JG-15805	John Grady	Corporate	Maidenhead	England	United Kingdom	...	North	TEC-PH-10003683	Technology	Phones	Motorola Audio Dock, VoIP	8
	301.466	1783	MX-2014-116267	03-01-2015	Second Class	EB-13975	Erica Bern	Corporate	São Paulo	São Paulo	Brazil	...	South	TEC-CO-10000137	Technology	Copiers	Canon Wireless Fax, Color	12

50867 rows × 22 columns

```
In [ ]: df.nunique()
```

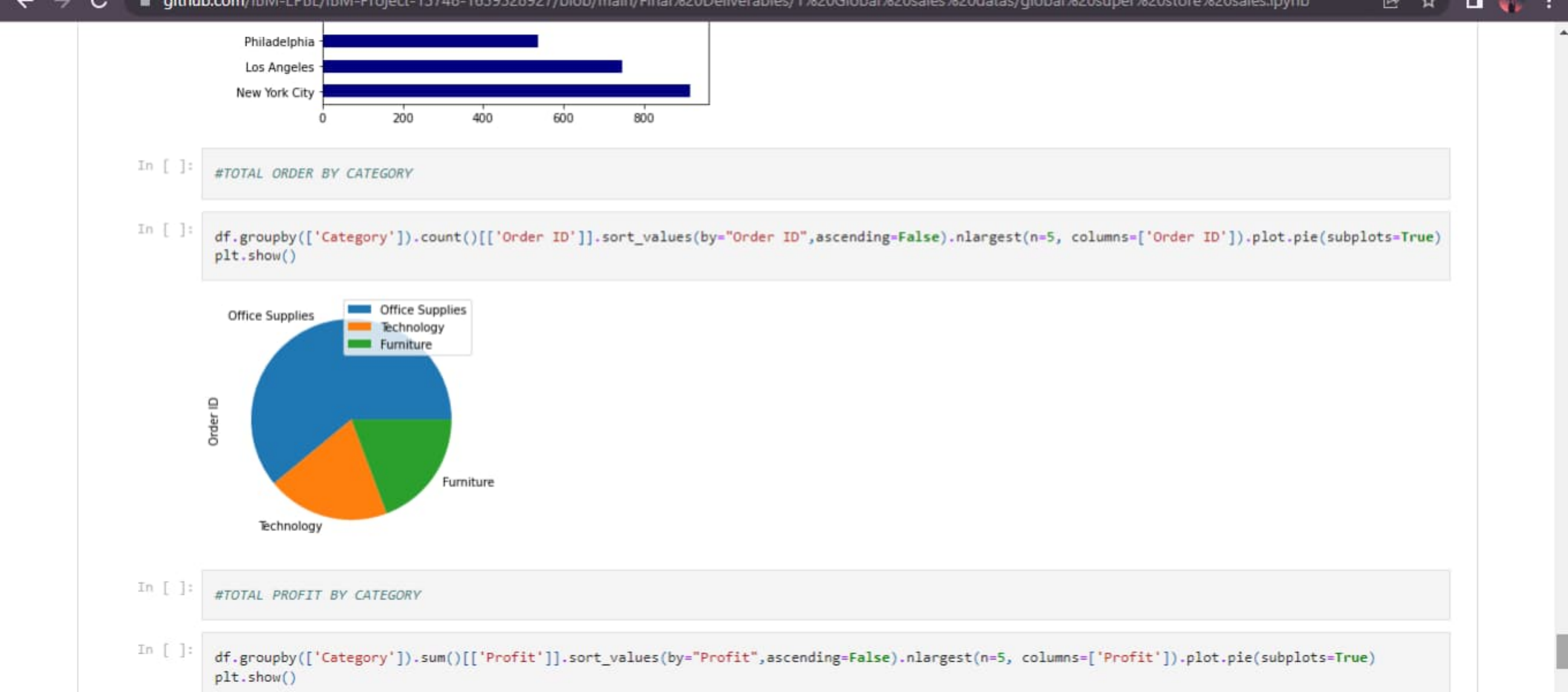

github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb
Canon imageCLASS 22
Cisco Sm
Motorola Sm
Sauder Classic B
Product Name

In [ ]:

```
#TOP 5 COUNTRY BY TOTAL PROFIT
```

In [ ]:

```
df.groupby(['Country']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.bar(color="green")
plt.show()
```

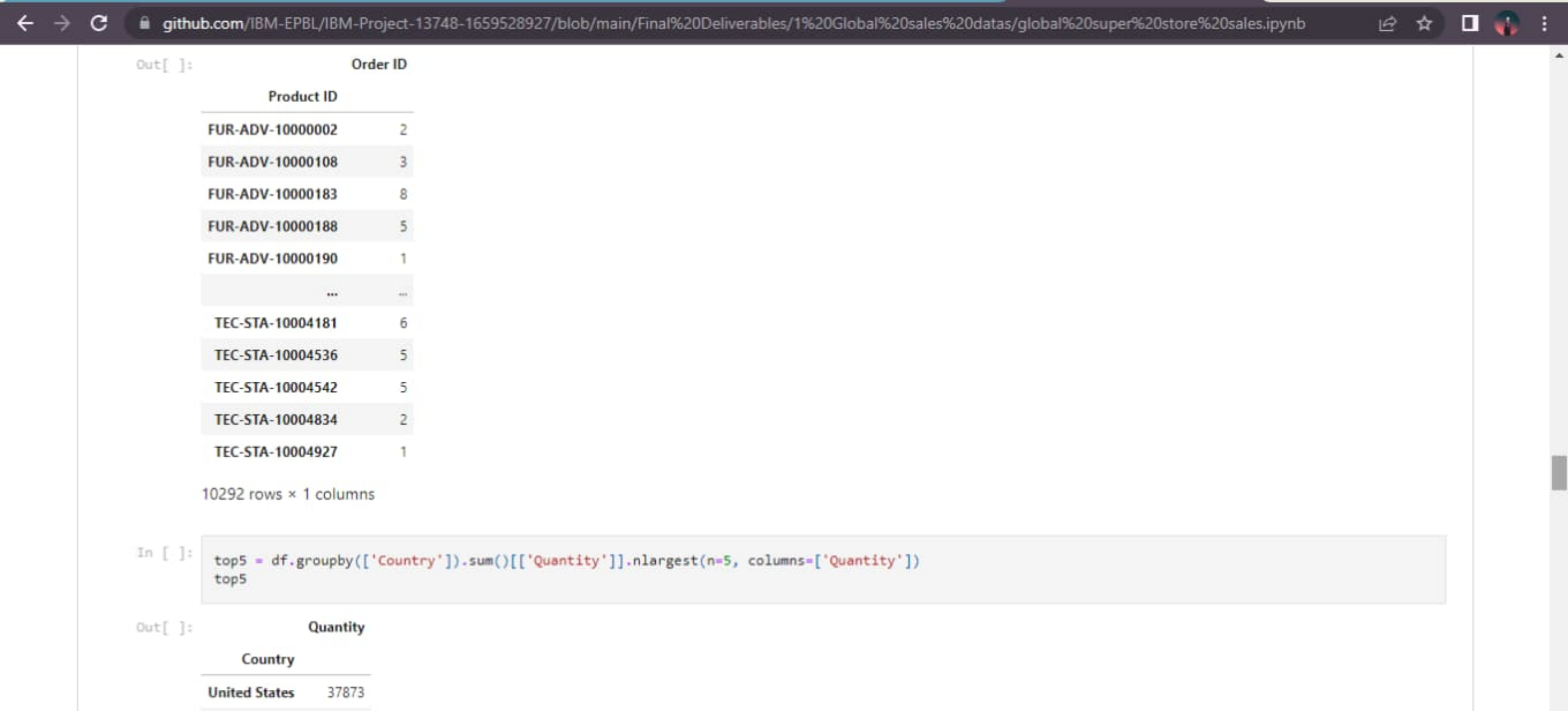


In [ ]:
```
#TOTAL ORDER BY CATEGORY
```

In [ ]:
```
df.groupby(['Category']).count()[['Order ID']].sort_values(by="Order ID",ascending=False).nlargest(n=5, columns=['Order ID']).plot.pie(subplots=True)
plt.show()
```

In [ ]:
```
#TOTAL PROFIT BY CATEGORY
```

In [ ]:
```
df.groupby(['Category']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.pie(subplots=True)
plt.show()
```


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

Out[ ]:

	Order ID
Product ID	
FUR-ADV-10000002	2
FUR-ADV-10000108	3
FUR-ADV-10000183	8
FUR-ADV-10000188	5
FUR-ADV-10000190	1
...	...
TEC-STA-10004181	6
TEC-STA-10004536	5
TEC-STA-10004542	5
TEC-STA-10004834	2
TEC-STA-10004927	1

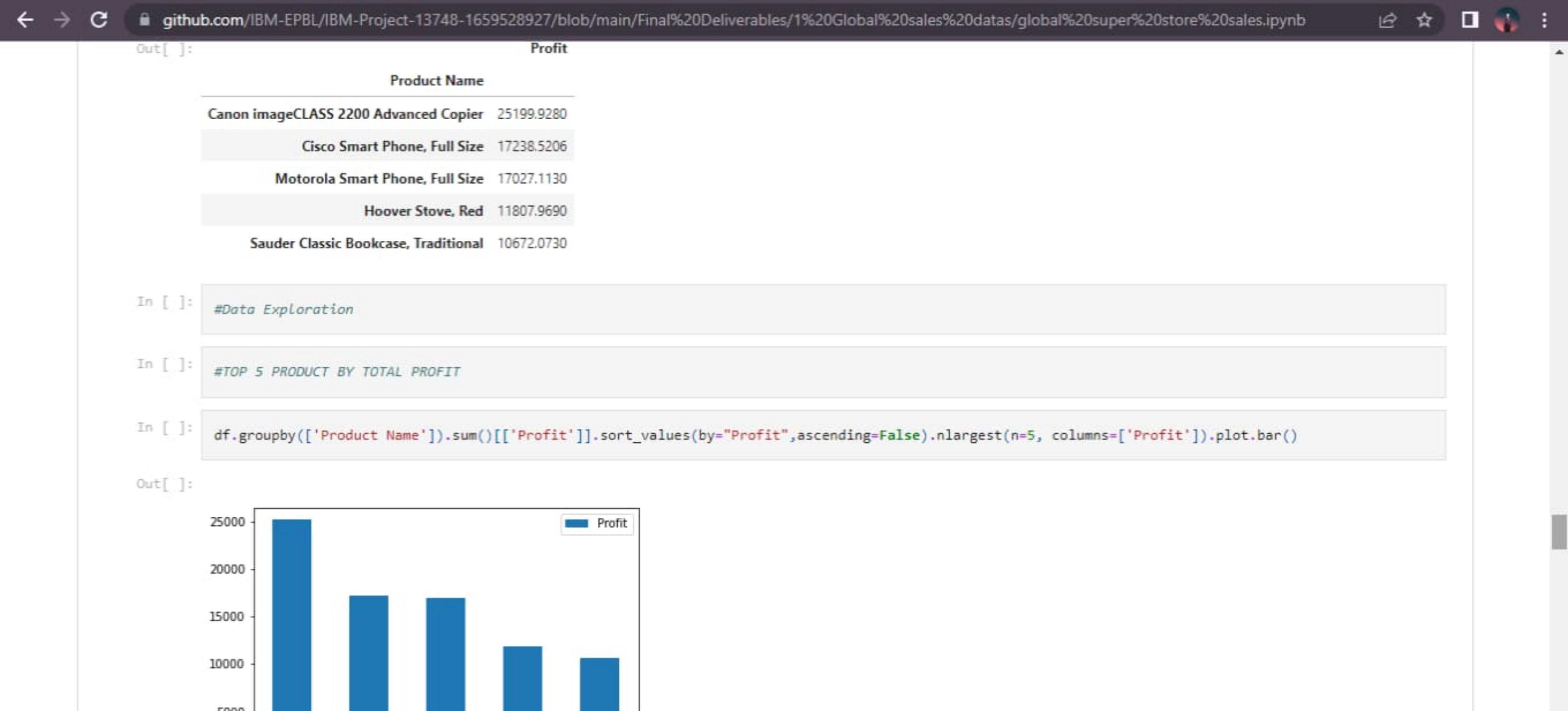
10292 rows × 1 columns

In [ ]:

```
top5 = df.groupby(['Country']).sum()[['Quantity']].nlargest(n=5, columns=['Quantity'])
top5
```

Out[ ]:

	Quantity
Country	
United States	37873



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

Out[ ]:

```
                                          Profit
                         Product Name
   Canon imageCLASS 2200 Advanced Copier  25199.9280
         Cisco Smart Phone, Full Size     17238.5206
      Motorola Smart Phone, Full Size     17027.1130
                   Hoover Stove, Red      11807.9690
     Sauder Classic Bookcase, Traditional 10672.0730
```

In [ ]:

```
#Data Exploration
```

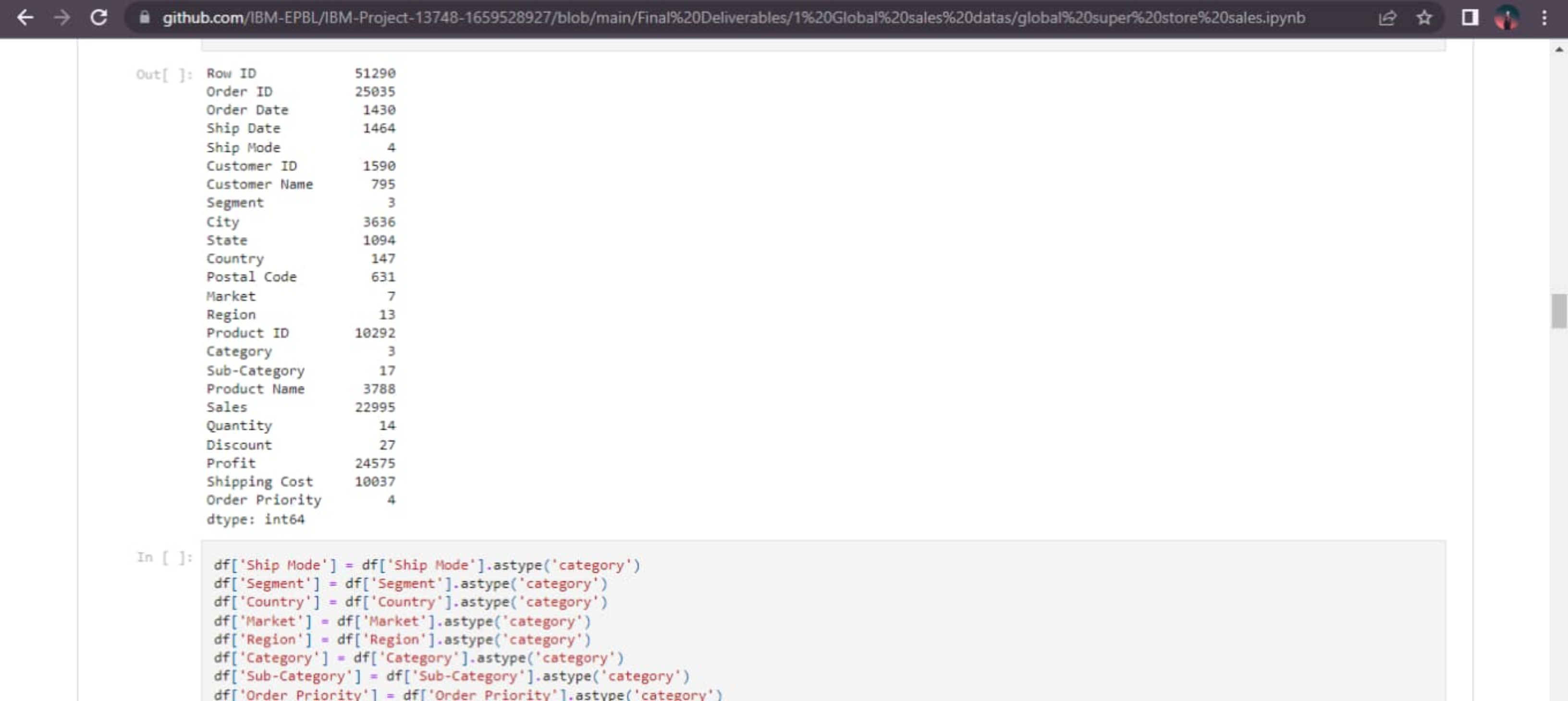
In [ ]:

```
#TOP 5 PRODUCT BY TOTAL PROFIT
```

In [ ]:

```
df.groupby(['Product Name']).sum()[['Profit']].sort_values(by="Profit",ascending=False).nlargest(n=5, columns=['Profit']).plot.bar()
```

Out[ ]:



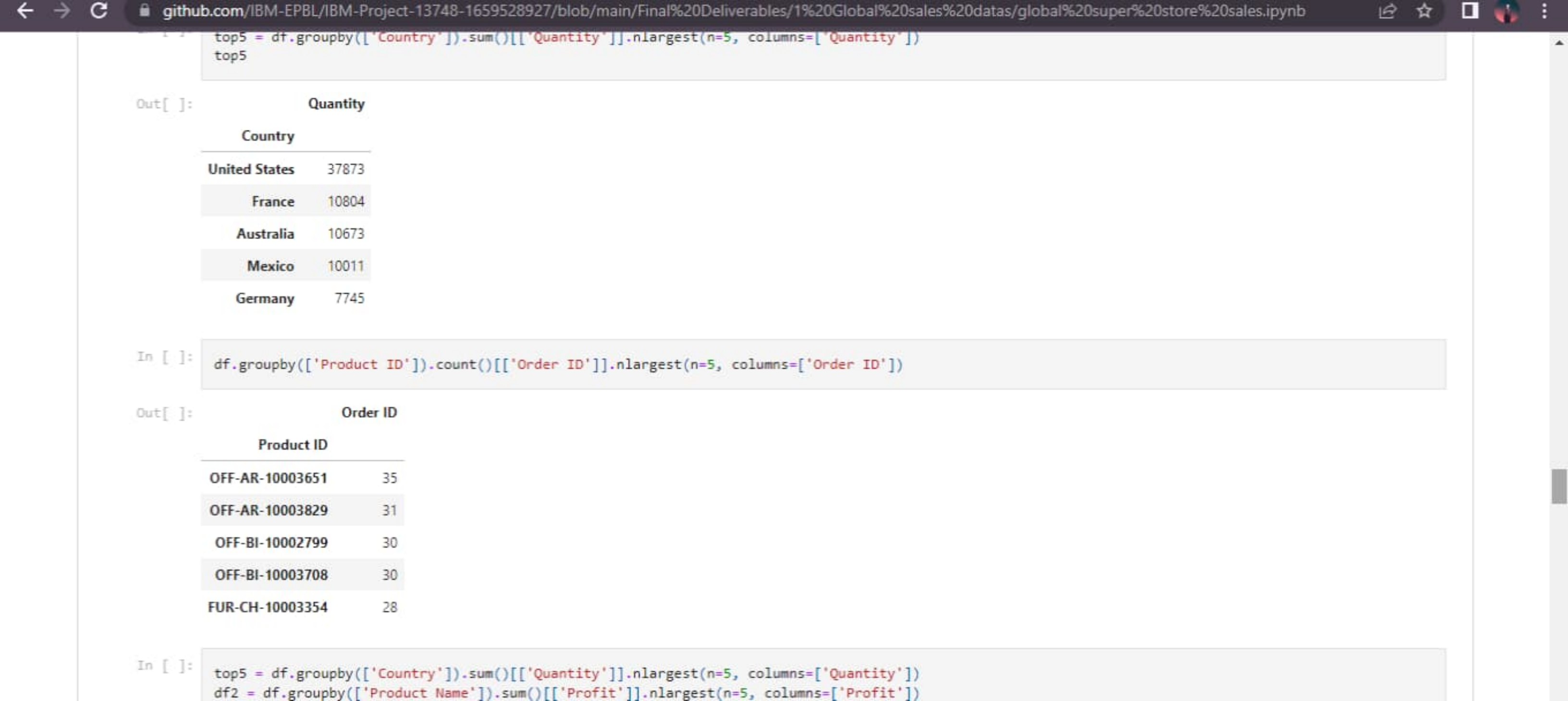
github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

Out[ ]:

```
Row ID            51290
Order ID          25035
Order Date         1430
Ship Date          1464
Ship Mode             4
Customer ID        1590
Customer Name       795
Segment               3
City               3636
State              1094
Country             147
Postal Code         631
Market                7
Region               13
Product ID        10292
Category              3
Sub-Category         17
Product Name       3788
Sales             22995
Quantity             14
Discount             27
Profit            24575
Shipping Cost     10037
Order Priority        4
dtype: int64
```

In [ ]:

```
df['Ship Mode'] = df['Ship Mode'].astype('category')
df['Segment'] = df['Segment'].astype('category')
df['Country'] = df['Country'].astype('category')
df['Market'] = df['Market'].astype('category')
df['Region'] = df['Region'].astype('category')
df['Category'] = df['Category'].astype('category')
df['Sub-Category'] = df['Sub-Category'].astype('category')
df['Order Priority'] = df['Order Priority'].astype('category')
```



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

```
top5 = df.groupby(['Country']).sum()[['Quantity']].nlargest(n=5, columns=['Quantity'])
top5
```

Out[ ]:

```
                Quantity
      Country
United States      37873
       France      10804
    Australia      10673
       Mexico      10011
      Germany       7745
```

In [ ]:

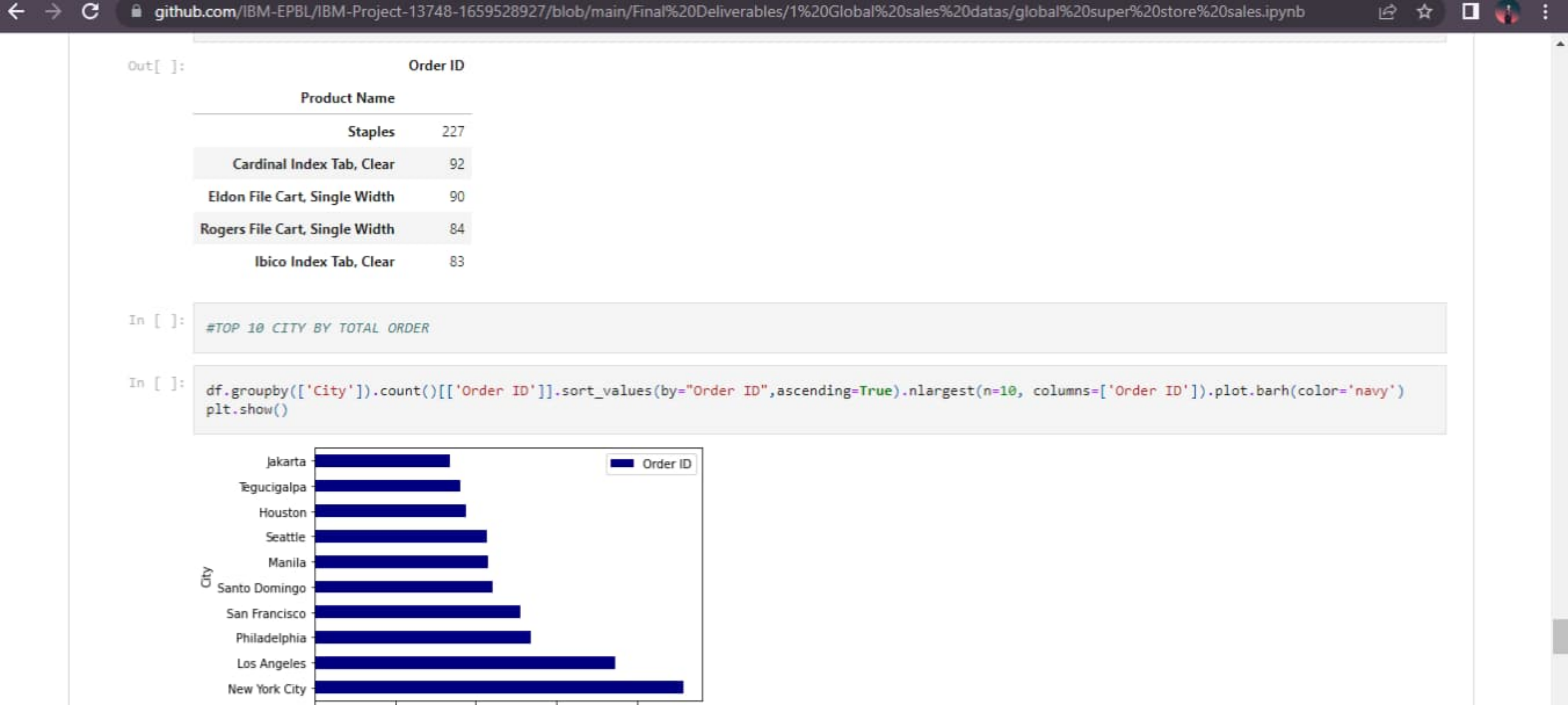
```
df.groupby(['Product ID']).count()[['Order ID']].nlargest(n=5, columns=['Order ID'])
```

Out[ ]:

```
                 Order ID
      Product ID
OFF-AR-10003651        35
OFF-AR-10003829        31
 OFF-BI-10002799       30
 OFF-BI-10003708       30
FUR-CH-10003354        28
```

In [ ]:

```
top5 = df.groupby(['Country']).sum()[['Quantity']].nlargest(n=5, columns=['Quantity'])
df2 = df.groupby(['Product Name']).sum()[['Profit']].nlargest(n=5, columns=['Profit'])
```



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

Out[ ]:

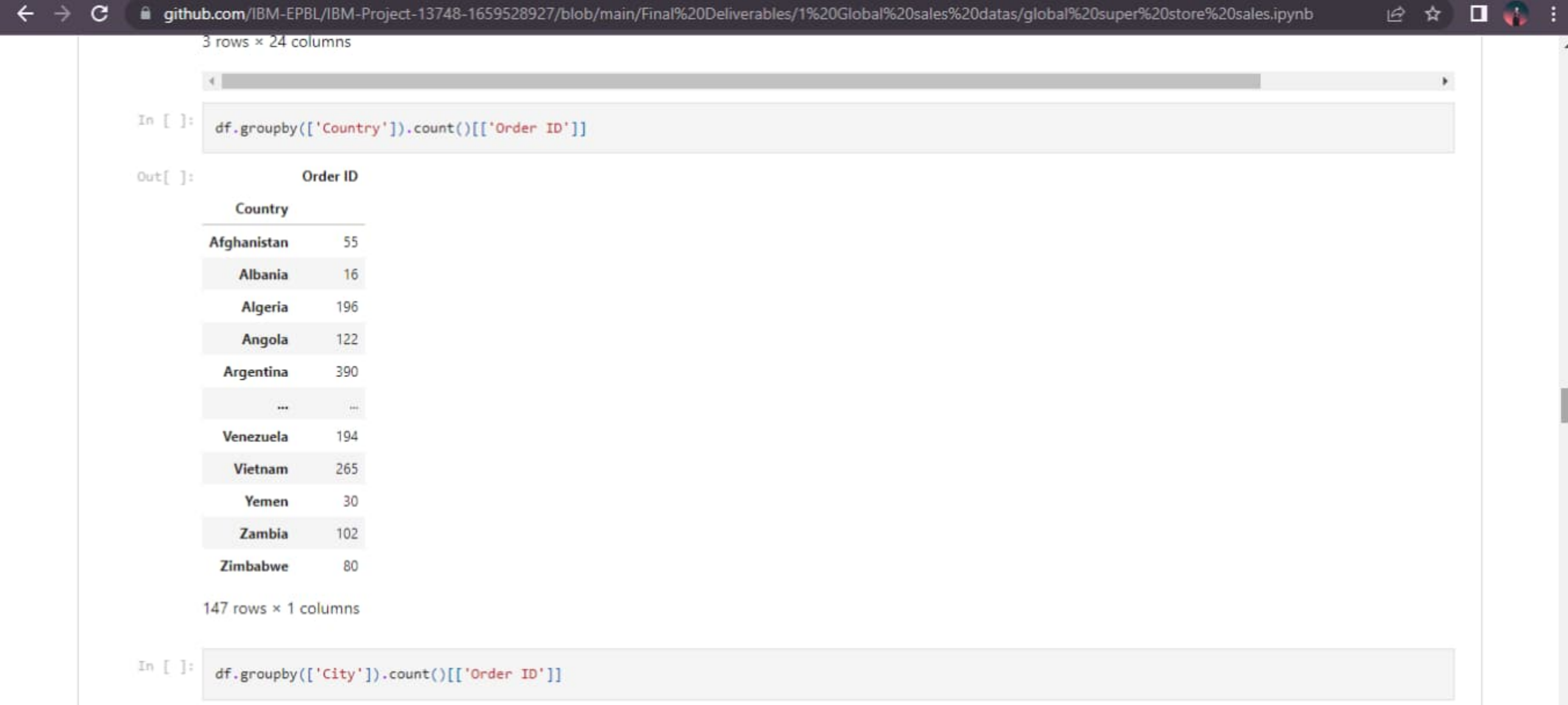
Product Name	Order ID
Staples	227
Cardinal Index Tab, Clear	92
Eldon File Cart, Single Width	90
Rogers File Cart, Single Width	84
Ibico Index Tab, Clear	83

In [ ]:

```
#TOP 10 CITY BY TOTAL ORDER
```

In [ ]:

```
df.groupby(['City']).count()[['Order ID']].sort_values(by="Order ID",ascending=True).nlargest(n=10, columns=['Order ID']).plot.barh(color='navy')
plt.show()
```



github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

3 rows × 24 columns

In [ ]:

```
df.groupby(['Country']).count()[['Order ID']]
```

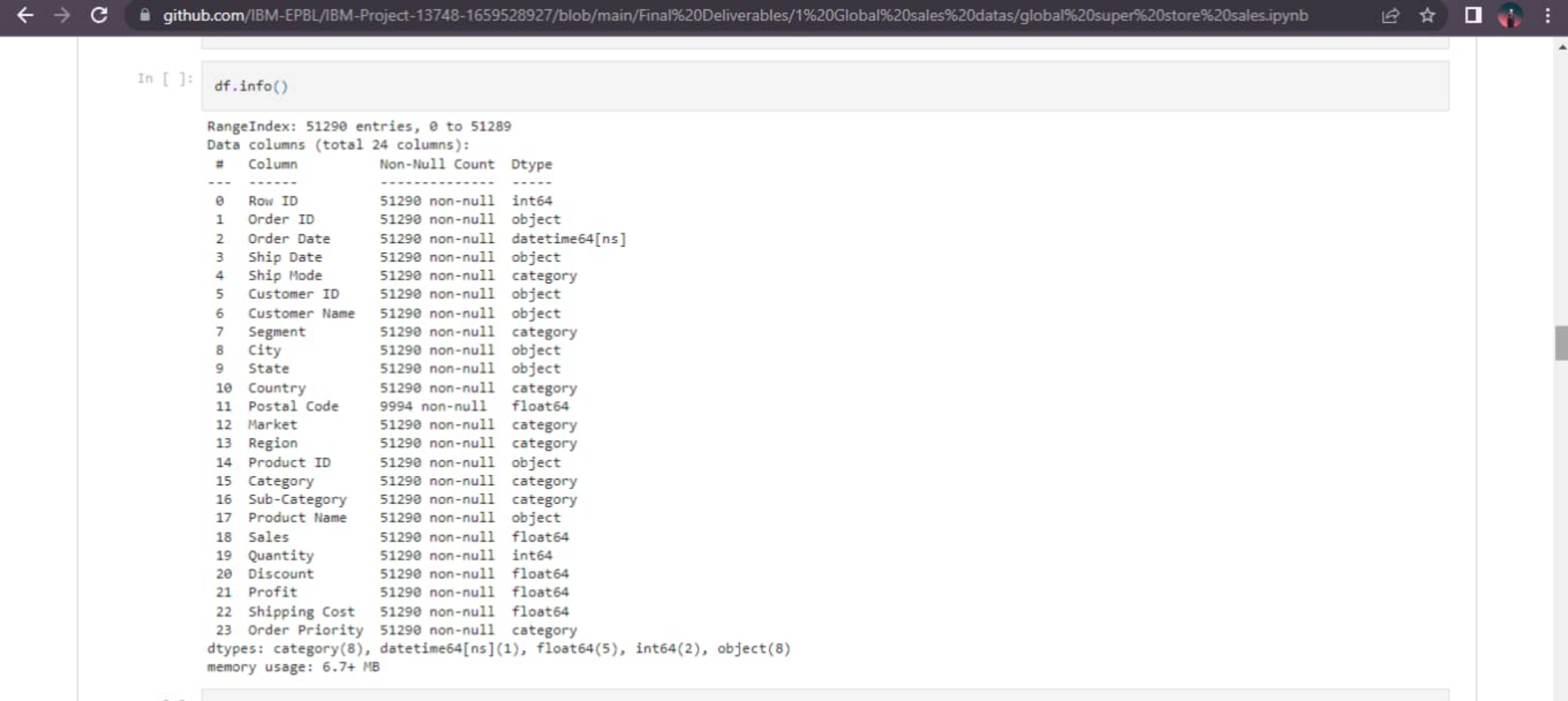
Out[ ]:

```
             Order ID
    Country
Afghanistan        55
    Albania        16
    Algeria       196
     Angola       122
  Argentina       390
        ...       ...
  Venezuela       194
    Vietnam       265
      Yemen        30
     Zambia       102
   Zimbabwe        80
```

147 rows × 1 columns

In [ ]:

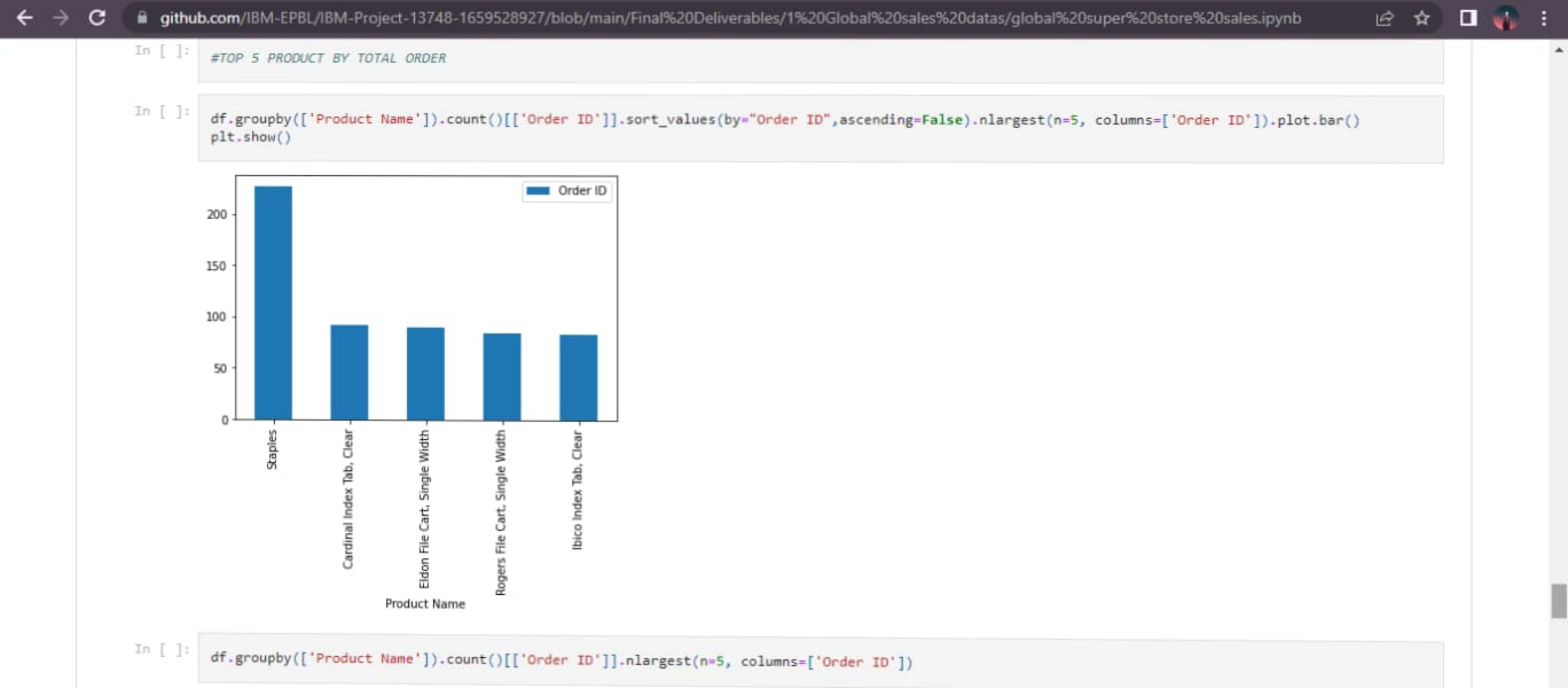
```
df.groupby(['City']).count()[['Order ID']]
```


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

In [ ]:

```
df.info()
```

```
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  object
 4   Ship Mode       51290 non-null  category
 5   Customer ID     51290 non-null  object
 6   Customer Name   51290 non-null  object
 7   Segment         51290 non-null  category
 8   City            51290 non-null  object
 9   State           51290 non-null  object
 10  Country         51290 non-null  category
 11  Postal Code     9994 non-null   float64
 12  Market          51290 non-null  category
 13  Region          51290 non-null  category
 14  Product ID      51290 non-null  object
 15  Category        51290 non-null  category
 16  Sub-Category    51290 non-null  category
 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  category
dtypes: category(8), datetime64[ns](1), float64(5), int64(2), object(8)
memory usage: 6.7+ MB
```


github.com/IBM-EPBL/IBM-Project-13748-1659528927/blob/main/Final%20Deliverables/1%20Global%20sales%20datas/global%20super%20store%20sales.ipynb

In [ ]:

```
#TOP 5 PRODUCT BY TOTAL ORDER
```

In [ ]:

```
df.groupby(['Product Name']).count()[['Order ID']].sort_values(by="Order ID",ascending=False).nlargest(n=5, columns=['Order ID']).plot.bar()
plt.show()
```

Order ID
200
150
100
50
0
Staples
Cardinal Index Tab, Clear
Eldon File Cart, Single Width
Rogers File Cart, Single Width
Ibico Index Tab, Clear
Product Name

In [ ]:

```
df.groupby(['Product Name']).count()[['Order ID']].nlargest(n=5, columns=['Order ID'])
```