Name: Muhammad Sheroz StudentID: 9852

Packages

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

Read All Create CSV By Own Structure

In [ ]:
```
DimCst = pd.read_csv("E:/Course/DataWareHousing/Final/Labs/DataSets/DW/DimCst.csv")
```

In [ ]:
```
DimLoc = pd.read_csv("E:/Course/DataWareHousing/Final/Labs/DataSets/DW/DimLoc.csv")
```

In [ ]:
```
DimPrd = pd.read_csv("E:/Course/DataWareHousing/Final/Labs/DataSets/DW/DimPrd.csv")
```

In [ ]:
```
DimDt = pd.read_csv("E:/Course/DataWareHousing/Final/Labs/DataSets/DW/DimDt.csv")
```

In [ ]:
```
SalesFct = pd.read_csv("E:/Course/DataWareHousing/Final/Labs/DataSets/DW/SalesFct.csv")
```

In [ ]:
```
SaleCstsFct = pd.read_csv("E:/Course/DataWareHousing/Final/Labs/DataSets/DW/SaleCstsFct
```

Top 3 Sales Product

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

Out[ ]:

	CustKey	Customer ID	Customer Name	Segment
0	1	AA-10315	Alex Avila	Consumer
1	2	AA-10375	Allen Armold	Consumer
2	3	AA-10480	Andrew Allen	Consumer
3	4	AA-10645	Anna Andreadi	Consumer
4	5	AB-10015	Aaron Bergman	Consumer

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

Out[ ]:

	CityKey	City	State	Country	Region
0	1	Aberdeen	South Dakota	United States	Central

	CityKey	City	State	Country	Region
1	2	Abilene	Texas	United States	Central
2	3	Akron	Ohio	United States	East
3	4	Albuquerque	New Mexico	United States	West
4	5	Alexandria	Virginia	United States	South

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

Out[ ]:

	Order Date	Day	Week	Month	Quarter	Year
0	2014-01-03	3	1	1	1	2014
1	2014-01-04	4	1	1	1	2014
2	2014-01-05	5	1	1	1	2014
3	2014-01-06	6	2	1	1	2014
4	2014-01-07	7	2	1	1	2014

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

Out[ ]:

	Order Date	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	CostAmt
0	2014-01-03	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	10.8968
1	2014-01-04	Naperville	OFF-BI-10004094	2	3.540	0.8	-5.4870	9.0270
2	2014-01-04	Naperville	OFF-LA-10003223	3	11.784	0.2	4.2717	7.5123
3	2014-01-04	Naperville	OFF-ST-10002743	3	272.736	0.2	-64.7748	337.5108
4	2014-01-05	Philadelphia	OFF-AR-10003478	3	19.536	0.2	4.8840	14.6520

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

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512
1	2014-01-04	PO-19195	Naperville	OFF-BI-10004094	2	3.540	0.8	-5.4870
2	2014-01-04	PO-19195	Naperville	OFF-LA-10003223	3	11.784	0.2	4.2717

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal
3	2014-01-04	PO-19195	Naperville	OFF-ST-10002743	3	272.736	0.2	-64.7748
4	2014-01-05	MB-18085	Philadelphia	OFF-AR-10003478	3	19.536	0.2	4.8840

In [ ]:
```
largerst = SaleCstsFct.nlargest(3,'SalesQtyTotal')
```

In [ ]:
```
largerst
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal
4369	2016-03-15	SJ-20125	Jackson	OFF-PA-10001954	16	365.440	0.0	168.1024
2807	2015-07-24	SS-20515	Louisville	OFF-BI-10000069	15	225.150	0.0	103.5690
112	2014-02-17	DL-13315	Chicago	OFF-AR-10003631	14	54.208	0.2	8.8088

In [ ]:
```
print(largerst['Product ID'])
```

```
4369    OFF-PA-10001954
2807    OFF-BI-10000069
112     OFF-AR-10003631
Name: Product ID, dtype: object
```

In [ ]:
```
top3 = SaleCstsFct.groupby('Product ID').agg(TotalQty = ('SalesQtyTotal' , 'sum'))
```

In [ ]:
```
top3.head()
```

Out[ ]:

Product ID	TotalQty
FUR-BO-10000112	9
FUR-BO-10000330	10
FUR-BO-10000362	14
FUR-BO-10000468	21
FUR-BO-10000711	12

In [ ]:
```
top3 = top3.nlargest(3,'TotalQty')
top3
```

Out[ ]:

Product ID	TotalQty
TEC-AC-10003832	75
OFF-PA-10001970	70
OFF-BI-10001524	67

In [ ]:
```
top3 = pd.merge(top3,DimPrd,on='Product ID')
```

In [ ]:
```
top3
```

Out[ ]:

	Product ID	TotalQty	Product Name	Sub-Category	Category
0	TEC-AC-10003832	75	Logitech P710e Mobile Speakerphone	Accessories	Technology
1	TEC-AC-10003832	75	Imation 16GB Mini TravelDrive USB 2.0 Flash Drive	Accessories	Technology
2	OFF-PA-10001970	70	Xerox 1908	Paper	Office Supplies
3	OFF-PA-10001970	70	Xerox 1881	Paper	Office Supplies
4	OFF-BI-10001524	67	GBC Premium Transparent Covers with Diagonal L...	Binders	Office Supplies

In [ ]:
```
top3 = top3[['Product ID','Product Name','TotalQty']]
top3
```

Out[ ]:

	Product ID	Product Name	TotalQty
0	TEC-AC-10003832	Logitech P710e Mobile Speakerphone	75
1	TEC-AC-10003832	Imation 16GB Mini TravelDrive USB 2.0 Flash Drive	75
2	OFF-PA-10001970	Xerox 1908	70
3	OFF-PA-10001970	Xerox 1881	70
4	OFF-BI-10001524	GBC Premium Transparent Covers with Diagonal L...	67

In [ ]:
```
top3_catgy = pd.merge(SaleCstsFct,DimPrd,on='Product ID')
top3_catgy.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	
									F X
1	2014-02-24	JH-15430	Medford	OFF-PA-10000174	4	32.896	0.2	11.1024	W F X
2	2014-09-07	LR-17035	Santa Clara	OFF-PA-10000174	2	20.560	0.0	9.6632	W F X
3	2014-09-26	DA-13450	Lancaster	OFF-PA-10000174	4	32.896	0.2	11.1024	W F X
4	2015-12-03	AB-10255	Philadelphia	OFF-PA-10000174	2	16.448	0.2	5.5512	W F X

In [ ]:
```
top3_catgy = top3_catgy.groupby('Product ID').agg(TotalQty = ('Category' , 'sum'))
```

In [ ]:
```
top3_catgy.head()
```

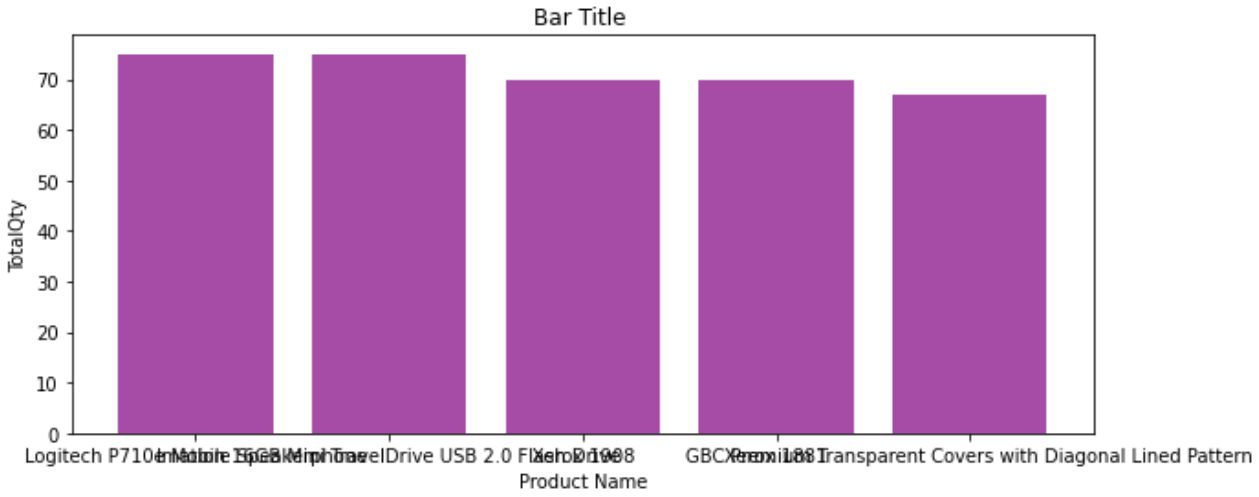
Out[ ]:

Product ID	TotalQty
FUR-BO-10000112	Furniture
FUR-BO-10000330	FurnitureFurnitureFurniture
FUR-BO-10000362	FurnitureFurnitureFurnitureFurnitureFurniture
FUR-BO-10000468	FurnitureFurnitureFurnitureFurnitureFurnitureF...
FUR-BO-10000711	FurnitureFurniture

In [ ]:
```
import matplotlib.pyplot as plt
```

In [ ]:
```
plt.figure(figsize=(10,4))
plt.title("Bar Title ")
plt.xlabel("Product Name")
plt.ylabel("TotalQty")
plt.bar(top3["Product Name"] , top3.TotalQty ,color='purple',alpha=0.7)
```

Out[ ]: <BarContainer object of 5 artists>



Top 3 Category

In [ ]:
```
ctg = pd.merge(SaleCstsFct,DimPrd,on='Product ID')
ctg.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	W F X
1	2014-02-24	JH-15430	Medford	OFF-PA-10000174	4	32.896	0.2	11.1024	W F X
2	2014-09-07	LR-17035	Santa Clara	OFF-PA-10000174	2	20.560	0.0	9.6632	W F X
3	2014-09-26	DA-13450	Lancaster	OFF-PA-10000174	4	32.896	0.2	11.1024	W F X
4	2015-12-03	AB-10255	Philadelphia	OFF-PA-10000174	2	16.448	0.2	5.5512	W F X

In [ ]:
```
ctg = ctg.groupby('Category').agg(TotalQty = ('SalesQtyTotal' , 'sum')).reset_index()
```

In [ ]:
```
ctg
```

Out[ ]:

	Category	TotalQty
0	Furniture	8379
1	Office Supplies	23518
2	Technology	7249

In [ ]:
```
ctg = ctg.nlargest(3,'TotalQty')
ctg
```

Out[ ]:

	Category	TotalQty
1	Office Supplies	23518
0	Furniture	8379
2	Technology	7249

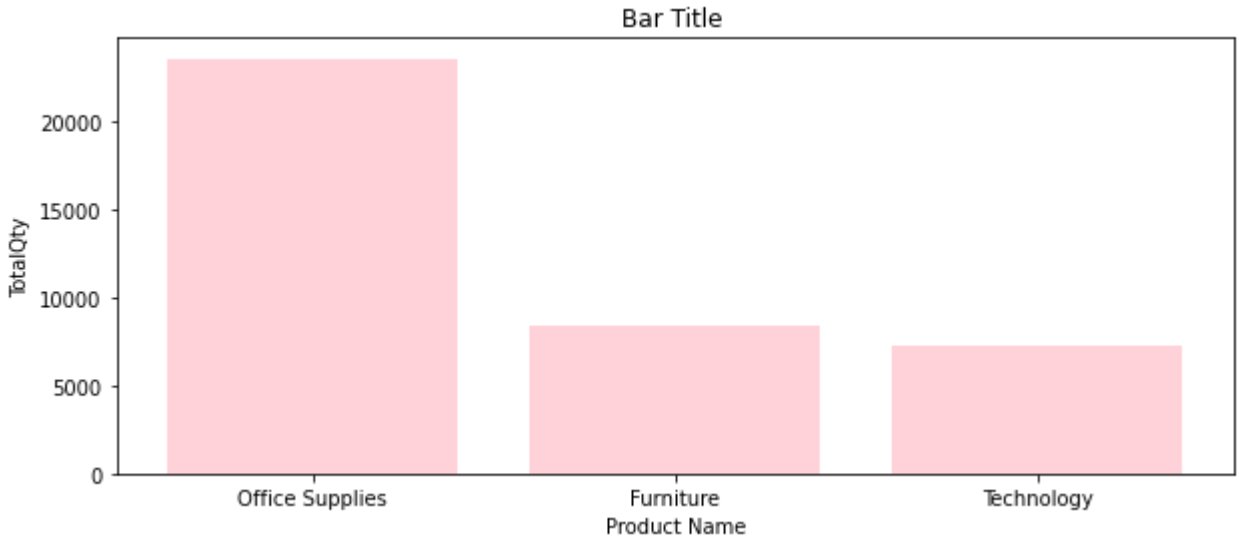
In [ ]:
```
ctg.sort_values(by = 'Category'  , ascending=False)
```

Out[ ]:

	Category	TotalQty
2	Technology	7249
1	Office Supplies	23518
0	Furniture	8379

In [ ]:
```
plt.figure(figsize=(10,4))
plt.title("Bar Title ")
plt.xlabel("Product Name")
plt.ylabel("TotalQty")
plt.bar(ctg.Category , ctg.TotalQty ,color='pink',alpha=0.7)
```

Out[ ]:
```
<BarContainer object of 3 artists>
```



Find top 3 sales Products lay on quarter 3

In [ ]:
```
date_Qtr = pd.merge(SaleCstsFct,DimDt ,on='Order Date')
date_Qtr.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	D
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	
1	2014-01-04	PO-19195	Naperville	OFF-BI-10004094	2	3.540	0.8	-5.4870	
2	2014-01-04	PO-19195	Naperville	OFF-LA-10003223	3	11.784	0.2	4.2717	
3	2014-01-04	PO-19195	Naperville	OFF-ST-10002743	3	272.736	0.2	-64.7748	
4	2014-01-05	MB-18085	Philadelphia	OFF-AR-10003478	3	19.536	0.2	4.8840	

In [ ]:
```
date_Qtr = pd.merge(date_Qtr,DimPrd ,on='Product ID')
date_Qtr.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	D
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	
1	2014-02-24	JH-15430	Medford	OFF-PA-10000174	4	32.896	0.2	11.1024	

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	D
2	2014-09-07	LR-17035	Santa Clara	OFF-PA-10000174	2	20.560	0.0	9.6632	
3	2014-09-26	DA-13450	Lancaster	OFF-PA-10000174	4	32.896	0.2	11.1024	
4	2015-12-03	AB-10255	Philadelphia	OFF-PA-10000174	2	16.448	0.2	5.5512	

In [ ]:
```
date_Qtr = date_Qtr[["Product ID" , "Product Name" ,"SalesQtyTotal" ,'Quarter']]
```

In [ ]:
```
date_Qtr = date_Qtr[date_Qtr.Quarter == 3]
```

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

Out[ ]:

	Product ID	Product Name	SalesQtyTotal	Quarter
2	OFF-PA-10000174	Message Book, Wirebound, Four 5 1/2" X 4" Form...	2	3
3	OFF-PA-10000174	Message Book, Wirebound, Four 5 1/2" X 4" Form...	4	3
5	OFF-PA-10000174	Message Book, Wirebound, Four 5 1/2" X 4" Form...	4	3
8	OFF-BI-10004094	GBC Standard Plastic Binding Systems Combs	2	3
9	OFF-BI-10004094	GBC Standard Plastic Binding Systems Combs	3	3

In [ ]:
```
date_Qtr = date_Qtr.groupby(['Product ID' , 'Product Name']).agg(TotalQty = ('SalesQtyT
```

In [ ]:
```
date_Qtr = date_Qtr.nlargest(3,['TotalQty'])
```

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

Out[ ]:

	Product ID	Product Name	TotalQty
551	OFF-BI-10002012	Wilson Jones Easy Flow II Sheet Lifters	32

	Product ID	Product Name	TotalQty
763	OFF-LA-10002762	Avery 485	32
972	OFF-PA-10004100	Xerox 216	32

Find top 3 Sub-Category sales product

In [ ]:
```
sub_ctg = pd.merge(SaleCstsFct,DimPrd,on='Product ID')
sub_ctg.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	W F X
1	2014-02-24	JH-15430	Medford	OFF-PA-10000174	4	32.896	0.2	11.1024	W F X
2	2014-09-07	LR-17035	Santa Clara	OFF-PA-10000174	2	20.560	0.0	9.6632	W F X
3	2014-09-26	DA-13450	Lancaster	OFF-PA-10000174	4	32.896	0.2	11.1024	W F X
4	2015-12-03	AB-10255	Philadelphia	OFF-PA-10000174	2	16.448	0.2	5.5512	W F X

In [ ]:
```
sub_ctg = sub_ctg.groupby('Sub-Category').agg(TotalQty = ('SalesQtyTotal' , 'sum')).res
```

In [ ]:
```
sub_ctg
```

Out[ ]:

	Sub-Category	TotalQty
0	Accessories	3138
1	Appliances	1758
2	Art	3030

	Sub-Category	TotalQty
3	Binders	6119
4	Bookcases	910
5	Chairs	2407
6	Copiers	234
7	Envelopes	906
8	Fasteners	914
9	Furnishings	3821
10	Labels	1400
11	Machines	462
12	Paper	5507
13	Phones	3415
14	Storage	3237
15	Supplies	647
16	Tables	1241

In [ ]:
```
sub_ctg = sub_ctg.nlargest(3, 'TotalQty')
sub_ctg
```

Out[ ]:

	Sub-Category	TotalQty
3	Binders	6119
12	Paper	5507
9	Furnishings	3821

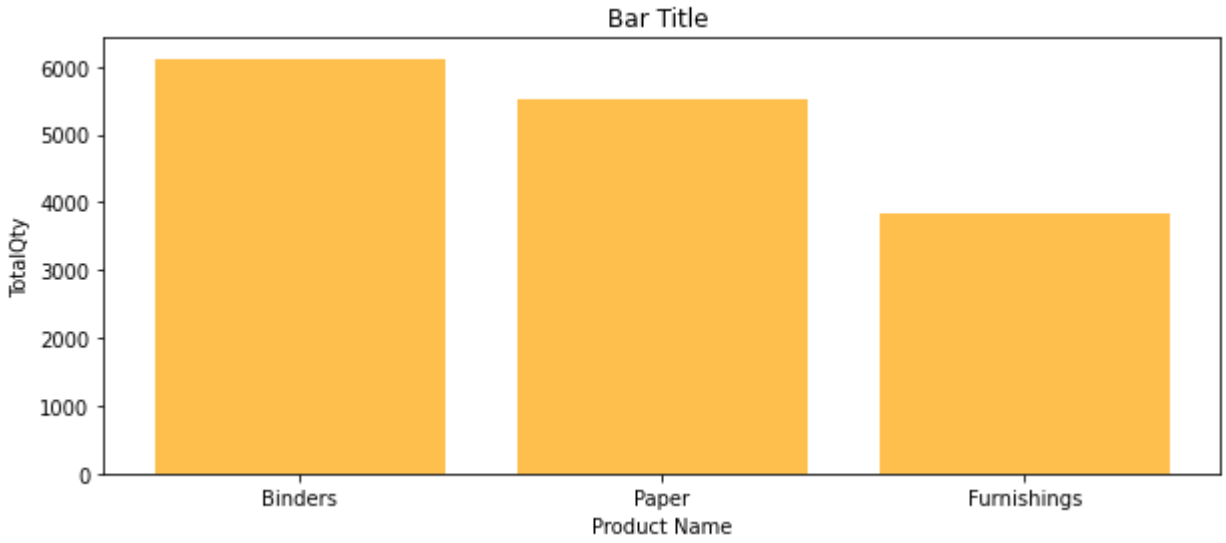
In [ ]:
```
sub_ctg.sort_values(by = "Sub-Category" , ascending=False)
```

Out[ ]:

	Sub-Category	TotalQty
12	Paper	5507
9	Furnishings	3821
3	Binders	6119

In [ ]:
```
plt.figure(figsize=(10,4))
plt.title("Bar Title ")
plt.xlabel("Product Name")
plt.ylabel("TotalQty")
plt.bar(sub_ctg['Sub-Category'] , sub_ctg.TotalQty ,color='orange',alpha=0.7)
```

Out[ ]:
```
<BarContainer object of 3 artists>
```



Find top 3 product lay on quarter 2

In [ ]:
```
sub_ctg_date_Qtr = pd.merge(SaleCstsFct,DimDt ,on='Order Date')
sub_ctg_date_Qtr.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	D
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	
1	2014-01-04	PO-19195	Naperville	OFF-BI-10004094	2	3.540	0.8	-5.4870	
2	2014-01-04	PO-19195	Naperville	OFF-LA-10003223	3	11.784	0.2	4.2717	
3	2014-01-04	PO-19195	Naperville	OFF-ST-10002743	3	272.736	0.2	-64.7748	
4	2014-01-05	MB-18085	Philadelphia	OFF-AR-10003478	3	19.536	0.2	4.8840	

In [ ]:
```
sub_ctg_date_Qtr = pd.merge(sub_ctg_date_Qtr,DimPrd ,on='Product ID')
sub_ctg_date_Qtr.head()
```

Out[ ]:

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	D
0	2014-01-03	DP-13000	Houston	OFF-PA-10000174	2	16.448	0.2	5.5512	
1	2014-02-24	JH-15430	Medford	OFF-PA-10000174	4	32.896	0.2	11.1024	

	Order Date	Customer ID	City	Product ID	SalesQtyTotal	SalesAmtTotal	DisAmtTotal	ProfAmtTotal	D
2	2014-09-07	LR-17035	Santa Clara	OFF-PA-10000174	2	20.560	0.0	9.6632	
3	2014-09-26	DA-13450	Lancaster	OFF-PA-10000174	4	32.896	0.2	11.1024	
4	2015-12-03	AB-10255	Philadelphia	OFF-PA-10000174	2	16.448	0.2	5.5512	

In [ ]:
```
sub_ctg_date_Qtr = sub_ctg_date_Qtr[["Product ID" , "Product Name" ,"SalesQtyTotal" ,'Q
```

In [ ]:
```
sub_ctg_date_Qtr = sub_ctg_date_Qtr[sub_ctg_date_Qtr.Quarter == 2]
```

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

Out[ ]:

	Product ID	Product Name	SalesQtyTotal	Quarter
16	OFF-LA-10003223	Avery 508	2	2
29	OFF-AR-10003478	Avery Hi-Liter EverBold Pen Style Fluorescent ...	6	2
30	OFF-AR-10003478	Avery Hi-Liter EverBold Pen Style Fluorescent ...	11	2
35	OFF-AR-10002399	Dixon Prang Watercolor Pencils, 10-Color Set w...	4	2
36	OFF-AR-10002399	Dixon Prang Watercolor Pencils, 10-Color Set w...	4	2

In [ ]:
```
sub_ctg_date_Qtr = sub_ctg_date_Qtr.groupby(['Product ID' , 'Product Name']).agg(TotalQ
```

In [ ]:
```
sub_ctg_date_Qtr = sub_ctg_date_Qtr.nlargest(3,['TotalQty'])
```

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

Out[ ]:

	Product ID	Product Name	TotalQty
586	OFF-BI-10003982	Wilson Jones Century Plastic Molded Ring Binders	28

	Product ID	Product Name	TotalQty
1090	TEC-AC-10003038	Kingston Digital DataTraveler 16GB USB 2.0	28
1104	TEC-AC-10003832	Imation 16GB Mini TravelDrive USB 2.0 Flash Drive	27

In [ ]:
```
sub_ctg_date_Qtr.sort_values(by = 'Product ID'  , ascending=False)
```

Out[ ]:

	Product ID	Product Name	TotalQty
1104	TEC-AC-10003832	Imation 16GB Mini TravelDrive USB 2.0 Flash Drive	27
1090	TEC-AC-10003038	Kingston Digital DataTraveler 16GB USB 2.0	28
586	OFF-BI-10003982	Wilson Jones Century Plastic Molded Ring Binders	28

In [ ]:
```

```