Sales Analysis

Importing libraries

```
In [1]: import pandas as pd
   import os
   from datetime import datetime as dt
   import matplotlib.pyplot as plt
```

Merging 12 months sales data

```
In [2]: files = [i for i in os.listdir('//Mac/Home/Downloads/Pandas-Data-Science-Tasks-mast
    df = pd.DataFrame()
    for file in files:
        df1 = pd.read_csv('//Mac/Home/Downloads/Pandas-Data-Science-Tasks-master/SalesA
        df = pd.concat([df,df1])

df.head()
```

Out[2]:		Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	0	295665	Macbook Pro Laptop	1	1700	12/30/19 00:01	136 Church St, New York City, NY 10001
	1	295666	LG Washing Machine	1	600.0	12/29/19 07:03	562 2nd St, New York City, NY 10001
	2	295667	USB-C Charging Cable	1	11.95	12/12/19 18:21	277 Main St, New York City, NY 10001
	3	295668	27in FHD Monitor	1	149.99	12/22/19 15:13	410 6th St, San Francisco, CA 94016

Data Cleaning

295669

USB-C Charging

Cable

```
In [3]: df.info()
```

1

11.95

12/18/19

12:38

43 Hill St, Atlanta, GA 30301

```
<class 'pandas.core.frame.DataFrame'>
        Int64Index: 186850 entries, 0 to 13621
        Data columns (total 6 columns):
             Column
                              Non-Null Count
                                               Dtype
            -----
                              -----
                                               ____
            Order ID
                              186305 non-null object
         0
             Product
                              186305 non-null object
         2
            Quantity Ordered 186305 non-null object
         3 Price Each
                              186305 non-null object
         4 Order Date
                               186305 non-null object
             Purchase Address 186305 non-null object
        dtypes: object(6)
        memory usage: 10.0+ MB
In [4]: df.isna().sum()
Out[4]: Order ID
                            545
        Product
                            545
        Quantity Ordered
                            545
        Price Each
                            545
        Order Date
                            545
        Purchase Address
                            545
        dtype: int64
In [5]: df.dropna(how='all',inplace=True)
In [6]: df.isna().sum()
Out[6]: Order ID
                            0
        Product
        Quantity Ordered
        Price Each
                            0
        Order Date
                            0
        Purchase Address
                            0
        dtype: int64
In [7]: df.value_counts()
```

Out[7]:	Order ID Product	Quantity Ordered	Price Each	Order Date
	Purchase Address			
	Order ID Product	Quantity Ordered	Price Each	Order Date
	Purchase Address	355		
	158236 AA Batteries (4-pack)	1	3.84	02/19/19 09:49
	319 West St, San Francisco, CA 94016	5 2		
	315204 Wired Headphones	1	11.99	12/12/19 12:41
	680 6th St, San Francisco, CA 94016	2		
	256196 USB-C Charging Cable	1	11.95	09/27/19 21:09
	253 6th St, Boston, MA 02215	2		
	256763 27in FHD Monitor	1	149.99	09/15/19 22:28
	23 11th St, San Francisco, CA 94016	2		
	•••			
	200687 Lightning Charging Cable	1	14.95	05/11/19 11:31
	878 7th St, Atlanta, GA 30301	1		
	200688 27in 4K Gaming Monitor	1	389.99	05/22/19 10:31
	731 Wilson St, Los Angeles, CA 90003	1 1		
	200689 USB-C Charging Cable	1	11.95	05/30/19 13:24
	804 13th St, Portland, ME 04101	1		
	200690 27in FHD Monitor	1	149.99	05/20/19 19:31
	781 Maple St, Los Angeles, CA 90001	1		
	230355 AA Batteries (4-pack)	2	3.84	07/03/19 16:37
	849 Maple St, Boston, MA 02215	1		
	Length: 185687, dtype: int64			

In [8]: df[df['Order ID']=='Order ID']

Out[8]:		Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	254	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	705	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	1101	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	2875	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	3708	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	•••						
	10443	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	10784	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	10813	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	11047	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
	13304	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address

355 rows × 6 columns

```
In [9]: df = df[~(df['Order ID']=='Order ID')]
In [10]: df.duplicated().sum()
```

```
Out[10]: 264
        df.drop_duplicates(inplace=True)
In [12]: df['Quantity Ordered'] = df['Quantity Ordered'].astype('int')
         df['Price Each'] = df['Price Each'].astype('float')
         df['Order Date'] = df['Order Date'].astype('datetime64')
In [13]: df.info()
         <class 'pandas.core.frame.DataFrame'>
         Int64Index: 185686 entries, 0 to 13621
         Data columns (total 6 columns):
             Column
                              Non-Null Count
                                               Dtype
            ____
                               -----
          0 Order ID
                              185686 non-null object
            Product
                              185686 non-null object
          1
          2 Quantity Ordered 185686 non-null int32
          3 Price Each
                             185686 non-null float64
                              185686 non-null datetime64[ns]
             Order Date
             Purchase Address 185686 non-null object
         dtypes: datetime64[ns](1), float64(1), int32(1), object(3)
         memory usage: 9.2+ MB
         What was the best month for sales? How much was earned that month?
In [14]: df['month'] = df['Order Date'].dt.strftime('%B')
         df['sales'] = df['Quantity Ordered']*df['Price Each']
```

In [15]: df.head()

Product Quantity Price Order Date Address month Sales			• • • • • • • • • • • • • • • • • • • •						
0 295665 Macbook Pro Laptop 1 1700.00 2019-12-30 00:01:00 New York City, NY 10001 December 1700.00 1 295666 LG Washing Machine 1 600.00 2019-12-29 07:03:00 562 2nd St, New York City, NY 10001 December 600.00 2 295667 Charging Cable 1 11.95 2019-12-12 18:21:00 New York City, NY 10001 December 11.95 3 295668 27in FHD Monitor 1 149.99 2019-12-22 15:13:00 410 6th St, San Francisco, CA 94016 December 149.99 4 295669 Charging 1 11.95 2019-12-18 12:38:00 43 Hill St, Atlanta, GA December 11.95	ut[15]:			Product	-			month	sales
1 295666 LG Washing Machine 1 600.00 2019-12-29 07:03:00 York City, NY 10001 December 600.00 2 295667 Charging Cable 1 11.95 2019-12-12 12 18:21:00 NY 10001 December 11.95 3 295668 27in FHD Monitor 1 149.99 2019-12-215:13:00 410 6th St, San Francisco, CA 94016 December 149.99 4 295669 Charging 1 11.95 2019-12-18 12:38:00 43 Hill St, Atlanta, GA December 11.95		0	295665		1	1700.00	New York City,	December	1700.00
2 295667 Charging Cable 1 11.95 2019-12- 12 18:21:00 New York City, NY 10001 3 295668 27in FHD Monitor 1 149.99 2019-12- 22 15:13:00 Francisco, CA 94016 USB-C 2019-12- 18 12:38:00 43 Hill St, Atlanta, GA December 11.95		1	295666	_	1	600.00	York City, NY	December	600.00
3 295668 Monitor 1 149.99 2019-12- 22 15:13:00 Francisco, CA December 149.99 94016 USB-C 2019-12- 4 295669 Charging 1 11.95 2019-12- 18 12:38:00 Atlanta, GA December 11.95		2	295667	Charging	1	11.95	New York City,	December	11.95
4 295669 Charging 1 11.95 2019-12- Atlanta, GA December 11.95		3	295668		1	149.99	Francisco, CA	December	149.99
		4	295669	Charging	1	11.95	Atlanta, GA	December	11.95

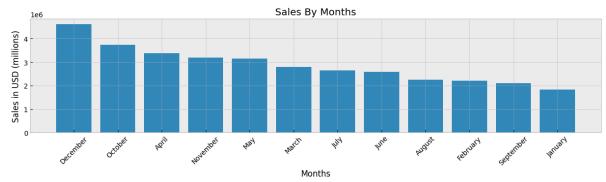
In [16]: (df.groupby('month').agg({'sales':'sum'}).sort_values('sales',ascending=False).rese

December

4608295.7

Out[16]: month

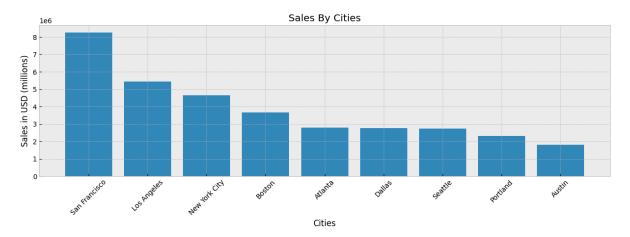
sales



Which City had the highest Sales?

```
In [18]: #df['city'] = df['Purchase Address'].str.split(',').str[1].str.strip() or
    df['city'] = df['Purchase Address'].apply(lambda x : x.split(',')[1].strip())
In [19]: df.head()
```

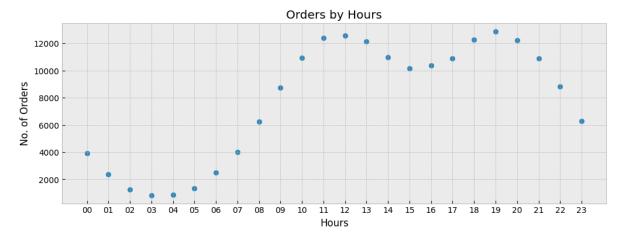
```
Out[19]:
                                  Quantity
                                                       Order
                                                                Purchase
               Order
                                              Price
                        Product
                                                                             month
                                                                                       sales
                                                                                                   city
                                  Ordered
                  ID
                                              Each
                                                        Date
                                                                 Address
                                                               136 Church
                                                    2019-12-
                       Macbook
                                                                                              New York
                                                                  St, New
                                                                          December 1700.00
             295665
                            Pro
                                           1700.00
                                         1
                                                          30
                                                                 York City,
                                                                                                   City
                         Laptop
                                                     00:01:00
                                                                NY 10001
                                                               562 2nd St,
                                                    2019-12-
                             LG
                                                                New York
                                                                                              New York
                                                                                      600.00
           1 295666
                        Washing
                                         1
                                             600.00
                                                          29
                                                                           December
                                                                  City, NY
                                                                                                   City
                                                     07:03:00
                        Machine
                                                                   10001
                                                                277 Main
                          USB-C
                                                    2019-12-
                                                                  St, New
                                                                                              New York
                                                                          December
                                                                                       11.95
           2 295667
                       Charging
                                         1
                                              11.95
                                                          12
                                                                 York City,
                                                                                                   City
                          Cable
                                                     18:21:00
                                                                NY 10001
                                                               410 6th St,
                                                    2019-12-
                       27in FHD
                                                                     San
                                                                                                   San
           3 295668
                                         1
                                             149.99
                                                          22
                                                                           December
                                                                                      149.99
                        Monitor
                                                                Francisco,
                                                                                              Francisco
                                                     15:13:00
                                                                CA 94016
                          USB-C
                                                    2019-12-
                                                                43 Hill St,
           4 295669
                                                               Atlanta, GA
                                                                                       11.95
                                                                                                Atlanta
                       Charging
                                         1
                                              11.95
                                                          18
                                                                          December
                          Cable
                                                     12:38:00
                                                                   30301
In [20]:
           df.groupby('city').agg({'sales':'sum'}).sort_values('sales',ascending=False).reset_
Out[20]: city
                     San Francisco
           sales
                        8254743.55
           Name: 0, dtype: object
In [21]: SalesSumByCity = df.groupby('city').agg({'sales':'sum'}).sort_values('sales',ascend
           City = SalesSumByCity['city']
           Sales = SalesSumByCity['sales']
           plt.figure(figsize=(15,4))
           plt.title('Sales By Cities')
           plt.xticks(rotation=45)
           plt.xlabel('Cities')
           plt.ylabel('Sales in USD (millions)')
           plt.bar(City,Sales)
           plt.show()
```



What time should we display advertisements to maximize likelihood of customer's buying product?

```
In [22]: SalesByHour = df.groupby(df['Order Date'].dt.strftime('%H')).agg({'Order ID':'count SalesByHour.rename(columns={'Order Date':'Hour'},inplace=True)
hour = SalesByHour['Hour']
OrderCount = SalesByHour['Order ID']

plt.figure(figsize=(12,4))
plt.scatter(hour,OrderCount)
plt.xlabel('Hours')
plt.ylabel('No. of Orders')
plt.title('Orders by Hours')
plt.show()
```



In [23]: #around 12noon and 7pm

What products are most often sold together?

```
In [24]: df.head()
```

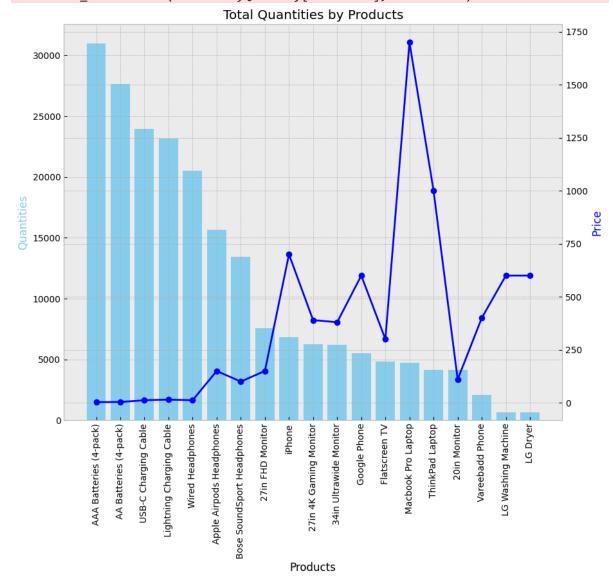
```
Out[24]:
               Order
                                  Quantity
                                              Price
                                                        Order
                                                                 Purchase
                        Product
                                                                              month
                                                                                        sales
                                                                                                    city
                  ID
                                  Ordered
                                               Each
                                                        Date
                                                                  Address
                                                               136 Church
                       Macbook
                                                     2019-12-
                                                                   St. New
                                                                                               New York
              295665
                            Pro
                                            1700.00
                                                                           December 1700.00
                                                           30
                                                                 York City,
                                                                                                    City
                         Laptop
                                                      00:01:00
                                                                 NY 10001
                                                                562 2nd St,
                                                     2019-12-
                             LG
                                                                 New York
                                                                                               New York
                                                                                       600.00
             295666
                        Washing
                                         1
                                             600.00
                                                           29
                                                                           December
                                                                                                    City
                                                                  City, NY
                        Machine
                                                      07:03:00
                                                                    10001
                                                                 277 Main
                          USB-C
                                                     2019-12-
                                                                   St, New
                                                                                               New York
                                                                                        11.95
                                                                           December
             295667
                        Charging
                                         1
                                              11.95
                                                          12
                                                                 York City,
                                                                                                    City
                          Cable
                                                      18:21:00
                                                                 NY 10001
                                                                410 6th St,
                                                     2019-12-
                       27in FHD
                                                                      San
                                                                                                    San
           3 295668
                                         1
                                             149.99
                                                          22
                                                                           December
                                                                                       149.99
                        Monitor
                                                                 Francisco,
                                                                                               Francisco
                                                      15:13:00
                                                                 CA 94016
                          USB-C
                                                     2019-12-
                                                                 43 Hill St,
                                                                                        11.95
             295669
                        Charging
                                         1
                                              11.95
                                                           18
                                                               Atlanta, GA
                                                                           December
                                                                                                 Atlanta
                          Cable
                                                      12:38:00
                                                                    30301
           df['Order ID'].duplicated().sum()
In [25]:
Out[25]: 7249
           ProductByOrder = df.groupby('Order ID').agg(Products=('Product',lambda x: ','.join(
                                                              ,ProductCount=('Product','count')).rese
           MultipleProductByOrder = ProductByOrder[ProductByOrder['ProductCount']>1][['Order I
           MultipleProductByOrder.groupby('Products').agg(CountOfOrders=('Order ID','count'))
                                     .sort_values('CountOfOrders',ascending=False).reset_index().i
                                         Products CountOfOrders
Out[26]:
           0
                     iPhone, Lightning Charging Cable
                                                              886
           1
                 Google Phone, USB-C Charging Cable
                                                              857
           2
                          iPhone, Wired Headphones
                                                              361
              Vareebadd Phone, USB-C Charging Cable
                                                              312
           4
                                                              303
                    Google Phone, Wired Headphones
In [27]: # iPhone with Lightning Charging Cable most often sold together (886 times)
           # followed by Google Phone with USB-C Charging Cable (857 times)
```

What product sold the most? Why do you think it sold the most?

```
In [28]: df.head()
```

Out[28]:		Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month	sales	city
	0	295665	Macbook Pro Laptop	1	1700.00	2019-12- 30 00:01:00	136 Church St, New York City, NY 10001	December	1700.00	New York City
	1	295666	LG Washing Machine	1	600.00	2019-12- 29 07:03:00	562 2nd St, New York City, NY 10001	December	600.00	New York City
	2	295667	USB-C Charging Cable	1	11.95	2019-12- 12 18:21:00	277 Main St, New York City, NY 10001	December	11.95	New York City
	3	295668	27in FHD Monitor	1	149.99	2019-12- 22 15:13:00	410 6th St, San Francisco, CA 94016	December	149.99	San Francisco
	4	295669	USB-C Charging Cable	1	11.95	2019-12- 18 12:38:00	43 Hill St, Atlanta, GA 30301	December	11.95	Atlanta
In [29]:							antity Orde .reset_inde			
In [29]: Out[29]:			es('TotalQ		,ascendi					
		rt_value	es('TotalQ	Quantity'	,ascendi					
	SO	rt_value	es('TotalQ	Quantity' Luct Total ack)	, ascendi					
	0 1 2	AAA B AA B USB-C	Procent (4-p) atteries (4-p) Charging Care	Quantity' Luct Total ack) ack) able	Quantity 30986					
	0 1 2 3	AAA B AA B USB-C	Procestic 'Total Control Contr	Quantity' Luct Total ack) ack) able able	30986 27615 23931 23169					
	0 1 2	AAA B AA B USB-C	Procent (4-p) atteries (4-p) Charging Care	Quantity' Luct Total ack) ack) able able	30986 27615 23931					
Out[29]:	0 1 2 3 4	AAA B AA B USB-C Lightning	Procent atteries (4-posteries (Quantity' Luct Total ack) ack) able able ones	30986 27615 23931 23169 20524	ng= False)		ex().iloc[uantity	
Out[29]:	0 1 2 3 4	AAA B AA B USB-C Lightning Win	Procent atteries (4-posteries (duct Total ack) ack) able able ones = df.grou	Quantity 30986 27615 23931 23169 20524	oduct').a	.reset_inde	ex().iloc[uantity	
Out[29]:	0 1 2 3 4 Pr fi ax ax ax ax	AAA B AA B USB-C Lightning Win oductsBy g,ax1 = 1.bar(Pr 1.set_x1 1.set_x1 1.set_y1	Process('TotalQ Process('TotalQ Process('-TotalQ Process('-Tota	duct Total ack) ack) able bnes = df.grou ots(figs: ducts') c(Products) antities'	Quantity 30986 27615 23931 23169 20524 upby('Product sByQuant: ,color='	oduct').a .si 8)) '], Produ ity['Prod skyblue')	gg(TotalQuant_values	ex().iloc[entity=('Q ('TotalQua	Quantity	ascending=
Out[29]:	o 1 2 3 4 Pr fi ax ax ax ax ax	AAA B AA B USB-C Lightning Win oductsBy g,ax1 = 1.bar(Pr 1.set_x1 1.set_x1 1.set_ti 2 = ax1. 2.plot(F	Process('TotalQ Process('TotalQ Process('	duct Total ack) ack) able bnes = df.grou ots(figs: ducts') c(Products) clarities' al Quantity	Quantity 30986 27615 23931 23169 20524 upby('Product sByQuant: color=': ties by	oduct').a .s 8)) '], Produ ity['Prod skyblue') Products'	gg(TotalQuant_values	ex().iloc[entity=('Q ('TotalQua ity['Total tion=90)	Quantity Quantity	ascending: y'], color
ut[29]:	o 1 2 3 4 Pr fi ax ax ax ax ax ax	AAA B AA B USB-C Lightning Win oductsBy g,ax1 = 1.bar(Pr 1.set_x1 1.set_x1 1.set_ti 2 = ax1. 2.plot(F	Process('TotalQ Process('TotalQ Process('	duct Total ack) ack) able bnes = df.grou ots(figs: ducts') c(Products) clarities' al Quantity	Quantity 30986 27615 23931 23169 20524 upby('Product sByQuant: color=': ties by	oduct').a .s 8)) '], Produ ity['Prod skyblue') Products'	gg(TotalQuant_values	ex().iloc[entity=('Q ('TotalQua ity['Total tion=90)	Quantity Quantity	ascending: y'], color

C:\Users\atish\AppData\Local\Temp\ipykernel_8984\81398722.py:8: UserWarning: Fixed
Formatter should only be used together with FixedLocator
 ax1.set_xticklabels(ProductsByQuantity['Product'],rotation=90)



In [31]: # AAA Batteries(4-pack) sold the most (total quantities : 30986)
One possible reason : Price is the lowest among all products