Sales Analysis

Importing necessary libraries

```
import os
import pandas as pd
import glob
```

Merging 12 months of sales Data

```
path_sales = r'C:\Users\Dell\Desktop\Sales_Data'
```

```
all_files_sales = glob.glob(path_sales + "/*.csv")
```

```
[]
```

```
all_data = pd.concat([pd.read_csv(f) for f in all_files_sales])
```

```
all_data.shape
```

```
(186850, 6)
```

```
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st Street, Los Angeles, CA 90001
1	NaN	NaN	NaN	NaN	NaN	NaN
2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	176561	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

```
all_data.tail()
```

```
[]
```

```
all_data.to_csv("all_data.csv")
```

```
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
NameError: name 'all_data' is not defined
```

```
all_data = pd.read_csv("all_data.csv",index_False)
```

```
all_data.shape
```

```
(186850, 7)
```

```
all_data.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st Street, Los Angeles, CA 90001
1	1	NaN	NaN	NaN	NaN	NaN	NaN
2	2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	4	176561	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

```
all_data.tail()
```

```
[]
```

Reading Merged Data File

```
df = pd.read_csv("merged_data.csv")
```

```
df.shape
```

```
(186850, 7)
```

```
df.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st Street, Los Angeles, CA 90001
1	1	NaN	NaN	NaN	NaN	NaN	NaN
2	2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	4	176561	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001

```
df.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 186850 entries, 0 to 186849
Data columns (total 7 columns):
# Column Unnamed: 0 Non-Null Count Dtype
---  ------ --------- -------------- -----
0  Unnamed: 0  186850 non-null  int64
1  Order ID    186305 non-null  object
2  Product     186305 non-null  object
3  Quantity Ordered 186305 non-null  object
4  Price Each  186305 non-null  object
5  Order Date  186305 non-null  object
6  Purchase Address 186305 non-null  object
dtypes: int64(1), object(6)
memory usage: 9.9 MB
```

Cleaning Data

Drop Rows of NaN

```
df = df.dropna(axis=0,inplace = 1)
```

```
df.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st Street, Los Angeles, CA 90001
2	2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	4	176561	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
5	5	176562	USB-C Charging Cable	1	11.95	04/29/19 13:33	333 Hickory St, New York City, NY 10001
6	6	176563	Bose SoundSport Headphones	1	99.99	04/18/19 20:30	733 Ash St, San Francisco, CA 94016
7	7	176564	USB-C Charging Cable	1	11.95	04/26/19 11:48	381 7th St, Boston, MA 02215
8	8	176565	Macbook Pro Laptop	1	1700	04/10/19 17:00	790 Ridge St, Atlanta, GA 30301
9	9	176566	Wired Headphones	1	11.99	04/18/19 01:18	915 Dogwood St, San Francisco, CA 94016
10	10	176567	Google Phone	1	600	04/05/19 14:39	861 Maple St, San Francisco, CA 94016
11	11	176568	USB-C Charging Cable	1	11.95	04/29/19 17:39	381 7th St, Boston, MA 02215
12	12	176569	Apple Airpods Headphones	1	150	04/07/19 12:04	668 Center St, Los Angeles, CA 90001
13	13	176570	USB-C Charging Cable	1	11.95	04/23/19 19:09	790 Ridge St, Atlanta, GA 30301
14	14	176571	Bose SoundSport Headphones	1	99.99	04/16/19 07:53	444 Maple St, San Francisco, CA 94016
15	15	176572	High Speed Internet	1	70	04/24/19 10:41	824 Sunset St, San Francisco, CA 94016
16	16	176573	USB-C Charging Cable	1	11.95	04/07/19 16:06	915 Dogwood St, San Francisco, CA 94016
17	17	176574	Google Phone	1	600	04/03/19 15:54	335 7th St, Los Angeles, CA 90001
18	18	176575	Wired Headphones	1	11.99	04/18/19 19:02	916 Main St, Los Angeles, CA 90001
19	19	176576	Apple Airpods Headphones	1	150	04/26/19 12:43	381 Hickory St, Seattle, WA 98101
20	20	176577	Flatscreen TV	1	300.00	04/10/19 20:31	775 River St, Dallas, TX 75001
21	21	176578	Bose SoundSport Headphones	1	99.99	04/21/19 14:32	494 Hickory St, San Francisco, CA 94016
22	22	176579	USB-C Charging Cable	1	11.95	04/17/19 19:44	916 Place St, Seattle, WA 98101
23	23	176580	Lightning Charging Cable	1	14.95	04/11/19 12:10	610 Chestnut St, Boston, MA 02215
24	24	176581	Flatscreen TV	1	300.00	04/29/19 07:35	771 Maple St, San Francisco, CA 94016
25	25	176582	Lightning Charging Cable	1	14.95	04/22/19 15:09	826 Elm St, San Francisco, CA 94016
26	26	176583	Lightning Charging Cable	1	14.95	04/30/19 09:00	843 Main St, Los Angeles, CA 90001
27	27	176584	Google Phone	1	600.00	04/04/19 08:00	548 Main St, Los Angeles, CA 90001
28	28	176585	Bose SoundSport Headphones	1	99.99	04/07/19 83:00	790 Ridge St, Los Angeles, CA 90001
29	29	176586	AAA Batteries (4-pack)	2	3.84	04/10/19 10:39	365 Dogwood St, San Francisco, CA 94016
30	30	176587	AAA Batteries (4-pack)	2	3.84	04/07/19 12:12	548 Main St, Los Angeles, CA 90001
31	31	176588	AA Batteries (4-pack)	1	3.84	04/19/19 08:44	917 1st St, Los Angeles, CA 90001

```
df.shape
```

```
(186305, 7)
```

```
df.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 186850 entries, 0 to 186849
Data columns (total 7 columns):
# Column Unnamed: 0 Non-Null Count Dtype
---  ------ --------- -------------- -----
0  Unnamed: 0  186850 non-null  int64
1  Order ID    186305 non-null  object
2  Product     186305 non-null  object
3  Quantity Ordered 186305 non-null  object
4  Price Each  186305 non-null  object
5  Order Date  186305 non-null  object
6  Purchase Address 186305 non-null  object
dtypes: int64(1), object(6)
memory usage: 9.9 MB
```

Removing Header Which get added multiple times while merging Months data

```
df = df[df['Order ID'] != df['Order ID']]
```

```
df.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st Street, Los Angeles, CA 90001
2	2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	4	176561	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
5	5	176562	USB-C Charging Cable	1	11.95	04/29/19 13:33	333 Hickory St, New York City, NY 10001
6	6	176563	Bose SoundSport Headphones	1	99.99	04/18/19 20:30	733 Ash St, San Francisco, CA 94016
7	7	176564	USB-C Charging Cable	1	11.95	04/26/19 11:48	381 7th St, Boston, MA 02215
8	8	176565	Macbook Pro Laptop	1	1700	04/10/19 17:00	790 Ridge St, Atlanta, GA 30301
9	9	176566	Wired Headphones	1	11.99	04/18/19 01:18	915 Dogwood St, San Francisco, CA 94016
10	10	176567	Google Phone	1	600	04/05/19 14:39	861 Maple St, San Francisco, CA 94016
11	11	176568	USB-C Charging Cable	1	11.95	04/29/19 17:39	381 7th St, Boston, MA 02215
12	12	176569	Apple Airpods Headphones	1	150	04/07/19 12:04	668 Center St, Los Angeles, CA 90001
13	13	176570	USB-C Charging Cable	1	11.95	04/23/19 19:09	790 Ridge St, Atlanta, GA 30301
14	14	176571	Bose SoundSport Headphones	1	99.99	04/16/19 07:53	444 Maple St, San Francisco, CA 94016
15	15	176572	High Speed Internet	1	70	04/24/19 10:41	824 Sunset St, San Francisco, CA 94016
16	16	176573	USB-C Charging Cable	1	11.95	04/07/19 16:06	915 Dogwood St, San Francisco, CA 94016
17	17	176574	Google Phone	1	600	04/03/19 15:54	335 7th St, Los Angeles, CA 90001
18	18	176575	Wired Headphones	1	11.99	04/18/19 19:02	916 Main St, Los Angeles, CA 90001
19	19	176576	Apple Airpods Headphones	1	150	04/26/19 12:43	381 Hickory St, Seattle, WA 98101
20	20	176577	Flatscreen TV	1	300.00	04/10/19 20:31	775 River St, Dallas, TX 75001
21	21	176578	Bose SoundSport Headphones	1	99.99	04/21/19 14:32	494 Hickory St, San Francisco, CA 94016
22	22	176579	USB-C Charging Cable	1	11.95	04/17/19 19:44	916 Place St, Seattle, WA 98101
23	23	176580	Lightning Charging Cable	1	14.95	04/11/19 12:10	610 Chestnut St, Boston, MA 02215
24	24	176581	Flatscreen TV	1	300.00	04/29/19 07:35	771 Maple St, San Francisco, CA 94016
25	25	176582	Lightning Charging Cable	1	14.95	04/22/19 15:09	826 Elm St, San Francisco, CA 94016
26	26	176583	Lightning Charging Cable	1	14.95	04/30/19 09:00	843 Main St, Los Angeles, CA 90001
27	27	176584	Google Phone	1	600.00	04/04/19 08:00	548 Main St, Los Angeles, CA 90001
28	28	176585	Bose SoundSport Headphones	1	99.99	04/07/19 83:00	790 Ridge St, Los Angeles, CA 90001
29	29	176586	AAA Batteries (4-pack)	2	3.84	04/10/19 10:39	365 Dogwood St, San Francisco, CA 94016
30	30	176587	AAA Batteries (4-pack)	2	3.84	04/07/19 12:12	548 Main St, Los Angeles, CA 90001
31	31	176588	AA Batteries (4-pack)	1	3.84	04/19/19 08:44	917 1st St, Los Angeles, CA 90001

```
df.shape
```

```
(186305, 7)
```

```
df.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 186850 entries, 0 to 186849
Data columns (total 7 columns):
# Column Unnamed: 0 Non-Null Count Dtype
---  ------ --------- -------------- -----
0  Unnamed: 0  186850 non-null  int64
1  Order ID    186305 non-null  object
2  Product     186305 non-null  object
3  Quantity Ordered 186305 non-null  object
4  Price Each  186305 non-null  object
5  Order Date  186305 non-null  object
6  Purchase Address 186305 non-null  object
dtypes: int64(1), object(6)
memory usage: 9.9 MB
```

Removing Duplicates

```
df.drop_duplicates(inplace=True) #df_dedupe = df.drop_duplicates()
```

```
df.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	0	176558	USB-C Charging Cable	2	11.95	04/19/19 08:46	917 1st Street, Los Angeles, CA 90001
2	2	176559	Bose SoundSport Headphones	1	99.99	04/07/19 22:30	682 Chestnut St, Boston, MA 02215
3	3	176560	Google Phone	1	600	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
4	4	176561	Wired Headphones	1	11.99	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001
5	5	176562	USB-C Charging Cable	1	11.95	04/29/19 13:33	333 Hickory St, New York City, NY 10001
6	6	176563	Bose SoundSport Headphones	1	99.99	04/18/19 20:30	733 Ash St, San Francisco, CA 94016
7	7	176564	USB-C Charging Cable	1	11.95	04/26/19 11:48	381 7th St, Boston, MA 02215
8	8	176565	Macbook Pro Laptop	1	1700	04/10/19 17:00	790 Ridge St, Atlanta, GA 30301
9	9	176566	Wired Headphones	1	11.99	04/18/19 01:18	915 Dogwood St, San Francisco, CA 94016
10	10	176567	Google Phone	1	600	04/05/19 14:39	861 Maple St, San Francisco, CA 94016
11	11	176568	USB-C Charging Cable	1	11.95	04/29/19 17:39	381 7th St, Boston, MA 02215
12	12	176569	Apple Airpods Headphones	1	150	04/07/19 12:04	668 Center St, Los Angeles, CA 90001
13	13	176570	USB-C Charging Cable	1	11.95	04/23/19 19:09	790 Ridge St, Atlanta, GA 30301
14	14	176571	Bose SoundSport Headphones	1	99.99	04/16/19 07:53	444 Maple St, San Francisco, CA 94016
15	15	176572	High Speed Internet	1	70	04/24/19 10:41	824 Sunset St, San Francisco, CA 94016
16	16	176573	USB-C Charging Cable	1	11.95	04/07/19 16:06	915 Dogwood St, San Francisco, CA 94016
17	17	176574	Google Phone	1	600	04/03/19 15:54	335 7th St, Los Angeles, CA 90001
18	18	176575	Wired Headphones	1	11.99	04/18/19 19:02	916 Main St, Los Angeles, CA 90001
19	19	176576	Apple Airpods Headphones	1	150	04/26/19 12:43	381 Hickory St, Seattle, WA 98101
20	20	176577	Flatscreen TV	1	300.00	04/10/19 20:31	775 River St, Dallas, TX 75001
21	21	176578	Bose SoundSport Headphones	1	99.99	04/21/19 14:32	494 Hickory St, San Francisco, CA 94016
22	22	176579	USB-C Charging Cable	1	11.95	04/17/19 19:44	916 Place St, Seattle, WA 98101
23	23	176580	Lightning Charging Cable	1	14.95	04/11/19 12:10	610 Chestnut St, Boston, MA 02215
24	24	176581	Flatscreen TV	1	300.00	04/29/19 07:35	771 Maple St, San Francisco, CA 94016
25	25	176582	Lightning Charging Cable	1	14.95	04/22/19 15:09	826 Elm St, San Francisco, CA 94016
26	26	176583	Lightning Charging Cable	1	14.95	04/30/19 09:00	843 Main St, Los Angeles, CA 90001
27	27	176584	Google Phone	1	600.00	04/04/19 08:00	548 Main St, Los Angeles, CA 90001
28	28	176585	Bose SoundSport Headphones	1	99.99	04/07/19 83:00	790 Ridge St, Los Angeles, CA 90001
29	29	176586	AAA Batteries (4-pack)	2	3.84	04/10/19 10:39	365 Dogwood St, San Francisco, CA 94016
30	30	176587	AAA Batteries (4-pack)	2	3.84	04/07/19 12:12	548 Main St, Los Angeles, CA 90001
31	31	176588	AA Batteries (4-pack)	1	3.84	04/19/19 08:44	917 1st St, Los Angeles, CA 90001

```
df.shape
```

```
(186305, 7)
```

```
df.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 186850 entries, 0 to 186849
Data columns (total 7 columns):
# Column Unnamed: 0 Non-Null Count Dtype
---  ------ --------- -------------- -----
0  Unnamed: 0  186850 non-null  int64
1  Order ID    186305 non-null  object
2  Product     186305 non-null  object
3  Quantity Ordered 186305 non-null  object
4  Price Each  186305 non-null  object
5  Order Date  186305 non-null  object
6  Purchase Address 186305 non-null  object
dtypes: int64(1), object(6)
memory usage: 9.9 MB
```

Converting Order Date Column to Date Format

```
df.loc[0, 'Order Date']
```

```
'04/19/19 08:46'
```

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

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_12480\1280938.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

```
df['Order Date'].dt.year
```

```
0  2019
2  2019
3  2019
4  2019
5  2019
6  2019
7  2019
8  2019
9  2019
10  2019
11  2019
12  2019
13  2019
14  2019
15  2019
16  2019
17  2019
18  2019
19  2019
20  2019
21  2019
22  2019
23  2019
24  2019
25  2019
26  2019
27  2019
28  2019
29  2019
30  2019
31  2019
Name: Order Date, dtype: int64
```

```
df.loc[0, 'Order Date']
```

```
Timestamp('2019-04-19 08:46:00')
```

Adding Month Columns

```
df['Month'] = df['Order Date'].dt.month_name()
```

```
df.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month
0	0	176558	USB-C Charging Cable	2	11.95	2019-04-19 08:46:00	917 1st Street, Los Angeles, CA 90001	January
2	2	176559	Bose SoundSport Headphones	1	99.99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	April
3	3	176560	Google Phone	1	600	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	April
4	4	176561	Wired Headphones	1	11.99	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	April
5	5	176562	USB-C Charging Cable	1	11.95	2019-04-29 13:33:00	333 Hickory St, New York City, NY 10001	April

Changing Data Type Of Other Columns for Numerical Computation

```
df.info()
```

```
<class'pandas.core.frame.DataFrame'>
RangeIndex: 186850 entries, 0 to 186849
Data columns (total 8 columns):
# Column Unnamed: 0 Non-Null Count Dtype
---  ------ --------- -------------- -----
0  Unnamed: 0  186850 non-null  int64
1  Order ID    186305 non-null  object
2  Product     186305 non-null  object
3  Quantity Ordered 186305 non-null  object
4  Price Each  186305 non-null  object
5  Order Date  186305 non-null  datetime64[ns]
6  Purchase Address 186305 non-null  object
7  Month       186305 non-null  object
dtypes: int64(1), datetime64[ns](1), object(6)
memory usage: 11.4 MB
```

Sorting As per Months

```
df.sort_values(by=['Month'], inplace=True) #df.sort_values('Month')
```

```
df['Month'].value_counts().sort_values(ascending=True)
```

```
Month
December  14069.0
November  17604.0
October   18744.0
September 18732.0
August    19024.0
July      19798.0
June      19237.0
May       19456.0
April     20689.0
March     17872.0
January   10345.0
Name: count, dtype: float64
```

```
df.groupby('Month')['Month'].count().sort_values(ascending = True)
```

```
Month
December  14069
November  17604
October   18744
September 18732
August    19024
July      19798
June      19237
May       19456
April     20689
March     17872
January   10345
Name: count, dtype: int64
```

Analysis

Finding which sale the best Month for sales and How much was Earned that Months

```
df['Price Each'] = df['Price Each'].astype(int)
```

```
df.groupby('Month')['Price Each'].sum()
```

```
Month
April     2.080000e+06
August    1.909000e+06
December  1.411000e+06
July      1.988000e+06
June      1.931000e+06
March     1.794000e+06
May       1.954000e+06
November  1.766000e+06
October   1.882000e+06
September 1.880000e+06
Name: Price Each, dtype: float64
```

December was the Best Month for Sales with sales of 1411000.0F and January was worst month with sales of 1022258.72

```
df.groupby('Month')['Price Each'].sum()
```

```
Month
April     2.080000e+06
August    1.909000e+06
December  1.411000e+06
July      1.988000e+06
June      1.931000e+06
March     1.794000e+06
May       1.954000e+06
November  1.766000e+06
October   1.882000e+06
September 1.880000e+06
Name: Price Each, dtype: float64
```

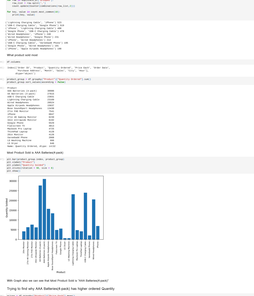
```
df['Price Each'] = df['Price Each'].astype(int)
```

```
df.groupby('Month')['Price Each'].sum()
```

```
Month
April     2.080000e+06
August    1.909000e+06
December  1.411000e+06
July      1.988000e+06
June      1.931000e+06
March     1.794000e+06
May       1.954000e+06
November  1.766000e+06
October   1.882000e+06
September 1.880000e+06
Name: Price Each, dtype: float64
```

```
import matplotlib.pyplot as plt
%matplotlib inline
plt.figure(figure=(10,6))
plt.title('Total Sales per Month')
plt.xlabel('Month')
plt.ylabel('Sales')
plt.show()
```

```
ERROR: Failed to execute line 3 in cell 'Total Sales per Month'
ERROR: TypeWarning: Ignoring unknown usage of %matplotlib at line 2 in cell 'Total Sales per Month'
```



With Graph we can also see that December has highest sales and January has Lowest Sales

Finding which City was the Highest Number of sales

Adding City Column

```
df['CityS(ST)'] = df['Purchase Address'].str.split(',').str[1] #df.head(2)
```

```
df['CityS(ST)'].value_counts()
```

```
CityS(ST)
Los Angeles  30261
New York City  25035
San Francisco  50239
Boston  20833
Atlanta  16602
Dallas  16730
Seattle  16553
Name: count, dtype: int64
```

```
df.groupby('CityS(ST)')['Price Each'].sum()
```

```
CityS(ST)
Atlanta  2.768000e+06
Boston  3.637000e+06
Dallas  2.916000e+06
Los Angeles  5.421000e+06
New York City  4.484000e+06
San Francisco  8.962000e+06
Seattle  2.916000e+06
Name: Price Each, dtype: float64
```

```
df.groupby('CityS(ST)')['Price Each', 'Month', 'Product', 'Quantity Ordered', 'Order Date', 'Purchase Address', 'Order ID', 'Unnamed: 0'].count()
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

San Francisco was the Best City for Sales with sales of 5062201.9T and Portland (ME) was worst City with sales of 169759.27

```
df.groupby('CityS(ST)')['Price Each'].sum()
```

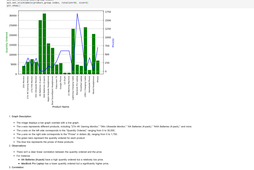
```
CityS(ST)
Atlanta  2.768000e+06
Boston  3.637000e+06
Dallas  2.916000e+06
Los Angeles  5.421000e+06
New York City  4.484000e+06
San Francisco  8.962000e+06
Seattle  2.916000e+06
Name: Price Each, dtype: float64
```

```
df.groupby('CityS(ST)')['Price Each'].sum()
```

```
CityS(ST)
Atlanta  2.768000e+06
Boston  3.637000e+06
Dallas  2.916000e+06
Los Angeles  5.421000e+06
New York City  4.484000e+06
San Francisco  8.962000e+06
Seattle  2.916000e+06
Name: Price Each, dtype: float64
```

```
import matplotlib.pyplot as plt
%matplotlib inline
plt.figure(figure=(10,6))
plt.title('Total Sales by City')
plt.xlabel('City')
plt.ylabel('Sales')
plt.show()
```

```
ERROR: Failed to execute line 3 in cell 'Total Sales by City'
ERROR: TypeWarning: Ignoring unknown usage of %matplotlib at line 2 in cell 'Total Sales by City'
```



With Graph we can also see that San Francisco(CA) has highest sales and Portland(ME) has Lowest Sales

Finding best time to display advertisements to maximize the likelihood of purchases

```
df['Order Date'].dt.hour
```

```
df['Order Date'].dt.minute
```

```
df['Count'] = df['Order ID'].count()
```

```
df.head()
```

	Unnamed: 0	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	CityS(ST)	Count
0	0	176558	USB-C Charging Cable	2	11	2019-04-19 08:46:00	917 1st Street, Los Angeles, CA 90001	January	Los Angeles	186305
2	2	176559	Bose SoundSport Headphones	1	99	2019-04-07 22:30:00	682 Chestnut St, Boston, MA 02215	April	Boston	186305
3	3	176560	Google Phone	1	600	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	April	Los Angeles	186305
4	4	176561	Wired Headphones	1	11	2019-04-12 14:38:00	669 Spruce St, Los Angeles, CA 90001	April	Los Angeles	186305
5	5	176562	USB-C Charging Cable	1	11	2019-04-29 13:33:00	333 Hickory St, New York City, NY 10001	April	New York City	186305

```
df.groupby('Product')['Price Each', 'Month', 'Quantity Ordered', 'Order Date', 'Purchase Address', 'Order ID', 'Unnamed: 0', 'Count'].count()
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

What Products are most often sold together

```
df[df['Order ID'] == 176558]
```

```
Not. Applicable as True except for the first occurrence
Not. Applicable as True except for the last occurrence
```

```
df[df.duplicated(keep=False)]
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

```
df[df.duplicated(keep=False)].sort_values(by=['Order ID'])
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

```
df[df.duplicated(keep=False)].sort_values(by=['Order ID'])
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

My recommendation is around 10 PM (22) or 7PM (19)

What Products are most often sold together

```
df[df['Order ID'] == 176558]
```

```
Not. Applicable as True except for the first occurrence
Not. Applicable as True except for the last occurrence
```

```
df[df.duplicated(keep=False)]
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

```
df[df.duplicated(keep=False)].sort_values(by=['Order ID'])
```

```
ERROR: KeyerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise ValueerrorTraceback (most recent call last)
File C:\Anaconda3\lib\site-packages\pandas\core\indexing.py:1073, in _getitem_tuple(self, tuple)
  1071 if isinstance(tuple[0], slice):
  1072 raise Value
```

```
df[df.duplicated(keep=False)].sort_values(by=['Order ID'])
```

```
ERROR: KeyerrorTraceback (most recent call
```