Name: Ayush Fating

PRN: 202201070127

Roll no. : 642

Division: F

```
import numpy as np
import pandas as pd
```

```
all_data=pd.read_csv("/content/1686715083343_all_data (4).csv")
```

```
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001

Clean up the data!

```
all_data.shape
```

```
(69, 6)
```

Drop rows of NAN

```
#Find NAN
nan_df=all_data[all_data.isna().any(axis=1)]
display(nan_df.head())
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
36	NaN	NaN	NaN	NaN	NaN	NaN
51	NaN	NaN	NaN	NaN	NaN	NaN

```
all_data.shape
```

```
(69, 6)
```

```
all_data=all_data.dropna(how='all')
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001

```
all_data.shape
```

Get rid of text in order date column

```
all_data=all_data[all_data['Order Date'].str[0:2]!='Or']
print(all_data)
```

```
    Order ID                    Product  Quantity Ordered  Price Each  \
0   176559.0  Bose SoundSport Headphones               1.0       99.99
1   176560.0                Google Phone               1.0      600.00
2   176560.0            Wired Headphones               1.0       11.99
3   176561.0            Wired Headphones               1.0       11.99
4   176562.0        USB-C Charging Cable               1.0       11.95
..       ...                         ...               ...         ...
64  259329.0    Lightning Charging Cable               1.0       14.95
65  259330.0       AA Batteries (4-pack)               2.0        3.84
66  259331.0    Apple Airpods Headphones               1.0      150.00
67  259332.0    Apple Airpods Headphones               1.0      150.00
68  259333.0  Bose SoundSport Headphones               1.0       99.99

          Order Date                      Purchase Address  Month  \
0   04-07-2019 22:30       682 Chestnut St, Boston, MA 02215      4
1   04-12-2019 14:38    669 Spruce St, Los Angeles, CA 90001      4
2   04-12-2019 14:38    669 Spruce St, Los Angeles, CA 90001      4
3      05/30/19 9:27       333 8th St, Los Angeles, CA 90001      5
4     04/29/19 13:03  381 Wilson St, San Francisco, CA 94016      4
..               ...                                     ...    ...
64  09-05-2019 19:00       480 Lincoln St, Atlanta, GA 30301      9
65    09/25/19 22:01    763 Washington St, Seattle, WA 98101      9
66     09/29/19 7:00     770 4th St, New York City, NY 10001      9
67    09/16/19 19:21          782 Lake St, Atlanta, GA 30301      9
68    09/19/19 18:03   347 Ridge St, San Francisco, CA 94016      9

                 City   Sales
0          Boston (MA)   99.99
1     Los Angeles (CA)  600.00
2     Los Angeles (CA)   11.99
3     Los Angeles (CA)   11.99
4   San Francisco (CA)   11.95
..                 ...     ...
64        Atlanta (GA)   14.95
65       Seattle (WA)    7.68
66  New York City (NY)  150.00
67        Atlanta (GA)  150.00
68  San Francisco (CA)   99.99

[67 rows x 9 columns]
```

Make columns correct type

```
all_data['Quantity Ordered']=pd.to_numeric(all_data['Quantity Ordered'])
all_data['Price Each']=pd.to_numeric(all_data['Price Each'])
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001

Augment data with additional columns

```
all_data['Month']=all_data['Order Date'].str[0:2]
all_data['Month']=all_data['Month'].astype('int32')
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	City	Sal
		Bose			04-07-	682 Chestnut			

Add city column

MA 02215

```
def get_city(address):
  return address.split(",")[1].strip(" ")

def get_state(address):
  return address.split(",")[2].split(" ")[1]

all_data['City']=all_data['Purchase Address'].apply(lambda x: f"{get_city(x)} ({get_state(x)})")
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	City
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215	4	Boston (MA)
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)

Data Exploration

Question 1.What was the best month for sales? How much was earned that month?

```
all_data['Sales']=all_data['Quantity Ordered'].astype('int')*all_data['Price Each'].astype('float')
all_data.groupby(['Month']).sum()
all_data.head()
```

```
<ipython-input-15-b2aa472d8a54>:2: FutureWarning: The default value of numeric_onl
  all_data.groupby(['Month']).sum()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	City	Sal
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215	4	Boston (MA)	99.
1	176560.0	Google	1.0	600.00	04-12-2019	669 Spruce St, Los	4	Los Angeles	600

Question 2. What city sold the most product ?

```
Dummycity=all_data.groupby(['City'])
print(Dummycity)
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f0d91377c40>
```

Question 4. What products are most often sold together?

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

df['Grouped']=df.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))
df2=df[['Order ID','Grouped']].drop_duplicates()
print(df['Grouped'])
```

```
1    Google Phone,Wired Headphones
2    Google Phone,Wired Headphones
Name: Grouped, dtype: object
<ipython-input-17-4df8b316003d>:3: 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-co
  df['Grouped']=df.groupby('Order ID')['Product'].transform(lambda x: ','.join(x))
```

```
from itertools import combinations
from collections import Counter

count=Counter()

for row in df2['Grouped']:
  row_list=row.split(',')
  count.update(Counter(combinations(row_list,2)))

for key,value in count.most_common(10):
  print(key,value)
```

```
('Google Phone', 'Wired Headphones') 1
```

Question 3 . What product sold the most? Why do you think it sold the most?

```
product_group=all_data.groupby('Product')
quantity_ordered=product_group.sum()['Quantity Ordered']
```

```
<ipython-input-19-11142b314e0e>:2: FutureWarning: The default value of numeric_only in DataFrameGroupBy.sum is deprecated. In a future v
  quantity_ordered=product_group.sum()['Quantity Ordered']
```

```
print(quantity_ordered)
```

```
Product
AA Batteries (4-pack)          64.0
AAA Batteries (4-pack)        109.0
Apple Airpods Headphones        3.0
Bose SoundSport Headphones      3.0
Google Phone                    1.0
Lightning Charging Cable        4.0
USB-C Charging Cable            8.0
Wired Headphones                7.0
Name: Quantity Ordered, dtype: float64
```

```
prices=all_data.groupby('Product').mean()['Price Each']
```

```
<ipython-input-21-1f4f73bca841>:1: FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In a future
  prices=all_data.groupby('Product').mean()['Price Each']
```

```
print(prices)
```

```
Product
AA Batteries (4-pack)           3.84
AAA Batteries (4-pack)          2.99
Apple Airpods Headphones      150.00
Bose SoundSport Headphones     99.99
Google Phone                  600.00
Lightning Charging Cable       14.95
USB-C Charging Cable           11.95
Wired Headphones               11.99
Name: Price Each, dtype: float64
```

✓ 0s completed at 15:03