Name: Rutuja Wanwe

Rollno: 668

PRN: 202201090170

Batch: F(F4)

EDS assignment 4:

Code:

```
import pandas as pd
all_data=pd.read_csv("/content/1686715083343_all_data (1).csv")
all_data
```

	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
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001
4	176562.0	USB-C Charging Cable	1.0	11.95	04/29/19 13:03	381 Wilson St, San Francisco, CA 94016
...	...	...	...	...	...	...
64	259329.0	Lightning Charging Cable	1.0	14.95	09-05-2019 19:00	480 Lincoln St, Atlanta, GA 30301
65	259330.0	AA Batteries (4-pack)	2.0	3.84	09/25/19 22:01	763 Washington St, Seattle, WA 98101
66	259331.0	Apple Airpods Headphones	1.0	150.00	09/29/19 7:00	770 4th St, New York City, NY 10001
67	259332.0	Apple Airpods Headphones	1.0	150.00	09/16/19 19:21	782 Lake St, Atlanta, GA 30301
68	259333.0	Bose SoundSport Headphones	1.0	99.99	09/19/19 18:03	347 Ridge St, San Francisco, CA 94016

69 rows × 6 columns

```
all_data.shape
```

```
(69, 6)
```

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

```
# Find NAN
nan_df = all_data[all_data.isna().any(axis=1)]
display(nan_df.head())
all_data = all_data.dropna(how='all')
all_data.head()
all_data.shape
```

```
(67, 9)
```

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

all_data['Month'] = all_data['Order Date'].str[0:2]
all_data['Month'] = all_data['Month'].astype('float32')
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month 2	Month
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215	4.0	4.0
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4.0	4.0
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4.0	4.0
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001	5.0	5.0
4	176562.0	USB-C Charging Cable	1.0	11.95	04/29/19 13:03	381 Wilson St, San Francisco, CA 94016	4.0	4.0

```
all_data['Month 2'] = pd.to_datetime(all_data['Order Date']).dt.month
all_data.head()
```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month 2	Month
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215	4.0	4.0
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4.0	4.0
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4.0	4.0
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001	5.0	5.0
4	176562.0	USB-C Charging Cable	1.0	11.95	04/29/19 13:03	381 Wilson St, San Francisco, CA 94016	4.0	4.0

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

def get_state(address):
  return address.split(",")[2].strip(" ")[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 2	Month	Sales	city
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215	4.0	4.0	99.989998	Boston A
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4.0	4.0	600.000000	Los Angeles A
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4.0	4.0	11.990000	Los Angeles A
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001	5.0	5.0	11.990000	Los Angeles A
4	176562.0	USB-C Charging Cable	1.0	11.95	04/29/19 13:03	381 Wilson St, San Francisco, CA 94016	4.0	4.0	11.950000	San Francisco A

```
#What was the best month for sales? How much was earned?
all_data['Sales']= all_data['Quantity
Ordered'].astype('float32')*all_data['Price Each'].astype('float32')
all_data['Sales']
```

```
0      99.989998
1     600.000000
2      11.990000
3      11.990000
4      11.950000
          ...
64     14.950000
65      7.680000
66    150.000000
67    150.000000
68     99.989998
Name: Sales, Length: 69, dtype: float32
```

```
#que.1 What was the best month for sales? How much was earned the month?
all_data['Sales'] = all_data['Quantity Ordered'].astype('float32') *
all_data['Price Each'].astype('float32')
all_data.groupby(['Month']).sum()
```

Month					
4.0	7335546.0	123.0	885.80	160.0	1210.760010
5.0	353124.0	2.0	111.98	10.0	111.979996
6.0	184076.0	1.0	14.95	6.0	14.950000
8.0	726962.0	9.0	23.92	32.0	50.829998
9.0	2378802.0	17.0	591.44	90.0	616.619995
10.0	550924.0	11.0	10.67	30.0	39.689999
11.0	740314.0	19.0	13.66	44.0	65.309998
12.0	550635.0	17.0	8.97	36.0	50.830002

```
#que.2 what city sold the most product?
Dummycity=all_data.groupby(['city'])
print(Dummycity)
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7fadcedfe260>
```

```
#que. 3 whatproduct sold the most? Why do you think it sold the most?
product_group=all_data.groupby('Product')
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']
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
```

```
#que.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-56-41b316017d16>:2: 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
```