

```

1 name- Pratik
2
3
4
5
6
7 import numpy as np
8 import pandas as pd
9 all_data=pd.read_csv("/content/1686715083343_all_data (7).csv")
10 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
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001 381 Wilcox St San Francisco CA

```
1
```

```

1 #clean up the data
2 all_data.shape

```

```
(69, 6)
```

```

1 # drop rows of nana
2 nan_df=all_data[all_data.isna().any(axis=1)]
3 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

```
1 all_data.shape
```

```
(69, 6)
```

```

1 all_data=all_data.dropna(how='all')
2 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

```
1 all_data.shape
```

```
(67, 6)
```

```
3 176561.0 1.0 11.99
```

```
1 #get rid of text order date column
```

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

```
3 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

```
[67 rows x 6 columns]
```

```
1 #make column correct type
```

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

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

```
4 all_data.head()
```

```

Order ID      Product      Quantity Ordered  Price Each  Order Date  Purchase Address
1 all_data['Month']= all_data['Order Date'].str[0:2]
2 all_data['Month']= all_data['Month'].astype('int32')
3 all_data.head()

```

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month
0	176559.0	Bose SoundSport Headphones	1.0	99.99	04-07-2019 22:30	682 Chestnut St, Boston, MA 02215	4
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4
2	176560.0	Wired Headphones	1.0	11.99	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001	5

```

1 #Add city column
2 def get_city(address):
3     return address.split(",")[1].strip(" ")
4 def get_state(address):
5     return address.split(",")[2].strip(" ")[1]
6
7 all_data['city']=all_data['Purchase Address'].apply(lambda x:f'{get_city(x)} ({get_state(x)})')
8 all_data.head()
9

```

	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 (A)
1	176560.0	Google Phone	1.0	600.00	04-12-2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	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	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	Los Angeles

```

1 #waht was the best month for sales?how much was earned that month?
2 all_data['Sales']=all_data['Quantity Ordered'].astype('int')*all_data['Price Each'].astype('float')

```

```

1
2
3
4
5
6
7 import numpy as np
8 import pandas as pd
9 all_data=pd.read_csv("/content/1686715083343_all_data (7).csv")
10 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
3	176561.0	Wired Headphones	1.0	11.99	05/30/19 9:27	333 8th St, Los Angeles, CA 90001 381 Wilcox St San Francisco CA

1

```

1 #clean up the data
2 all_data.shape

```

```
(69, 6)
```

```

1 # drop rows of nana
2 nan_df=all_data[all_data.isna().any(axis=1)]
3 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

```
1 all_data.shape
```

```
(69, 6)
```

```

1 all_data=all_data.dropna(how='all')
2 all_data.head()

```

6/22/23, 2:47 PM

Untitled6.ipynb - Colaboratory

```
12
13

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

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

<ipython-input-20-11142b314e0e>:2: FutureWarning: The default value of numeric_only in DataFrameGroupBy.sum is deprecated
  quantity_ordered=product_group.sum()['Quantity Ordered']
```

```
1 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
```

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

```
<ipython-input-22-1f4f73bca841>:1: FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is deprecated
  prices=all_data.groupby('Product').mean()['Price Each']
```

```
1 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
```