In [9]:
```
import numpy as np
import pandas as pd

all_data=pd.read_csv("/content/sample_data/613_order.csv")
```

In [31]:
```
all_data.head()
```

Out [31]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
0	176559.0	Bose SoundSport Headphones	1.0	99.99	4/7/2019 22:30	682 Chestnut St, Boston, MA 02215	4	Boston (MA)	99.99
1	176560.0	Google Phone	1.0	600.00	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	600.00
2	176560.0	Wired Headphones	1.0	11.99	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	11.99
3	176561.0	Wired Headphones	1.0	11.99	5/30/2019 9:27	333 8th St, Los Angeles, CA 90001	5	Los Angeles (CA)	11.99
4	176562.0	USB-C Charging Cable	1.0	11.95	4/29/2019 13:03	381 Wilson St, San Francisco, CA 94016	4	San Francisco (CA)	11.95

Clean up the data

In [32]:
```
all_data.shape
```

Out [32]: (67, 9)

Drop rows of NAN

In [33]:
```
#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()
```

Out [33]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
0	176559.0	Bose SoundSport Headphones	1.0	99.99	4/7/2019 22:30	682 Chestnut	4	Boston (MA)	99.99

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
						St, Boston, MA 02215			
1	176560.0	Google Phone	1.0	600.00	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	600.00
2	176560.0	Wired Headphones	1.0	11.99	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	11.99
3	176561.0	Wired Headphones	1.0	11.99	5/30/2019 9:27	333 8th St, Los Angeles, CA 90001	5	Los Angeles (CA)	11.99
4	176562.0	USB-C Charging Cable	1.0	11.95	4/29/2019 13:03	381 Wilson St, San Francisco, CA 94016	4	San Francisco (CA)	11.95

Get rid of text in order data column

In [34]:
```
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 2  \
0    4/7/2019 22:30       682 Chestnut St, Boston, MA 02215        4
1   4/12/2019 14:38    669 Spruce St, Los Angeles, CA 90001        4
2   4/12/2019 14:38    669 Spruce St, Los Angeles, CA 90001        4
3    5/30/2019 9:27       333 8th St, Los Angeles, CA 90001        5
4   4/29/2019 13:03  381 Wilson St, San Francisco, CA 94016        4
..              ...                                     ...      ...
64   9/5/2019 19:00       480 Lincoln St, Atlanta, GA 30301        9
65  9/25/2019 22:01    763 Washington St, Seattle, WA 98101        9
66   9/29/2019 7:00     770 4th St, New York City, NY 10001        9
67  9/16/2019 19:21          782 Lake St, Atlanta, GA 30301        9
68  9/19/2019 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 column correct type

In [35]:
```
all_data['Quantity Ordered'] = pd.to_numeric(all_data['Quantity Ordered'])
all_data['Price Each'] = pd.to_numeric(all_data['Price Each'])
```

Augment data with addityional coloumns

Add month column

In [36]:
```
all_data['month 2'] = pd.to_datetime(all_data['Order Date']).dt.month
all_data.head()
```

Out [36]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
0	176559.0	Bose SoundSport Headphones	1.0	99.99	4/7/2019 22:30	682 Chestnut St, Boston, MA 02215	4	Boston (MA)	99.99
1	176560.0	Google Phone	1.0	600.00	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	600.00
2	176560.0	Wired Headphones	1.0	11.99	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	11.99
3	176561.0	Wired Headphones	1.0	11.99	5/30/2019 9:27	333 8th St, Los Angeles, CA 90001	5	Los Angeles (CA)	11.99
4	176562.0	USB-C Charging Cable	1.0	11.95	4/29/2019 13:03	381 Wilson St, San Francisco, CA 94016	4	San Francisco (CA)	11.95

Add city column

In [37]:
```
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)
all_data.head()
```

Out [37]:

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
0	176559.0	Bose SoundSport Headphones	1.0	99.99	4/7/2019 22:30	682 Chestnut St, Boston, MA 02215	4	Boston (MA)	99.99

	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month 2	City	sales
1	176560.0	Google Phone	1.0	600.00	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	600.00
2	176560.0	Wired Headphones	1.0	11.99	4/12/2019 14:38	669 Spruce St, Los Angeles, CA 90001	4	Los Angeles (CA)	11.99
3	176561.0	Wired Headphones	1.0	11.99	5/30/2019 9:27	333 8th St, Los Angeles, CA 90001	5	Los Angeles (CA)	11.99
4	176562.0	USB-C Charging Cable	1.0	11.95	4/29/2019 13:03	381 Wilson St, San Francisco, CA 94016	4	San Francisco (CA)	11.95

Data Exploration!

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

In [51]:
```
all_data.groupby(['month 2']).sum()
#print(all_data)
```

```
<ipython-input-51-5bcdf2b31d8e>:1: FutureWarning: The default value of numeric_only in
DataFrameGroupBy.sum is deprecated. In a future version, numeric_only will default to False.
Either specify numeric_only or select only columns which should be valid for the function.
  all_data.groupby(['month 2']).sum()
```

Out [51]:

month 2	Order ID	Quantity Ordered	Price Each	sales
4	7335546.0	123.0	885.80	1210.76
5	353124.0	2.0	111.98	111.98
6	184076.0	1.0	14.95	14.95
8	726962.0	9.0	23.92	50.83
9	2378802.0	17.0	591.44	616.62
10	550924.0	11.0	10.67	39.69
11	740314.0	19.0	13.66	65.31
12	550635.0	17.0	8.97	50.83

In [53]:
```
all_data.groupby(['month 2']).sum()
```

```
<ipython-input-53-9a232da6fa68>:1: FutureWarning: The default value of numeric_only in
DataFrameGroupBy.sum is deprecated. In a future version, numeric_only will default to False.
Either specify numeric_only or select only columns which should be valid for the function.
  all_data.groupby(['month 2']).sum()
```

Out [53]:

month 2	Order ID	Quantity Ordered	Price Each	sales
4	7335546.0	123.0	885.80	1210.76

month 2	Order ID	Quantity Ordered	Price Each	sales
5	353124.0	2.0	111.98	111.98
6	184076.0	1.0	14.95	14.95
8	726962.0	9.0	23.92	50.83
9	2378802.0	17.0	591.44	616.62
10	550924.0	11.0	10.67	39.69
11	740314.0	19.0	13.66	65.31
12	550635.0	17.0	8.97	50.83

#Question 2 : What city sold the most product ?

In [45]:
```
Dummy = all_data.groupby(['City'])
print(Dummy)
city_max=all_data.groupby(['City']).sum
print(city_max)
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f6886dcdf90>
<bound method GroupBy.sum of <pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f6886dceef0>>
```

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

In [46]:
```
product_group = all_data.groupby('Product')
quantity_ordered = product_group.sum()['Quantity Ordered']
```

```
<ipython-input-46-4815a60ac30b>:2: FutureWarning: The default value of numeric_only in
DataFrameGroupBy.sum is deprecated. In a future version, numeric_only will default to False.
Either specify numeric_only or select only columns which should be valid for the function.
  quantity_ordered = product_group.sum()['Quantity Ordered']
```