CLEANING OF DATA

df=pd.read_csv("all_merged_data.csv")

df

	Order ID		Product	Quantity O	dered Price	
Each \	176558	l	JSB-C Charging Cable		2	
11.95 1	NaN		NaN		NaN	
NaN 2	176559	Bose So	oundSport Headphones		1	
99.99 3	176560		Google Phone		1	
600 4 11.99	176560		Wired Headphones		1	
			• • •			
186845 2.99	259353	AAA	A Batteries (4-pack)		3	
186846	259354		iPhone 1			
700 186847	259355	iPhone 1				
700 186848	259356	34:	34in Ultrawide Monitor 1			
379.99 186849 11.95	259357	l	JSB-C Charging Cable		1	
0 1 2 3 4	0rde 04/19/19 04/07/19 04/12/19 04/12/19	NaN 22:30 14:38	917 1st S 682 Chestnut S 669 Spruce St, Lo 669 Spruce St, Lo	t, Dallas, t, Boston, s Angeles,	NaN MA 02215 CA 90001	
186845 186846 186847	09/17/19 09/01/19	20:56 16:00 07:39 17:30	840 Highland St, Lo 216 Dogwood St, San 220 12th St, San 511 Forest St, San 250 Meadow St, San	s Angeles, Francisco, Francisco, Francisco,	CA 90001 CA 94016 CA 94016 CA 94016	
[186850	rows x 6	columns	5]			
<pre>nan_df=df[df.isna().any(axis=1)] nan_df</pre>						

```
Order ID Product Quantity Ordered Price Each Order Date \
1
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
             NaN
356
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
735
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
1433
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
1553
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
             . . .
                                                     . . .
185176
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
185438
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
186042
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
186548
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
186826
             NaN
                      NaN
                                        NaN
                                                     NaN
                                                                 NaN
       Purchase Address
1
                      NaN
356
                      NaN
735
                      NaN
1433
                      NaN
1553
                      NaN
                      . . .
185176
                      NaN
                      NaN
185438
                      NaN
186042
186548
                      NaN
                      NaN
186826
[545 rows x 6 columns]
df=df.dropna()
df
       Order ID
                                       Product Quantity Ordered Price
Each \
          176558
                         USB-C Charging Cable
                                                                 2
11.95
                  Bose SoundSport Headphones
2
          176559
                                                                 1
99.99
          176560
                                  Google Phone
                                                                 1
3
600
          176560
                             Wired Headphones
                                                                 1
11.99
          176561
                             Wired Headphones
                                                                 1
11.99
. . .
186845
                       AAA Batteries (4-pack)
                                                                 3
          259353
2.99
186846
          259354
                                        iPhone
                                                                 1
700
186847
         259355
                                        iPhone
                                                                 1
```

```
700
                     34in Ultrawide Monitor
186848
         259356
                                                            1
379.99
186849
         259357
                       USB-C Charging Cable
                                                            1
11.95
                                                Purchase Address
            Order Date
        04/19/19 08:46
                                   917 1st St, Dallas, TX 75001
0
2
                              682 Chestnut St, Boston, MA 02215
        04/07/19 22:30
3
        04/12/19 14:38
                           669 Spruce St, Los Angeles, CA 90001
4
        04/12/19 14:38
                           669 Spruce St, Los Angeles, CA 90001
5
        04/30/19 09:27
                              333 8th St, Los Angeles, CA 90001
                         840 Highland St, Los Angeles, CA 90001
186845
       09/17/19 20:56
                        216 Dogwood St, San Francisco, CA 94016
186846 09/01/19 16:00
                           220 12th St, San Francisco, CA 94016
186847
        09/23/19 07:39
                         511 Forest St, San Francisco, CA 94016
        09/19/19 17:30
186848
186849
        09/30/19 00:18
                         250 Meadow St, San Francisco, CA 94016
[186305 rows x 6 columns]
df=df[df['Order Date'].str[0:2]!='Or']
df
       Order ID
                                    Product Quantity Ordered Price
Each
                       USB-C Charging Cable
                                                            2
         176558
0
11.95
         176559 Bose SoundSport Headphones
                                                            1
99.99
3
         176560
                               Google Phone
                                                            1
600
                           Wired Headphones
                                                            1
         176560
11.99
5
         176561
                           Wired Headphones
                                                            1
11.99
. . .
186845
         259353
                     AAA Batteries (4-pack)
                                                            3
2.99
186846
         259354
                                     iPhone
                                                            1
700
186847
         259355
                                     iPhone
                                                            1
700
                     34in Ultrawide Monitor
186848
         259356
                                                            1
379.99
186849
         259357
                       USB-C Charging Cable
                                                            1
11.95
```

Order Date

Purchase Address

```
0
        04/19/19 08:46
                                    917 1st St, Dallas, TX 75001
2
        04/07/19 22:30
                              682 Chestnut St, Boston, MA 02215
3
        04/12/19 14:38
                           669 Spruce St, Los Angeles, CA 90001
4
        04/12/19 14:38
                           669 Spruce St, Los Angeles, CA 90001
5
        04/30/19 09:27
                              333 8th St, Los Angeles, CA 90001
                         840 Highland St. Los Angeles, CA 90001
       09/17/19 20:56
186845
186846 09/01/19 16:00
                        216 Dogwood St, San Francisco, CA 94016
186847
        09/23/19 07:39
                           220 12th St, San Francisco, CA 94016
186848 09/19/19 17:30
                         511 Forest St, San Francisco, CA 94016
186849 09/30/19 00:18
                         250 Meadow St, San Francisco, CA 94016
[185950 rows x 6 columns]
df["month"]=df["Order Date"].str[0:2]
df["month"]=df["month"].astype("int32")
df
C:\Users\shuvodeep\AppData\Local\Temp\ipykernel 4872\964920890.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["month"]=df["Order Date"].str[0:2]
C:\Users\shuvodeep\AppData\Local\Temp\ipykernel 4872\964920890.py: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
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user guide/indexing.html#
returning-a-view-versus-a-copy
  df["month"]=df["month"].astype("int32")
       Order ID
                                    Product Quantity Ordered Price
Each
     \
                       USB-C Charging Cable
                                                            2
0
         176558
11.95
                 Bose SoundSport Headphones
2
         176559
                                                            1
99.99
3
         176560
                               Google Phone
                                                            1
600
         176560
                           Wired Headphones
                                                            1
11.99
         176561
                           Wired Headphones
                                                            1
11.99
. . .
                                                          . . .
            . . .
                                         . . .
```

```
186845
                     AAA Batteries (4-pack)
         259353
                                                            3
2.99
186846
         259354
                                     iPhone
                                                            1
700
186847
         259355
                                     iPhone
                                                            1
700
186848
         259356
                     34in Ultrawide Monitor
                                                            1
379.99
186849
         259357
                       USB-C Charging Cable
                                                            1
11.95
            Order Date
                                                Purchase Address
                                                                  month
                                   917 1st St, Dallas, TX 75001
0
        04/19/19 08:46
                                                                      4
2
        04/07/19 22:30
                              682 Chestnut St, Boston, MA 02215
                                                                      4
3
        04/12/19 14:38
                           669 Spruce St, Los Angeles, CA 90001
                                                                      4
        04/12/19 14:38
                           669 Spruce St, Los Angeles, CA 90001
4
                                                                      4
        04/30/19 09:27
5
                              333 8th St, Los Angeles, CA 90001
                                                                      4
                         840 Highland St, Los Angeles, CA 90001
186845
        09/17/19 20:56
                                                                      9
        09/01/19 16:00
                        216 Dogwood St, San Francisco, CA 94016
186846
                                                                      9
       09/23/19 07:39
                           220 12th St, San Francisco, CA 94016
                                                                      9
186847
186848 09/19/19 17:30
                         511 Forest St. San Francisco, CA 94016
                                                                      9
                         250 Meadow St, San Francisco, CA 94016
186849 09/30/19 00:18
                                                                      9
[185950 rows x 7 columns]
df['Quantity Ordered']=pd.to numeric(df['Quantity Ordered'])
df['Price Each']=pd.to numeric(df['Price Each'])
df
C:\Users\shuvodeep\AppData\Local\Temp\ipykernel 4872\2567832521.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['Quantity Ordered']=pd.to_numeric(df['Quantity Ordered'])

C:\Users\shuvodeep\AppData\Local\Temp\ipykernel_4872\2567832521.py: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

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy

df['Price Each']=pd.to numeric(df['Price Each'])

El.	Order ID			Produc	ct (Quantity (0rde	red F	rice
0	\ 176558	USB-	C Chargir	ng Cabl	le			2	
11.95	176559	Bose Sound	Sport Hea	adphone	es			1	
99.99	176560		Googl	Le Phor	ne			1	
600.00	176560		Wired Hea	adphone	es			1	
11.99	176561		Wired Hea	adphone	es			1	
11.99									
186845	259353	AAA Ba	tteries ((4-pack	<)			3	
2.99 186846			iPhone						
700.00 186847 700.00	259355		iPhone					1	
186848	259356	34in U	ltrawide	Monito	or			1	
379.99 186849 11.95	259357	USB-C Charging Cable 1							
	0rde	r Date				Purchase	e Ad	dress	month
0	04/19/19	08:46	Q)17 1st	t St	, Dallas,	TX	75001	4
2	04/07/19	22:30	682 Cł	nestnut	t St	, Boston,	MA (92215	4
3	04/12/19	14:38	669 Sprud	e St,	Los	Angeles,	CA S	90001	4
4	04/12/19	14:38	669 Sprud	e St,	Los	Angeles,	CA S	90001	4
5	04/30/19	09:27	333 81	h St,	Los	Angeles,	CA S	90001	4

... ...

186845	09/17/19 20:56	840 Highland St, Los Angeles, CA 90001	9
186846	09/01/19 16:00	216 Dogwood St, San Francisco, CA 94016	9
186847	09/23/19 07:39	220 12th St, San Francisco, CA 94016	9
186848	09/19/19 17:30	511 Forest St, San Francisco, CA 94016	9
186849	09/30/19 00:18	250 Meadow St, San Francisco, CA 94016	9

[185950 rows x 7 columns]

df['sale']=df['Quantity Ordered']*df['Price Each']
df

C:\Users\shuvodeep\AppData\Local\Temp\ipykernel_4872\2650576186.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['sale']=df['Quantity Ordered']*df['Price Each']

	Order ID	Product	Quantity Ordered	Price
Each 0	\ 176558	USB-C Charging Cable	2	
11.95 2	176559	Bose SoundSport Headphones	1	
99.99	176560	Google Phone	1	
600.00	176560	Wired Headphones	1	
11.99	176561	Wired Headphones	1	
11.99				
186845 2.99	259353	AAA Batteries (4-pack)	3	
186846 700.00	259354	iPhone	1	
186847 700.00	259355	iPhone	1	
186848 379.99	259356	34in Ultrawide Monitor	1	

[185950 rows x 8 columns]

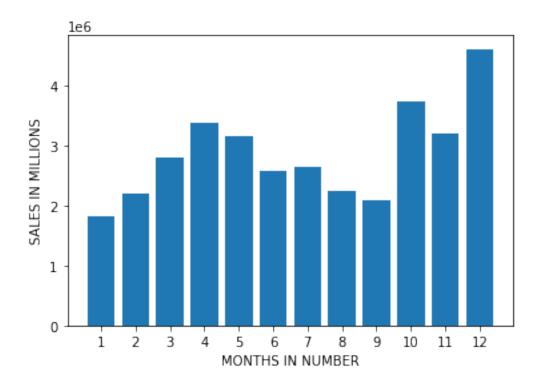
df.groupby('month').sum()

	Quantity	0rdered	Price Each	sale
month				
1		10903	1811768.38	1822256.73
2		13449	2188884.72	2202022.42
3		17005	2791207.83	2807100.38
4		20558	3367671.02	3390670.24
5		18667	3135125.13	3152606.75
6		15253	2562025.61	2577802.26
7		16072	2632539.56	2647775.76
8		13448	2230345.42	2244467.88
9		13109	2084992.09	2097560.13
10		22703	3715554.83	3736726.88
11		19798	3180600.68	3199603.20
12		28114	4588415.41	4613443.34

visualising of Data

import matplotlib.pyplot as plt

```
result=df.groupby('month').sum()
months=range(1,13)
plt.bar(months,result['sale'])
plt.xticks(months)
plt.ylabel("SALES IN MILLIONS")
plt.xlabel("MONTHS IN NUMBER")
plt.show()
```



which city has the most number of sales

```
df.to_csv("cleaneddata.csv")

df.nd_read_sev("cleaneddata.csv")
```

df=pd.read_csv("cleaneddata.csv")
df

Quantity	Product	Order ID	: 0		Ordorod
	USB-C Charging Cable	176558	0	· (Ordered 0 2
	Bose SoundSport Headphones	176559	2		1
	Google Phone	176560	3		2
	Wired Headphones	176560	4		3
	Wired Headphones	176561	5		1 4
					1

```
. . .
                . . .
                          . . .
185945
            186845
                       259353
                                    AAA Batteries (4-pack)
                                                     iPhone
185946
            186846
                       259354
185947
            186847
                       259355
                                                     iPhone
185948
            186848
                       259356
                                    34in Ultrawide Monitor
1
185949
            186849
                       259357
                                      USB-C Charging Cable
1
        Price Each
                         Order Date
                                                               Purchase
Address
                                                  917 1st St, Dallas, TX
              11.95
                     04/19/19 08:46
75001
             99.99
                     04/07/19 22:30
                                             682 Chestnut St, Boston, MA
1
02215
            600.00
                     04/12/19 14:38
                                         669 Spruce St, Los Angeles, CA
2
90001
             11.99
                                         669 Spruce St, Los Angeles, CA
3
                     04/12/19 14:38
90001
                                             333 8th St, Los Angeles, CA
              11.99
                     04/30/19 09:27
90001
. . .
                . . .
                                       840 Highland St, Los Angeles, CA
185945
              2.99
                     09/17/19 20:56
90001
            700.00
                     09/01/19 16:00
                                      216 Dogwood St, San Francisco, CA
185946
94016
185947
            700.00
                     09/23/19 07:39
                                         220 12th St, San Francisco, CA
94016
185948
            379.99
                     09/19/19 17:30
                                       511 Forest St, San Francisco, CA
94016
185949
                     09/30/19 00:18
                                       250 Meadow St, San Francisco, CA
             11.95
94016
        month
                  sale
                 23.90
0
            4
1
            4
                 99.99
2
            4
                600.00
3
            4
                 11.99
4
            4
                 11.99
            9
                  8.97
185945
185946
            9
                700.00
185947
            9
                700.00
            9
                379.99
185948
185949
                 11.95
```

we have to add a city column with the help of purchase addresss column

df["city"]=df["Purchase Address"].apply(lambda x: x.split(',')[1])
df

u 1			
0	Unnamed: 0	Order ID	Product Quantity
Ordered 0 2	0	176558	USB-C Charging Cable
1 1	2	176559	Bose SoundSport Headphones
2	3	176560	Google Phone
3 1	4	176560	Wired Headphones
4 1	5	176561	Wired Headphones
			• • •
185945 3	186845	259353	AAA Batteries (4-pack)
185946 1	186846	259354	iPhone
185947	186847	259355	iPhone
1 185948 1	186848	259356	34in Ultrawide Monitor
185949 1	186849	259357	USB-C Charging Cable
Addross	Price Each	0rder	r Date Purchase
Address 0 75001	11.95	04/19/19	08:46 917 1st St, Dallas, TX
1 02215	99.99	04/07/19	22:30 682 Chestnut St, Boston, MA
2 90001	600.00	04/12/19	14:38 669 Spruce St, Los Angeles, CA
3 90001	11.99	04/12/19	14:38 669 Spruce St, Los Angeles, CA
90001 4 90001	11.99	04/30/19	09:27 333 8th St, Los Angeles, CA
			•••
185945	2.99	09/17/19	20:56 840 Highland St, Los Angeles, CA

```
90001
            700.00
                    09/01/19 16:00
                                      216 Dogwood St, San Francisco, CA
185946
94016
185947
            700.00
                    09/23/19 07:39
                                         220 12th St, San Francisco, CA
94016
185948
            379.99
                    09/19/19 17:30
                                       511 Forest St, San Francisco, CA
94016
185949
             11.95 09/30/19 00:18
                                       250 Meadow St, San Francisco, CA
94016
        month
                  sale
                                   city
                23.90
0
            4
                                Dallas
1
            4
                99.99
                                Boston
2
            4
               600.00
                           Los Angeles
3
            4
                 11.99
                           Los Angeles
4
                11.99
            4
                           Los Angeles
. . .
            9
                 8.97
185945
                           Los Angeles
185946
            9
               700.00
                         San Francisco
185947
            9
               700.00
                         San Francisco
            9
185948
               379.99
                         San Francisco
185949
                11.95
            9
                         San Francisco
[185950 rows x 10 columns]
df["state"]=df["Purchase Address"].apply(lambda y: y.split(",")[2])
df
                    Order ID
        Unnamed: 0
                                                   Product
                                                             Quantity
0rdered
         \
                                      USB-C Charging Cable
                 0
                       176558
2
1
                 2
                       176559 Bose SoundSport Headphones
1
2
                 3
                       176560
                                              Google Phone
1
3
                 4
                       176560
                                          Wired Headphones
1
                 5
                                          Wired Headphones
4
                       176561
1
. . .
                          . . .
                . . .
185945
            186845
                       259353
                                   AAA Batteries (4-pack)
                                                     iPhone
185946
            186846
                       259354
1
185947
            186847
                       259355
                                                     iPhone
1
                                   34in Ultrawide Monitor
185948
            186848
                       259356
```

```
USB-C Charging Cable
185949
            186849
                       259357
1
        Price Each
                         Order Date
                                                              Purchase
Address
             11.95
                    04/19/19 08:46
                                                 917 1st St, Dallas, TX
75001
                                            682 Chestnut St, Boston, MA
1
             99.99
                    04/07/19 22:30
02215
            600.00
                    04/12/19 14:38
                                        669 Spruce St, Los Angeles, CA
90001
             11.99
                    04/12/19 14:38
                                        669 Spruce St, Los Angeles, CA
90001
                    04/30/19 09:27
                                           333 8th St, Los Angeles, CA
             11.99
90001
. . .
               . . .
. . .
                                      840 Highland St, Los Angeles, CA
185945
              2.99
                    09/17/19 20:56
90001
            700.00
                    09/01/19 16:00
                                     216 Dogwood St, San Francisco, CA
185946
94016
185947
            700.00
                    09/23/19 07:39
                                        220 12th St, San Francisco, CA
94016
                                      511 Forest St, San Francisco, CA
185948
            379.99 09/19/19 17:30
94016
             11.95 09/30/19 00:18
                                      250 Meadow St, San Francisco, CA
185949
94016
        month
                 sale
                                  city
                                             state
0
                23.90
                                Dallas
                                         TX 75001
            4
1
            4
                99.99
                                Boston
                                         MA 02215
2
               600.00
                           Los Angeles
                                         CA 90001
                                         CA 90001
3
            4
                11.99
                           Los Angeles
4
            4
                11.99
                           Los Angeles
                                         CA 90001
185945
            9
                 8.97
                           Los Angeles
                                         CA 90001
            9
               700.00
                         San Francisco
                                         CA 94016
185946
185947
            9
               700.00
                         San Francisco
                                         CA 94016
            9
               379.99
185948
                         San Francisco
                                         CA 94016
                                         CA 94016
185949
            9
                11.95
                         San Francisco
[185950 rows x 11 columns]
df=df.drop(columns='state')
df
        Unnamed: 0
                   Order ID
                                                   Product
                                                            Quantity
0rdered
                                     USB-C Charging Cable
                 0
                       176558
```

```
2
                 2
                               Bose SoundSport Headphones
1
                       176559
1
2
                 3
                       176560
                                              Google Phone
1
3
                 4
                       176560
                                         Wired Headphones
1
4
                 5
                       176561
                                         Wired Headphones
1
                                   AAA Batteries (4-pack)
185945
            186845
                       259353
185946
            186846
                       259354
                                                    iPhone
185947
            186847
                       259355
                                                    iPhone
185948
            186848
                       259356
                                   34in Ultrawide Monitor
185949
                                     USB-C Charging Cable
            186849
                       259357
1
                         Order Date
        Price Each
                                                             Purchase
Address
             11.95
                    04/19/19 08:46
                                                 917 1st St, Dallas, TX
75001
                                            682 Chestnut St, Boston, MA
1
             99.99
                    04/07/19 22:30
02215
                                        669 Spruce St, Los Angeles, CA
            600.00
                    04/12/19 14:38
90001
             11.99
                    04/12/19 14:38
                                        669 Spruce St, Los Angeles, CA
90001
             11.99
                    04/30/19 09:27
                                            333 8th St, Los Angeles, CA
90001
. . .
              2.99
                    09/17/19 20:56
                                      840 Highland St, Los Angeles, CA
185945
90001
185946
            700.00
                    09/01/19 16:00
                                     216 Dogwood St, San Francisco, CA
94016
                                        220 12th St, San Francisco, CA
185947
            700.00
                    09/23/19 07:39
94016
185948
            379.99
                    09/19/19 17:30
                                      511 Forest St, San Francisco, CA
94016
185949
             11.95 09/30/19 00:18
                                      250 Meadow St, San Francisco, CA
94016
        month
                  sale
                                  city
0
                23.90
                                Dallas
            4
1
                99.99
                                Boston
```

```
600.00
                           Los Angeles
2
            4
3
                11.99
            4
                           Los Angeles
                           Los Angeles
4
            4
                11.99
185945
            9
                 8.97
                           Los Angeles
185946
            9
               700.00
                         San Francisco
            9
185947
               700.00
                         San Francisco
            9
               379.99
                         San Francisco
185948
185949
            9
                11.95
                         San Francisco
[185950 rows x 10 columns]
df["state"]=df["Purchase Address"].apply(lambda y: y.split(",")
[2].split(" ")[1])
df
        Unnamed: 0
                   Order ID
                                                   Product
                                                             Quantity
Ordered \
                 0
                       176558
                                      USB-C Charging Cable
2
1
                 2
                       176559
                               Bose SoundSport Headphones
1
2
                 3
                       176560
                                              Google Phone
1
3
                 4
                                          Wired Headphones
                       176560
1
4
                 5
                                          Wired Headphones
                       176561
1
. . .
                          . . .
185945
            186845
                       259353
                                   AAA Batteries (4-pack)
185946
            186846
                       259354
                                                    iPhone
185947
                                                    iPhone
            186847
                       259355
                                   34in Ultrawide Monitor
185948
            186848
                       259356
185949
            186849
                       259357
                                      USB-C Charging Cable
1
        Price Each
                         Order Date
                                                              Purchase
Address
             11.95
                     04/19/19 08:46
                                                 917 1st St, Dallas, TX
75001
             99.99
                                            682 Chestnut St, Boston, MA
                     04/07/19 22:30
1
02215
            600.00
                    04/12/19 14:38
                                         669 Spruce St, Los Angeles, CA
90001
             11.99
                    04/12/19 14:38
                                         669 Spruce St, Los Angeles, CA
```

```
90001
              11.99
                     04/30/19 09:27
                                            333 8th St, Los Angeles, CA
4
90001
                . . .
. . .
185945
               2.99
                     09/17/19 20:56
                                      840 Highland St, Los Angeles, CA
90001
                     09/01/19 16:00 216 Dogwood St, San Francisco, CA
            700.00
185946
94016
185947
            700.00
                    09/23/19 07:39
                                         220 12th St, San Francisco, CA
94016
                                       511 Forest St, San Francisco, CA
185948
            379.99
                     09/19/19 17:30
94016
             11.95 09/30/19 00:18
                                       250 Meadow St, San Francisco, CA
185949
94016
        month
                  sale
                                   city state
                 23.90
                                 Dallas
0
            4
                                           TX
            4
                 99.99
1
                                 Boston
                                           MA
2
            4
               600.00
                           Los Angeles
                                           CA
3
                 11.99
                           Los Angeles
                                           CA
4
            4
                 11.99
                           Los Angeles
                                           CA
                  8.97
            9
                           Los Angeles
                                           \mathsf{CA}
185945
            9
               700.00
                         San Francisco
185946
                                           CA
            9
               700.00
185947
                         San Francisco
                                           CA
185948
            9
                379.99
                         San Francisco
                                           \mathsf{CA}
                         San Francisco
185949
            9
                 11.95
                                           CA
[185950 rows x 11 columns]
df['City']=df['city']+ " "+df['state']
df
        Unnamed: 0 Order ID
                                                    Product
                                                              Quantity
Ordered \
                  0
                       176558
                                      USB-C Charging Cable
2
1
                  2
                       176559
                               Bose SoundSport Headphones
1
2
                       176560
                                               Google Phone
                  3
1
3
                  4
                       176560
                                          Wired Headphones
1
                  5
                                          Wired Headphones
4
                       176561
1
. . .
                          . . .
```

185945

186845

259353

AAA Batteries (4-pack)

185946	186846	259354			iPhone		
1 185947	186847	259355			iPhone		
1 185948	186848	259356	34in U	Jltrawide	Monitor		
1 185949 1	186849	259357	USB-	·C Chargi	ng Cable		
	Price Each	0rde:	r Date			Purchase	9
Address 0	11.95	04/19/19	08:46	9	17 1st St,	Dallas,	TX
75001 1	99.99	04/07/19	22:30	682 Ch	estnut St,	Boston,	MA
02215	600.00	04/12/19	14:38 6	669 Spruc	e St, Los	Angeles,	CA
90001	11.99	04/12/19	14:38 6	669 Spruc	e St, Los	Angeles,	CA
90001	11.99	04/30/19	09:27	333 8t	h St, Los	Angeles,	CA
90001							
185945	2.99	09/17/19	20:56 840	Highlan	d St, Los	Angeles,	CA
90001 185946	700.00	09/01/19	16:00 216	Dogwood S	St, San Fr	ancisco,	CA
94016 185947	700.00	09/23/19	07:39 2	220 12th	St, San Fr	ancisco,	CA
94016 185948	379.99	09/19/19	17:30 511	L Forest :	St, San Fr	ancisco,	CA
94016 185949 94016	11.95	09/30/19	00:18 256	Meadow 9	St, San Fr	ancisco,	CA
0 1 2 3 4	month sa 4 23. 4 99. 4 600. 4 11. 4 11.	90 99 00 Los 99 Los 99 Los	city st Dallas Boston s Angeles s Angeles s Angeles	TX MA CA CA	Dalla Bosto Los Angele Los Angele Los Angele	on MA es CA es CA	
185945 185946 185947 185948 185949		00 San F 00 San F 99 San F	Angeles Francisco Francisco Francisco Francisco	CA Sai CA Sai CA Sai	Los Angele n Francisc n Francisc n Francisc n Francisc	co CA co CA co CA	

[185950 rows x 12 columns]

now which city has the highest number of sales
result=df.groupby("City").sum()
result

Each \ City	Unnamed:	0	Order ID	Quantity	Ordered	Price
Atlanta GA	138917051	18	3423838407		16602	
2779908.20 Austin TX	92433183	31	2280982185		11153	
1809873.61 Boston MA	187355203	39	4598265261		22528	
3637409.77 Dallas TX 2752627.82	139614395	59	3415643578		16730	
Los Angeles CA 5421435.23	276852816	95	6811084693		33289	
New York City NY 4635370.83	231488509	97	5736333884		27932	
Portland ME 447189.25	22713872	21	563266345		2750	
Portland OR 1860558.22	935485289		2305594747		11303	
San Francisco CA 8211461.74	416432590	96	10304443952		50239	
Seattle WA 2733296.01	137949471	12	3406693974		16553	
	month		sale			
City Atlanta GA Austin TX Boston MA Dallas TX Los Angeles CA New York City NY Portland ME Portland OR San Francisco CA Seattle WA dff=pd.read csv("c	141112 3 104620 2 208325 5 175741 4 17144 70621 1 315520 8 104941 2	1819 3661 2767 5452 4664 449 1876 3262	9581.75 1642.01 7975.40 2570.80 4317.43 9758.27 9732.34 2203.91 7755.48			
dff			- ,			
Unnamed: 0	Order II)		Pro	duct Ous	ntity

	Unnamed:	0	Order ID	Product	Quantity
0rdered	\				
Θ		0	176558	USB-C Charging Cable	
2					
1		2	176559	Bose SoundSport Headphones	

```
1
2
                                              Google Phone
                  3
                       176560
1
3
                  4
                       176560
                                          Wired Headphones
1
4
                  5
                       176561
                                          Wired Headphones
1
                          . . .
. . .
185945
            186845
                       259353
                                    AAA Batteries (4-pack)
185946
            186846
                       259354
                                                     iPhone
185947
            186847
                       259355
                                                     iPhone
185948
            186848
                       259356
                                    34in Ultrawide Monitor
                                      USB-C Charging Cable
185949
            186849
                       259357
        Price Each
                         Order Date
                                                              Purchase
Address
                     04/19/19 08:46
                                                  917 1st St, Dallas, TX
             11.95
75001
             99.99
                     04/07/19 22:30
                                            682 Chestnut St, Boston, MA
1
02215
                                         669 Spruce St, Los Angeles, CA
2
            600.00
                     04/12/19 14:38
90001
                                         669 Spruce St, Los Angeles, CA
             11.99
                     04/12/19 14:38
90001
             11.99
                     04/30/19 09:27
                                            333 8th St, Los Angeles, CA
90001
. . .
                . . .
                                       840 Highland St, Los Angeles, CA
              2.99
                     09/17/19 20:56
185945
90001
185946
            700.00
                     09/01/19 16:00
                                      216 Dogwood St, San Francisco, CA
94016
185947
            700.00
                     09/23/19 07:39
                                         220 12th St, San Francisco, CA
94016
                                       511 Forest St, San Francisco, CA
185948
            379.99
                     09/19/19 17:30
94016
                                       250 Meadow St, San Francisco, CA
185949
             11.95
                     09/30/19 00:18
94016
        month
                  sale
                 23.90
0
            4
1
            4
                 99.99
2
                600.00
            4
3
                 11.99
```

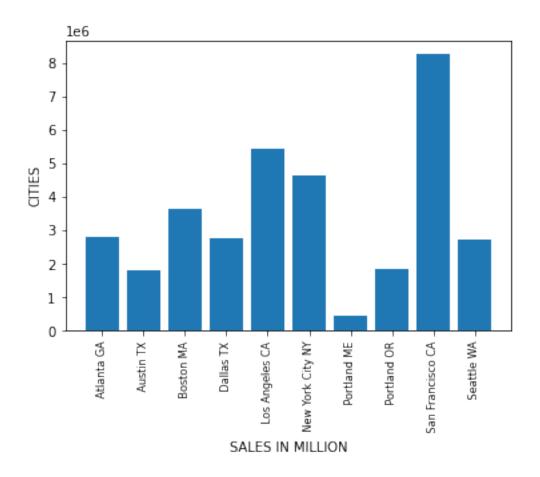
```
4
            4
                11.99
185945
            9
                  8.97
185946
            9
               700.00
185947
               700.00
185948
            9
                379.99
            9
185949
                11.95
[185950 rows x 9 columns]
df
        Unnamed: 0
                     Order ID
                                                    Product
                                                             Quantity
0rdered
                                      USB-C Charging Cable
                  0
                       176558
0
2
                  2
                               Bose SoundSport Headphones
1
                       176559
1
2
                  3
                       176560
                                              Google Phone
1
3
                  4
                       176560
                                          Wired Headphones
1
4
                  5
                       176561
                                          Wired Headphones
1
. . .
                          . . .
                                    AAA Batteries (4-pack)
185945
            186845
                       259353
                       259354
                                                     iPhone
185946
            186846
185947
            186847
                       259355
                                                     iPhone
1
                                    34in Ultrawide Monitor
185948
            186848
                       259356
                                      USB-C Charging Cable
185949
            186849
                       259357
                         Order Date
                                                              Purchase
        Price Each
Address
                                                 917 1st St, Dallas, TX
             11.95
                     04/19/19 08:46
75001
                                            682 Chestnut St, Boston, MA
             99.99
                     04/07/19 22:30
02215
            600.00
                     04/12/19 14:38
                                         669 Spruce St, Los Angeles, CA
90001
             11.99
                                         669 Spruce St, Los Angeles, CA
                     04/12/19 14:38
90001
```

333 8th St, Los Angeles, CA

11.99 04/30/19 09:27

90001

```
2.99
                     09/17/19 20:56
                                       840 Highland St, Los Angeles, CA
185945
90001
185946
            700.00
                     09/01/19 16:00
                                      216 Dogwood St, San Francisco, CA
94016
185947
            700.00
                     09/23/19 07:39
                                         220 12th St, San Francisco, CA
94016
185948
            379.99
                     09/19/19 17:30
                                       511 Forest St, San Francisco, CA
94016
185949
              11.95
                     09/30/19 00:18
                                       250 Meadow St, San Francisco, CA
94016
                  sale
        month
                                   city state
                                                             City
0
                 23.90
                                 Dallas
                                           TX
                                                        Dallas TX
                                                        Boston MA
1
            4
                 99.99
                                 Boston
                                           MA
2
            4
               600.00
                           Los Angeles
                                           CA
                                                   Los Angeles CA
3
                 11.99
                                                   Los Angeles CA
            4
                           Los Angeles
                                           CA
4
            4
                 11.99
                           Los Angeles
                                                   Los Angeles CA
                                           \mathsf{CA}
                                          . . .
. . .
           . . .
185945
            9
                  8.97
                           Los Angeles
                                           CA
                                                   Los Angeles CA
                                                 San Francisco CA
185946
            9
               700.00
                         San Francisco
                                           CA
                                                 San Francisco CA
            9
185947
               700.00
                         San Francisco
                                           CA
            9
                                                 San Francisco CA
185948
                379.99
                         San Francisco
                                           CA
            9
                                                 San Francisco CA
185949
                 11.95
                         San Francisco
                                           CA
[185950 rows x 12 columns]
from importlib import reload
plt=reload(plt)
cities=[cityy for cityy,dff in df.groupby("City")]
plt.bar(cities, result['sale'])
plt.xticks(cities, rotation="vertical", size=8)
plt.xlabel("SALES IN MILLION")
plt.ylabel("CITIES")
plt.show()
```

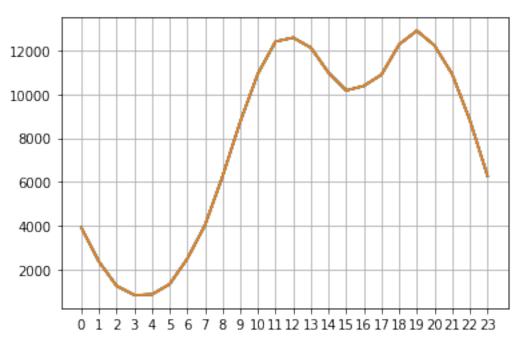


what time should we display advertisement to maximize likelihood of customers buying product

```
df=df.drop(columns='city')
df['Order Date']=pd.to_datetime(df["Order Date"])
df["Hour"]=df["Order Date"].dt.hour
df["minute"]=df["Order Date"].dt.minute
df.head()
```

,	Unnamed: 0	Order ID	Product	Quantity Ordered
0	0	176558	USB-C Charging Cable	2
1	2	176559	Bose SoundSport Headphones	1
2	3	176560	Google Phone	1
3	4	176560	Wired Headphones	1
4	5	176561	Wired Headphones	1

```
Order Date
                                                         Purchase
   Price Each
Address \
        11.95 2019-04-19 08:46:00
                                            917 1st St, Dallas, TX
75001
        99.99 2019-04-07 22:30:00
                                       682 Chestnut St, Boston, MA
02215
2
       600.00 2019-04-12 14:38:00
                                    669 Spruce St, Los Angeles, CA
90001
        11.99 2019-04-12 14:38:00
                                    669 Spruce St, Los Angeles, CA
90001
        11.99 2019-04-30 09:27:00
                                       333 8th St, Los Angeles, CA
90001
   month
            sale state
                                    City
                                          Hour
                                                minute
0
           23.90
                               Dallas TX
                                             8
                                                     46
       4
                    TX
           99.99
                               Boston MA
1
       4
                    MA
                                            22
                                                     30
2
       4
          600.00
                    CA
                         Los Angeles CA
                                            14
                                                     38
3
                         Los Angeles CA
       4
           11.99
                    CA
                                            14
                                                     38
4
           11.99
                         Los Angeles CA
                                                     27
                    CA
                                             9
hours=[hour for hour,dff in df.groupby("Hour")]
res=df.groupby(['Hour']).count()
plt.plot(hours,res)
plt.xticks(hours)
plt.grid()
plt.show()
```



so from the graph we can say that the peak times to advertise are 11am and 7pm onwards

qs.... what products most often sold together? df.head()

,	Unnamed: 0	Order ID	Product Quantity Ordered
0	0	176558	USB-C Charging Cable 2
1	2	176559	Bose SoundSport Headphones 1
2	3	176560	Google Phone 1
3	4	176560	Wired Headphones 1
4	5	176561	Wired Headphones 1
0 75 1 02 2 96 3 96 4 96 1 2 3 4 d1	ddress \	2019-04-07 2019-04-12 2019-04-12 2019-04-30 ale state 90 TX 99 CA 99 CA 99 CA cder ID'].de	08:46:00 917 1st St, Dallas, TX 22:30:00 682 Chestnut St, Boston, MA 14:38:00 669 Spruce St, Los Angeles, CA 14:38:00 669 Spruce St, Los Angeles, CA 09:27:00 333 8th St, Los Angeles, CA
	f2.head()		
\	Unnamed: 0		•
2		3 176560	•
3		176560	·
17	7 18	3 176574	Google Phone 1

```
29
            30
                   176585
                           Bose SoundSport Headphones
    Price Each
                         Order Date
                                                           Purchase
Address
        600.00 2019-04-12 14:38:00
                                      669 Spruce St, Los Angeles, CA
90001
         11.99 2019-04-12 14:38:00
                                      669 Spruce St, Los Angeles, CA
90001
        600.00 2019-04-03 19:42:00
                                         20 Hill St, Los Angeles, CA
17
90001
         11.95 2019-04-03 19:42:00
                                         20 Hill St, Los Angeles, CA
18
90001
29
         99.99 2019-04-07 11:31:00
                                         823 Highland St, Boston, MA
02215
    month
             sale state
                                      City
                                            Hour
                                                  minute
2
        4
           600.00
                     CA
                           Los Angeles CA
                                              14
                                                      38
                           Los Angeles CA
3
            11.99
        4
                      CA
                                              14
                                                      38
17
        4
           600.00
                      \mathsf{CA}
                           Los Angeles CA
                                              19
                                                      42
            11.95
                           Los Angeles CA
                                              19
                                                      42
18
        4
                      CA
29
            99.99
                      MΑ
                                Boston MA
                                              11
                                                      31
                                                Grouped
2
                         Google Phone, Wired Headphones
3
                         Google Phone, Wired Headphones
                     Google Phone, USB-C Charging Cable
17
                     Google Phone, USB-C Charging Cable
18
    Bose SoundSport Headphones, Bose SoundSport Hea...
29
```

USB-C Charging Cable

1

1

here we can say that there are multiple copied rows as we grouped it by order id

```
df2=df2[['Order ID','Grouped']].drop_duplicates()
df2.head()
```

	Order ID	Grouped
2	176560	Google Phone,Wired Headphones
17	176574	Google Phone,USB-C Charging Cable
29	176585	Bose SoundSport Headphones, Bose SoundSport Hea
31	176586	AAA Batteries (4-pack),Google Phone
118	176672	Lightning Charging Cable, USB-C Charging Cable

what product sold the most and why do you think it sold the most?

```
product_group=df.groupby('Product')
product group.sum()
```

18

19

176574

```
quantity_ordered=product_group.sum()['Quantity Ordered']
products=[product for product,df in product_group]
plt.bar(products,quantity_ordered)
plt.xlabel('Product')
plt.xticks(products,rotation='vertical',size=8)
plt.show()
```

