```
import pandas as pd
```

CLEANING OF DATA

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

```
df
```

```
       Order ID                     Product Quantity Ordered Price
Each  \
0        176558        USB-C Charging Cable                2
11.95
1           NaN                         NaN              NaN
NaN
2        176559  Bose SoundSport Headphones                1
99.99
3        176560                Google Phone                1
600
4        176560            Wired Headphones                1
11.99
...         ...                         ...              ...        ..
.
186845   259353      AAA Batteries (4-pack)                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
0       04/19/19 08:46             917 1st St, Dallas, TX 75001
1                  NaN                                      NaN
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
...                ...                                      ...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001
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

[186850 rows x 6 columns]
```

```
nan_df=df[df.isna().any(axis=1)]
nan_df
```

```
       Order ID Product Quantity Ordered Price Each Order Date  \
1           NaN     NaN               NaN        NaN        NaN
356         NaN     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
185438              NaN
186042              NaN
186548              NaN
186826              NaN

[545 rows x 6 columns]
```

```
df=df.dropna()
df
```

```
       Order ID                     Product Quantity Ordered Price
Each  \
0        176558        USB-C Charging Cable                2
11.95
2        176559  Bose SoundSport Headphones                1
99.99
3        176560                Google Phone                1
600
4        176560            Wired Headphones                1
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
186848   259356      34in Ultrawide Monitor                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
...                ...                                     ...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001
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

[186305 rows x 6 columns]
```

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

```
       Order ID                    Product Quantity Ordered Price
Each  \
0        176558        USB-C Charging Cable                2
11.95
2        176559  Bose SoundSport Headphones                1
99.99
3        176560                Google Phone                1
600
4        176560            Wired Headphones                1
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
186848   259356      34in Ultrawide Monitor                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
...               ...                                       ...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001
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  \
0        176558       USB-C Charging Cable                2
11.95
2        176559  Bose SoundSport Headphones                1
99.99
3        176560               Google Phone                1
600
4        176560           Wired Headphones                1
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
186848   259356      34in Ultrawide Monitor                1
379.99
186849   259357        USB-C Charging Cable                1
11.95
```

	Order Date	Purchase Address	month
0	04/19/19 08:46	917 1st St, Dallas, TX 75001	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
4	04/12/19 14:38	669 Spruce St, Los Angeles, CA 90001	4
5	04/30/19 09:27	333 8th St, Los Angeles, CA 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['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'])

       Order ID                     Product  Quantity Ordered  Price
Each  \
0        176558        USB-C Charging Cable                 2
11.95
2        176559  Bose SoundSport Headphones                 1
99.99
3        176560                Google Phone                 1
600.00
4        176560            Wired Headphones                 1
11.99
5        176561            Wired Headphones                 1
11.99
...         ...                         ...               ...
...
186845   259353      AAA Batteries (4-pack)                 3
2.99
186846   259354                      iPhone                 1
700.00
186847   259355                      iPhone                 1
700.00
186848   259356      34in Ultrawide Monitor                 1
379.99
186849   259357        USB-C Charging Cable                 1
11.95

            Order Date                        Purchase Address  month

0       04/19/19 08:46             917 1st St, Dallas, TX 75001      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

4       04/12/19 14:38     669 Spruce St, Los Angeles, CA 90001      4

5       04/30/19 09:27        333 8th St, Los Angeles, CA 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
3        176560                Google Phone                 1
600.00
4        176560            Wired Headphones                 1
11.99
5        176561            Wired Headphones                 1
11.99
...         ...                         ...               ...
...
186845   259353      AAA Batteries (4-pack)                 3
2.99
186846   259354                      iPhone                 1
700.00
186847   259355                      iPhone                 1
700.00
186848   259356      34in Ultrawide Monitor                 1
379.99
```

```
186849   259357        USB-C Charging Cable                1
11.95

            Order Date                          Purchase Address  month
sale
0       04/19/19 08:46             917 1st St, Dallas, TX 75001      4
23.90
2       04/07/19 22:30        682 Chestnut St, Boston, MA 02215      4
99.99
3       04/12/19 14:38     669 Spruce St, Los Angeles, CA 90001      4
600.00
4       04/12/19 14:38     669 Spruce St, Los Angeles, CA 90001      4
11.99
5       04/30/19 09:27        333 8th St, Los Angeles, CA 90001      4
11.99
...                ...                                       ...    ...
...
186845  09/17/19 20:56   840 Highland St, Los Angeles, CA 90001      9
8.97
186846  09/01/19 16:00  216 Dogwood St, San Francisco, CA 94016      9
700.00
186847  09/23/19 07:39     220 12th St, San Francisco, CA 94016      9
700.00
186848  09/19/19 17:30   511 Forest St, San Francisco, CA 94016      9
379.99
186849  09/30/19 00:18   250 Meadow St, San Francisco, CA 94016      9
11.95

[185950 rows x 8 columns]
```

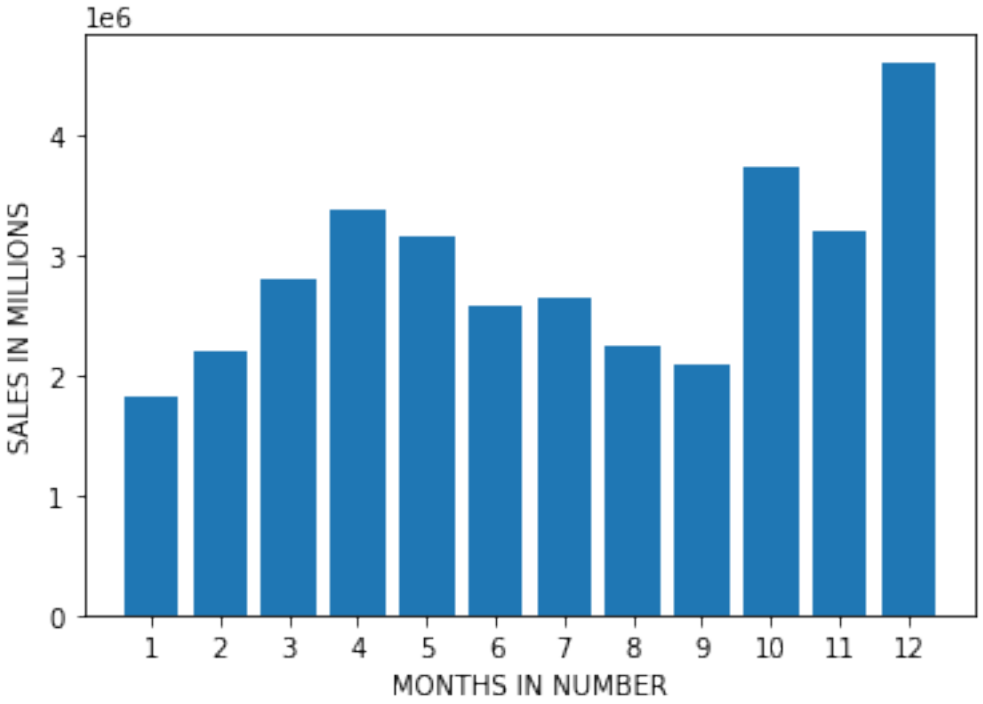
```
df.groupby('month').sum()
```

```
       Quantity Ordered  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=pd.read_csv("cleaneddata.csv")
df
```

```
        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
```

```
...            ...       ...                          ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale
0           4   23.90
1           4   99.99
2           4  600.00
3           4   11.99
4           4   11.99
...       ...     ...
185945      9    8.97
185946      9  700.00
185947      9  700.00
185948      9  379.99
185949      9   11.95
```

```
[185950 rows x 9 columns]
```

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
```

```
        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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
```

```
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

        month    sale            city
0           4   23.90          Dallas
1           4   99.99          Boston
2           4  600.00     Los Angeles
3           4   11.99     Los Angeles
4           4   11.99     Los Angeles
...       ...     ...             ...
185945      9    8.97     Los Angeles
185946      9  700.00   San Francisco
185947      9  700.00   San Francisco
185948      9  379.99   San Francisco
185949      9   11.95   San Francisco

[185950 rows x 10 columns]
```

```
df["state"]=df["Purchase Address"].apply(lambda y: y.split(",")[2])
df
```

```
        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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
```

```
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale            city      state
0           4   23.90          Dallas   TX 75001
1           4   99.99          Boston   MA 02215
2           4  600.00     Los Angeles   CA 90001
3           4   11.99     Los Angeles   CA 90001
4           4   11.99     Los Angeles   CA 90001
...       ...     ...             ...        ...
185945      9    8.97     Los Angeles   CA 90001
185946      9  700.00   San Francisco   CA 94016
185947      9  700.00   San Francisco   CA 94016
185948      9  379.99   San Francisco   CA 94016
185949      9   11.95   San Francisco   CA 94016

[185950 rows x 11 columns]

df=df.drop(columns='state')

df

        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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale           city
0           4   23.90         Dallas
1           4   99.99         Boston
```

```
2           4  600.00     Los Angeles
3           4   11.99     Los Angeles
4           4   11.99     Los Angeles
...       ...     ...             ...
185945      9    8.97     Los Angeles
185946      9  700.00   San Francisco
185947      9  700.00   San Francisco
185948      9  379.99   San Francisco
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                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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
```

```
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...              ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale            city state
0           4   23.90          Dallas    TX
1           4   99.99          Boston    MA
2           4  600.00     Los Angeles    CA
3           4   11.99     Los Angeles    CA
4           4   11.99     Los Angeles    CA
...       ...     ...             ...   ...
185945      9    8.97     Los Angeles    CA
185946      9  700.00   San Francisco    CA
185947      9  700.00   San Francisco    CA
185948      9  379.99   San Francisco    CA
185949      9   11.95   San Francisco    CA

[185950 rows x 11 columns]
```

```
df['City']=df['city']+ " "+df['state']
df
```

```
        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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
```

```
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale            city state               City
0           4   23.90          Dallas    TX          Dallas TX
1           4   99.99          Boston    MA          Boston MA
2           4  600.00     Los Angeles    CA     Los Angeles CA
3           4   11.99     Los Angeles    CA     Los Angeles CA
4           4   11.99     Los Angeles    CA     Los Angeles CA
...       ...     ...             ...   ...                ...
185945      9    8.97     Los Angeles    CA     Los Angeles CA
185946      9  700.00   San Francisco    CA   San Francisco CA
185947      9  700.00   San Francisco    CA   San Francisco CA
185948      9  379.99   San Francisco    CA   San Francisco CA
185949      9   11.95   San Francisco    CA   San Francisco CA

[185950 rows x 12 columns]
```

now which city has the highest number of sales

```
result=df.groupby("City").sum()
result
```

```
                  Unnamed: 0     Order ID  Quantity Ordered  Price
Each  \
City

 Atlanta GA        1389170518   3423838407             16602
2779908.20
 Austin TX          924331831   2280982185             11153
1809873.61
 Boston MA         1873552039   4598265261             22528
3637409.77
 Dallas TX         1396143959   3415643578             16730
2752627.82
 Los Angeles CA    2768528105   6811084693             33289
5421435.23
 New York City NY  2314885097   5736333884             27932
4635370.83
 Portland ME        227138721    563266345              2750
447189.25
 Portland OR        935485289   2305594747             11303
1860558.22
 San Francisco CA  4164325906  10304443952             50239
8211461.74
 Seattle WA        1379494712   3406693974             16553
2733296.01

                   month        sale
City
 Atlanta GA        104794  2795498.58
 Austin TX          69829  1819581.75
 Boston MA         141112  3661642.01
 Dallas TX         104620  2767975.40
 Los Angeles CA    208325  5452570.80
 New York City NY  175741  4664317.43
 Portland ME        17144   449758.27
 Portland OR        70621  1870732.34
 San Francisco CA  315520  8262203.91
 Seattle WA        104941  2747755.48
```

```
dff=pd.read_csv("cleaneddata.csv")
dff
```

```
        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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale
0           4   23.90
1           4   99.99
2           4  600.00
3           4   11.99
```

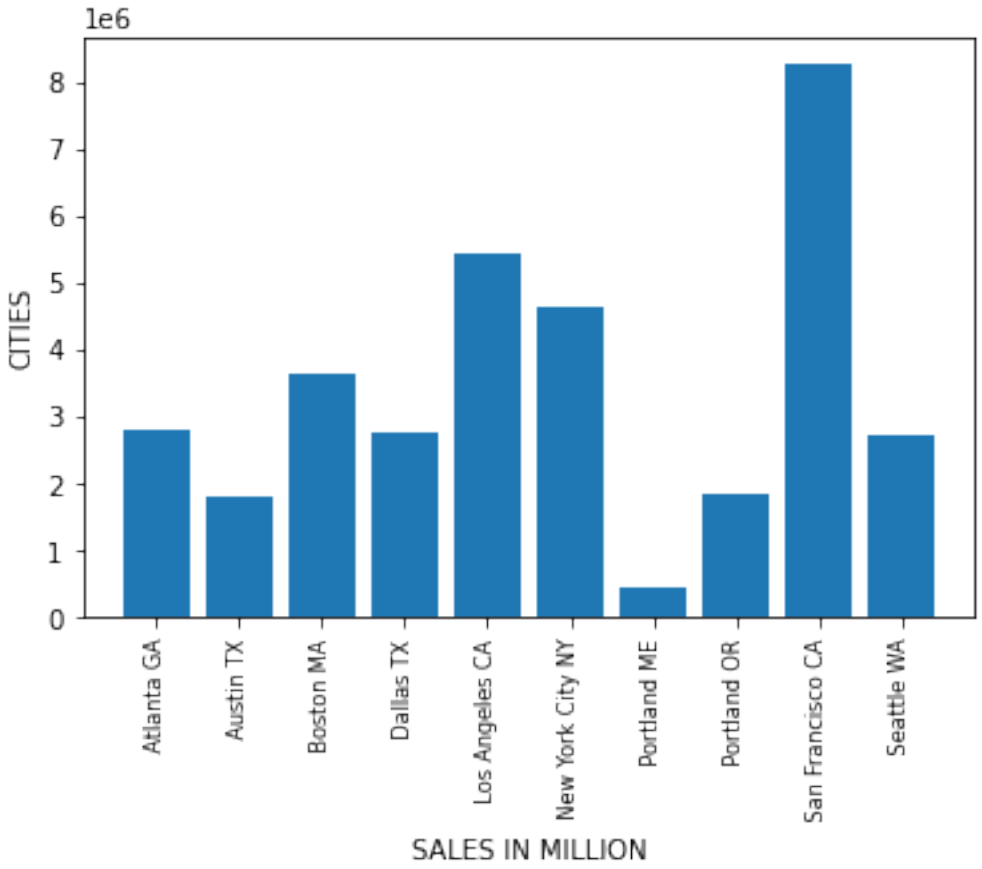
```
4           4   11.99
...       ...     ...
185945      9    8.97
185946      9  700.00
185947      9  700.00
185948      9  379.99
185949      9   11.95

[185950 rows x 9 columns]
```

```
df
```

```
        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
...            ...       ...                         ...
...
185945      186845    259353      AAA Batteries (4-pack)
3
185946      186846    259354                      iPhone
1
185947      186847    259355                      iPhone
1
185948      186848    259356      34in Ultrawide Monitor
1
185949      186849    259357        USB-C Charging Cable
1

        Price Each      Order Date                        Purchase
Address  \
0            11.95  04/19/19 08:46             917 1st St, Dallas, TX
75001
1            99.99  04/07/19 22:30        682 Chestnut St, Boston, MA
02215
2           600.00  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
3            11.99  04/12/19 14:38     669 Spruce St, Los Angeles, CA
90001
4            11.99  04/30/19 09:27        333 8th St, Los Angeles, CA
90001
...            ...             ...
```

```
...
185945        2.99  09/17/19 20:56   840 Highland St, Los Angeles, CA
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

        month    sale            city state              City
0           4   23.90          Dallas    TX          Dallas TX
1           4   99.99          Boston    MA          Boston MA
2           4  600.00     Los Angeles    CA     Los Angeles CA
3           4   11.99     Los Angeles    CA     Los Angeles CA
4           4   11.99     Los Angeles    CA     Los Angeles CA
...       ...     ...             ...   ...                ...
185945      9    8.97     Los Angeles    CA     Los Angeles CA
185946      9  700.00   San Francisco    CA   San Francisco CA
185947      9  700.00   San Francisco    CA   San Francisco CA
185948      9  379.99   San Francisco    CA   San Francisco CA
185949      9   11.95   San Francisco    CA   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

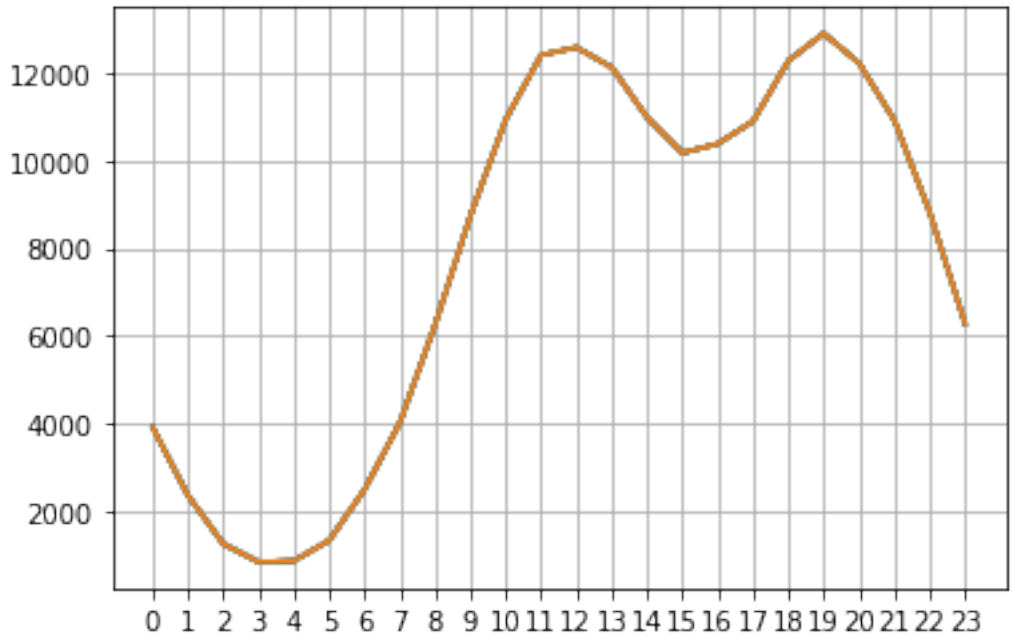
```
df=df.drop(columns='city')

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
```

```
   Price Each          Order Date                        Purchase
Address  \
0       11.95 2019-04-19 08:46:00          917 1st St, Dallas, TX
75001
1       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
3       11.99 2019-04-12 14:38:00  669 Spruce St, Los Angeles, CA
90001
4       11.99 2019-04-30 09:27:00     333 8th St, Los Angeles, CA
90001
```

	month	sale	state	City	Hour	minute
0	4	23.90	TX	Dallas TX	8	46
1	4	99.99	MA	Boston MA	22	30
2	4	600.00	CA	Los Angeles CA	14	38
3	4	11.99	CA	Los Angeles CA	14	38
4	4	11.99	CA	Los Angeles CA	9	27

```
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


   Price Each          Order Date                      Purchase
Address  \
0       11.95 2019-04-19 08:46:00          917 1st St, Dallas, TX
75001
1       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
3       11.99 2019-04-12 14:38:00  669 Spruce St, Los Angeles, CA
90001
4       11.99 2019-04-30 09:27:00     333 8th St, Los Angeles, CA
90001

   month    sale state             City  Hour  minute
0      4   23.90    TX        Dallas TX     8      46
1      4   99.99    MA        Boston MA    22      30
2      4  600.00    CA   Los Angeles CA    14      38
3      4   11.99    CA   Los Angeles CA    14      38
4      4   11.99    CA   Los Angeles CA     9      27
```

```
df2=df2[df['Order ID'].duplicated(keep=False)]
df2['Grouped']=df.groupby("Order ID")['Product'].transform(lambda
x:",".join(x))
df2.head()
```

```
    Unnamed: 0  Order ID                     Product  Quantity Ordered
\
2            3    176560                Google Phone                 1

3            4    176560            Wired Headphones                 1

17          18    176574                Google Phone                 1
```

```
18          19    176574        USB-C Charging Cable                1

29          30    176585  Bose SoundSport Headphones                1


    Price Each          Order Date                       Purchase
Address  \
2       600.00 2019-04-12 14:38:00  669 Spruce St, Los Angeles, CA
90001
3        11.99 2019-04-12 14:38:00  669 Spruce St, Los Angeles, CA
90001
17      600.00 2019-04-03 19:42:00     20 Hill St, Los Angeles, CA
90001
18       11.95 2019-04-03 19:42:00     20 Hill St, Los Angeles, CA
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
3       4   11.99    CA   Los Angeles CA    14      38
17      4  600.00    CA   Los Angeles CA    19      42
18      4   11.95    CA   Los Angeles CA    19      42
29      4   99.99    MA         Boston MA    11      31

                                          Grouped
2                      Google Phone,Wired Headphones
3                      Google Phone,Wired Headphones
17                  Google Phone,USB-C Charging Cable
18                  Google Phone,USB-C Charging Cable
29  Bose SoundSport Headphones,Bose SoundSport Hea...
```

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()
```

```
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()
```