```
import pandas as pd
df=pd.read_csv("pollution.csv")
print(df.head(3))
```

```
               location  month  year  SO2 µg/l  NO2µg/l  PM10 µg/l  \
0  CLOCK TOWER-DEHRADUN      1  2012     27.33    30.33     193.28
1  CLOCK TOWER-DEHRADUN      2  2012     25.68    25.80     173.77
2  CLOCK TOWER-DEHRADUN      3  2012     29.64    27.50     211.35

   PM2.5 µ g/l  CO µg/l  O3 µ g/l 8 HR  NH3  µ g/l     AQI Air Quality

0         60.0        2            100         400  162.19    Moderate

1         60.0        2            100         400  149.18    Moderate

2         60.0        2            100         400  174.23    Moderate
```

```
df.columns
df
```

```
                  location  month  year  SO2 µg/l  NO2µg/l  PM10
µg/l  \
0    CLOCK TOWER-DEHRADUN      1  2012     27.33    30.33     193.28

1    CLOCK TOWER-DEHRADUN      2  2012     25.68    25.80     173.77

2    CLOCK TOWER-DEHRADUN      3  2012     29.64    27.50     211.35

3    CLOCK TOWER-DEHRADUN      4  2012     28.64    26.81     230.76

4    CLOCK TOWER-DEHRADUN      5  2012     31.09    29.30     310.73

..                    ...    ...   ...       ...      ...        ...

955              RUDRAPUR      8  2021     80.00    80.00     121.28

956              RUDRAPUR      9  2021     18.73    20.93      92.96

957              RUDRAPUR     10  2021     80.00    80.00     100.00

958              RUDRAPUR     11  2021     80.00    80.00     100.00

959              RUDRAPUR     12  2021     80.00    80.00     100.00


     PM2.5 µ g/l  CO µg/l  O3 µ g/l 8 HR  NH3  µ g/l     AQI   Air
Quality
0          60.00        2            100         400  162.19
Moderate
```

```
1          60.00        2            100         400  149.18
Moderate
2          60.00        2            100         400  174.23
Moderate
3          60.00        2            100         400  187.17
Moderate
4          60.00        2            100         400  260.73
Poor
..           ...      ...            ...         ...     ...
...
955       208.44        2            100         400  368.03     Very
Poor
956       153.75        2            100         400  325.96     Very
Poor
957        60.00        2            100         400  100.00
Satisfactory
958        60.00        2            100         400  100.00
Satisfactory
959        60.00        2            100         400  100.00
Satisfactory

[960 rows x 12 columns]
```

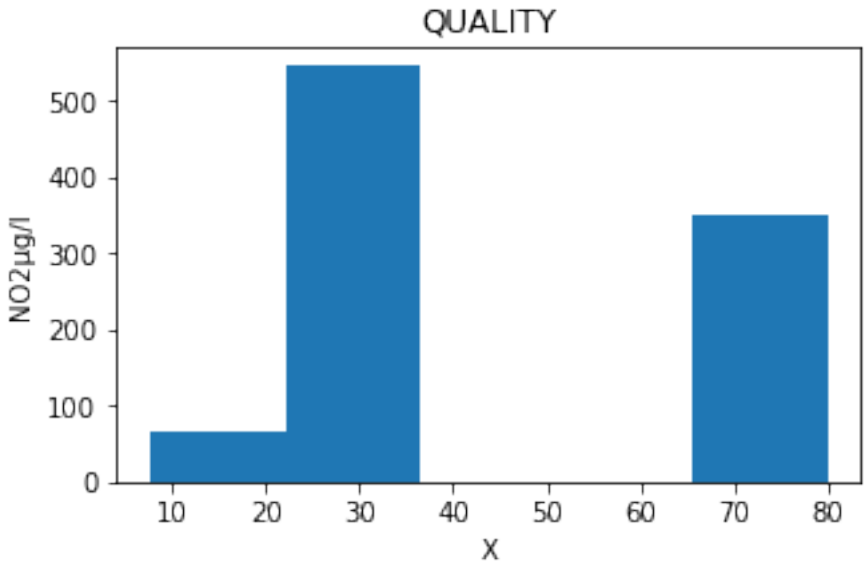
```
#look for missing values
df.isna().sum()
```

```
location         0
month            0
year             0
SO2 µg/l         0
NO2µg/l          0
PM10 µg/l        0
PM2.5 µ g/l      0
CO µg/l          0
O3 µ g/l 8 HR    0
NH3  µ g/l       0
AQI              0
Air Quality      0
dtype: int64
```

histogram

```
import matplotlib.pyplot as plt

fig, ax = plt.subplots(figsize=(5,3))
plt.hist(df['NO2µg/l'],bins=5)
plt.title('QUALITY')
plt.xlabel('X')
plt.ylabel('NO2µg/l')
plt.show()
```



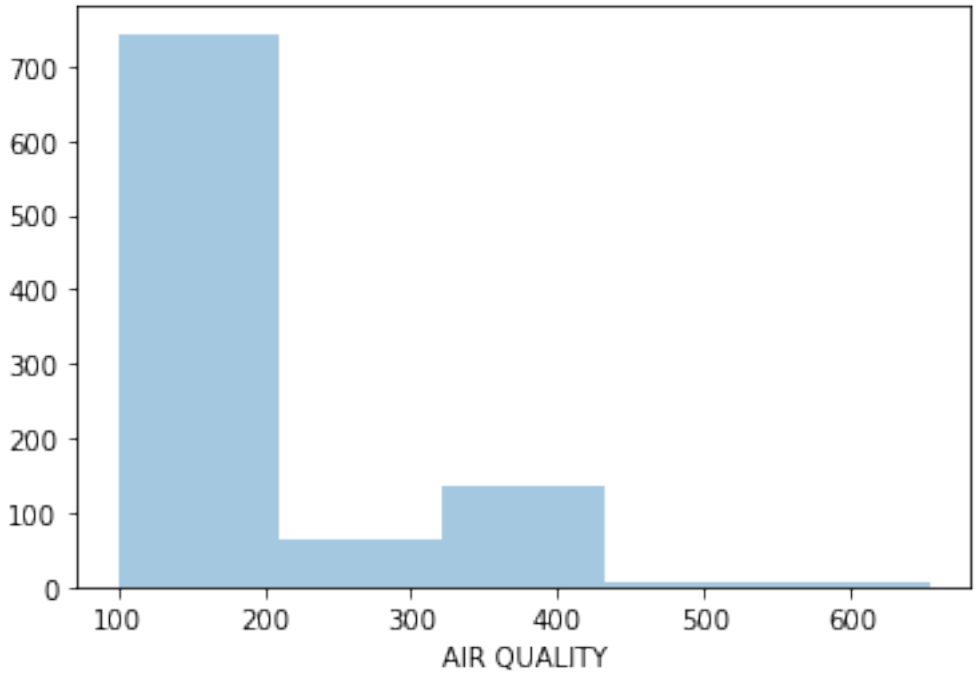
using seaborn

```
import seaborn as sns

sns.distplot(df['AQI'],kde=False,bins=5)
plt.xlabel('AIR QUALITY')
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\
distributions.py:2557: FutureWarning: `distplot` is a deprecated
function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar
flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)

Text(0.5, 0, 'AIR QUALITY')
```

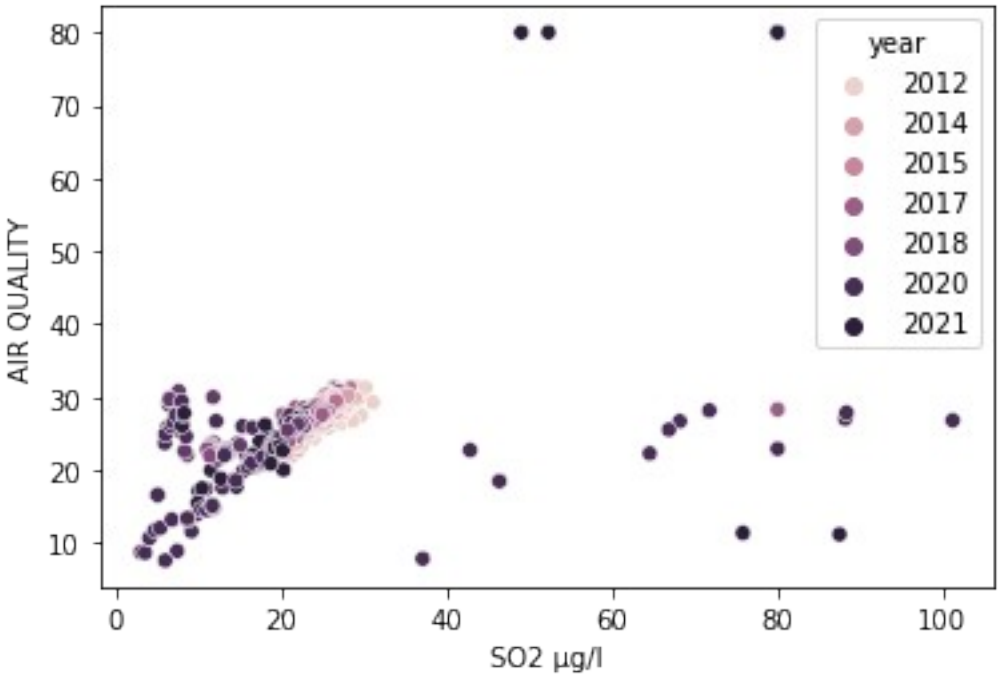


SCATTER PLOT

```
sns.scatterplot(df['SO2 µg/l'],df['NO2µg/l'], hue=df['year'])

plt.ylabel("AIR QUALITY")
plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36:
FutureWarning: Pass the following variables as keyword args: x, y.
From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in
an error or misinterpretation.
  warnings.warn(
```

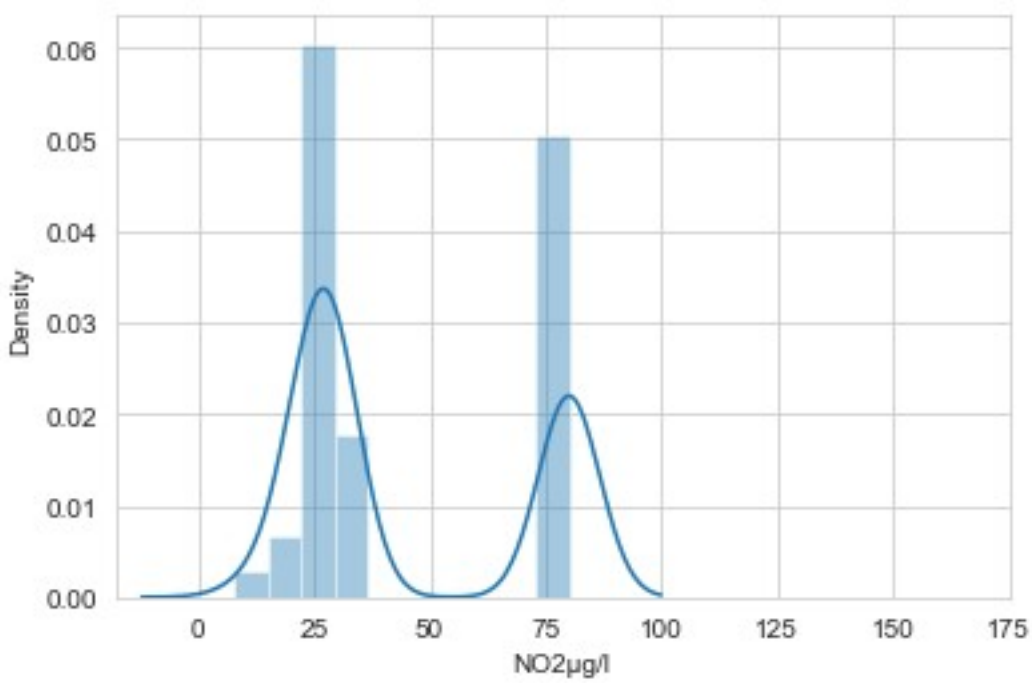


box plot

```
# Box Plot visualization MSSubClass with Seaborn
import seaborn as sns

sns.distplot(df['NO2µg/l'],bins=10)
plt.xticks(range(0,200,25))
plt.show()
```

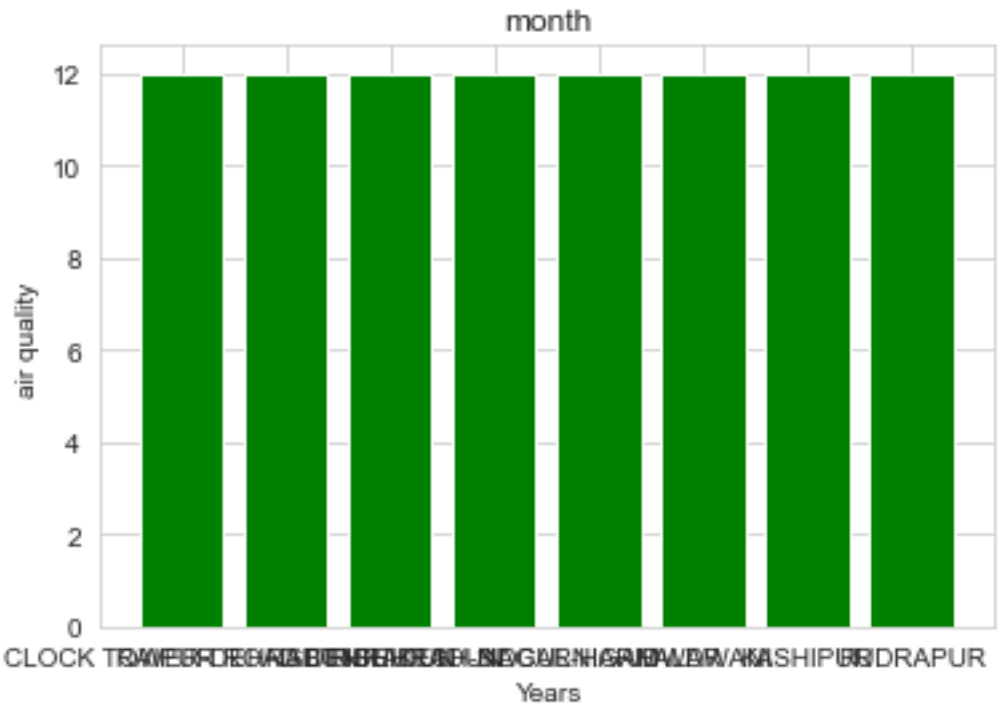
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\
distributions.py:2557: FutureWarning: `distplot` is a deprecated
function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar
flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



BAR GRAPH

```
X = list(df.iloc[:, 0])
Y = list(df.iloc[:, 1])

# Plot the data using bar() method
plt.bar(X, Y, color='g')
plt.title("month")
plt.xlabel("Years")
plt.ylabel("air quality")
figsize=(10,10)
# Show the plot
plt.show()
```


month
12
10
8
6
4
2
0
air quality
CLOCK
SHIPUR
DRAPUR
Years