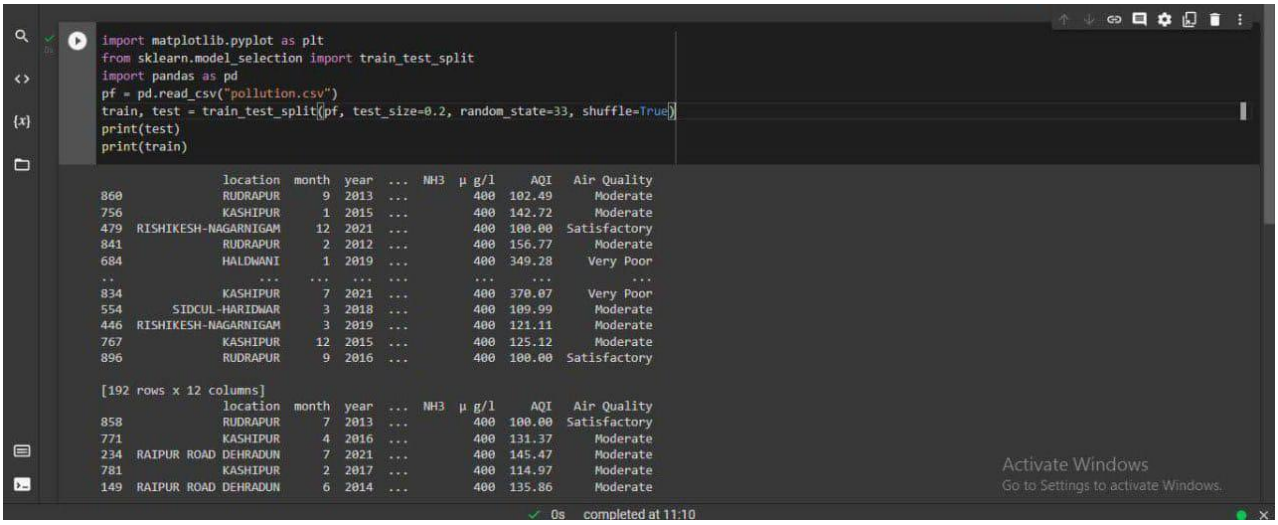
NAME- DHANA KORANGA

MCA-3B

ROLL NO -2001056

1.



```
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
import pandas as pd
pf = pd.read_csv("pollution.csv")
train, test = train_test_split(pf, test_size=0.2, random_state=33, shuffle=True)
print(test)
print(train)
```

```
                location  month  year  ...  NH3  µ g/l    AQI   Air Quality
860             RUDRAPUR      9  2013  ...         400  102.49      Moderate
756             KASHIPUR      1  2015  ...         400  142.72      Moderate
479  RISHIKESH-NAGARNIGAM     12  2021  ...         400  100.00  Satisfactory
841             RUDRAPUR      2  2012  ...         400  156.77      Moderate
684             HALDWANI      1  2019  ...         400  349.28     Very Poor
..                   ...    ...   ...  ...         ...     ...           ...
834             KASHIPUR      7  2021  ...         400  370.07     Very Poor
554      SIDCUL-HARIDWAR      3  2018  ...         400  109.99      Moderate
446  RISHIKESH-NAGARNIGAM      3  2019  ...         400  121.11      Moderate
767             KASHIPUR     12  2015  ...         400  125.12      Moderate
896             RUDRAPUR      9  2016  ...         400  100.00  Satisfactory

[192 rows x 12 columns]
                location  month  year  ...  NH3  µ g/l    AQI   Air Quality
858             RUDRAPUR      7  2013  ...         400  100.00  Satisfactory
771             KASHIPUR      4  2016  ...         400  131.37      Moderate
234  RAIPUR ROAD DEHRADUN      7  2021  ...         400  145.47      Moderate
781             KASHIPUR      2  2017  ...         400  114.97      Moderate
149  RAIPUR ROAD DEHRADUN      6  2014  ...         400  135.86      Moderate
```

Activate Windows
Go to Settings to activate Windows.



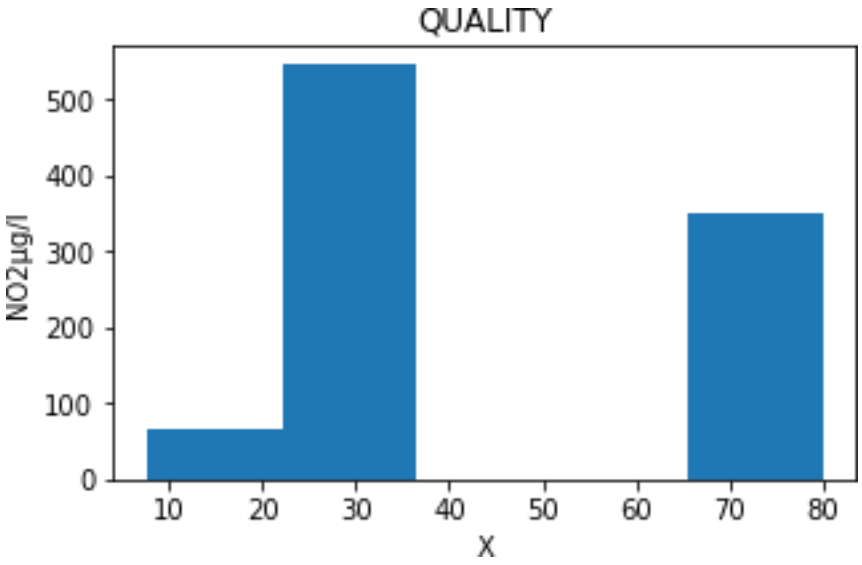
*#look for missing values*

```
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
```

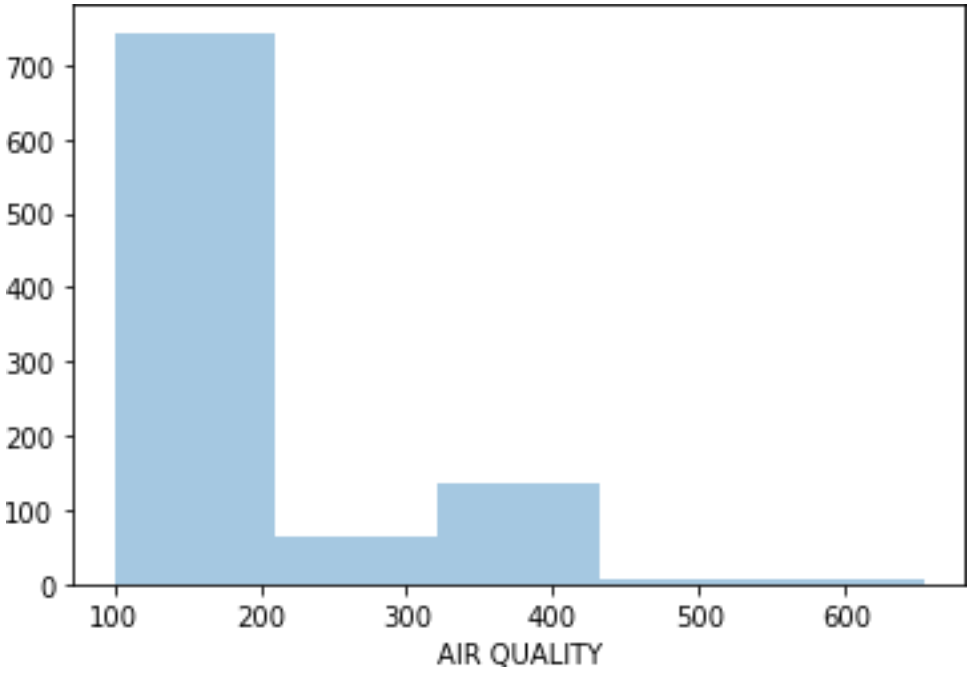
## histogram

```
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()
```



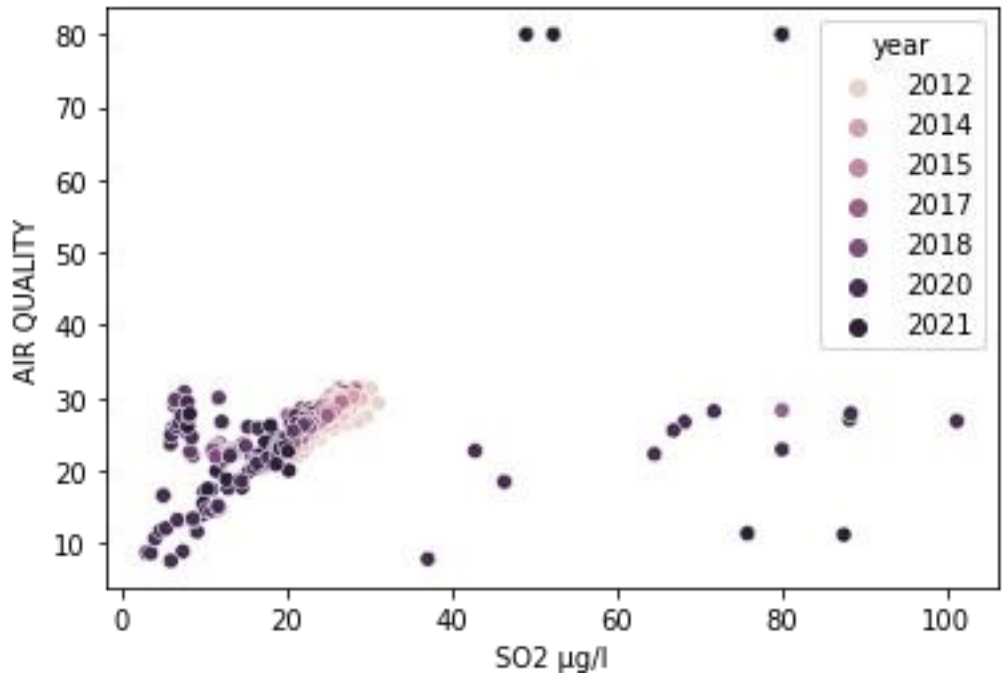

## using seaborn

```
import seaborn as sns

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




## . SCATTER PLOT

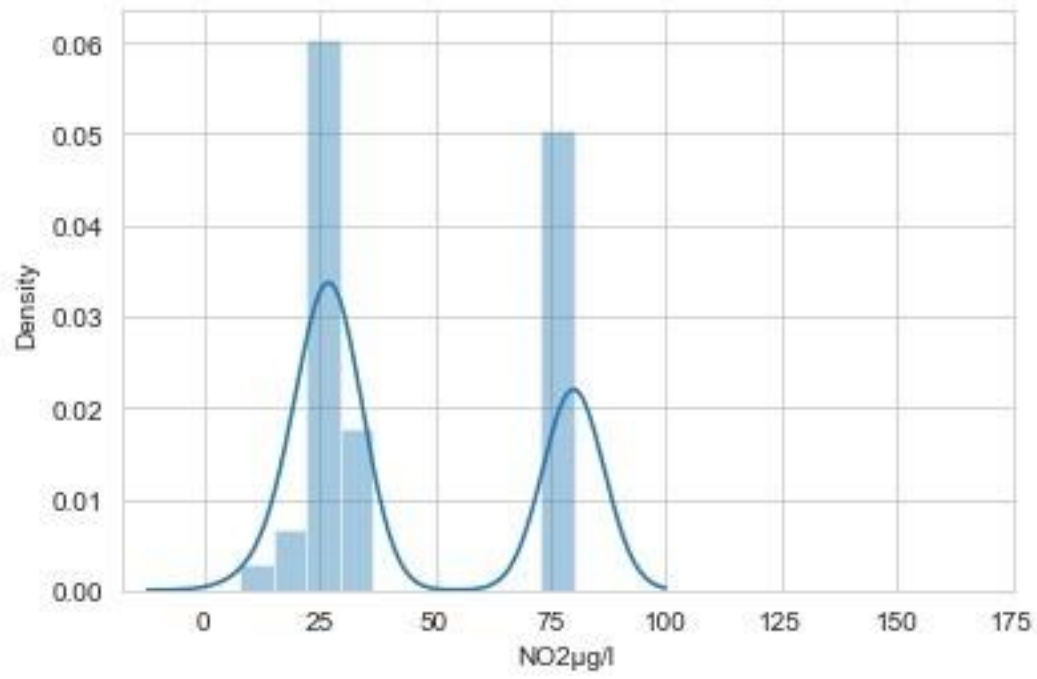
```
scatterplot(df['SO2 µg/l'],df['NO2µg/l'], hue=df['year'])

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

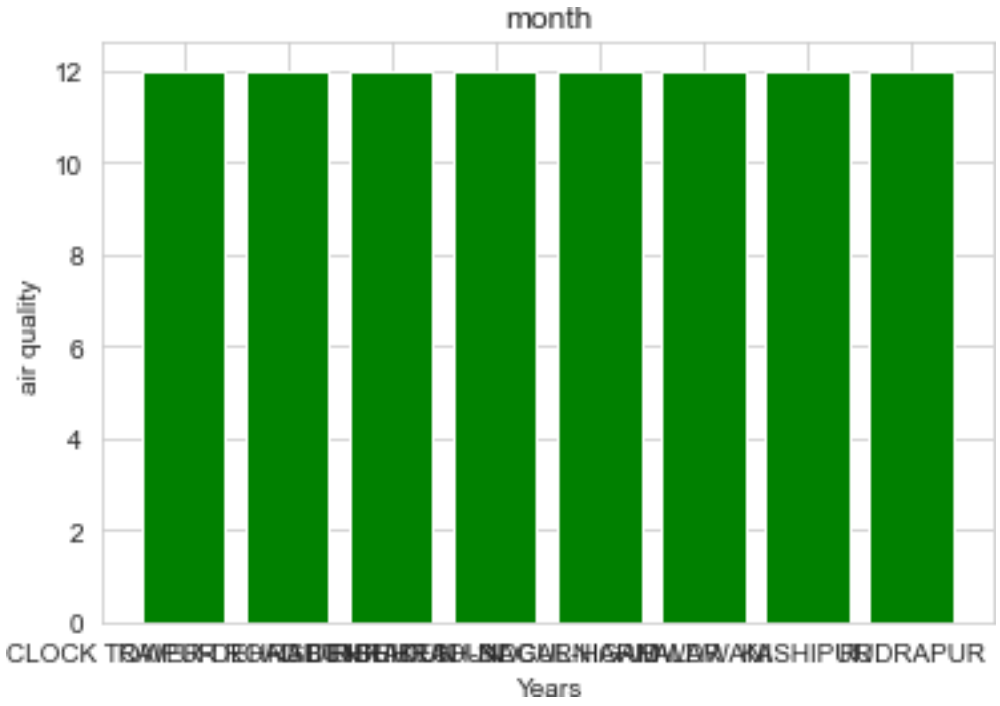



## box plot

*# 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()
```




## BAR GRAPH

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

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

# Show the plot

plt.show()
```