

```
In [1]: import numpy as np
import pandas as pd
```

```
In [3]: df= pd.read_csv("Desktop/pollution.csv")
```

```
In [4]: df.head(10)
```

Out[4]:

	location	month	year	SO2 μg/l	NO2μg/l	PM10 μg/l	PM2.5 μ g/l	CO μg/l	O3 μ g/l 8 HR	NH3 μ g/l	AQI	Air Quality
0	CLOCK TOWER- DEHRADUN	1	2012	27.33	30.33	193.28	60.0	2	100	400	162.19	Moderate
1	CLOCK TOWER- DEHRADUN	2	2012	25.68	25.80	173.77	60.0	2	100	400	149.18	Moderate
2	CLOCK TOWER- DEHRADUN	3	2012	29.64	27.50	211.35	60.0	2	100	400	174.23	Moderate
3	CLOCK TOWER- DEHRADUN	4	2012	28.64	26.81	230.76	60.0	2	100	400	187.17	Moderate
4	CLOCK TOWER- DEHRADUN	5	2012	31.09	29.30	310.73	60.0	2	100	400	260.73	Poor
5	CLOCK TOWER- DEHRADUN	6	2012	28.73	30.62	200.61	60.0	2	100	400	167.07	Moderate
6	CLOCK TOWER- DEHRADUN	7	2012	27.55	30.06	129.22	60.0	2	100	400	119.48	Moderate
7	CLOCK TOWER- DEHRADUN	8	2012	23.04	26.00	78.19	60.0	2	100	400	100.00	Satisfactory
8	CLOCK TOWER- DEHRADUN	9	2012	22.22	25.40	108.37	60.0	2	100	400	105.58	Moderate
9	CLOCK TOWER- DEHRADUN	10	2012	24.40	28.79	21.83	60.0	2	100	400	100.00	Satisfactory

```
In [5]: df.tail()
```

Out[5]:

	location	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3		NH3 µ g/l	AQI	Air Quality
									µ g/l	8 HR			
955	RUDRAPUR	8	2021	80.00	80.00	121.28	208.44	2	100	400	368.03	Very Poor	
956	RUDRAPUR	9	2021	18.73	20.93	92.96	153.75	2	100	400	325.96	Very Poor	
957	RUDRAPUR	10	2021	80.00	80.00	100.00	60.00	2	100	400	100.00	Satisfactory	
958	RUDRAPUR	11	2021	80.00	80.00	100.00	60.00	2	100	400	100.00	Satisfactory	
959	RUDRAPUR	12	2021	80.00	80.00	100.00	60.00	2	100	400	100.00	Satisfactory	

In [9]: `df.describe()`

Out[9]:

	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3 µ g/l 8 HR	NH3 µ g/l
count	960.000000	960.000000	960.000000	960.000000	960.000000	960.000000	960.0	960.0	960.0
mean	6.500000	2016.500000	43.621531	45.821833	142.745323	86.342937	2.0	100.0	400.0
std	3.453852	2.873778	28.571588	26.109628	53.281255	58.664746	0.0	0.0	0.0
min	1.000000	2012.000000	3.030000	7.620000	11.080000	37.100000	2.0	100.0	400.0
25%	3.750000	2014.000000	22.410000	26.422500	110.057500	60.000000	2.0	100.0	400.0
50%	6.500000	2016.500000	26.130000	29.285000	127.790000	60.000000	2.0	100.0	400.0
75%	9.250000	2019.000000	80.000000	80.000000	161.150000	60.000000	2.0	100.0	400.0
max	12.000000	2021.000000	101.170000	80.000000	483.180000	452.920000	2.0	100.0	400.0

◀ ▶

In [10]: `df.shape`

Out[10]: (960, 12)

In [8]: `df.shape`

Out[8]: (960, 12)

In [12]: `df.unique()`

Out[12]:

location	8
month	12
year	10
SO2 µg/l	486
NO2µg/l	455
PM10 µg/l	860
PM2.5 µ g/l	232

```
CO µg/l      1  
O3 µ g/l 8 HR  1  
NH3  µ g/l     1  
AQI          785  
Air Quality    5  
dtype: int64
```

In [13]: `df.isnull().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
```

In []:

In [15]: `df['Air Quality'].value_counts()`

```
Moderate      597  
Satisfactory  138  
Very Poor     130  
Poor           72  
Severe          23  
Name: Air Quality, dtype: int64
```

In [16]: `df['location'].value_counts()`

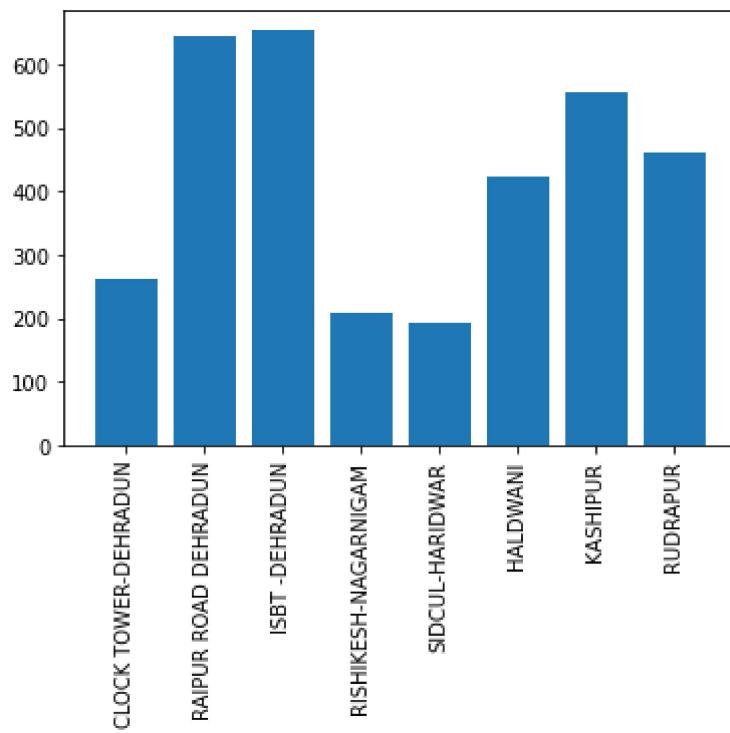
```
CLOCK TOWER-DEHRADUN 120  
RAIPUR ROAD DEHRADUN 120  
ISBT -DEHRADUN       120  
RISHIKESH-NAGARNIGAM 120  
SIDCUL-HARIDWAR      120  
HALDWANI             120  
KASHIPUR              120  
RUDRAPUR              120  
Name: location, dtype: int64
```

In [17]: `import matplotlib.pyplot as plt`

In [18]: `%matplotlib inline`

In [32]: `#distribution for different locations`
`plt.xticks(rotation='vertical')`
`plt.bar(df['location'],df['AQI'])`

Out[32]: <BarContainer object of 960 artists>



In [24]: `loca=pd.pivot_table(df,index='location',values='AQI',aggfunc='count')`

In [25]: `loca`

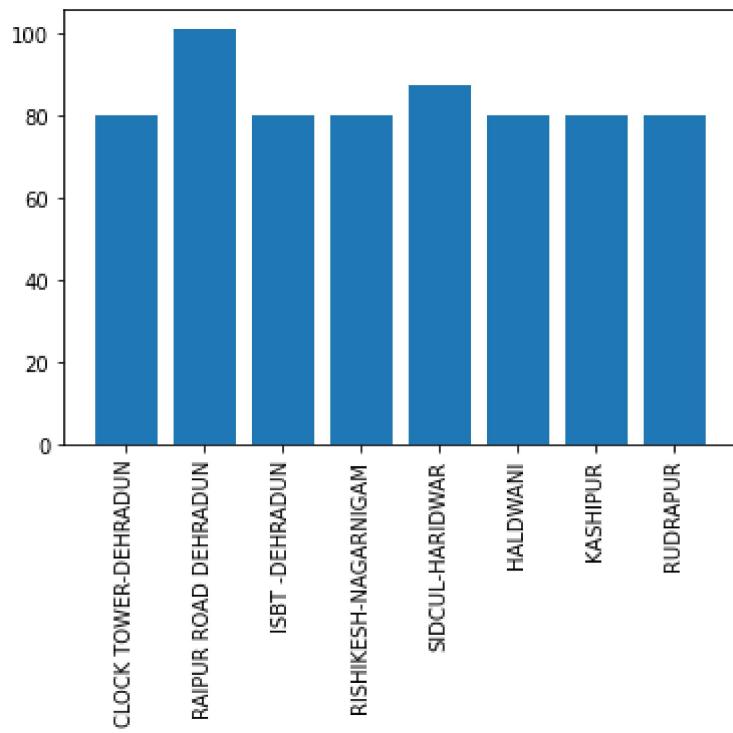
Out[25]: **AQI**

location	
CLOCK TOWER-DEHRADUN	120
HALDWANI	120
ISBT -DEHRADUN	120
KASHIPUR	120
RAIPUR ROAD DEHRADUN	120
RISHIKESH-NAGARNIGAM	120
RUDRAPUR	120
SIDCUL-HARIDWAR	120

In [26]: *#concentration of so2 in different Location*

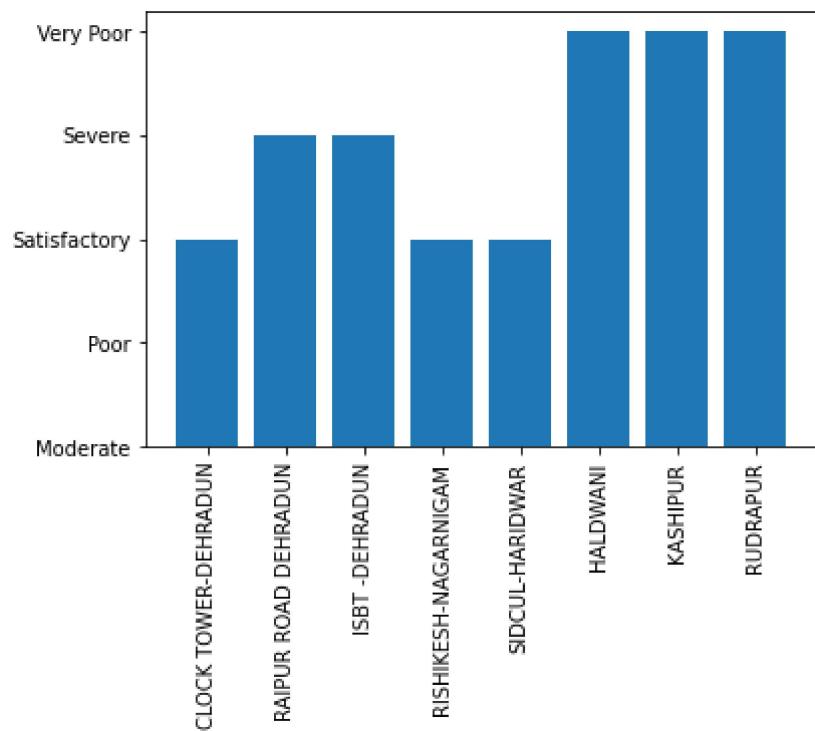
In [31]: `plt.xticks(rotation='vertical')
plt.bar(df['location'],df['S02 µg/l'])`

Out[31]: <BarContainer object of 960 artists>



```
In [34]:  
plt.xticks(rotation='vertical')  
plt.bar(df['location'],df['Air Quality'])
```

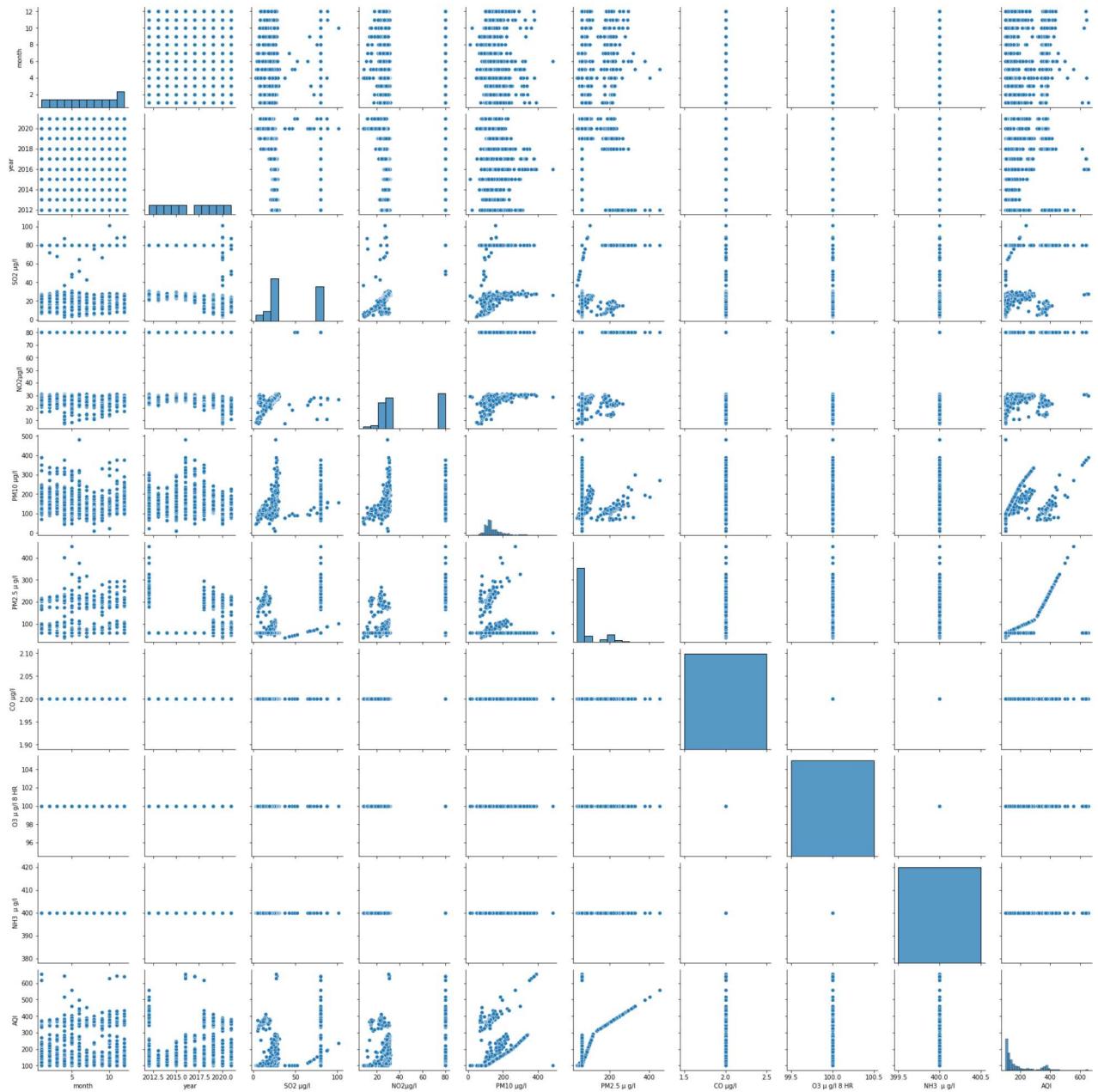
```
Out[34]: <BarContainer object of 960 artists>
```



```
In [35]: import seaborn as sns
```

```
In [36]: sns.pairplot(df)
```

Out[36]: <seaborn.axisgrid.PairGrid at 0x1b45986d400>

In [41]:
data_clock_tower= df.iloc[0:12]In [42]:
data_clock_tower

Out[42]:

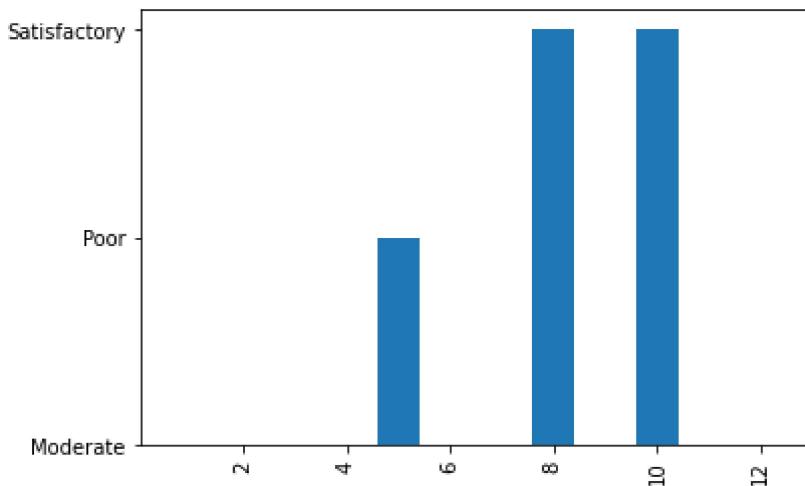
	location	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3 µ g/l	NH3 µ g/l	AQI	Air Quality
0	CLOCK TOWER- DEHRADUN	1	2012	27.33	30.33	193.28	60.0	2	100	400	162.19	Moderate

	location	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3		NH3 µ g/l	AQI	Air Quality
									µ g/l	8 HR			
1	CLOCK TOWER-DEHRADUN	2	2012	25.68	25.80	173.77	60.0	2	100	400	149.18	Moderate	
2	CLOCK TOWER-DEHRADUN	3	2012	29.64	27.50	211.35	60.0	2	100	400	174.23	Moderate	
3	CLOCK TOWER-DEHRADUN	4	2012	28.64	26.81	230.76	60.0	2	100	400	187.17	Moderate	
4	CLOCK TOWER-DEHRADUN	5	2012	31.09	29.30	310.73	60.0	2	100	400	260.73	Poor	
5	CLOCK TOWER-DEHRADUN	6	2012	28.73	30.62	200.61	60.0	2	100	400	167.07	Moderate	
6	CLOCK TOWER-DEHRADUN	7	2012	27.55	30.06	129.22	60.0	2	100	400	119.48	Moderate	
7	CLOCK TOWER-DEHRADUN	8	2012	23.04	26.00	78.19	60.0	2	100	400	100.00	Satisfactory	
8	CLOCK TOWER-DEHRADUN	9	2012	22.22	25.40	108.37	60.0	2	100	400	105.58	Moderate	
9	CLOCK TOWER-DEHRADUN	10	2012	24.40	28.79	21.83	60.0	2	100	400	100.00	Satisfactory	
10	CLOCK TOWER-DEHRADUN	11	2012	25.15	29.23	153.67	60.0	2	100	400	135.78	Moderate	
11	CLOCK TOWER-DEHRADUN	12	2012	25.40	29.24	214.04	60.0	2	100	400	176.03	Moderate	

In [48]:

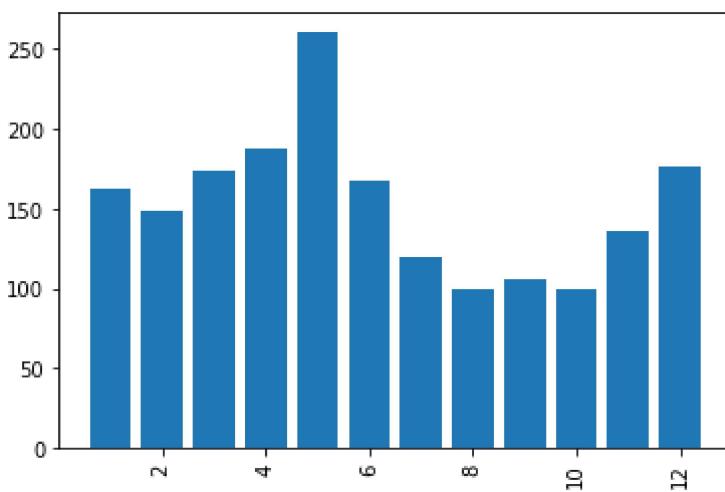
```
#DISTRIBUTION AMONG THE MONTHS IN CLOCK_TOWER DEHRADUN
plt.xticks(rotation='vertical')
plt.bar(data_clock_tower['month'],data_clock_tower['Air Quality'])
```

Out[48]: <BarContainer object of 12 artists>



```
In [49]: plt.xticks(rotation='vertical')
plt.bar(data_clock_tower['month'],data_clock_tower['AQI'])
```

Out[49]: <BarContainer object of 12 artists>



```
In [58]: data_jan= df[df['month']==1]
```

In [59]: data_jan

Out[59]:

	location	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3 µ g/l	NH3 µ g/l	AQI	Air Quality
0	CLOCK TOWER- DEHRADUN	1	2012	27.33	30.33	193.28	60.00	2	100	400	162.19	Moderate
12	CLOCK TOWER- DEHRADUN	1	2013	80.00	80.00	100.00	60.00	2	100	400	100.00	Satisfactory

		location	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3 µ g/l 8 HR	NH3 µ g/l	AQI	Air Quality
24	CLOCK TOWER- DEHRADUN		1	2014	25.86	29.61	183.05	60.00	2	100	400	155.37	Moderate
36	CLOCK TOWER- DEHRADUN		1	2015	26.40	31.43	196.19	60.00	2	100	400	164.13	Moderate
48	CLOCK TOWER- DEHRADUN		1	2016	25.38	29.38	226.92	60.00	2	100	400	184.61	Moderate
...
900	RUDRAPUR		1	2017	80.00	80.00	185.92	60.00	2	100	400	157.28	Moderate
912	RUDRAPUR		1	2018	80.00	80.00	108.15	207.30	2	100	400	367.15	Very Poor
924	RUDRAPUR		1	2019	13.28	22.52	118.14	217.30	2	100	400	374.85	Very Poor
936	RUDRAPUR		1	2020	13.32	22.39	121.09	212.51	2	100	400	371.16	Very Poor
948	RUDRAPUR		1	2021	18.84	22.30	123.58	212.51	2	100	400	371.16	Very Poor

80 rows × 12 columns

In [60]:

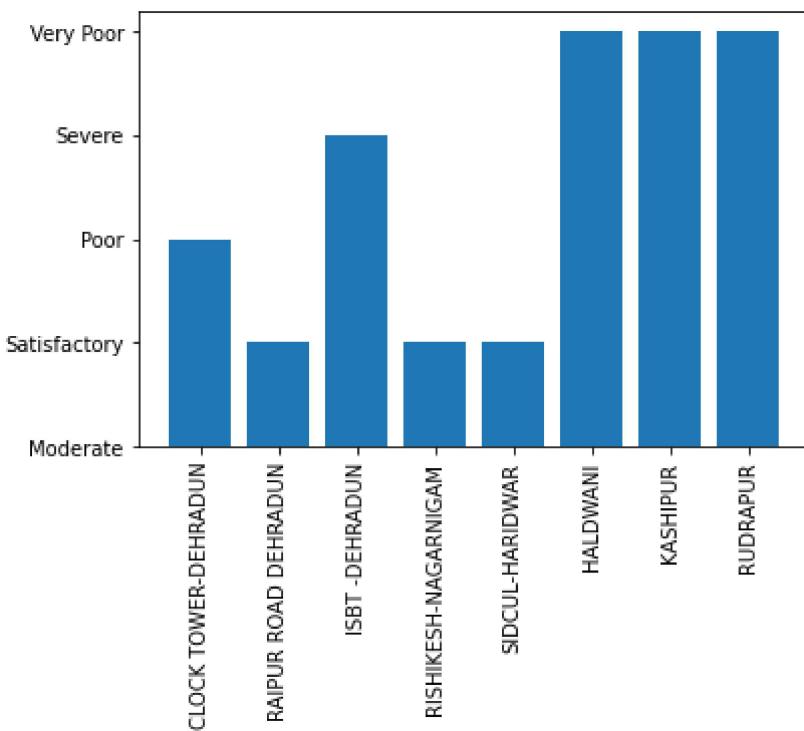
data_jan.head()

Out[60]:

		location	month	year	SO2 µg/l	NO2µg/l	PM10 µg/l	PM2.5 µ g/l	CO µg/l	O3 µ g/l 8 HR	NH3 µ g/l	AQI	Air Quality
0	CLOCK TOWER- DEHRADUN		1	2012	27.33	30.33	193.28	60.0	2	100	400	162.19	Moderate
12	CLOCK TOWER- DEHRADUN		1	2013	80.00	80.00	100.00	60.0	2	100	400	100.00	Satisfactory
24	CLOCK TOWER- DEHRADUN		1	2014	25.86	29.61	183.05	60.0	2	100	400	155.37	Moderate
36	CLOCK TOWER- DEHRADUN		1	2015	26.40	31.43	196.19	60.0	2	100	400	164.13	Moderate
48	CLOCK TOWER- DEHRADUN		1	2016	25.38	29.38	226.92	60.0	2	100	400	184.61	Moderate

```
In [62]: #distribution during the month of jan in differnt location
plt.xticks(rotation='vertical')
plt.bar(data_jan['location'],data_jan['Air Quality'])
```

Out[62]: <BarContainer object of 80 artists>



```
In [63]: #split the dataset into test train
#independent variable
independent = df.iloc[1:,3:10].values
```

In [64]: independent

```
Out[64]: array([[ 25.68,  25.8 , 173.77, ... , 2. , 100. , 400. ],
   [ 29.64,  27.5 , 211.35, ... , 2. , 100. , 400. ],
   [ 28.64,  26.81, 230.76, ... , 2. , 100. , 400. ],
   ... ,
   [ 80. ,  80. , 100. , ... , 2. , 100. , 400. ],
   [ 80. ,  80. , 100. , ... , 2. , 100. , 400. ],
   [ 80. ,  80. , 100. , ... , 2. , 100. , 400. ]])
```

```
In [65]: #dependent variable
dep= df.iloc[1:,11].values
```

In [66]: dep

```
Out[66]: array(['Moderate', 'Moderate', 'Moderate', 'Poor', 'Moderate', 'Moderate',
   'Satisfactory', 'Moderate', 'Satisfactory', 'Moderate', 'Moderate',
   'Satisfactory', 'Satisfactory', 'Satisfactory', 'Moderate',
   'Moderate', 'Moderate', 'Satisfactory', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Moderate', 'Moderate'],
  dtype='|S11')
```



```
'Moderate', 'Moderate', 'Very Poor', 'Very Poor', 'Very Poor',  
'Very Poor', 'Very Poor', 'Very Poor', 'Very Poor', 'Very Poor',  
'Very Poor', 'Satisfactory', 'Satisfactory', 'Satisfactory'],  
dtype=object)
```

```
In [67]: from sklearn.model_selection import train_test_split
```

```
In [69]: x_train, x_test,y_train, y_test=train_test_split(independent,dep,test_size=0.2,random_s
```

In [70]: x_train

```
Out[70]: array([[ 80. ,  80. , 129.36, ... , 2. , 100. , 400. ],
   [ 20.91, 24.6 , 101.02, ... , 2. , 100. , 400. ],
   [ 27.8 , 29.93, 188.52, ... , 2. , 100. , 400. ],
   ... ,
   [ 6.02, 24.94, 105.89, ... , 2. , 100. , 400. ],
   [ 26.51, 31.1 , 117.8 , ... , 2. , 100. , 400. ],
   [ 80. , 80. , 172.62, ... , 2. , 100. , 400. ]])
```

In [71]: x test

```
Out[71]: array([[ 80. ,  80. , 133.23, ...],  
 [ 25.27, 28.25, 164.99, ...],  
 [ 27.14, 30.49, 153.92, ...],  
 ...,  
 [ 80. ,  80. , 137.68, ...],  
 [ 80. ,  80. ,  86.85, ...],  
 [  6.55, 29.83, 107.94, ...]])
```

```
In [72]: y_train
```



```
'Moderate', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
'Moderate', 'Moderate', 'Moderate', 'Moderate', 'Poor', 'Moderate',
'Moderate', 'Moderate', 'Satisfactory', 'Moderate', 'Moderate',
'Moderate', 'Moderate', 'Moderate', 'Moderate', 'Very Poor',
'Moderate', 'Moderate', 'Poor', 'Moderate', 'Moderate', 'Moderate',
'Moderate', 'Satisfactory', 'Satisfactory', 'Moderate', 'Moderate',
'Satisfactory', 'Very Poor', 'Moderate', 'Moderate']], dtype=object)
```

In [73]: `y_test`

```
Out[73]: array(['Moderate', 'Moderate', 'Moderate', 'Very Poor', 'Very Poor',
   'Satisfactory', 'Poor', 'Very Poor', 'Severe', 'Moderate',
   'Satisfactory', 'Moderate', 'Moderate', 'Satisfactory', 'Moderate',
   'Satisfactory', 'Moderate', 'Moderate', 'Very Poor',
   'Satisfactory', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Poor', 'Very Poor', 'Moderate', 'Very Poor', 'Moderate',
   'Moderate', 'Poor', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Poor', 'Moderate',
   'Satisfactory', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Satisfactory', 'Satisfactory',
   'Very Poor', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Poor', 'Moderate', 'Satisfactory', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Poor',
   'Very Poor', 'Satisfactory', 'Satisfactory', 'Very Poor',
   'Moderate', 'Severe', 'Moderate', 'Moderate', 'Satisfactory',
   'Moderate', 'Moderate', 'Moderate', 'Very Poor', 'Satisfactory',
   'Very Poor', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Satisfactory', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Satisfactory', 'Very Poor',
   'Poor', 'Severe', 'Moderate', 'Moderate', 'Satisfactory',
   'Moderate', 'Satisfactory', 'Moderate', 'Satisfactory', 'Severe',
   'Satisfactory', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Satisfactory', 'Moderate', 'Very Poor', 'Moderate', 'Poor',
   'Moderate', 'Moderate', 'Severe', 'Satisfactory', 'Moderate',
   'Satisfactory', 'Poor', 'Satisfactory', 'Very Poor', 'Moderate',
   'Moderate', 'Satisfactory', 'Moderate', 'Very Poor', 'Moderate',
   'Poor', 'Satisfactory', 'Moderate', 'Moderate', 'Severe',
   'Moderate', 'Satisfactory', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Very Poor', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Very Poor', 'Moderate', 'Moderate', 'Moderate',
   'Poor', 'Poor', 'Poor', 'Satisfactory', 'Satisfactory', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Very Poor', 'Moderate',
   'Moderate', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Moderate', 'Satisfactory', 'Moderate', 'Very Poor',
   'Very Poor', 'Moderate', 'Moderate', 'Moderate', 'Moderate',
   'Moderate', 'Satisfactory', 'Moderate', 'Moderate', 'Very Poor',
   'Moderate', 'Moderate', 'Moderate', 'Satisfactory', 'Moderate',
   'Moderate', 'Satisfactory', 'Very Poor'], dtype=object)
```

In [74]: `#knn`
`from sklearn.neighbors import KNeighborsClassifier`
`knn=KNeighborsClassifier(n_neighbors=7)`
`knn.fit(x_train,y_train)`
`pred= knn.predict(x_test)`

In []:

```
In [75]: from sklearn.metrics import confusion_matrix  
print(confusion_matrix(pred,y_test))
```

```
[[115  3  0  0  0]  
 [ 1  10  0  1  0]  
 [ 3  0  32  0  0]  
 [ 0  0  0  4  0]  
 [ 0  0  0  1  22]]
```

```
In [77]: accuracy_knn=knn.score(x_test,y_test)*100  
print(accuracy_knn)
```

```
95.3125
```

```
In [78]: #decision tree  
from sklearn.tree import DecisionTreeClassifier  
d_tree= DecisionTreeClassifier(criterion='entropy',random_state=0)  
d_tree.fit(x_train,y_train)  
y_pred=d_tree.predict(x_test)
```

```
In [79]: from sklearn.metrics import confusion_matrix  
print(confusion_matrix(y_pred,y_test))
```

```
[[118  0  0  0  0]  
 [ 0  13  0  0  0]  
 [ 1  0  32  0  0]  
 [ 0  0  0  5  0]  
 [ 0  0  0  1  22]]
```

```
In [82]: accuracy_d_tree= d_tree.score(x_test,y_test)*100  
print("Decision tree accuracy is: ",accuracy_d_tree)
```

```
Decision tree accuracy is:  98.95833333333334
```

```
In [84]: #SVM (support vector machine)  
from sklearn.svm import SVC  
svm= SVC(kernel="linear",random_state=0)  
svm.fit(x_train,y_train)  
y_pred=svm.predict(x_test)
```

```
In [85]: from sklearn.metrics import confusion_matrix  
print(confusion_matrix(y_pred,y_test))
```

```
[[116  0  0  0  0]  
 [ 0  13  0  0  0]  
 [ 3  0  32  0  0]  
 [ 0  0  0  5  0]  
 [ 0  0  0  1  22]]
```

```
In [86]: accuracy_svm= svm.score(x_test,y_test)*100  
print("Accuracy of SVM is: ",accuracy_svm)
```

Accuracy of SVM is: 97.91666666666666

In []: