In [53]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [54]:

```
df = pd.read_csv(r"C:\Users\Vinayak\Downloads\dsbda\B_ass2\city_day.csv",sep=",")
df
```

Out[54]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	
0	Ahmedabad	2015-01-01	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	13
1	Ahmedabad	2015-01-02	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	3
2	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	3
3	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	3
4	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	3
5	Ahmedabad	2015-01-06	NaN	NaN	45.41	38.48	81.50	NaN	45.41	45.76	4
6	Ahmedabad	2015-01-07	NaN	NaN	112.16	40.62	130.77	NaN	112.16	32.28	3
7	Ahmedabad	2015-01-08	NaN	NaN	80.87	36.74	96.75	NaN	80.87	38.54	3
8	Ahmedabad	2015-01-09	NaN	NaN	29.16	31.00	48.00	NaN	29.16	58.68	2
9	Ahmedabad	2015-01-10	NaN	NaN	NaN	7.04	0.00	NaN	NaN	8.29	
10	Ahmedabad	2015-01-11	NaN	NaN	132.07	55.80	24.53	NaN	132.07	25.03	
11	Ahmedabad	2015-01-12	NaN	NaN	52.04	40.67	90.24	NaN	52.04	51.84	4
12	Ahmedabad	2015-01-13	NaN	NaN	48.82	44.20	87.09	NaN	48.82	68.21	3
13	Ahmedabad	2015-01-14	NaN	NaN	19.20	27.86	33.05	NaN	19.20	52.65	2
14	Ahmedabad	2015-01-15	NaN	NaN	0.60	16.96	16.60	NaN	0.60	28.89	4
15	Ahmedabad	2015-01-16	NaN	NaN	1.63	21.72	22.86	NaN	1.63	38.27	4
16	Ahmedabad	2015-01-17	NaN	NaN	11.44	24.73	34.75	NaN	11.44	49.50	5
17	Ahmedabad	2015-01-18	NaN	NaN	6.10	25.77	29.57	NaN	6.10	48.43	5
18	Ahmedabad	2015-01-19	NaN	NaN	2.51	26.88	27.45	NaN	2.51	50.03	4
19	Ahmedabad	2015-01-20	NaN	NaN	7.92	26.80	32.40	NaN	7.92	58.87	5
20	Ahmedabad	2015-01-21	NaN	NaN	9.52	33.56	39.28	NaN	9.52	106.93	4
21	Ahmedabad	2015-01-22	NaN	NaN	9.05	17.51	22.33	NaN	9.05	23.71	4
22	Ahmedabad	2015-01-23	NaN	NaN	22.53	27.96	47.79	NaN	22.53	39.19	3
23	Ahmedabad	2015-01-24	NaN	NaN	2.03	20.39	21.40	NaN	2.03	40.07	3

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	
24	Ahmedabad	2015-01-25	NaN	NaN	1.42	20.43	20.19	NaN	1.42	58.41	3
25	Ahmedabad	2015-01-26	NaN	NaN	2.27	21.16	21.81	NaN	2.27	43.73	3
26	Ahmedabad	2015-01-27	NaN	NaN	2.19	21.70	23.26	NaN	2.19	43.28	4
27	Ahmedabad	2015-01-28	73.24	NaN	5.72	21.11	25.84	NaN	5.72	36.52	6
28	Ahmedabad	2015-01-29	83.13	NaN	6.93	28.71	33.72	NaN	6.93	49.52	5
29	Ahmedabad	2015-01-30	79.84	NaN	13.85	28.68	41.08	NaN	13.85	48.49	9
...	...	...	...	...	...	...	...	...	...	...	
29501	Visakhapatnam	2020-06-02	9.91	43.30	2.81	19.91	12.60	9.16	0.47	8.78	1
29502	Visakhapatnam	2020-06-03	17.94	102.58	5.60	30.32	20.59	9.49	0.58	6.48	2
29503	Visakhapatnam	2020-06-04	19.14	77.06	4.12	35.57	22.18	8.82	0.68	7.19	2
29504	Visakhapatnam	2020-06-05	29.71	112.53	2.98	44.77	24.69	10.51	0.85	10.64	2
29505	Visakhapatnam	2020-06-06	21.66	74.97	3.38	38.42	22.56	8.63	0.64	7.02	2
29506	Visakhapatnam	2020-06-07	29.69	101.97	2.84	35.51	20.66	9.85	0.57	5.32	3
29507	Visakhapatnam	2020-06-08	31.14	125.15	6.55	41.56	27.26	9.71	0.64	6.61	3
29508	Visakhapatnam	2020-06-09	31.84	85.41	4.69	50.11	30.35	10.01	0.88	5.06	3
29509	Visakhapatnam	2020-06-10	32.55	79.28	4.78	44.59	27.38	9.34	0.69	5.90	3
29510	Visakhapatnam	2020-06-11	37.71	98.53	9.97	52.10	35.79	9.91	0.70	5.59	2
29511	Visakhapatnam	2020-06-12	37.67	76.27	12.17	53.59	35.52	9.85	0.97	5.74	1
29512	Visakhapatnam	2020-06-13	22.75	86.77	3.85	38.23	22.60	9.24	0.67	10.70	2
29513	Visakhapatnam	2020-06-14	21.33	83.23	2.81	34.56	20.43	9.78	0.58	9.37	2
29514	Visakhapatnam	2020-06-15	24.66	100.53	3.66	37.93	22.97	9.22	0.71	7.02	2
29515	Visakhapatnam	2020-06-16	21.06	80.51	6.67	40.33	26.81	8.81	0.63	11.09	2
29516	Visakhapatnam	2020-06-17	17.05	73.22	6.08	36.31	24.23	6.69	0.54	11.08	2
29517	Visakhapatnam	2020-06-18	21.16	88.80	4.94	42.96	26.79	3.22	0.68	11.05	2
29518	Visakhapatnam	2020-06-19	14.41	66.81	6.22	40.06	26.10	7.51	1.07	13.23	2

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	
29519	Visakhapatnam	2020-06-20	26.67	86.29	3.78	37.36	22.46	15.05	0.74	11.76	4
29520	Visakhapatnam	2020-06-21	25.96	91.75	5.65	40.22	25.78	13.32	0.71	10.10	3
29521	Visakhapatnam	2020-06-22	33.17	108.22	5.58	42.45	27.06	13.70	0.73	13.65	3
29522	Visakhapatnam	2020-06-23	25.40	83.38	2.76	34.09	19.92	13.13	0.54	10.40	4
29523	Visakhapatnam	2020-06-24	34.36	90.90	1.22	23.38	13.12	14.45	0.56	10.92	3
29524	Visakhapatnam	2020-06-25	13.45	58.54	2.30	21.60	13.09	12.27	0.41	8.19	2
29525	Visakhapatnam	2020-06-26	7.63	32.27	5.91	23.27	17.19	11.15	0.46	6.87	1
29526	Visakhapatnam	2020-06-27	15.02	50.94	7.68	25.06	19.54	12.47	0.47	8.55	2
29527	Visakhapatnam	2020-06-28	24.38	74.09	3.42	26.06	16.53	11.99	0.52	12.72	3
29528	Visakhapatnam	2020-06-29	22.91	65.73	3.45	29.53	18.33	10.71	0.48	8.42	3
29529	Visakhapatnam	2020-06-30	16.64	49.97	4.05	29.26	18.80	10.03	0.52	9.84	2
29530	Visakhapatnam	2020-07-01	15.00	66.00	0.40	26.85	14.05	5.20	0.59	2.10	1

29531 rows × 16 columns

In [11]:

```
df1.describe()
```

Out[11]:

	NO	NO2	NOx	NH3	CO	SO2
count	25949.000000	25946.000000	25346.000000	19203.000000	27472.000000	25677.000000
mean	17.574730	28.560659	32.309123	23.483476	2.248598	14.531977
std	22.785846	24.474746	31.646011	25.684275	6.962884	18.133775
min	0.020000	0.010000	0.000000	0.010000	0.000000	0.010000
25%	5.630000	11.750000	12.820000	8.580000	0.510000	5.670000
50%	9.890000	21.690000	23.520000	15.850000	0.890000	9.160000
75%	19.950000	37.620000	40.127500	30.020000	1.450000	15.220000
max	390.680000	362.210000	467.630000	352.890000	175.810000	193.860000

In [ ]:

```
##data cleaning
```

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
City             0
Date             0
PM2.5         4598
PM10         11140
NO            3582
NO2           3585
NOx           4185
NH3          10328
CO            2059
SO2           3854
O3            4022
Benzene       5623
Toluene       8041
Xylene       18109
AQI           4681
AQI_Bucket    4681
dtype: int64
```

In [9]:

```
#removing unwanted columns
select = ['City', 'Date', 'NO', 'NO2', 'NOx', 'NH3', 'CO', 'SO2', 'O3', 'Benzene', 'Tolu
df1 = df[select].copy()
df1
```

Out[9]:

	City	Date	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene
0	Ahmedabad	2015-01-01	0.92	18.22	17.15	NaN	0.92	27.64	133.36	0.00
1	Ahmedabad	2015-01-02	0.97	15.69	16.46	NaN	0.97	24.55	34.06	3.68
2	Ahmedabad	2015-01-03	17.40	19.30	29.70	NaN	17.40	29.07	30.70	6.80
3	Ahmedabad	2015-01-04	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4.43
4	Ahmedabad	2015-01-05	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7.01
5	Ahmedabad	2015-01-06	45.41	38.48	81.50	NaN	45.41	45.76	46.51	5.42
6	Ahmedabad	2015-01-07	112.16	40.62	130.77	NaN	112.16	32.28	33.47	0.00
7	Ahmedabad	2015-01-08	80.87	36.74	96.75	NaN	80.87	38.54	31.89	0.00
8	Ahmedabad	2015-01-09	29.16	31.00	48.00	NaN	29.16	58.68	25.75	0.00
9	Ahmedabad	2015-01-10	NaN	7.04	0.00	NaN	NaN	8.29	4.55	0.00
10	Ahmedabad	2015-01-11	132.07	55.80	24.53	NaN	132.07	25.03	6.79	0.00
11	Ahmedabad	2015-01-12	52.04	40.67	90.24	NaN	52.04	51.84	45.89	2.41
12	Ahmedabad	2015-01-13	48.82	44.20	87.09	NaN	48.82	68.21	35.16	9.45
13	Ahmedabad	2015-01-14	19.20	27.86	33.05	NaN	19.20	52.65	20.96	2.16
14	Ahmedabad	2015-01-15	0.60	16.96	16.60	NaN	0.60	28.89	47.63	0.14
15	Ahmedabad	2015-01-16	1.63	21.72	22.86	NaN	1.63	38.27	46.03	0.35
16	Ahmedabad	2015-01-17	11.44	24.73	34.75	NaN	11.44	49.50	52.24	0.68
17	Ahmedabad	2015-01-18	6.10	25.77	29.57	NaN	6.10	48.43	53.49	0.74
18	Ahmedabad	2015-01-19	2.51	26.88	27.45	NaN	2.51	50.03	49.48	0.26
19	Ahmedabad	2015-01-20	7.92	26.80	32.40	NaN	7.92	58.87	56.37	0.24
20	Ahmedabad	2015-01-21	9.52	33.56	39.28	NaN	9.52	106.93	48.75	0.33
21	Ahmedabad	2015-01-22	9.05	17.51	22.33	NaN	9.05	23.71	42.22	0.00
22	Ahmedabad	2015-01-23	22.53	27.96	47.79	NaN	22.53	39.19	32.92	0.39
23	Ahmedabad	2015-01-24	2.03	20.39	21.40	NaN	2.03	40.07	32.49	0.47

	City	Date	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene
24	Ahmedabad	2015-01-25	1.42	20.43	20.19	NaN	1.42	58.41	39.26	0.01
25	Ahmedabad	2015-01-26	2.27	21.16	21.81	NaN	2.27	43.73	39.83	0.06
26	Ahmedabad	2015-01-27	2.19	21.70	23.26	NaN	2.19	43.28	41.10	0.02
27	Ahmedabad	2015-01-28	5.72	21.11	25.84	NaN	5.72	36.52	62.42	0.03
28	Ahmedabad	2015-01-29	6.93	28.71	33.72	NaN	6.93	49.52	59.76	0.02
29	Ahmedabad	2015-01-30	13.85	28.68	41.08	NaN	13.85	48.49	97.07	0.04
...	...	...	...	...	...	...	...	...	...	...
29501	Visakhapatnam	2020-06-02	2.81	19.91	12.60	9.16	0.47	8.78	19.68	1.37
29502	Visakhapatnam	2020-06-03	5.60	30.32	20.59	9.49	0.58	6.48	23.82	2.65
29503	Visakhapatnam	2020-06-04	4.12	35.57	22.18	8.82	0.68	7.19	25.56	3.06
29504	Visakhapatnam	2020-06-05	2.98	44.77	24.69	10.51	0.85	10.64	21.94	4.50
29505	Visakhapatnam	2020-06-06	3.38	38.42	22.56	8.63	0.64	7.02	28.86	2.57
29506	Visakhapatnam	2020-06-07	2.84	35.51	20.66	9.85	0.57	5.32	31.71	3.07
29507	Visakhapatnam	2020-06-08	6.55	41.56	27.26	9.71	0.64	6.61	36.65	3.47
29508	Visakhapatnam	2020-06-09	4.69	50.11	30.35	10.01	0.88	5.06	33.86	4.07
29509	Visakhapatnam	2020-06-10	4.78	44.59	27.38	9.34	0.69	5.90	35.06	2.96
29510	Visakhapatnam	2020-06-11	9.97	52.10	35.79	9.91	0.70	5.59	26.59	5.29
29511	Visakhapatnam	2020-06-12	12.17	53.59	35.52	9.85	0.97	5.74	17.35	4.77
29512	Visakhapatnam	2020-06-13	3.85	38.23	22.60	9.24	0.67	10.70	27.24	3.06
29513	Visakhapatnam	2020-06-14	2.81	34.56	20.43	9.78	0.58	9.37	29.44	2.29
29514	Visakhapatnam	2020-06-15	3.66	37.93	22.97	9.22	0.71	7.02	28.63	3.13
29515	Visakhapatnam	2020-06-16	6.67	40.33	26.81	8.81	0.63	11.09	21.69	3.77
29516	Visakhapatnam	2020-06-17	6.08	36.31	24.23	6.69	0.54	11.08	26.09	2.96
29517	Visakhapatnam	2020-06-18	4.94	42.96	26.79	3.22	0.68	11.05	27.65	4.38
29518	Visakhapatnam	2020-06-19	6.22	40.06	26.10	7.51	1.07	13.23	27.34	2.55

	City	Date	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene
29519	Visakhapatnam	2020-06-20	3.78	37.36	22.46	15.05	0.74	11.76	41.97	3.14
29520	Visakhapatnam	2020-06-21	5.65	40.22	25.78	13.32	0.71	10.10	38.27	3.10
29521	Visakhapatnam	2020-06-22	5.58	42.45	27.06	13.70	0.73	13.65	34.85	3.99
29522	Visakhapatnam	2020-06-23	2.76	34.09	19.92	13.13	0.54	10.40	43.27	2.88
29523	Visakhapatnam	2020-06-24	1.22	23.38	13.12	14.45	0.56	10.92	35.12	2.99
29524	Visakhapatnam	2020-06-25	2.30	21.60	13.09	12.27	0.41	8.19	29.38	1.28
29525	Visakhapatnam	2020-06-26	5.91	23.27	17.19	11.15	0.46	6.87	19.90	1.45
29526	Visakhapatnam	2020-06-27	7.68	25.06	19.54	12.47	0.47	8.55	23.30	2.24
29527	Visakhapatnam	2020-06-28	3.42	26.06	16.53	11.99	0.52	12.72	30.14	0.74
29528	Visakhapatnam	2020-06-29	3.45	29.53	18.33	10.71	0.48	8.42	30.96	0.01
29529	Visakhapatnam	2020-06-30	4.05	29.26	18.80	10.03	0.52	9.84	28.30	0.00
29530	Visakhapatnam	2020-07-01	0.40	26.85	14.05	5.20	0.59	2.10	17.05	NaN

29531 rows × 14 columns

In [19]:

```
#here we remove rows with missing values of AQI and AQI_Bucket columns
df = df.dropna(subset=['AQI'])
df = df.dropna(subset=['AQI_Bucket'])
df
```

Out[19]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	C
28	Ahmedabad	2015-01-29	83.13	NaN	6.93	28.71	33.72	NaN	6.93	49.52	59.
29	Ahmedabad	2015-01-30	79.84	NaN	13.85	28.68	41.08	NaN	13.85	48.49	97.
30	Ahmedabad	2015-01-31	94.52	NaN	24.39	32.66	52.61	NaN	24.39	67.39	111.
31	Ahmedabad	2015-02-01	135.99	NaN	43.48	42.08	84.57	NaN	43.48	75.23	102.
32	Ahmedabad	2015-02-02	178.33	NaN	54.56	35.31	72.80	NaN	54.56	55.04	107.
33	Ahmedabad	2015-02-03	139.70	NaN	30.61	28.40	56.73	NaN	30.61	33.79	73.
34	Ahmedabad	2015-02-04	80.65	NaN	2.37	22.83	24.00	NaN	2.37	25.73	47.
35	Ahmedabad	2015-02-05	58.36	NaN	2.60	21.39	23.31	NaN	2.60	32.66	53.
36	Ahmedabad	2015-02-06	79.29	NaN	1.16	26.94	26.83	NaN	1.16	67.41	59.
37	Ahmedabad	2015-02-07	88.70	NaN	7.29	31.32	37.73	NaN	7.29	80.09	44.
38	Ahmedabad	2015-02-08	74.28	NaN	8.92	27.30	33.42	NaN	8.92	54.28	47.
39	Ahmedabad	2015-02-09	113.93	NaN	4.32	24.27	26.86	NaN	4.32	48.73	39.
40	Ahmedabad	2015-02-10	105.39	NaN	1.41	18.21	18.75	NaN	1.41	35.91	56.
41	Ahmedabad	2015-02-11	66.52	NaN	6.34	23.80	28.24	NaN	6.34	66.58	53.
42	Ahmedabad	2015-02-12	65.04	NaN	14.19	30.10	44.21	NaN	14.19	65.91	31.
43	Ahmedabad	2015-02-13	103.36	NaN	18.18	39.56	57.33	NaN	18.18	80.43	40.
44	Ahmedabad	2015-02-14	177.33	NaN	37.49	47.58	77.85	NaN	37.49	99.72	36.
45	Ahmedabad	2015-02-15	113.25	NaN	15.55	30.50	44.87	NaN	15.55	69.71	33.
46	Ahmedabad	2015-02-16	99.70	NaN	19.85	28.10	47.31	NaN	19.85	73.23	30.
47	Ahmedabad	2015-02-17	80.61	NaN	15.96	21.04	35.67	NaN	15.96	54.70	36.
48	Ahmedabad	2015-02-18	100.79	NaN	16.24	25.93	41.91	NaN	16.24	43.11	37.
49	Ahmedabad	2015-02-19	107.99	NaN	11.25	30.09	40.47	NaN	11.25	84.86	31.
50	Ahmedabad	2015-02-20	140.29	NaN	33.84	36.14	68.85	NaN	33.84	101.53	40.
51	Ahmedabad	2015-02-21	239.96	NaN	61.95	46.25	106.92	NaN	61.95	84.87	33.

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	C
52	Ahmedabad	2015-02-22	142.96	NaN	45.45	48.21	92.32	NaN	45.45	62.96	34.7
53	Ahmedabad	2015-02-23	238.41	NaN	77.70	55.07	131.32	NaN	77.70	89.07	37.3
54	Ahmedabad	2015-02-24	118.72	NaN	12.56	32.01	42.56	NaN	12.56	61.54	38.8
55	Ahmedabad	2015-02-25	52.99	NaN	2.23	10.45	10.39	NaN	2.23	32.76	67.2
60	Ahmedabad	2015-03-02	60.78	NaN	9.07	18.21	27.34	NaN	9.07	29.48	34.7
61	Ahmedabad	2015-03-03	91.83	NaN	12.99	28.71	40.76	NaN	12.99	39.71	38.4
...	...	...	...	...	...	...	...	...	...	...	
29501	Visakhapatnam	2020-06-02	9.91	43.30	2.81	19.91	12.60	9.16	0.47	8.78	19.6
29502	Visakhapatnam	2020-06-03	17.94	102.58	5.60	30.32	20.59	9.49	0.58	6.48	23.8
29503	Visakhapatnam	2020-06-04	19.14	77.06	4.12	35.57	22.18	8.82	0.68	7.19	25.5
29504	Visakhapatnam	2020-06-05	29.71	112.53	2.98	44.77	24.69	10.51	0.85	10.64	21.9
29505	Visakhapatnam	2020-06-06	21.66	74.97	3.38	38.42	22.56	8.63	0.64	7.02	28.8
29506	Visakhapatnam	2020-06-07	29.69	101.97	2.84	35.51	20.66	9.85	0.57	5.32	31.7
29507	Visakhapatnam	2020-06-08	31.14	125.15	6.55	41.56	27.26	9.71	0.64	6.61	36.6
29508	Visakhapatnam	2020-06-09	31.84	85.41	4.69	50.11	30.35	10.01	0.88	5.06	33.8
29509	Visakhapatnam	2020-06-10	32.55	79.28	4.78	44.59	27.38	9.34	0.69	5.90	35.6
29510	Visakhapatnam	2020-06-11	37.71	98.53	9.97	52.10	35.79	9.91	0.70	5.59	26.5
29511	Visakhapatnam	2020-06-12	37.67	76.27	12.17	53.59	35.52	9.85	0.97	5.74	17.3
29512	Visakhapatnam	2020-06-13	22.75	86.77	3.85	38.23	22.60	9.24	0.67	10.70	27.2
29513	Visakhapatnam	2020-06-14	21.33	83.23	2.81	34.56	20.43	9.78	0.58	9.37	29.4
29514	Visakhapatnam	2020-06-15	24.66	100.53	3.66	37.93	22.97	9.22	0.71	7.02	28.6
29515	Visakhapatnam	2020-06-16	21.06	80.51	6.67	40.33	26.81	8.81	0.63	11.09	21.6
29516	Visakhapatnam	2020-06-17	17.05	73.22	6.08	36.31	24.23	6.69	0.54	11.08	26.6
29517	Visakhapatnam	2020-06-18	21.16	88.80	4.94	42.96	26.79	3.22	0.68	11.05	27.6
29518	Visakhapatnam	2020-06-19	14.41	66.81	6.22	40.06	26.10	7.51	1.07	13.23	27.3

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	C
29519	Visakhapatnam	2020-06-20	26.67	86.29	3.78	37.36	22.46	15.05	0.74	11.76	41.9
29520	Visakhapatnam	2020-06-21	25.96	91.75	5.65	40.22	25.78	13.32	0.71	10.10	38.2
29521	Visakhapatnam	2020-06-22	33.17	108.22	5.58	42.45	27.06	13.70	0.73	13.65	34.8
29522	Visakhapatnam	2020-06-23	25.40	83.38	2.76	34.09	19.92	13.13	0.54	10.40	43.2
29523	Visakhapatnam	2020-06-24	34.36	90.90	1.22	23.38	13.12	14.45	0.56	10.92	35.1
29524	Visakhapatnam	2020-06-25	13.45	58.54	2.30	21.60	13.09	12.27	0.41	8.19	29.3
29525	Visakhapatnam	2020-06-26	7.63	32.27	5.91	23.27	17.19	11.15	0.46	6.87	19.9
29526	Visakhapatnam	2020-06-27	15.02	50.94	7.68	25.06	19.54	12.47	0.47	8.55	23.3
29527	Visakhapatnam	2020-06-28	24.38	74.09	3.42	26.06	16.53	11.99	0.52	12.72	30.1
29528	Visakhapatnam	2020-06-29	22.91	65.73	3.45	29.53	18.33	10.71	0.48	8.42	30.9
29529	Visakhapatnam	2020-06-30	16.64	49.97	4.05	29.26	18.80	10.03	0.52	9.84	28.3
29530	Visakhapatnam	2020-07-01	15.00	66.00	0.40	26.85	14.05	5.20	0.59	2.10	17.0

24850 rows × 16 columns

In [ ]:

```
#data integration
```

In [30]:

```
sub1 =df[['City','AQI_Bucket']]
sub1
```

Out[30]:

	City	AQI_Bucket
28	Ahmedabad	Poor
29	Ahmedabad	Very Poor
30	Ahmedabad	Severe
31	Ahmedabad	Severe
32	Ahmedabad	Severe
33	Ahmedabad	Severe
34	Ahmedabad	Poor
35	Ahmedabad	Moderate
36	Ahmedabad	Moderate
37	Ahmedabad	Poor
38	Ahmedabad	Very Poor
39	Ahmedabad	Very Poor
40	Ahmedabad	Poor
41	Ahmedabad	Very Poor
42	Ahmedabad	Poor
43	Ahmedabad	Severe
44	Ahmedabad	Severe
45	Ahmedabad	Severe
46	Ahmedabad	Severe
47	Ahmedabad	Severe
48	Ahmedabad	Severe
49	Ahmedabad	Severe
50	Ahmedabad	Severe
51	Ahmedabad	Severe
52	Ahmedabad	Severe
53	Ahmedabad	Severe
54	Ahmedabad	Severe
55	Ahmedabad	Poor
60	Ahmedabad	Moderate
61	Ahmedabad	Severe
...	...	...
29501	Visakhapatnam	Good
29502	Visakhapatnam	Satisfactory
29503	Visakhapatnam	Satisfactory
29504	Visakhapatnam	Satisfactory
29505	Visakhapatnam	Moderate
29506	Visakhapatnam	Satisfactory

	City	AQI_Bucket
29507	Visakhapatnam	Moderate
29508	Visakhapatnam	Satisfactory
29509	Visakhapatnam	Satisfactory
29510	Visakhapatnam	Satisfactory
29511	Visakhapatnam	Satisfactory
29512	Visakhapatnam	Satisfactory
29513	Visakhapatnam	Satisfactory
29514	Visakhapatnam	Satisfactory
29515	Visakhapatnam	Satisfactory
29516	Visakhapatnam	Satisfactory
29517	Visakhapatnam	Satisfactory
29518	Visakhapatnam	Moderate
29519	Visakhapatnam	Satisfactory
29520	Visakhapatnam	Satisfactory
29521	Visakhapatnam	Satisfactory
29522	Visakhapatnam	Satisfactory
29523	Visakhapatnam	Satisfactory
29524	Visakhapatnam	Satisfactory
29525	Visakhapatnam	Good
29526	Visakhapatnam	Good
29527	Visakhapatnam	Satisfactory
29528	Visakhapatnam	Satisfactory
29529	Visakhapatnam	Satisfactory
29530	Visakhapatnam	Good

24850 rows × 2 columns

In [31]:

```
sub2 =df[['CO','AQI']]
sub2
```

Out[31]:

	CO	AQI
28	6.93	209.0
29	13.85	328.0
30	24.39	514.0
31	43.48	782.0
32	54.56	914.0
33	30.61	660.0
34	2.37	294.0
35	2.60	149.0
36	1.16	190.0
37	7.29	247.0
38	8.92	379.0
39	4.32	341.0
40	1.41	256.0
41	6.34	388.0
42	14.19	288.0
43	18.18	510.0
44	37.49	761.0
45	15.55	475.0
46	19.85	536.0
47	15.96	479.0
48	16.24	592.0
49	11.25	427.0
50	33.84	588.0
51	61.95	1141.0
52	45.45	669.0
53	77.70	1247.0
54	12.56	411.0
55	2.23	292.0
60	9.07	189.0
61	12.99	408.0
...	...	...
29501	0.47	50.0
29502	0.58	88.0
29503	0.68	71.0
29504	0.85	93.0
29505	0.64	101.0
29506	0.57	82.0

	CO	AQI
29507	0.64	115.0
29508	0.88	96.0
29509	0.69	92.0
29510	0.70	90.0
29511	0.97	84.0
29512	0.67	81.0
29513	0.58	88.0
29514	0.71	92.0
29515	0.63	83.0
29516	0.54	90.0
29517	0.68	71.0
29518	1.07	110.0
29519	0.74	74.0
29520	0.71	92.0
29521	0.73	95.0
29522	0.54	100.0
29523	0.56	86.0
29524	0.41	77.0
29525	0.46	47.0
29526	0.47	41.0
29527	0.52	70.0
29528	0.48	68.0
29529	0.52	54.0
29530	0.59	50.0

24850 rows × 2 columns

In [37]:

```
merge = pd.concat([sub1, sub2],sort=False)
merge
```

Out[37]:

	City	AQI_Bucket	CO	AQI
28	Ahmedabad	Poor	NaN	NaN
29	Ahmedabad	Very Poor	NaN	NaN
30	Ahmedabad	Severe	NaN	NaN
31	Ahmedabad	Severe	NaN	NaN
32	Ahmedabad	Severe	NaN	NaN
33	Ahmedabad	Severe	NaN	NaN
34	Ahmedabad	Poor	NaN	NaN
35	Ahmedabad	Moderate	NaN	NaN
36	Ahmedabad	Moderate	NaN	NaN
37	Ahmedabad	Poor	NaN	NaN
38	Ahmedabad	Very Poor	NaN	NaN
39	Ahmedabad	Very Poor	NaN	NaN
40	Ahmedabad	Poor	NaN	NaN
41	Ahmedabad	Very Poor	NaN	NaN
42	Ahmedabad	Poor	NaN	NaN
43	Ahmedabad	Severe	NaN	NaN
44	Ahmedabad	Severe	NaN	NaN
45	Ahmedabad	Severe	NaN	NaN
46	Ahmedabad	Severe	NaN	NaN
47	Ahmedabad	Severe	NaN	NaN
48	Ahmedabad	Severe	NaN	NaN
49	Ahmedabad	Severe	NaN	NaN
50	Ahmedabad	Severe	NaN	NaN
51	Ahmedabad	Severe	NaN	NaN
52	Ahmedabad	Severe	NaN	NaN
53	Ahmedabad	Severe	NaN	NaN
54	Ahmedabad	Severe	NaN	NaN
55	Ahmedabad	Poor	NaN	NaN
60	Ahmedabad	Moderate	NaN	NaN
61	Ahmedabad	Severe	NaN	NaN
...	...	...	...	...
29501	NaN	NaN	0.47	50.0
29502	NaN	NaN	0.58	88.0
29503	NaN	NaN	0.68	71.0
29504	NaN	NaN	0.85	93.0
29505	NaN	NaN	0.64	101.0
29506	NaN	NaN	0.57	82.0

	City	AQI_Bucket	CO	AQI
29507	NaN	NaN	0.64	115.0
29508	NaN	NaN	0.88	96.0
29509	NaN	NaN	0.69	92.0
29510	NaN	NaN	0.70	90.0
29511	NaN	NaN	0.97	84.0
29512	NaN	NaN	0.67	81.0
29513	NaN	NaN	0.58	88.0
29514	NaN	NaN	0.71	92.0
29515	NaN	NaN	0.63	83.0
29516	NaN	NaN	0.54	90.0
29517	NaN	NaN	0.68	71.0
29518	NaN	NaN	1.07	110.0
29519	NaN	NaN	0.74	74.0
29520	NaN	NaN	0.71	92.0
29521	NaN	NaN	0.73	95.0
29522	NaN	NaN	0.54	100.0
29523	NaN	NaN	0.56	86.0
29524	NaN	NaN	0.41	77.0
29525	NaN	NaN	0.46	47.0
29526	NaN	NaN	0.47	41.0
29527	NaN	NaN	0.52	70.0
29528	NaN	NaN	0.48	68.0
29529	NaN	NaN	0.52	54.0
29530	NaN	NaN	0.59	50.0

49700 rows × 4 columns

In [ ]:

```
#Data transformation
```

In [38]:

```
df['Date']
```

Out[38]:

In [40]:
```
#sorting it date wise
df['Year'] = pd.DatetimeIndex(df['Date']).year
df['month'] = pd.DatetimeIndex(df['Date']).month
df
```

36	Ahmedabad	2015-02-06	79.29	NaN	1.16	26.94	26.83	NaN	1.16	67.41	59.30	0.00
37	Ahmedabad	2015-02-07	88.70	NaN	7.29	31.32	37.73	NaN	7.29	80.09	44.76	0.00
38	Ahmedabad	2015-02-08	74.28	NaN	8.92	27.30	33.42	NaN	8.92	54.28	47.42	0.00
39	Ahmedabad	2015-02-09	113.93	NaN	4.32	24.27	26.86	NaN	4.32	48.73	39.94	0.02
40	Ahmedabad	2015-02-10	105.39	NaN	1.41	18.21	18.75	NaN	1.41	35.91	56.15	0.00
41	Ahmedabad	2015-02-11	66.52	NaN	6.34	23.80	28.24	NaN	6.34	66.58	53.14	9.70
42	Ahmedabad	2015-02-12	65.04	NaN	14.19	30.10	44.21	NaN	14.19	65.91	31.88	7.72
43	Ahmedabad	2015-02-13	103.36	NaN	18.18	39.56	57.33	NaN	18.18	80.43	40.11	11.29

In [ ]:
```
# Error Correcting
```

```
29530    2020-07-01
Name: Date, Length: 24850, dtype: object
```

In [42]:

```
# replacing ahmedabad to pune
df['City'].replace('Ahmedabad', 'Pune', inplace=True)
df
```

Out[42]:

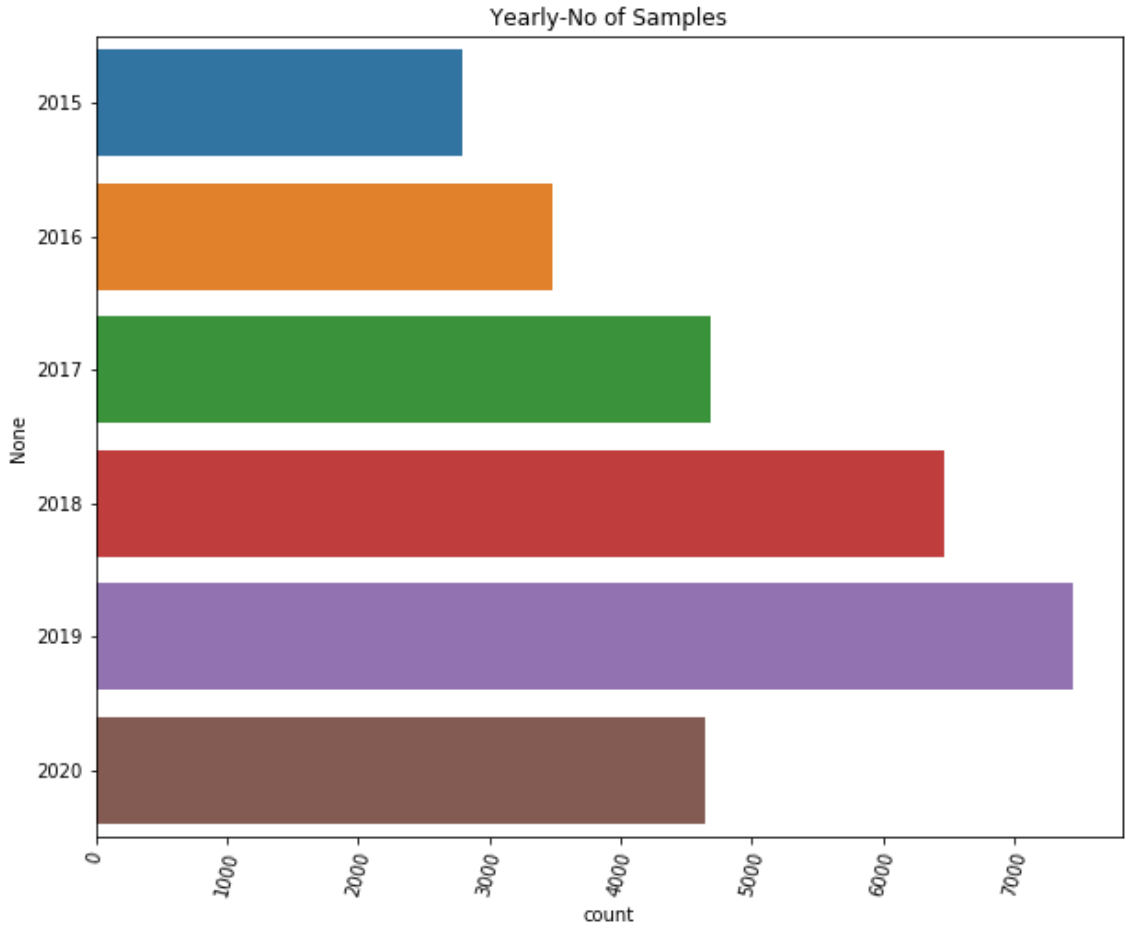
	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene
28	Pune	2015-01-29	83.13	NaN	6.93	28.71	33.72	NaN	6.93	49.52	59.76	0.02
29	Pune	2015-01-30	79.84	NaN	13.85	28.68	41.08	NaN	13.85	48.49	97.07	0.04
30	Pune	2015-01-31	94.52	NaN	24.39	32.66	52.61	NaN	24.39	67.39	111.33	0.24
31	Pune	2015-02-01	135.99	NaN	43.48	42.08	84.57	NaN	43.48	75.23	102.70	0.40
32	Pune	2015-02-02	178.33	NaN	54.56	35.31	72.80	NaN	54.56	55.04	107.38	0.46
33	Pune	2015-02-03	139.70	NaN	30.61	28.40	56.73	NaN	30.61	33.79	73.60	0.17

In [ ]:

```
# Data model building
```

In [67]:

```
fig, ax = plt.subplots(figsize=(10, 8))
sns.countplot(y=df.Year)
plt.xticks(rotation=75);
plt.title('Yearly-No of Samples')
plt.ylabel(None);
```



In [64]:

```
df['Year'] = pd.DatetimeIndex(df['Date']).year
```

In [65]:

```
year2017 = df[df.Year == 2017]
year2017
```

Out[65]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3
731	Ahmedabad	2017-01-01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
732	Ahmedabad	2017-01-02	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
733	Ahmedabad	2017-01-03	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
734	Ahmedabad	2017-01-04	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
735	Ahmedabad	2017-01-05	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
736	Ahmedabad	2017-01-06	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
737	Ahmedabad	2017-01-07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
738	Ahmedabad	2017-01-08	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
739	Ahmedabad	2017-01-09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
740	Ahmedabad	2017-01-10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
741	Ahmedabad	2017-01-11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
742	Ahmedabad	2017-01-12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
743	Ahmedabad	2017-01-13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
744	Ahmedabad	2017-01-14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
745	Ahmedabad	2017-01-15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
746	Ahmedabad	2017-01-16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
747	Ahmedabad	2017-01-17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
748	Ahmedabad	2017-01-18	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
749	Ahmedabad	2017-01-19	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
750	Ahmedabad	2017-01-20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
751	Ahmedabad	2017-01-21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
752	Ahmedabad	2017-01-22	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
753	Ahmedabad	2017-01-23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
754	Ahmedabad	2017-01-24	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3
755	Ahmedabad	2017-01-25	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
756	Ahmedabad	2017-01-26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
757	Ahmedabad	2017-01-27	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
758	Ahmedabad	2017-01-28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
759	Ahmedabad	2017-01-29	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
760	Ahmedabad	2017-01-30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...
28588	Visakhapatnam	2017-12-02	71.27	124.88	6.08	27.71	19.62	11.75	0.94	7.28	122.71
28589	Visakhapatnam	2017-12-03	69.02	114.81	1.30	29.42	16.66	11.20	0.93	6.61	110.02
28590	Visakhapatnam	2017-12-04	75.65	123.44	6.35	34.31	23.34	12.19	1.10	9.75	106.51
28591	Visakhapatnam	2017-12-05	98.32	184.65	19.59	55.94	45.60	13.31	1.23	9.07	81.64
28592	Visakhapatnam	2017-12-06	101.28	178.50	18.76	55.34	44.64	13.61	1.30	10.87	77.64
28593	Visakhapatnam	2017-12-07	83.65	141.02	6.82	36.78	24.99	20.20	1.09	12.56	96.71
28594	Visakhapatnam	2017-12-08	60.98	112.99	3.86	44.64	26.84	12.06	1.04	11.66	77.73
28595	Visakhapatnam	2017-12-09	55.61	111.86	10.14	44.59	31.93	11.12	1.06	10.45	53.68
28596	Visakhapatnam	2017-12-10	77.81	160.80	16.77	54.45	42.50	13.01	1.14	11.55	37.76
28597	Visakhapatnam	2017-12-11	72.49	126.78	8.91	47.58	32.49	15.85	1.11	10.08	46.58
28598	Visakhapatnam	2017-12-12	76.95	128.31	21.61	43.73	40.70	17.37	1.16	11.13	36.79
28599	Visakhapatnam	2017-12-13	80.72	149.08	37.79	41.51	52.74	15.23	1.26	7.11	36.13
28600	Visakhapatnam	2017-12-14	66.31	97.69	1.73	24.11	14.13	14.84	0.93	6.70	64.81
28601	Visakhapatnam	2017-12-15	66.12	97.87	6.82	36.82	25.01	13.13	1.04	5.84	71.93
28602	Visakhapatnam	2017-12-16	83.31	163.67	37.77	67.69	66.71	12.29	1.47	14.22	43.04
28603	Visakhapatnam	2017-12-17	92.11	181.98	23.36	56.71	49.16	14.04	1.26	14.12	52.25
28604	Visakhapatnam	2017-12-18	72.64	139.11	18.69	46.04	39.63	13.05	1.09	26.03	63.46
28605	Visakhapatnam	2017-12-19	55.81	121.39	4.86	43.94	27.22	11.15	0.91	9.21	77.78

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3
28606	Visakhapatnam	2017-12-20	72.83	137.31	10.25	54.01	37.01	12.09	1.12	6.59	69.33
28607	Visakhapatnam	2017-12-21	67.70	110.88	4.24	41.12	25.28	12.53	1.10	6.93	82.73
28608	Visakhapatnam	2017-12-22	89.92	178.94	29.44	52.59	51.86	13.92	1.39	10.78	57.23
28609	Visakhapatnam	2017-12-23	113.04	195.51	27.63	61.96	55.38	16.66	1.47	40.38	48.63
28610	Visakhapatnam	2017-12-24	117.20	183.31	19.69	45.40	40.00	16.56	1.32	13.44	65.61
28611	Visakhapatnam	2017-12-25	115.56	168.70	6.88	39.23	26.44	16.31	1.21	8.55	82.60
28612	Visakhapatnam	2017-12-26	123.90	202.38	16.37	60.49	45.40	17.16	1.45	15.18	66.62
28613	Visakhapatnam	2017-12-27	101.93	176.88	15.87	55.93	42.58	14.37	1.20	16.39	69.82
28614	Visakhapatnam	2017-12-28	88.29	195.79	28.26	60.83	55.30	11.48	1.28	16.54	57.78
28615	Visakhapatnam	2017-12-29	52.25	103.40	9.07	52.56	35.20	11.33	1.08	6.26	52.57
28616	Visakhapatnam	2017-12-30	58.56	100.61	2.43	32.45	19.08	12.22	0.93	5.44	80.89
28617	Visakhapatnam	2017-12-31	67.75	113.01	1.62	23.54	13.63	12.58	0.93	8.38	112.64

4689 rows × 17 columns

In [66]:

```
plt.figure(figsize=(12, 12))
sns.barplot(x = 'AQI',
            y = 'City',
           data = year2017)
plt.title("City VS AQI(year:2017)");
plt.xlabel('sum of AQI');
```

```
C:\Users\Vinayak\Anaconda3\lib\site-packages\scipy\stats\stats.py:1713: Fu
tureWarning: Using a non-tuple sequence for multidimensional indexing is d
eprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this
will be interpreted as an array index, `arr[np.array(seq)]`, which will re
sult either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```