In [ ]:
```
NAME: MANE SHIVRAJ PANDURANG
COURSE: CL I
CLASS: BE AI&DS
```

In [27]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seabornn as sns
```

In [4]:
```
df=pd.read_csv(r"C:\Users\Admin\Downloads\city_day.csv")
```

In [5]:
```
df.head(10)
```

Out[5]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	B(
0	Ahmedabad	2015-01-01	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	
1	Ahmedabad	2015-01-02	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	
2	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	
3	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	
4	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	
5	Ahmedabad	2015-01-06	NaN	NaN	45.41	38.48	81.50	NaN	45.41	45.76	46.51	
6	Ahmedabad	2015-01-07	NaN	NaN	112.16	40.62	130.77	NaN	112.16	32.28	33.47	
7	Ahmedabad	2015-01-08	NaN	NaN	80.87	36.74	96.75	NaN	80.87	38.54	31.89	
8	Ahmedabad	2015-01-09	NaN	NaN	29.16	31.00	48.00	NaN	29.16	58.68	25.75	
9	Ahmedabad	2015-01-10	NaN	NaN	NaN	7.04	0.00	NaN	NaN	8.29	4.55	
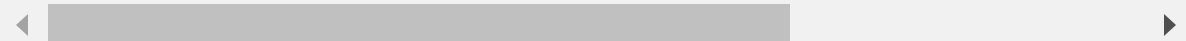
In [6]:
```
df.isnull()
```

Out[6]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	T
0	False	False	True	True	False	False	False	True	False	False	False	False	
1	False	False	True	True	False	False	False	True	False	False	False	False	
2	False	False	True	True	False	False	False	True	False	False	False	False	
3	False	False	True	True	False	False	False	True	False	False	False	False	
4	False	False	True	True	False	False	False	True	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	
29526	False	False	False	False	False	False	False	False	False	False	False	False	
29527	False	False	False	False	False	False	False	False	False	False	False	False	
29528	False	False	False	False	False	False	False	False	False	False	False	False	
29529	False	False	False	False	False	False	False	False	False	False	False	False	
29530	False	False	False	False	False	False	False	False	False	False	False	True	

29531 rows × 16 columns

In [8]:
```python
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 [10]:
```python
df["AQI"].fillna(df["AQI"].mean(),inplace=True)
```

In [11]:
```python
df.head(5)
```

Out[11]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benz
0	Ahmedabad	2015-01-01	NaN	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	(
1	Ahmedabad	2015-01-02	NaN	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	3
2	Ahmedabad	2015-01-03	NaN	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	6
3	Ahmedabad	2015-01-04	NaN	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	4
4	Ahmedabad	2015-01-05	NaN	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	7

In [12]:
```python
df["PM2.5"].fillna(df["PM2.5"].mean(),inplace=True)
```

In [13]:
```python
df.head(5)
```

Out[13]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	B
0	Ahmedabad	2015-01-01	67.450578	NaN	0.92	18.22	17.15	NaN	0.92	27.64	133.36	
1	Ahmedabad	2015-01-02	67.450578	NaN	0.97	15.69	16.46	NaN	0.97	24.55	34.06	
2	Ahmedabad	2015-01-03	67.450578	NaN	17.40	19.30	29.70	NaN	17.40	29.07	30.70	
3	Ahmedabad	2015-01-04	67.450578	NaN	1.70	18.48	17.97	NaN	1.70	18.59	36.08	
4	Ahmedabad	2015-01-05	67.450578	NaN	22.10	21.42	37.76	NaN	22.10	39.33	39.31	

In [14]:
```python
df["PM10"].fillna(df["PM10"].mean(),inplace=True)
```

In [15]:
```python
df.head(5)
```

Out[15]:

	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O
0	Ahmedabad	2015-01-01	67.450578	118.127103	0.92	18.22	17.15	NaN	0.92	27.64	133.3
1	Ahmedabad	2015-01-02	67.450578	118.127103	0.97	15.69	16.46	NaN	0.97	24.55	34.0
2	Ahmedabad	2015-01-03	67.450578	118.127103	17.40	19.30	29.70	NaN	17.40	29.07	30.7
3	Ahmedabad	2015-01-04	67.450578	118.127103	1.70	18.48	17.97	NaN	1.70	18.59	36.0
4	Ahmedabad	2015-01-05	67.450578	118.127103	22.10	21.42	37.76	NaN	22.10	39.33	39.3

In [19]:
```python
df['Date'] = pd.to_datetime(df['Date'])

plt.figure(figsize=(10, 6))
plt.plot(df['Date'], df['AQI'], marker='o', linestyle='-')
plt.title('AQI Trend over Time')
plt.xlabel('Date')
plt.ylabel('AQI')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
```
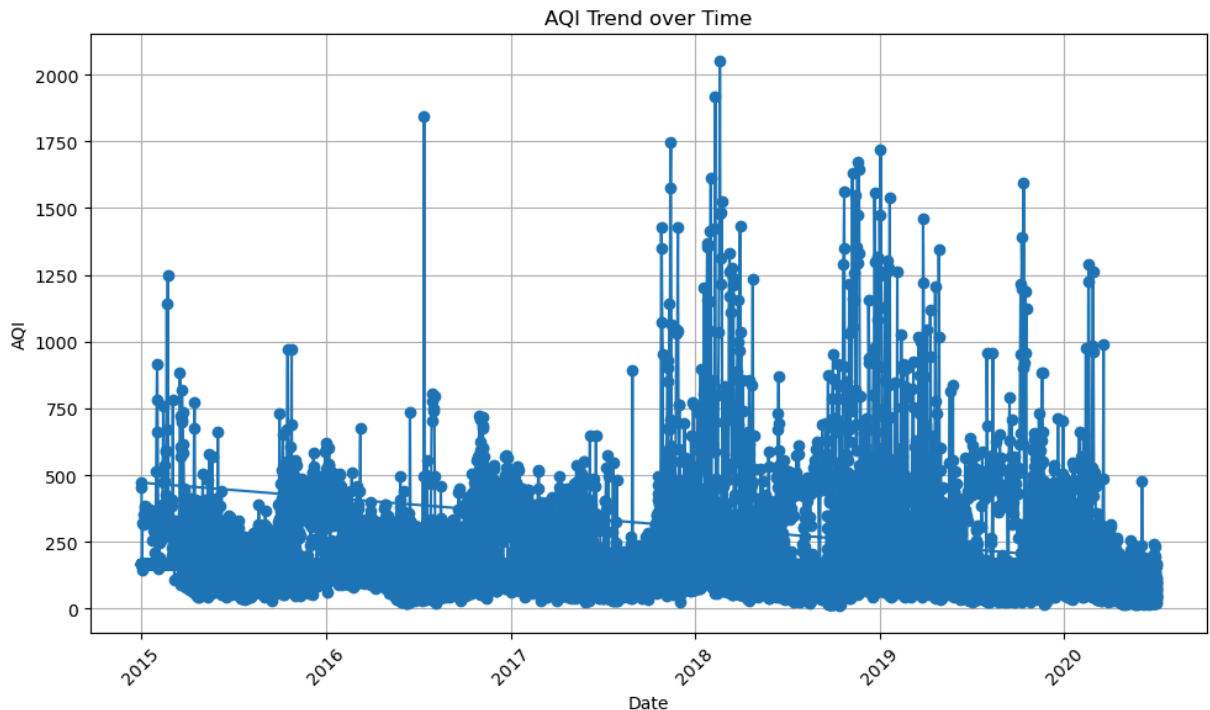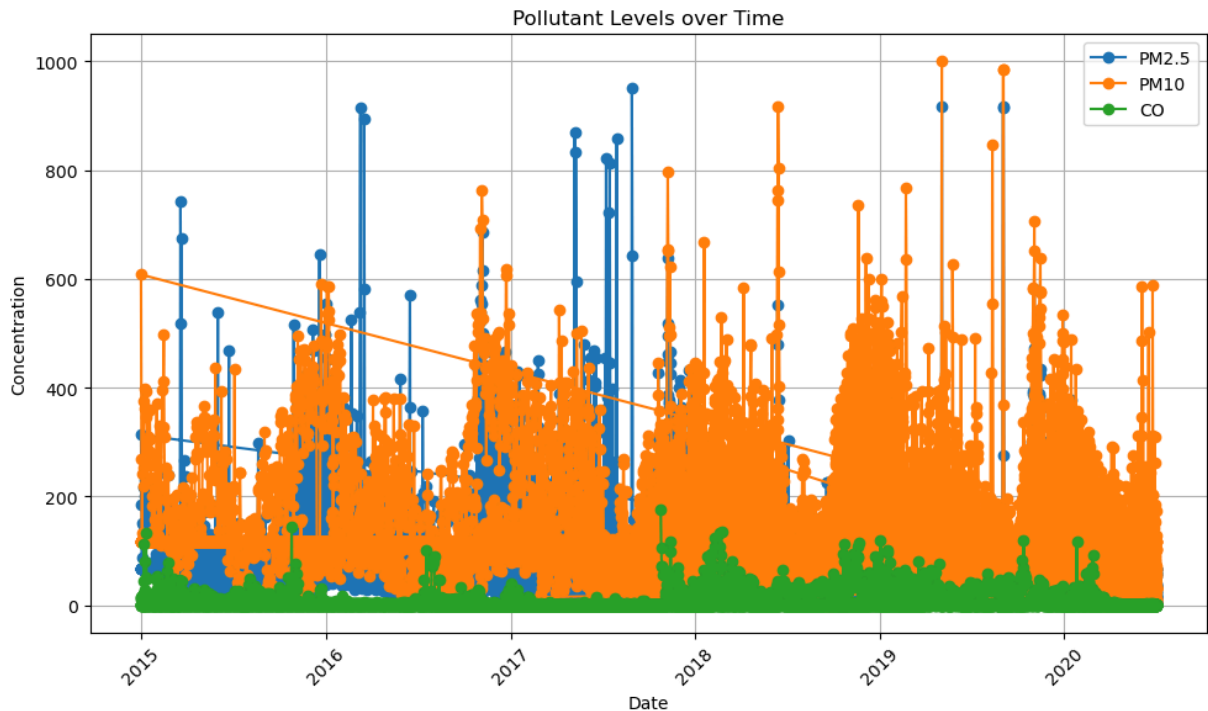

In [20]:
```python
plt.figure(figsize=(10, 6))
plt.plot(df['Date'], df['PM2.5'], label='PM2.5', marker='o')
plt.plot(df['Date'], df['PM10'], label='PM10', marker='o')
plt.plot(df['Date'], df['CO'], label='CO', marker='o')
```

```python
plt.title('Pollutant Levels over Time')
plt.xlabel('Date')
plt.ylabel('Concentration')
plt.legend()
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



```python
In [21]:  mean_aqi = df.groupby('Date')['AQI'].mean()


          plt.figure(figsize=(12, 6))
          plt.bar(mean_aqi.index, mean_aqi.values, width=0.5)
          plt.title('Mean AQI across Dates')
          plt.xlabel('Date')
          plt.ylabel('Mean AQI')
          plt.xticks(rotation=45)
          plt.tight_layout()
          plt.show()
```

Mean AQI across Dates
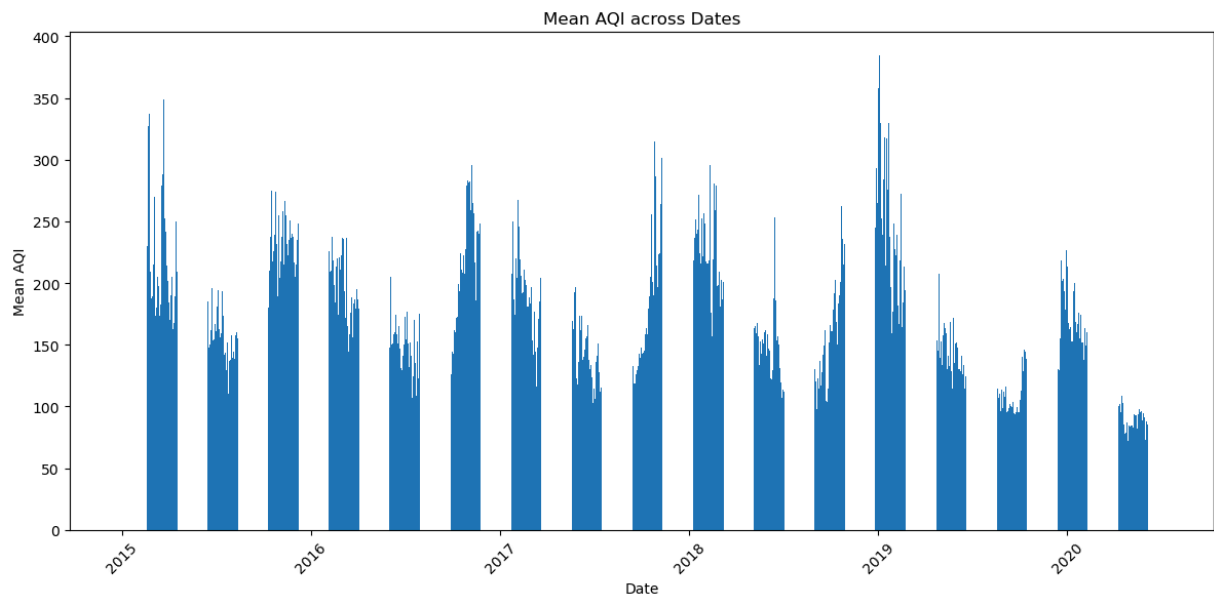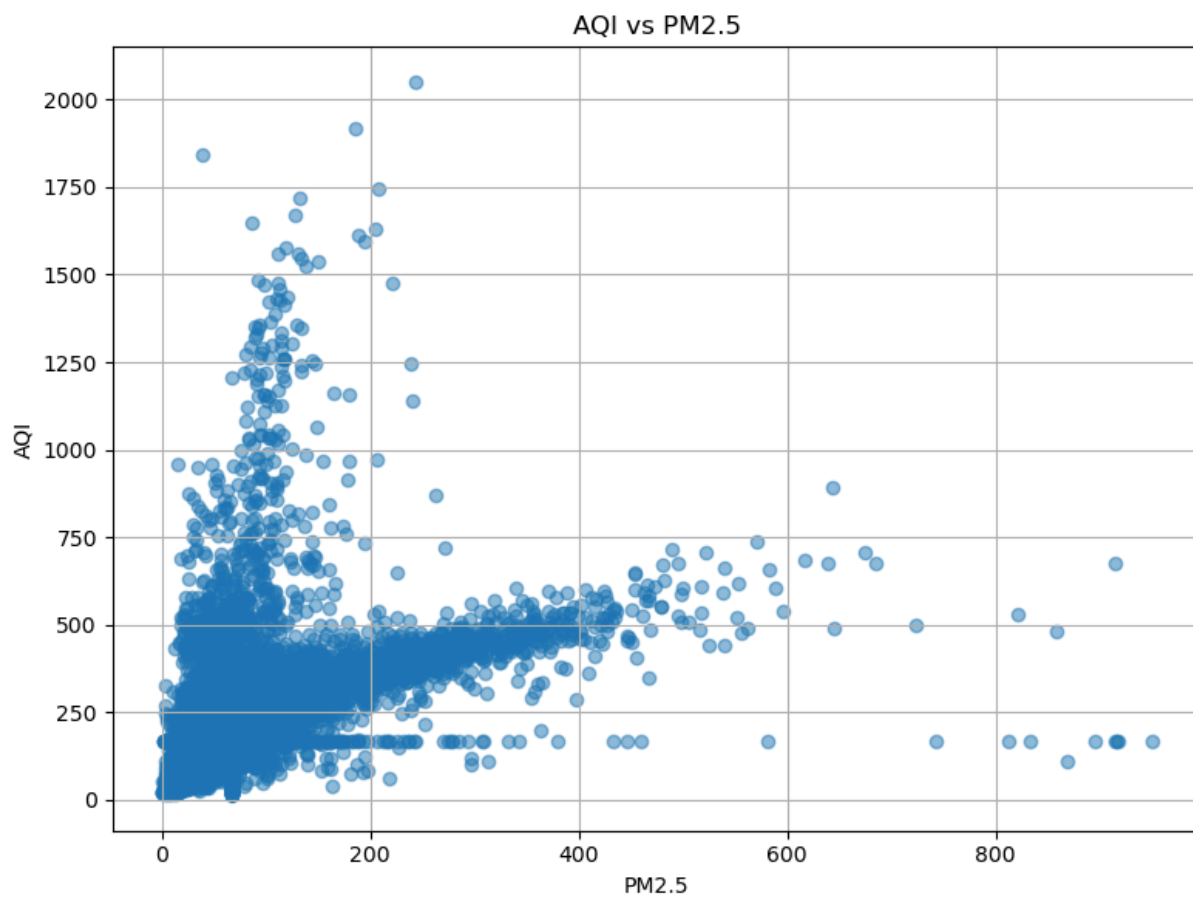


```
In [31]: plt.figure(figsize=(10, 6))
         sns.boxplot(data=df[['PM2.5', 'PM10', 'CO']], orient='h')
         plt.title('AQI Distribution by Pollutant')
         plt.xlabel('AQI')
         plt.ylabel('Pollutant')
         plt.tight_layout()
         plt.show()
```

AQI Distribution by Pollutant
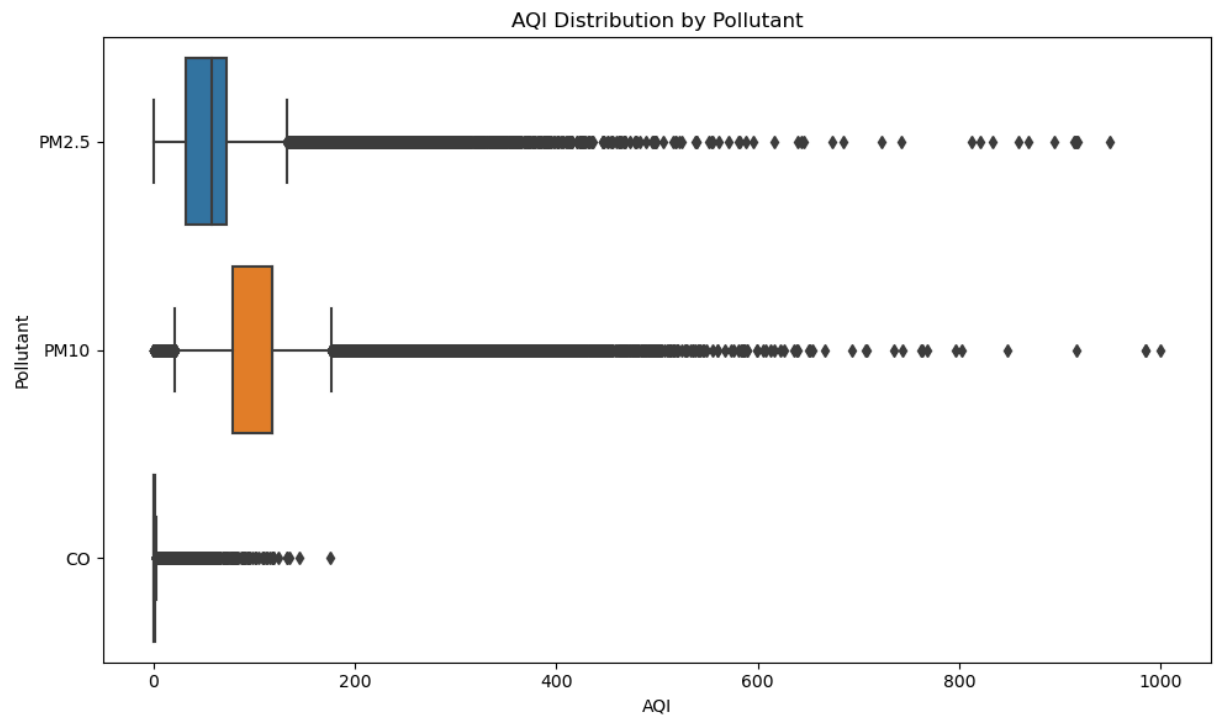

```
In [32]: plt.figure(figsize=(8, 6))
         plt.scatter(df['PM2.5'], df['AQI'], alpha=0.5)
         plt.title('AQI vs PM2.5')
         plt.xlabel('PM2.5')
         plt.ylabel('AQI')
         plt.grid(True)
```

```
plt.tight_layout()
plt.show()
```


In [ ]: