```
In [6]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```
In [8]: df = pd.read_csv('AQI Data set.csv', parse_dates=['Mounths'])
        df.head()
```

C:\Users\Vaishnavi\AppData\Local\Temp\ipykernel_14264\4022485709.py:1: Use
rWarning: Could not infer format, so each element will be parsed individua
lly, falling back to `dateutil`. To ensure parsing is consistent and as-ex
pected, please specify a format.
  df = pd.read_csv('AQI Data set.csv', parse_dates=['Mounths'])

Out[8]:

| | Id | Mounths | PM10 in æg/m3 | SO2 in æg/m3 | NOx in æg/m3 | PM2.5 in æg/m3 | Ammonia - NH3 in æg/m3 | O3 in æg/m3 | CO in mg/m3 | Benzene in æg/m3 | AQI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | Jan-17 | 174.0 | 26.4 | 35.0 | 79 | 25.0 | 107.6 | 0.9 | 0.7 | 149.0 |
| 1 | 2 | Feb-17 | 143.0 | 35.1 | 40.3 | 75 | 31.0 | 103.0 | 0.9 | 0.9 | 129.0 |
| 2 | 3 | Mar-17 | 142.0 | 32.1 | 30.9 | 59 | 26.0 | 80.7 | 0.8 | 0.5 | 128.0 |
| 3 | 4 | Apr-17 | 117.0 | 50.9 | 36.3 | 75 | 36.0 | 79.5 | 0.9 | 0.7 | 111.0 |
| 4 | 5 | May-17 | NaN | 41.6 | 25.2 | 53 | 28.0 | 70.0 | 0.5 | 0.5 | NaN |

```
In [9]: df.columns
```

Out[9]: Index(['Id', 'Mounths', 'PM10 in æg/m3', 'SO2 in æg/m3', 'NOx  in æg/m3',
               ' PM2.5  in æg/m3', 'Ammonia - NH3  in æg/m3', 'O3   in æg/m3',
               'CO  in mg/m3', ' Benzene  in æg/m3', 'AQI'],
              dtype='object')

```
In [10]: column_names = ['Id','Months','PM10','SO2','NOx','PM25','NH3','O3','CO','Be
         df.columns =column_names
         df.head()
```

Out[10]:

| | Id | Months | PM10 | SO2 | NOx | PM25 | NH3 | O3 | CO | Benzene | AQI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | Jan-17 | 174.0 | 26.4 | 35.0 | 79 | 25.0 | 107.6 | 0.9 | 0.7 | 149.0 |
| 1 | 2 | Feb-17 | 143.0 | 35.1 | 40.3 | 75 | 31.0 | 103.0 | 0.9 | 0.9 | 129.0 |
| 2 | 3 | Mar-17 | 142.0 | 32.1 | 30.9 | 59 | 26.0 | 80.7 | 0.8 | 0.5 | 128.0 |
| 3 | 4 | Apr-17 | 117.0 | 50.9 | 36.3 | 75 | 36.0 | 79.5 | 0.9 | 0.7 | 111.0 |
| 4 | 5 | May-17 | NaN | 41.6 | 25.2 | 53 | 28.0 | 70.0 | 0.5 | 0.5 | NaN |

```
In [12]: df.isna().sum()
```

```
Out[12]: Id          0
         Months      0
         PM10        6
         SO2         1
         NOx         2
         PM25        0
         NH3         0
         O3          0
         CO          0
         Benzene     0
         AQI         5
         dtype: int64
```

```
In [14]: df.dropna(inplace=True)
         df.isna().sum()
```

```
Out[14]: Id          0
         Months      0
         PM10        0
         SO2         0
         NOx         0
         PM25        0
         NH3         0
         O3          0
         CO          0
         Benzene     0
         AQI         0
         dtype: int64
```

```
In [15]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 66 entries, 0 to 71
Data columns (total 11 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Id       66 non-null     int64
 1   Months   66 non-null     object
 2   PM10     66 non-null     float64
 3   SO2      66 non-null     float64
 4   NOx      66 non-null     float64
 5   PM25     66 non-null     int64
 6   NH3      66 non-null     float64
 7   O3       66 non-null     float64
 8   CO       66 non-null     float64
 9   Benzene  66 non-null     float64
 10  AQI      66 non-null     float64
dtypes: float64(8), int64(2), object(1)
memory usage: 6.2+ KB
```

```
In [16]: df.describe()
```

Out[16]:

| | Id | PM10 | SO2 | NOx | PM25 | NH3 | O3 | |
|---|---|---|---|---|---|---|---|---|
| count | 66.000000 | 66.000000 | 66.000000 | 66.000000 | 66.000000 | 66.000000 | 66.000000 | 66.000 |
| mean | 38.500000 | 109.393939 | 16.093939 | 30.263636 | 46.393939 | 24.072727 | 25.350000 | 0.55 |
| std | 20.417376 | 25.271376 | 9.265218 | 3.947838 | 20.261277 | 5.960474 | 21.426413 | 0.24 |
| min | 1.000000 | 76.000000 | 4.000000 | 18.400000 | 12.000000 | 11.000000 | 2.400000 | 0.200 |
| 25% | 22.250000 | 90.000000 | 9.850000 | 28.125000 | 27.500000 | 20.250000 | 12.025000 | 0.400 |
| 50% | 38.500000 | 104.000000 | 13.700000 | 29.750000 | 46.500000 | 23.000000 | 18.750000 | 0.500 |
| 75% | 55.750000 | 128.000000 | 17.150000 | 32.550000 | 62.750000 | 28.000000 | 31.575000 | 0.640 |
| max | 72.000000 | 178.000000 | 50.900000 | 40.300000 | 87.000000 | 37.000000 | 107.600000 | 1.520 |

```
In [22]: plt.figure(figsize=(12,6))
         plt.plot(df['Months'].values, df['AQI'].values)
         plt.xlabel('Date')
         plt.ylabel('AQI')
         plt.title('Air Quality Index (AQT) Trend over Time')
         plt.xticks(rotation=75)
         plt.show()
```



Air Quality Index (AQT) Trend over Time
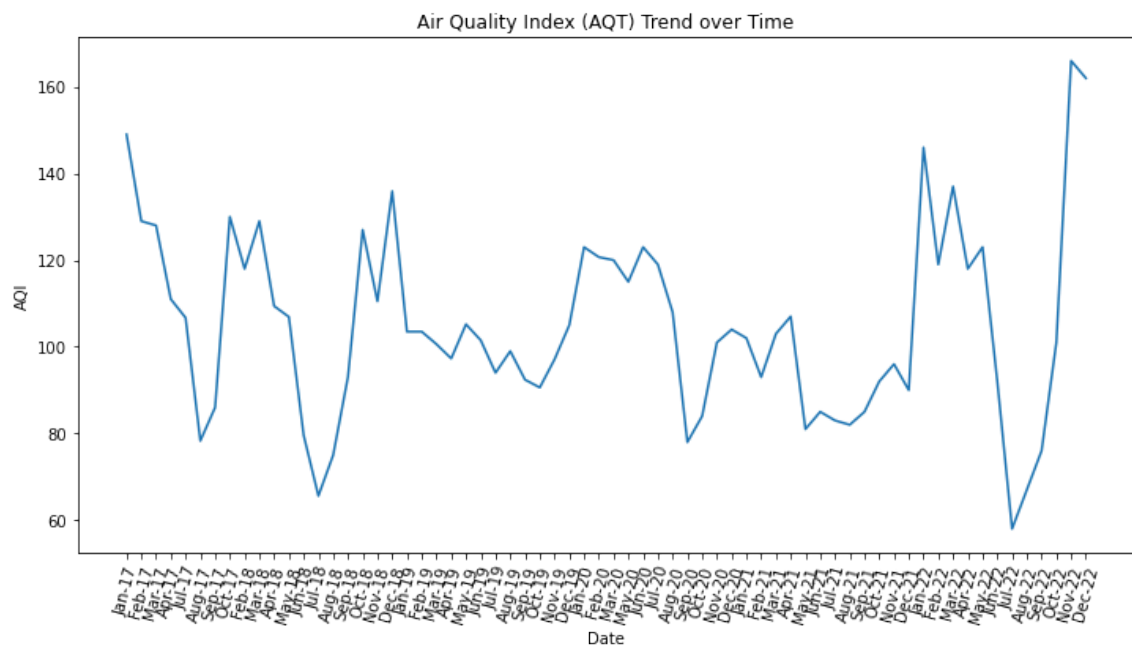
```
In [23]: df.columns
```

Out[23]: Index(['Id', 'Months', 'PM10', 'SO2', 'NOx', 'PM25', 'NH3', 'O3', 'CO',
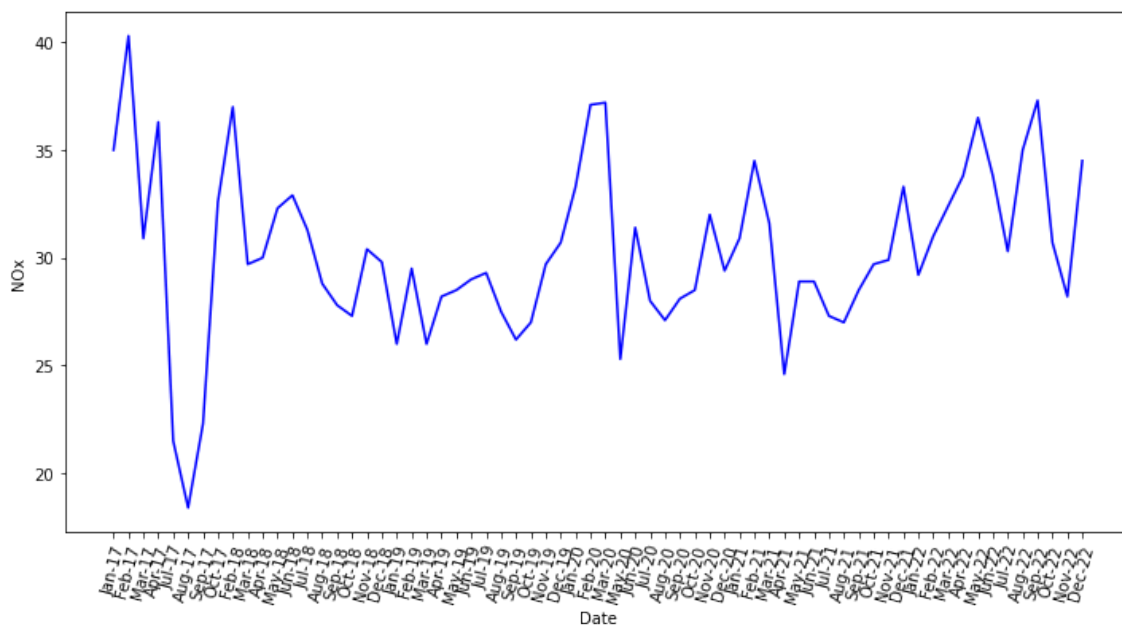               'Benzene', 'AQI'],
              dtype='object')

```python
plt.figure(figsize=(12,6))
plt.plot(df['Months'].values, df['PM10'].values, color='red')
plt.xlabel('Date')
plt.ylabel('PM10')
plt.xticks(rotation=75)
plt.show()
```
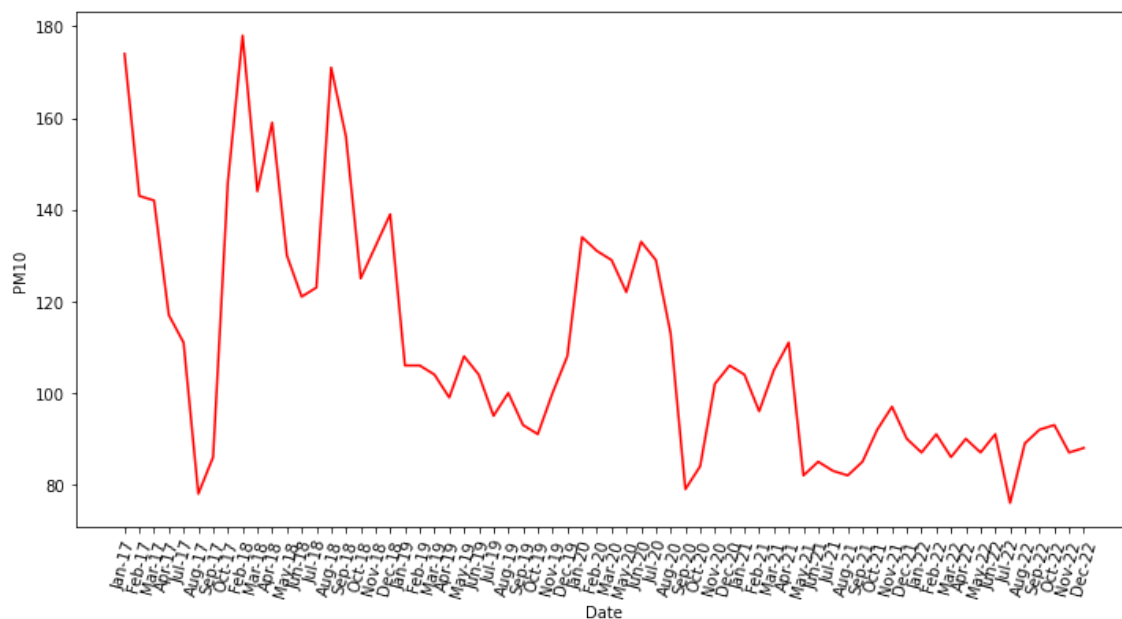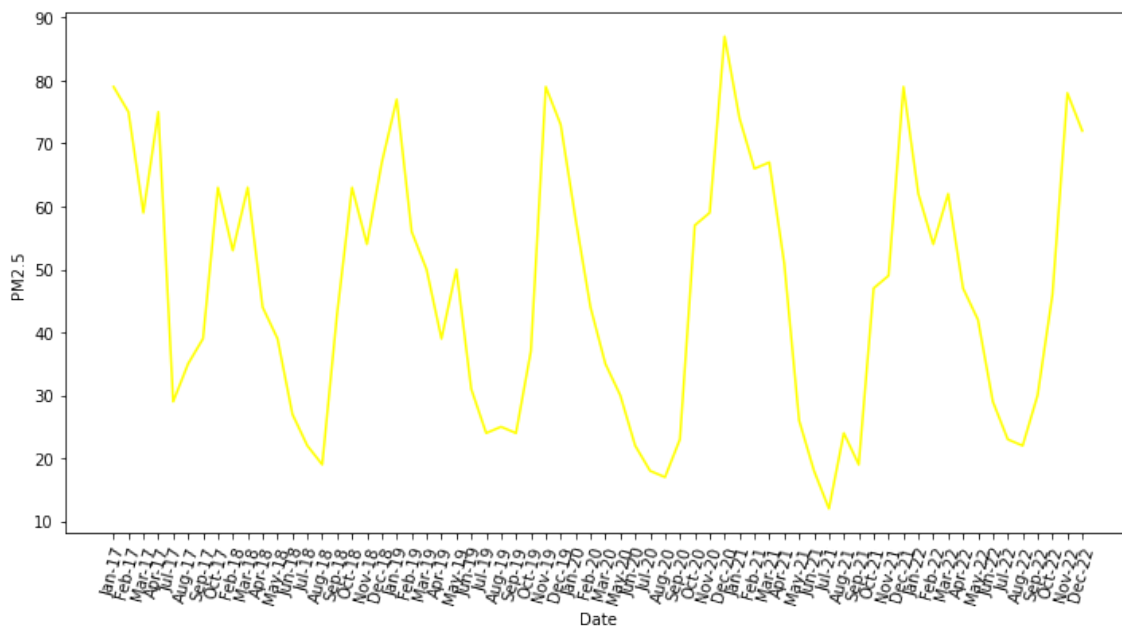
```python
plt.figure(figsize=(12,6))
plt.plot(df['Months'].values,df['SO2'].values, color='purple')
plt.xlabel('Date')
plt.ylabel('SO2')
plt.xticks(rotation=75)
plt.show()
```
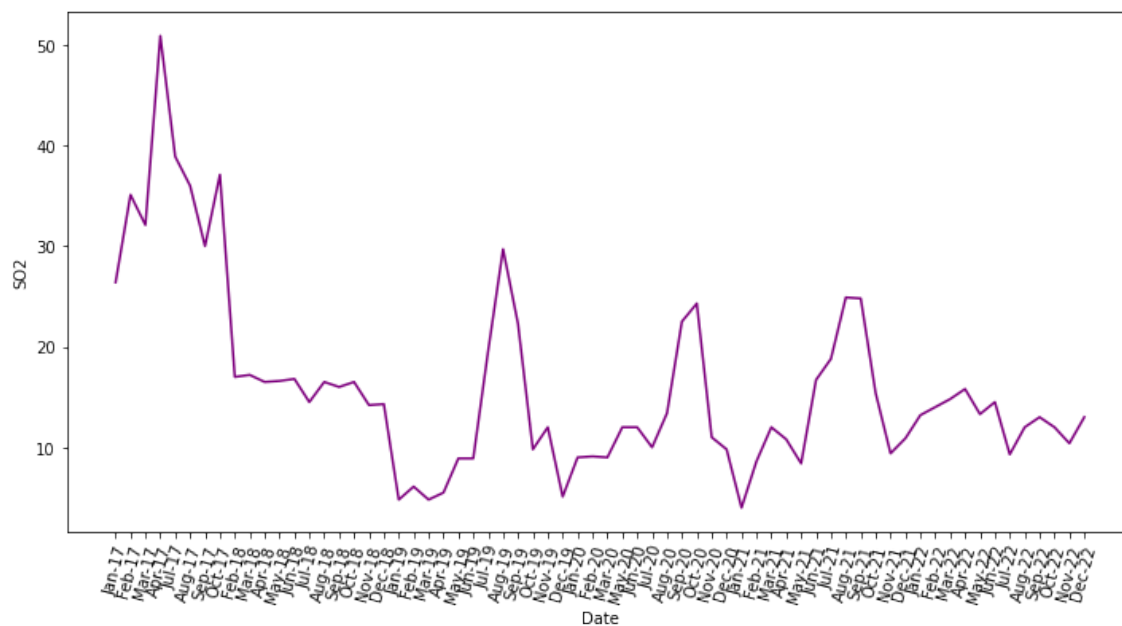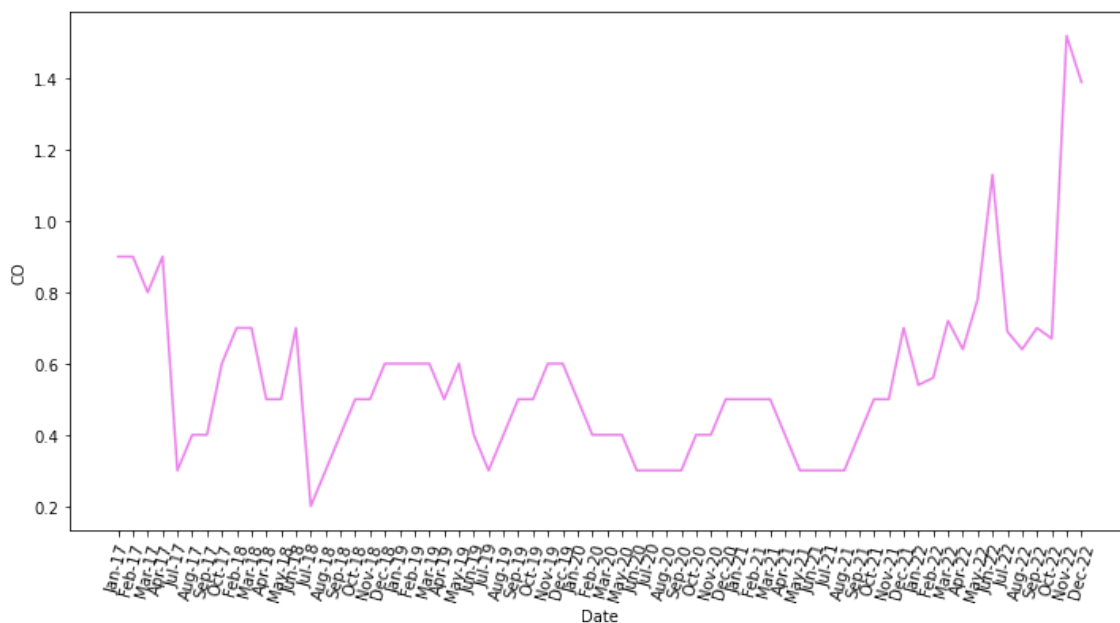
```
In [35]: plt.figure(figsize=(12,6))
         plt.plot(df['Months'].values,df['NOx'].values, color='blue')
         plt.xlabel('Date')
         plt.ylabel('NOx')
         plt.xticks(rotation=75)
         plt.show()
```



```
In [36]: plt.figure(figsize=(12,6))
         plt.plot(df['Months'].values,df['PM25'].values,color='yellow')
         plt.xlabel('Date')
         plt.ylabel('PM2.5')
         plt.xticks(rotation=75)
         plt.show()
```

```python
plt.figure(figsize=(12,6))
plt.plot(df['Months'].values,df['CO'].values,color='violet')
plt.xlabel('Date')
plt.ylabel('CO')
plt.xticks(rotation=75)
plt.show()
```
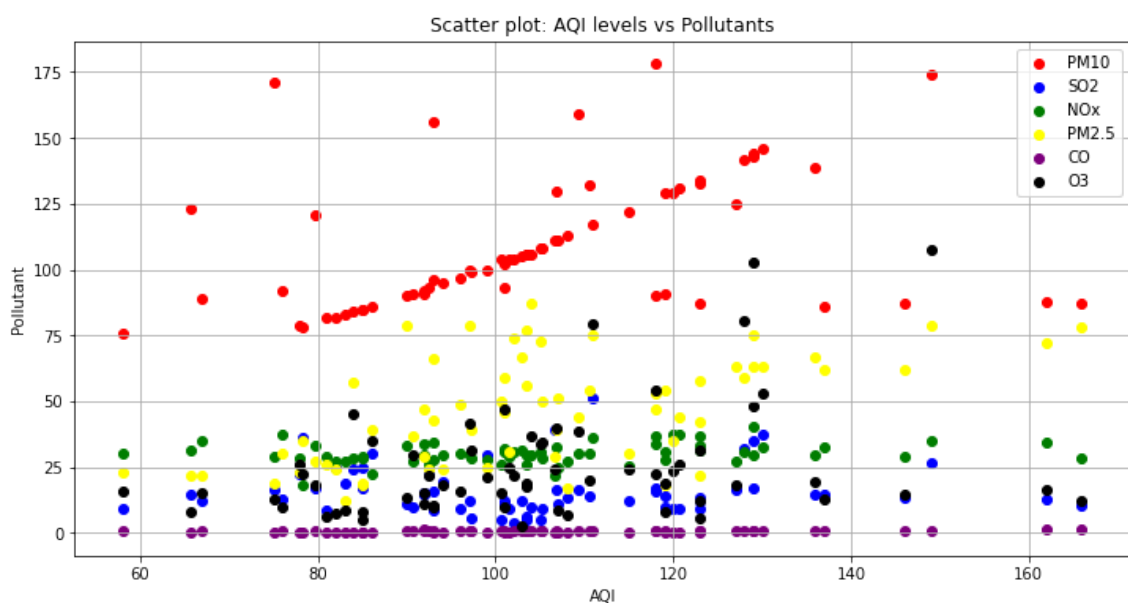
```python
plt.figure(figsize=(12,6))
plt.bar(df['Months'].values,df['AQI'].values,color='skyblue',label='AQI')
plt.title('AQI Comparison across Months')
plt.xlabel('Months')
plt.ylabel('AQI Values')
plt.xticks(rotation=75)
plt.legend()
plt.show()
```
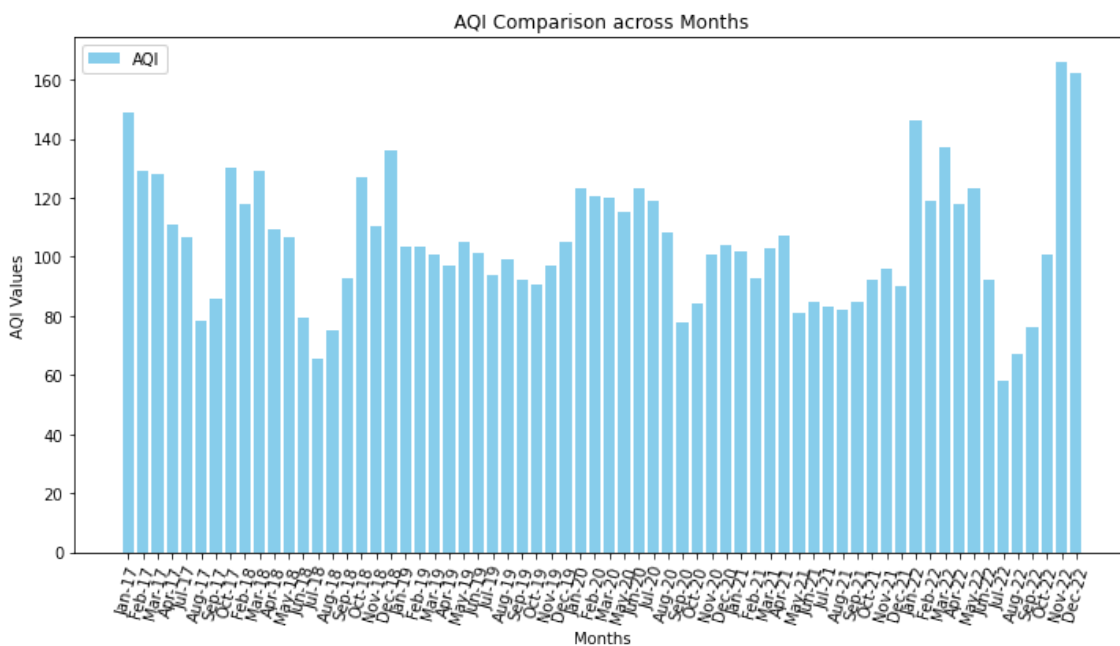
```
In [42]: plt.figure(figsize=(12,6))
         df.boxplot(color='blue')
         plt.title('Distribution of AQI values for Pollutants')
         plt.show()
```



Distribution of AQI values for Pollutants
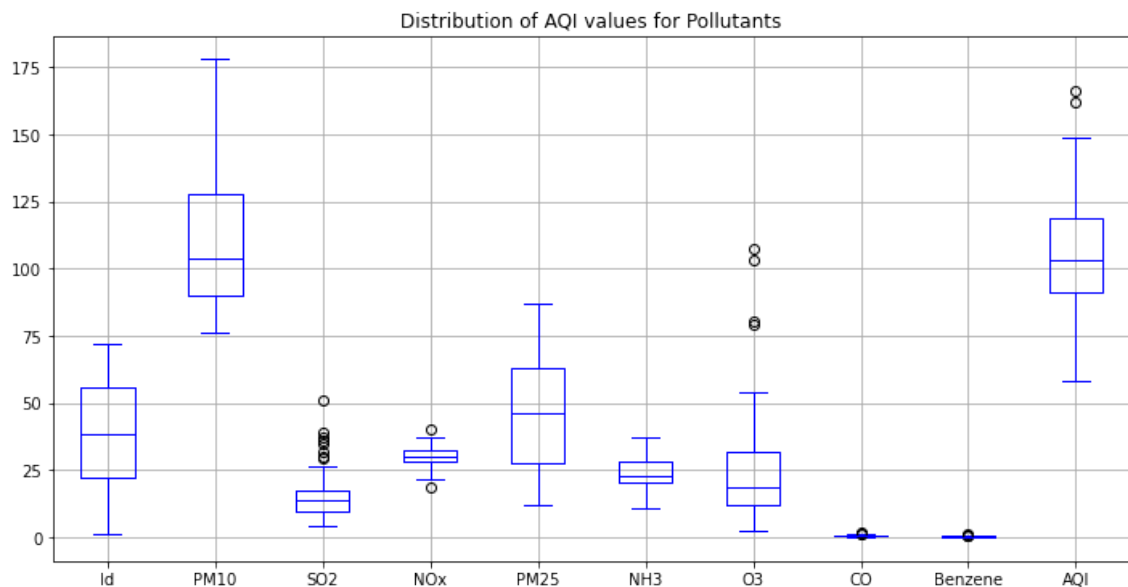
```
In [50]: plt.figure(figsize=(12,6))
         plt.scatter(df['AQI'],df['PM10'],c='red',label='PM10')
         plt.scatter(df['AQI'],df['SO2'],c='blue',label='SO2')
         plt.scatter(df['AQI'],df['NOx'],c='green',label='NOx')
         plt.scatter(df['AQI'],df['PM25'],c='yellow',label='PM2.5')
         plt.scatter(df['AQI'],df['CO'],c='purple',label='CO')
         plt.scatter(df['AQI'],df['O3'],c='black',label='O3')
         plt.title('Scatter plot: AQI levels vs Pollutants')
         plt.xlabel('AQI')
         plt.ylabel('Pollutant')
         plt.legend()
         plt.grid(True)
         plt.show()
```



Scatter plot: AQI levels vs Pollutants

In [ ]: