dmv5

November 3, 2024

1 Data Visualization using matplotlib Problem Statement: Analyzing Air Quality Index (AQI) Trends in a City. The goal is to use the matplotlib library to create visualizations that effectively represent the AQI trends and patterns for different pollutants in the city.

[2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
df= pd.read_csv(r"C:\Users\dell\Desktop\DMV and ML\DMV Datasets\AQI Data Set.
↪csv")
```

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

[4]:

	Id	Mounths	PM10 in æg/m3	SO2 in æg/m3	NOx in æg/m3	PM2.5 in æg/m3	\
0	1	Jan-17	174.0	26.4	35.0	79	
1	2	Feb-17	143.0	35.1	40.3	75	
2	3	Mar-17	142.0	32.1	30.9	59	
3	4	Apr-17	117.0	50.9	36.3	75	
4	5	May-17	NaN	41.6	25.2	53	

	Ammonia - NH3 in æg/m3	O3 in æg/m3	CO in mg/m3	Benzene in æg/m3	\
0	25.0	107.6	0.9	0.7	
1	31.0	103.0	0.9	0.9	
2	26.0	80.7	0.8	0.5	
3	36.0	79.5	0.9	0.7	
4	28.0	70.0	0.5	0.5	

	AQI
0	149.0
1	129.0
2	128.0
3	111.0
4	NaN

[5]:
```
df.columns
```

[5]:
```
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')
```

[6]:
```
column_names= ['Id', 'Months', 'PM10', 'SO2', 'NOx', 'PM25', 'NH3', 'O3','CO',␣
 ↪' Benzene', 'AQI']
df.columns=column_names
df.head()
```

[6]:

	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

[7]:
```
df.isna().sum()
```

[7]:
```
Id           0
Months       0
PM10         6
SO2          1
NOx          2
PM25         0
NH3          0
O3           0
CO           0
 Benzene     0
AQI          5
dtype: int64
```

[8]:
```
df.dropna(inplace=True)
df.isna().sum()
```

[8]:
```
Id           0
Months       0
PM10         0
SO2          0
NOx          0
PM25         0
NH3          0
O3           0
CO           0
 Benzene     0
```

```
AQI         0
dtype: int64
```

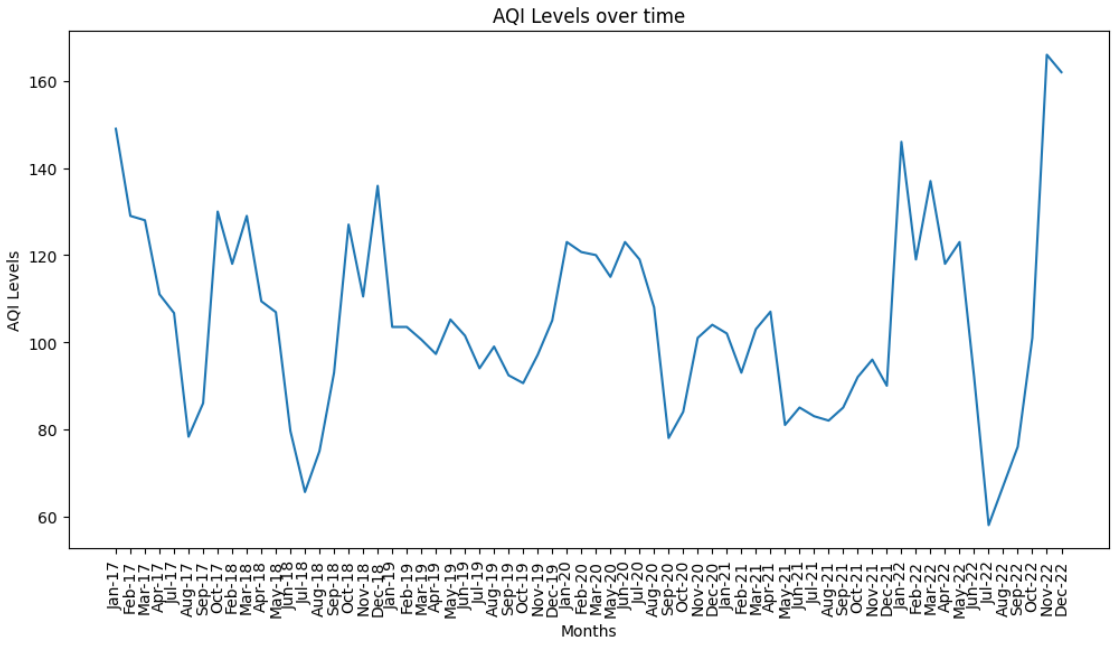
[9]:
```
df.describe()
```

[9]:
```
                Id        PM10        SO2        NOx       PM25        NH3  \
count   66.000000   66.000000  66.000000  66.000000  66.000000  66.000000
mean    38.500000  109.393939  16.093939  30.263636  46.393939  24.072727
std     20.417376   25.271376   9.265218   3.947838  20.261277   5.960474
min      1.000000   76.000000   4.000000  18.400000  12.000000  11.000000
25%     22.250000   90.000000   9.850000  28.125000  27.500000  20.250000
50%     38.500000  104.000000  13.700000  29.750000  46.500000  23.000000
75%     55.750000  128.000000  17.150000  32.550000  62.750000  28.000000
max     72.000000  178.000000  50.900000  40.300000  87.000000  37.000000

                O3         CO    Benzene         AQI
count   66.000000  66.000000  66.000000   66.000000
mean    25.350000   0.551212   0.213636  104.807576
std     21.426413   0.241550   0.190922   22.054250
min      2.400000   0.200000   0.000000   58.000000
25%     12.025000   0.400000   0.100000   90.950000
50%     18.750000   0.500000   0.150000  103.250000
75%     31.575000   0.640000   0.300000  119.000000
max    107.600000   1.520000   0.900000  166.000000
```

[10]:
```
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
```

[11]:
```
plt.figure(figsize=(12,6))
plt.plot(df["Months"],df["AQI"])
plt.xlabel("Months")
plt.ylabel("AQI Levels")
plt.title("AQI Levels over time")
plt.xticks(rotation=90)
plt.show()
```



[12]:
```
df.columns
```

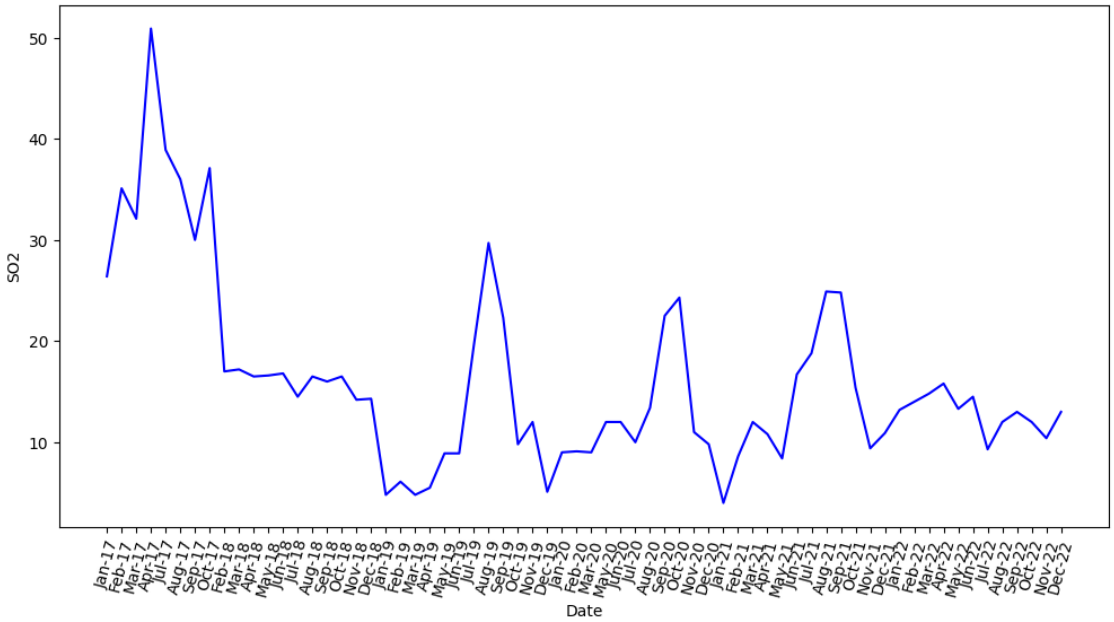
[12]:
```
Index(['Id', 'Months', 'PM10', 'SO2', 'NOx', 'PM25', 'NH3', 'O3', 'CO',
       ' Benzene', 'AQI'],
      dtype='object')
```

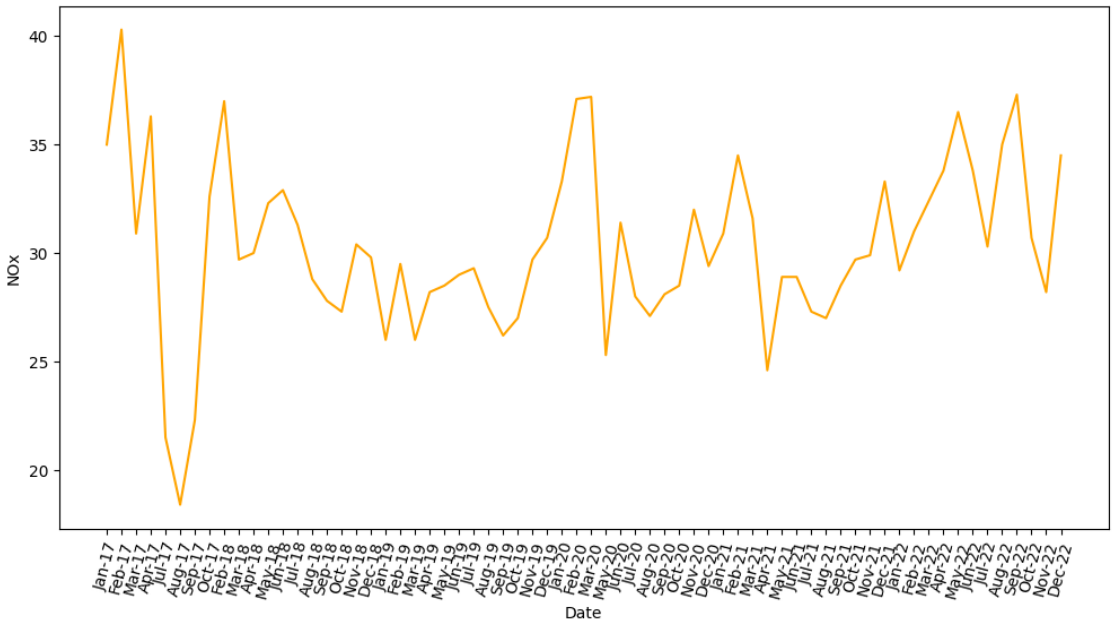
[13]:
```
plt.figure(figsize=(12,6))
plt.plot(df['Months'], df['PM10'], color='red')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('PM10')


plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['SO2'], color='blue')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('SO2')
```

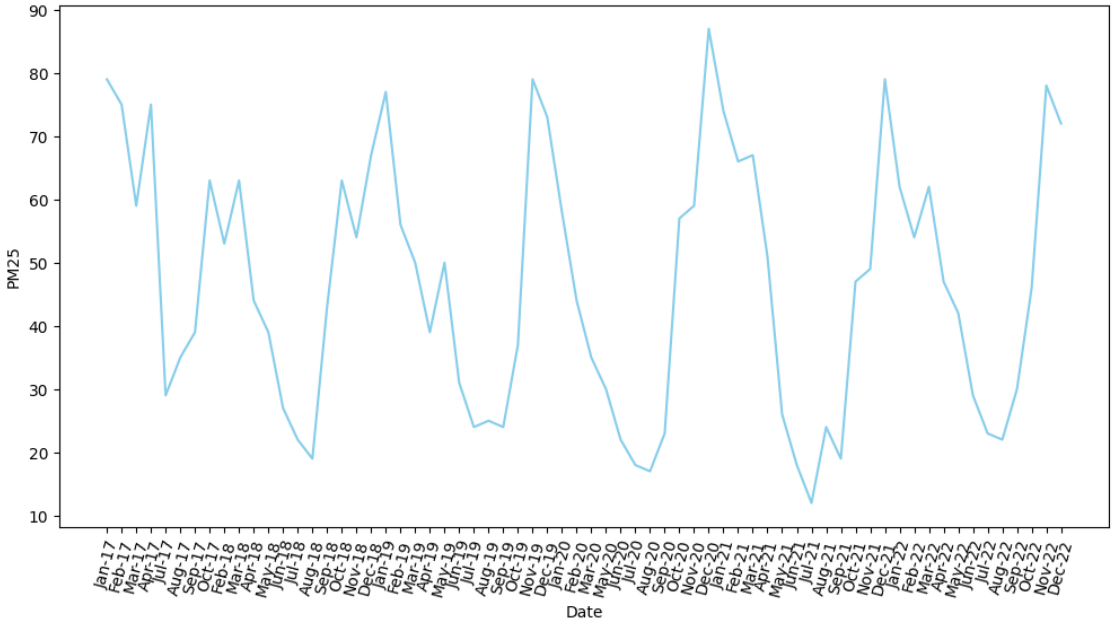
```
plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['NOx'], color='orange')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('NOx')

plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['PM25'], color='skyblue')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('PM25')


plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['NH3'], color='purple')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('NH3')


plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['O3'], color='green')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('O3')


plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['CO'], color='crimson')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('CO')
```

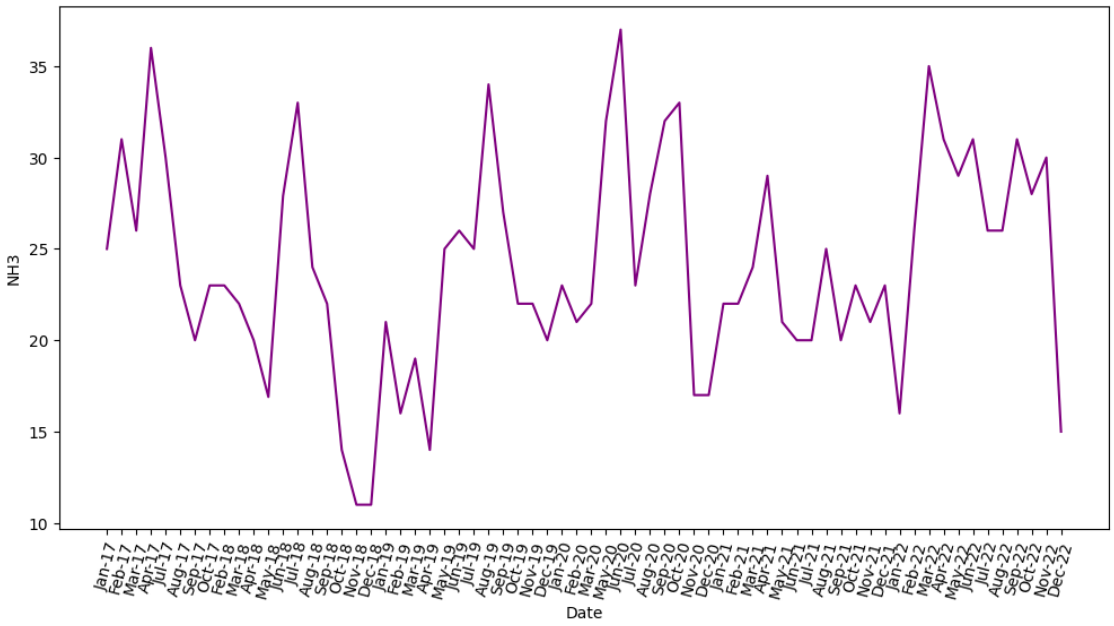
[13]:
```
Text(0, 0.5, 'CO')
```

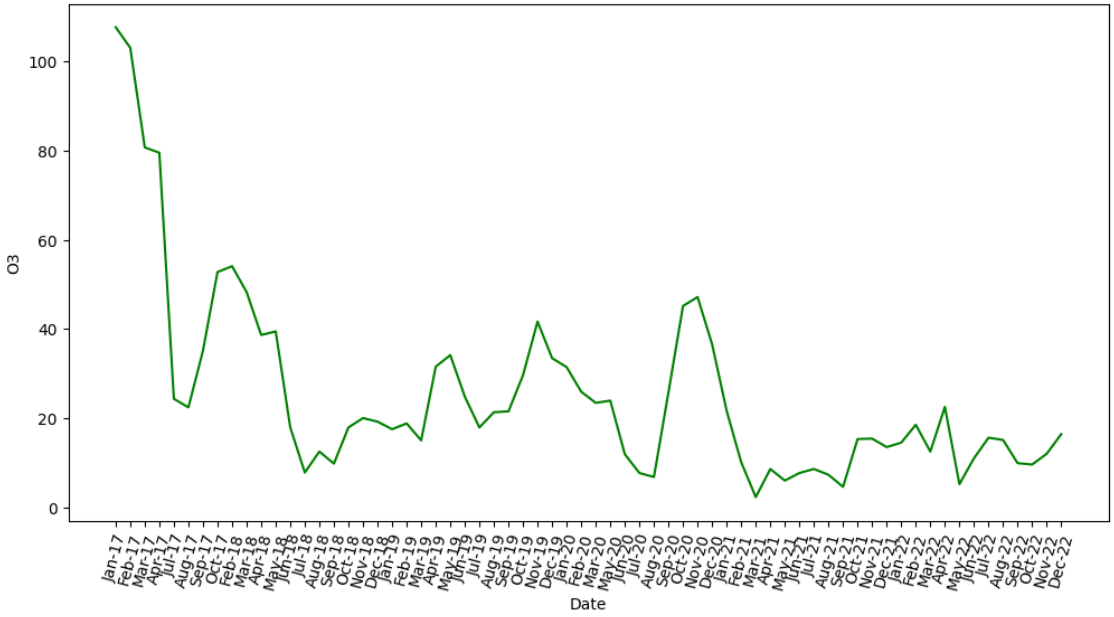

PM10
180
160
140
120
100
80
Jan-17 Feb-17 Mar-17 Apr-17 Jul-17 Aug-17 Sep-17 Oct-17 Feb-18 Mar-18 Apr-18 May-18 Jun-18 Jul-18 Aug-18 Sep-18 Oct-18 Nov-18 Dec-18 Jan-19 Feb-19 Mar-19 Apr-19 May-19 Jun-19 Jul-19 Aug-19 Sep-19 Oct-19 Nov-19 Dec-19 Jan-20 Feb-20 Mar-20 May-20 Jun-20 Jul-20 Aug-20 Sep-20 Oct-20 Nov-20 Dec-20 Jan-21 Feb-21 Mar-21 Apr-21 May-21 Jun-21 Jul-21 Aug-21 Sep-21 Oct-21 Nov-21 Dec-21 Jan-22 Feb-22 Mar-22 Apr-22 May-22 Jun-22 Jul-22 Aug-22 Sep-22 Oct-22 Nov-22 Dec-22
Date

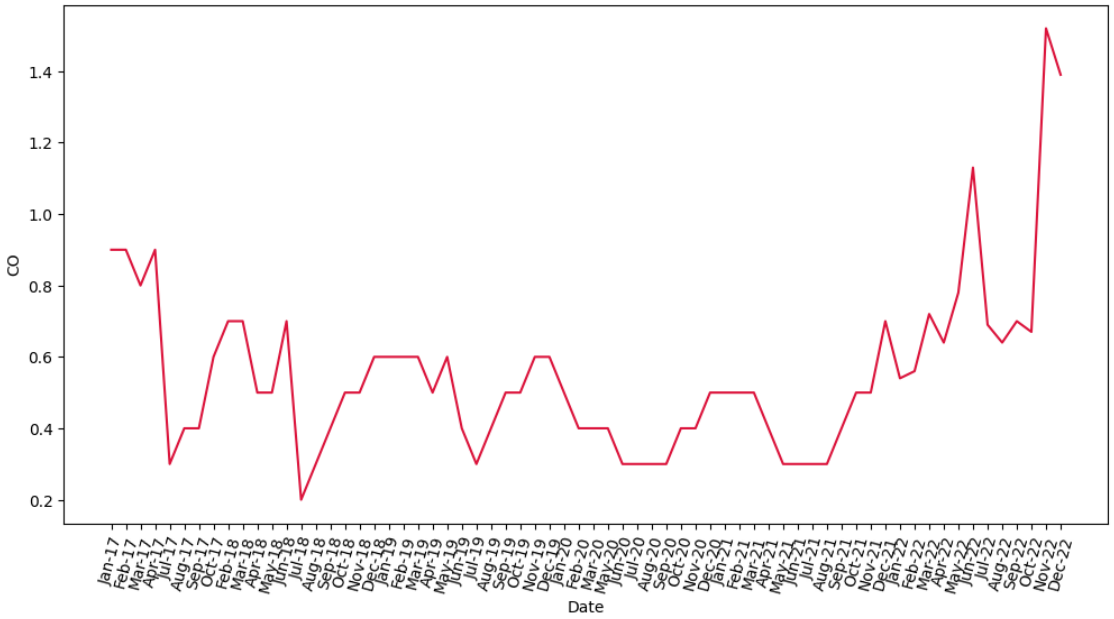

SO2
50
40
30
20
10
Jan-17 Feb-17 Mar-17 Apr-17 Jul-17 Aug-17 Sep-17 Oct-17 Feb-18 Mar-18 Apr-18 May-18 Jun-18 Jul-18 Aug-18 Sep-18 Oct-18 Nov-18 Dec-18 Jan-19 Feb-19 Mar-19 Apr-19 May-19 Jun-19 Jul-19 Aug-19 Sep-19 Oct-19 Nov-19 Dec-19 Jan-20 Feb-20 Mar-20 May-20 Jun-20 Jul-20 Aug-20 Sep-20 Oct-20 Nov-20 Dec-20 Jan-21 Feb-21 Mar-21 Apr-21 May-21 Jun-21 Jul-21 Aug-21 Sep-21 Oct-21 Nov-21 Dec-21 Jan-22 Feb-22 Mar-22 Apr-22 May-22 Jun-22 Jul-22 Aug-22 Sep-22 Oct-22 Nov-22 Dec-22
Date


NOx
40
35
30
25
20
Jan-17
Feb-17
Mar-17
Apr-17
Jul-17
Aug-17
Sep-17
Oct-17
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Jan-20
Feb-20
Mar-20
May-20
Jun-20
Jul-20
Aug-20
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21
May-21
Jun-21
Jul-21
Aug-21
Sep-21
Oct-21
Nov-21
Dec-21
Jan-22
Feb-22
Mar-22
Apr-22
May-22
Jun-22
Jul-22
Aug-22
Sep-22
Oct-22
Nov-22
Dec-22
Date

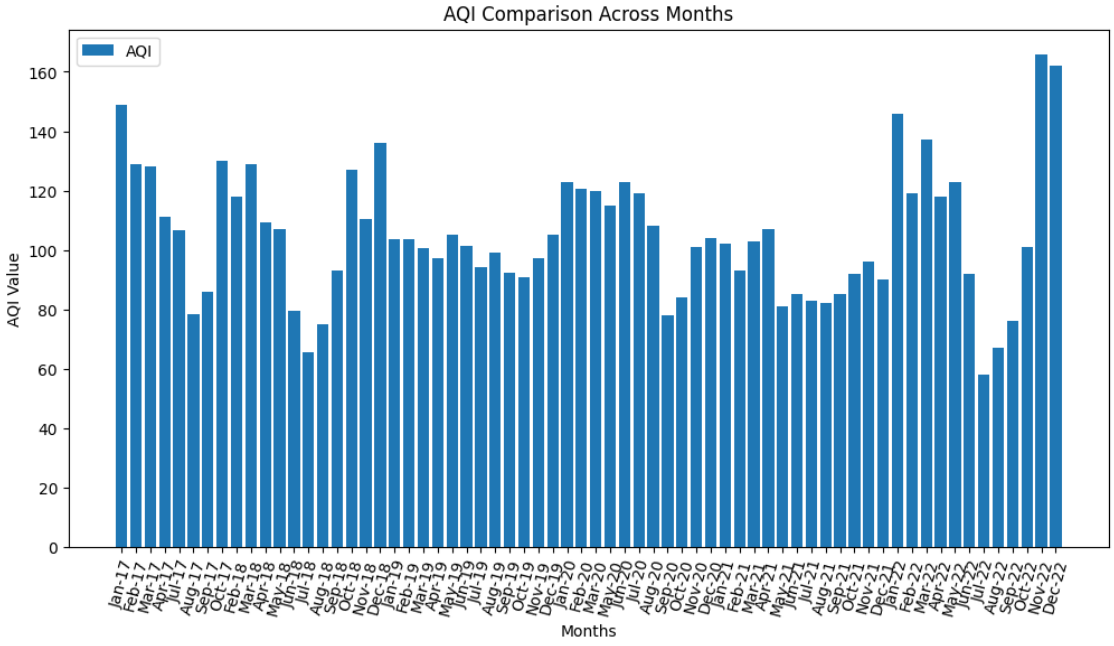

PM25
90
80
70
60
50
40
30
20
10
Jan-17
Feb-17
Mar-17
Apr-17
Jul-17
Aug-17
Sep-17
Oct-17
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Aug-18
Sep-18
Oct-18
Nov-18
Dec-18
Jan-19
Feb-19
Mar-19
Apr-19
May-19
Jun-19
Jul-19
Aug-19
Sep-19
Oct-19
Nov-19
Dec-19
Jan-20
Feb-20
Mar-20
May-20
Jun-20
Jul-20
Aug-20
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21
May-21
Jun-21
Jul-21
Aug-21
Sep-21
Oct-21
Nov-21
Dec-21
Jan-22
Feb-22
Mar-22
Apr-22
May-22
Jun-22
Jul-22
Aug-22
Sep-22
Oct-22
Nov-22
Dec-22
Date


NH3
Date

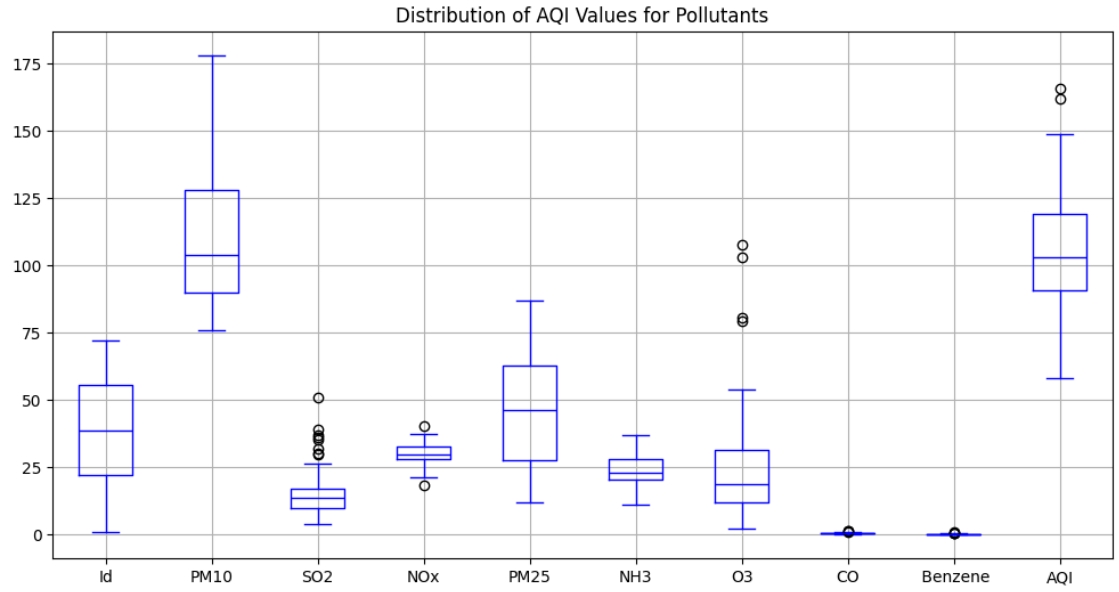

O3
Date



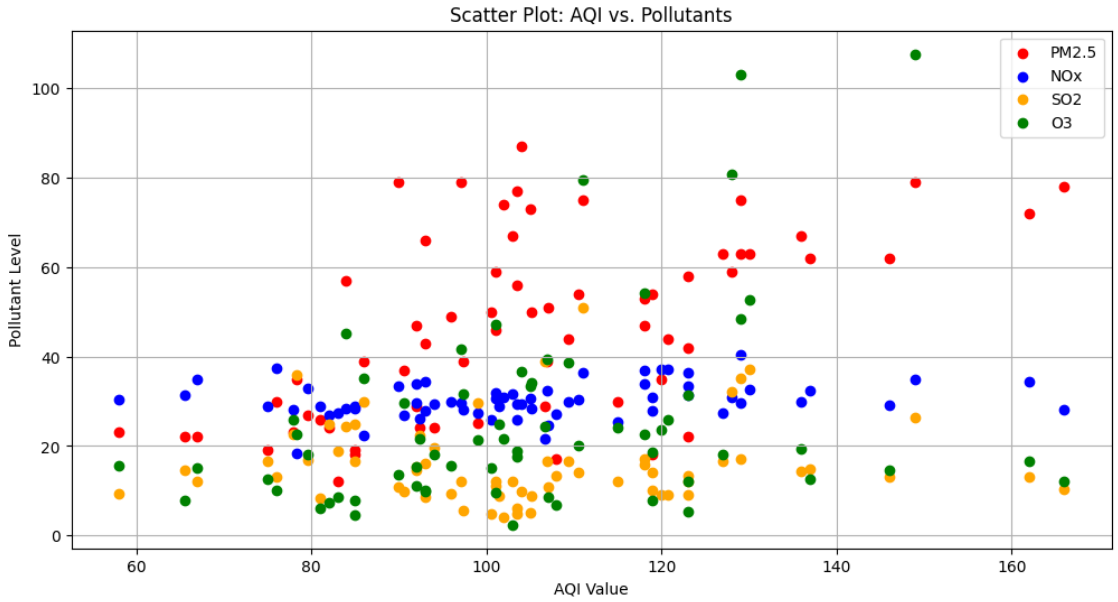
[14]:
```
plt.figure(figsize=(12, 6))
plt.bar(df["Months"], df["AQI"], label="AQI")
plt.title("AQI Comparison Across Months")
plt.xlabel("Months")
plt.ylabel("AQI Value")
plt.legend()
plt.xticks(rotation=75)
plt.show()
```



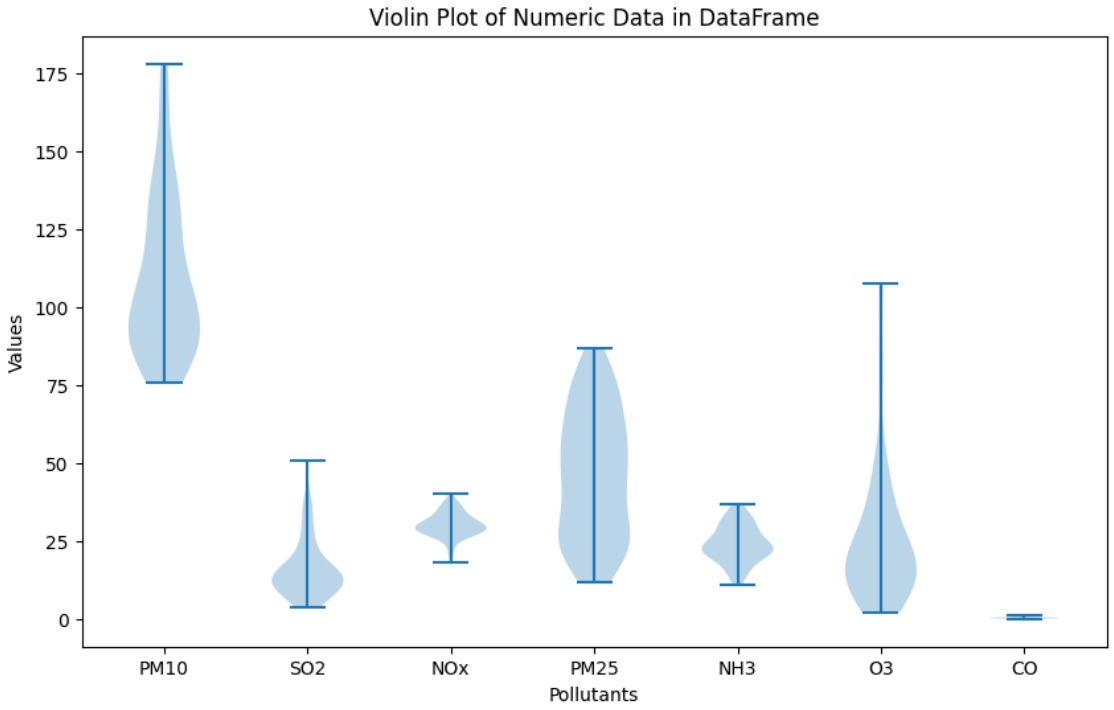
[15]:
```
plt.figure(figsize=(12, 6))
df.boxplot(color='blue')
plt.title("Distribution of AQI Values for Pollutants")
plt.show()
```



[16]:
```
#Scatter plots to explore the relationship between AQI values and pollutant␣
↪levels
plt.figure(figsize=(12, 6))
plt.scatter(df["AQI"], df["PM25"], c="red", label="PM2.5")
plt.scatter(df["AQI"], df["NOx"], c="blue", label="NOx")
plt.scatter(df["AQI"], df["SO2"], c="orange", label="SO2")
plt.scatter(df["AQI"], df["O3"], c="green", label="O3")
plt.title("Scatter Plot: AQI vs. Pollutants")
plt.xlabel("AQI Value")
plt.ylabel("Pollutant Level")
plt.legend()
plt.grid(True)
plt.show()
```



[18]:
```
pollutants=['PM10', 'SO2', 'NOx', 'PM25', 'NH3', 'O3', 'CO']
plt.figure(figsize=(10, 6))
plt.violinplot(df[pollutants])
plt.xlabel("Pollutants")
plt.xticks(range(1, len(pollutants) + 1),pollutants)
plt.ylabel("Values")
plt.title("Violin Plot of Numeric Data in DataFrame")
plt.show()
```


Violin Plot of Numeric Data in DataFrame
Values
175
150
125
100
75
50
25
0
PM10
SO2
NOx
PM25
NH3
O3
CO
Pollutants

[ ]: