In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns

df = pd.read_csv("AQI Data Set.csv",  parse_dates=['Mounths'])
df.head()
```

Out[1]:

	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 [2]:
```
df.columns
```

Out[2]:
```
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 [3]:
```
column_names = ['Id', 'Months', 'PM10', 'SO2', 'NOx',
       'PM25', 'NH3', 'O3','CO', ' Benzene', 'AQI']

df.columns = column_names
df.head()
```

Out[3]:

	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 [4]:
```
df.isna().sum()
```

Out[4]:
```
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 [5]:
```
df.dropna(inplace=True)
df.isna().sum()
```

Out[5]:
```
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 [6]:
```
df.describe()
```

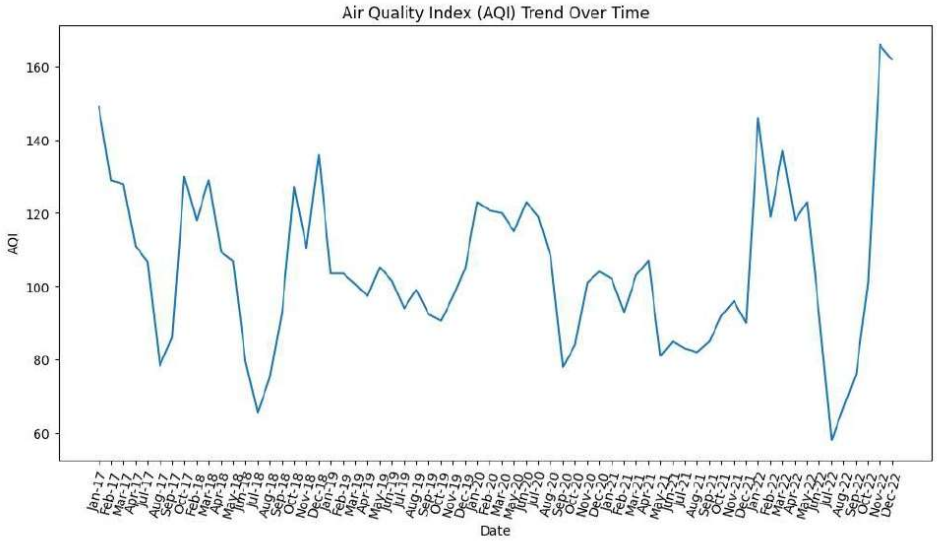
Out[6]:

	Id	PM10	SO2	NOx	PM25	NH3	O3	CO	Benzene	AQI
count	66.000000	66.000000	66.000000	66.000000	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	25.350000	0.551212	0.213636	104.807576
std	20.417376	25.271376	9.265218	3.947838	20.261277	5.960474	21.426413	0.241550	0.190922	22.054250
min	1.000000	76.000000	4.000000	18.400000	12.000000	11.000000	2.400000	0.200000	0.000000	58.000000
25%	22.250000	90.000000	9.850000	28.125000	27.500000	20.250000	12.025000	0.400000	0.100000	90.950000
50%	38.500000	104.000000	13.700000	29.750000	46.500000	23.000000	18.750000	0.500000	0.150000	103.250000
75%	55.750000	128.000000	17.150000	32.550000	62.750000	28.000000	31.575000	0.640000	0.300000	119.000000
max	72.000000	178.000000	50.900000	40.300000	87.000000	37.000000	107.600000	1.520000	0.900000	166.000000

In [7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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 [8]:
```
plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['AQI'])
plt.xlabel('Date')
plt.ylabel('AQI')
plt.xticks(rotation=75)
plt.title('Air Quality Index (AQI) Trend Over Time')
plt.show()
```

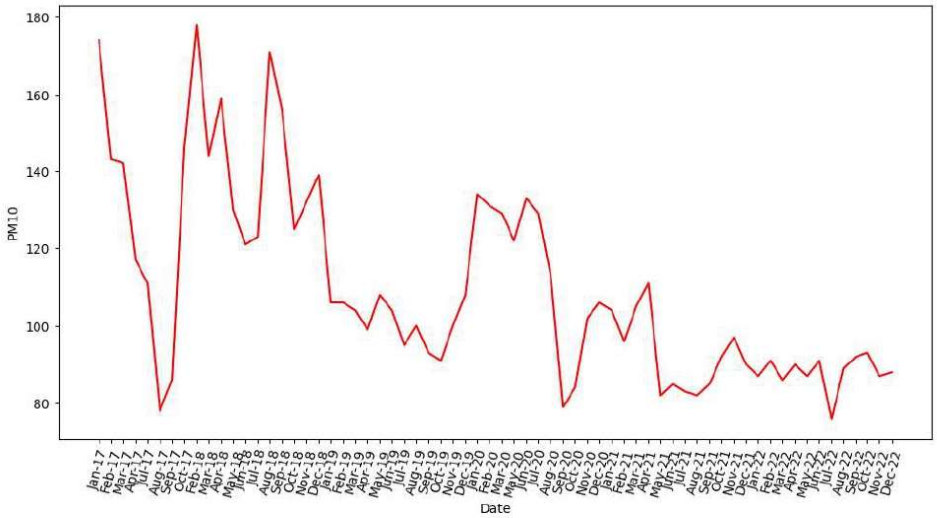


In [9]:
```
df.columns
```

Out[9]:
```
Index(['Id', 'Months', 'PM10', 'SO2', 'NOx', 'PM25', 'NH3', 'O3', 'CO',
       ' Benzene', 'AQI'],
      dtype='object')
```

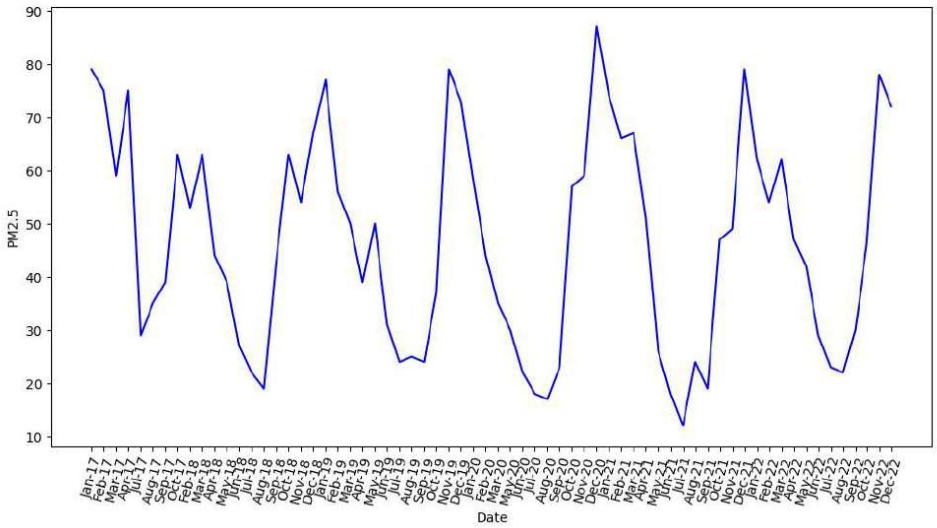
```
In [10]: plt.figure(figsize=(12, 6))
         plt.plot(df['Months'], df['PM10'], color='red')
         plt.xlabel('Date')
         plt.xticks(rotation=75)
         plt.ylabel('PM10')
```

```
Out[10]: Text(0, 0.5, 'PM10')
```



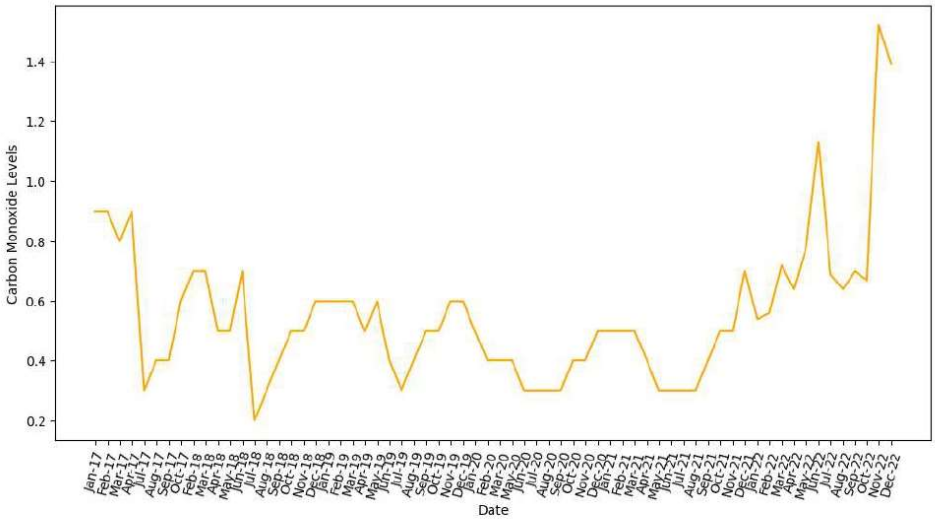

```
In [11]: plt.figure(figsize=(12, 6))
         plt.plot(df['Months'], df['PM25'], color='blue')
         plt.xlabel('Date')
         plt.xticks(rotation=75)
         plt.ylabel('PM2.5')
```

```
Out[11]: Text(0, 0.5, 'PM2.5')
```

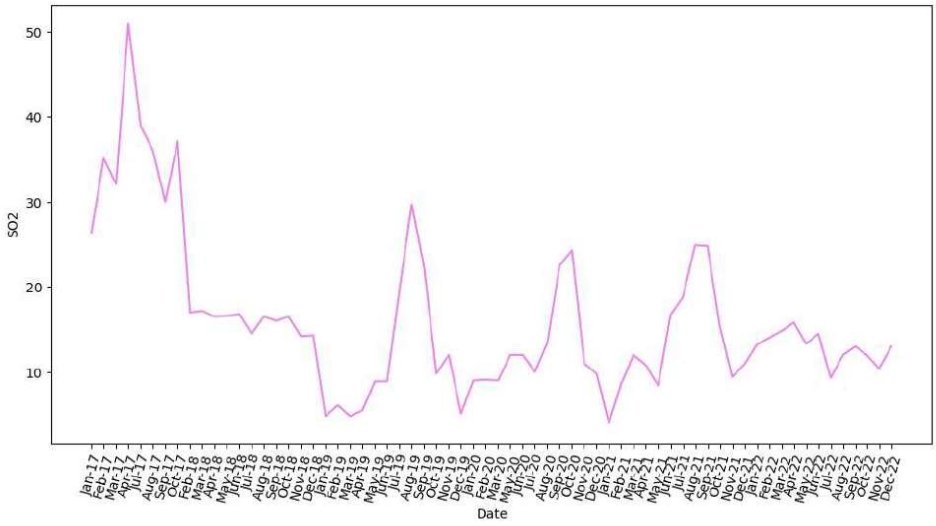


In [12]:
```
plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['CO'], label='CO', color='orange')
plt.xlabel('Date')
plt.xticks(rotation=75)
plt.ylabel('Carbon Monoxide Levels')
```

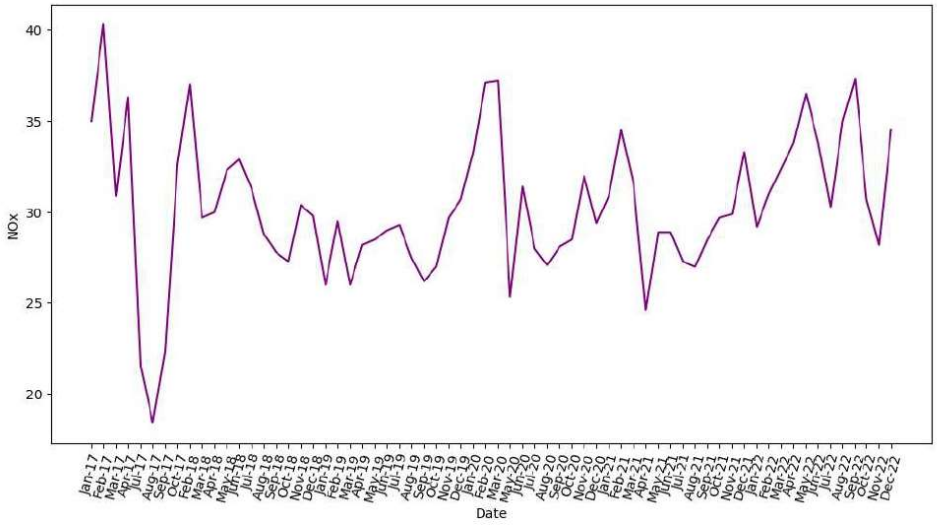
Out[12]: Text(0, 0.5, 'Carbon Monoxide Levels')



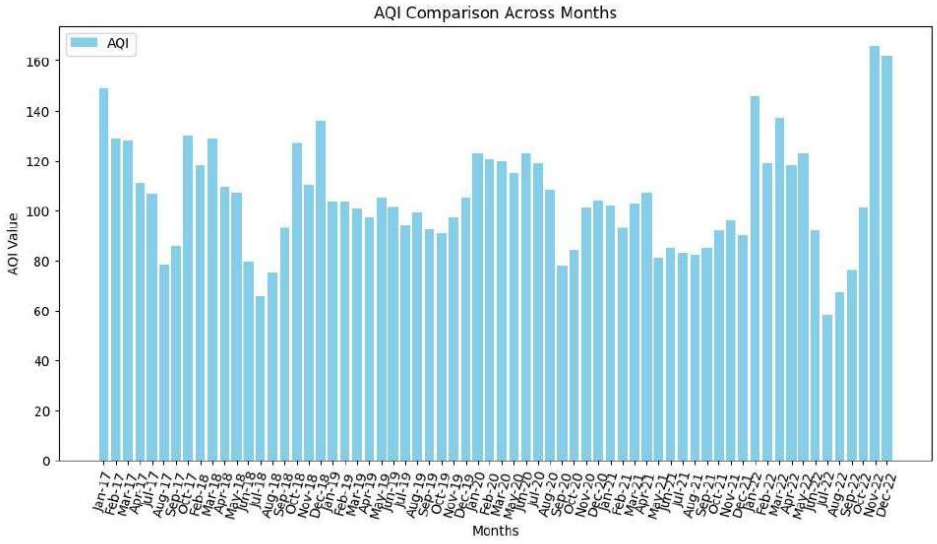
In [13]:
```
plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['SO2'], label='SO2', color='violet')
plt.xlabel('Date')
plt.ylabel('SO2')
plt.xticks(rotation=75)
plt.show()
```



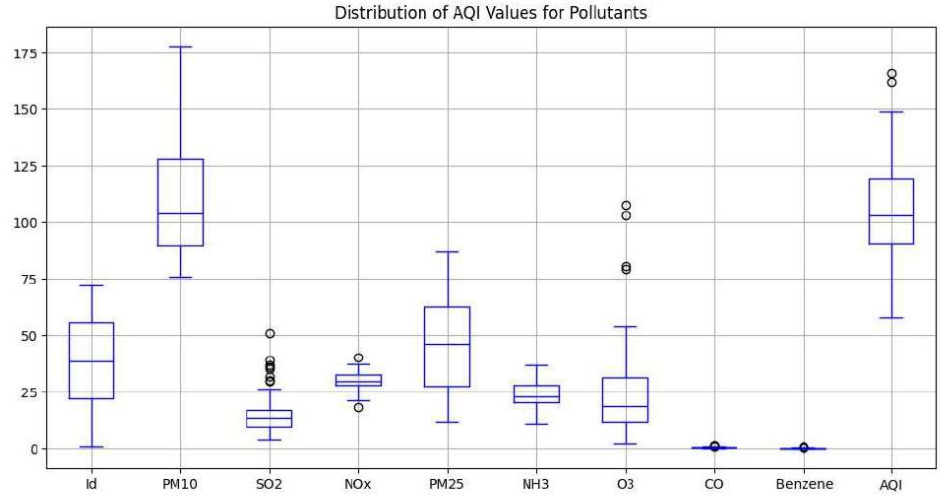
In [14]:
```
plt.figure(figsize=(12, 6))
plt.plot(df['Months'], df['NOx'], label='NOx', color='purple')
plt.xlabel('Date')
plt.ylabel('NOx')
plt.xticks(rotation=75)
plt.show()
```



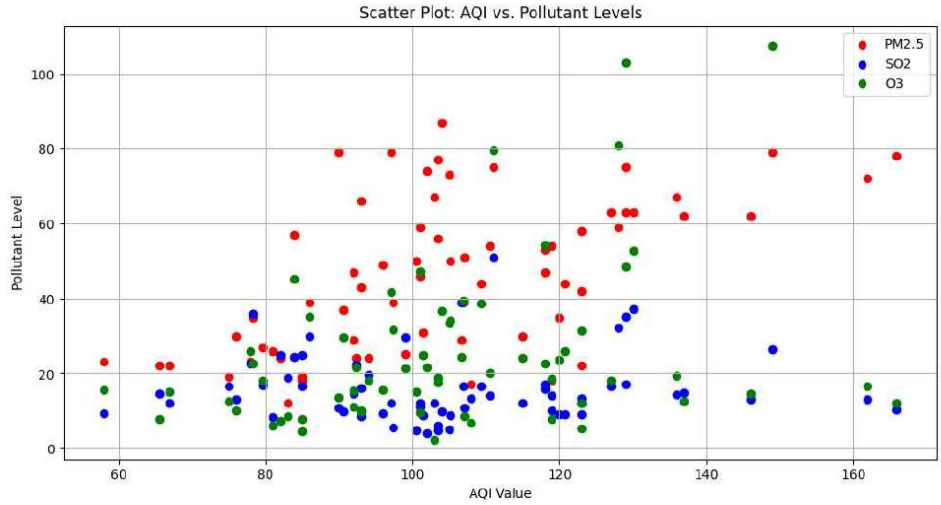
In [15]:
```
# Bar plots to compare AQI values across different dates or time periods
plt.figure(figsize=(12, 6))
plt.bar(df["Months"], df["AQI"], color="skyblue", label="AQI")
plt.title("AQI Comparison Across Months")
plt.xlabel("Months")
plt.ylabel("AQI Value")
plt.legend()
plt.xticks(rotation=75)
plt.show()
```



In [16]:
```
#Box plots to analyze the distribution of AQI values for different pollutant categories
plt.figure(figsize=(12, 6))
df.boxplot(color='blue')
plt.title("Distribution of AQI Values for Pollutants")
plt.show()
```



In [17]:
```
#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["SO2"], c="blue", label="SO2")
plt.scatter(df["AQI"], df["O3"], c="green", label="O3")
plt.title("Scatter Plot: AQI vs. Pollutant Levels")
plt.xlabel("AQI Value")
plt.ylabel("Pollutant Level")
plt.legend()
plt.grid(True)
plt.show()
```



In [ ]: