DATE:17-10-2023 TEAM ID:3886 PROJECT TITLE:Air Quality Analysis in TamilNadu

IMPORTING LIBRARIES

In [244]:

```
import numpy as np
import pandas as pd
import os
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import warnings
```

LOADING THE DATASET

In [245]:

```
dataset= pd.read_csv("C:\\Users\\SANTH\\Downloads\\cpcb_dly_aq_tamil_nadu-2014.csv")
```

DATA EXPLORATION:

In [246]:

```
dataset
```

Out[246]:

	Stn Code	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10	PM 2.5
0	38	01-02-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
1	38	01-07-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
2	38	21-01-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN
3	38	23-01-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0	NaN
4	38	28-01-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...
2874	773	12-03-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	18.0	102.0	NaN
2875	773	12-10-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	12.0	14.0	91.0	NaN
2876	773	17-12-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	19.0	22.0	100.0	NaN
2877	773	24-12-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	17.0	95.0	NaN
2878	773	31-12-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	14.0	16.0	94.0	NaN

2879 rows × 11 columns

In [247]:
```
dataset.drop('Stn Code', axis=1, inplace=True)
dataset.drop('PM 2.5', axis=1, inplace=True)
dataset
```

Out[247]:

	Sampling Date	State	City/Town/Village/Area	Location of Monitoring Station	Agency	Type of Location	SO2	NO2	RSPM/PM10
0	01-02-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0
1	01-07-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0
2	21-01-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0
3	23-01-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	16.0	46.0
4	28-01-2014	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0
...	...	...	...	...	...	...	...	...	...
2874	12-03-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	18.0	102.0
2875	12-10-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	12.0	14.0	91.0
2876	17-12-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	19.0	22.0	100.0
2877	24-12-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	15.0	17.0	95.0
2878	31-12-2014	Tamil Nadu	Trichy	Central Bus Stand, Trichy	Tamilnadu State Pollution Control Board	Residential, Rural and other Areas	14.0	16.0	94.0

2879 rows × 9 columns

In [248]:
```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2879 entries, 0 to 2878
Data columns (total 9 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Sampling Date                   2879 non-null   object 
 1   State                           2879 non-null   object 
 2   City/Town/Village/Area          2879 non-null   object 
 3   Location of Monitoring Station  2879 non-null   object 
 4   Agency                          2879 non-null   object 
 5   Type of Location                2879 non-null   object 
 6   SO2                             2868 non-null   float64
 7   NO2                             2866 non-null   float64
 8   RSPM/PM10                       2875 non-null   float64
dtypes: float64(3), object(6)
memory usage: 202.6+ KB
```

In [249]:
```
dataset.describe()
```

Out[249]:

	SO2	NO2	RSPM/PM10
count	2868.000000	2866.000000	2875.000000
mean	11.503138	22.136776	62.494261
std	5.051702	7.128694	31.368745
min	2.000000	5.000000	12.000000
25%	8.000000	17.000000	41.000000
50%	12.000000	22.000000	55.000000
75%	15.000000	25.000000	78.000000
max	49.000000	71.000000	269.000000

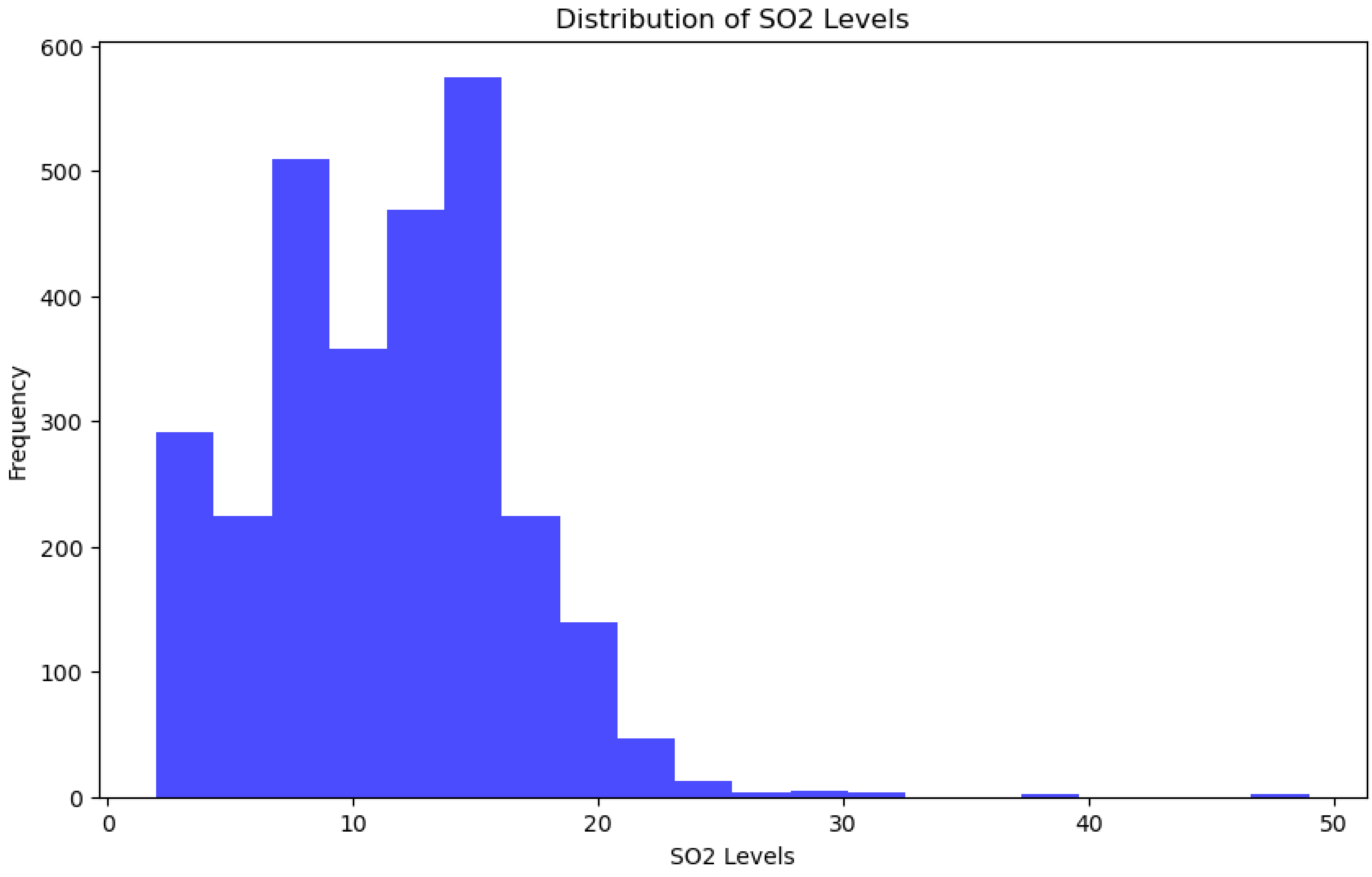
In [250]:
```
dataset.columns
```

Out[250]:
```
Index(['Sampling Date', 'State', 'City/Town/Village/Area',
       'Location of Monitoring Station', 'Agency', 'Type of Location', 'SO2',
       'NO2', 'RSPM/PM10'],
      dtype='object')
```

DATA VISUALIZATION:

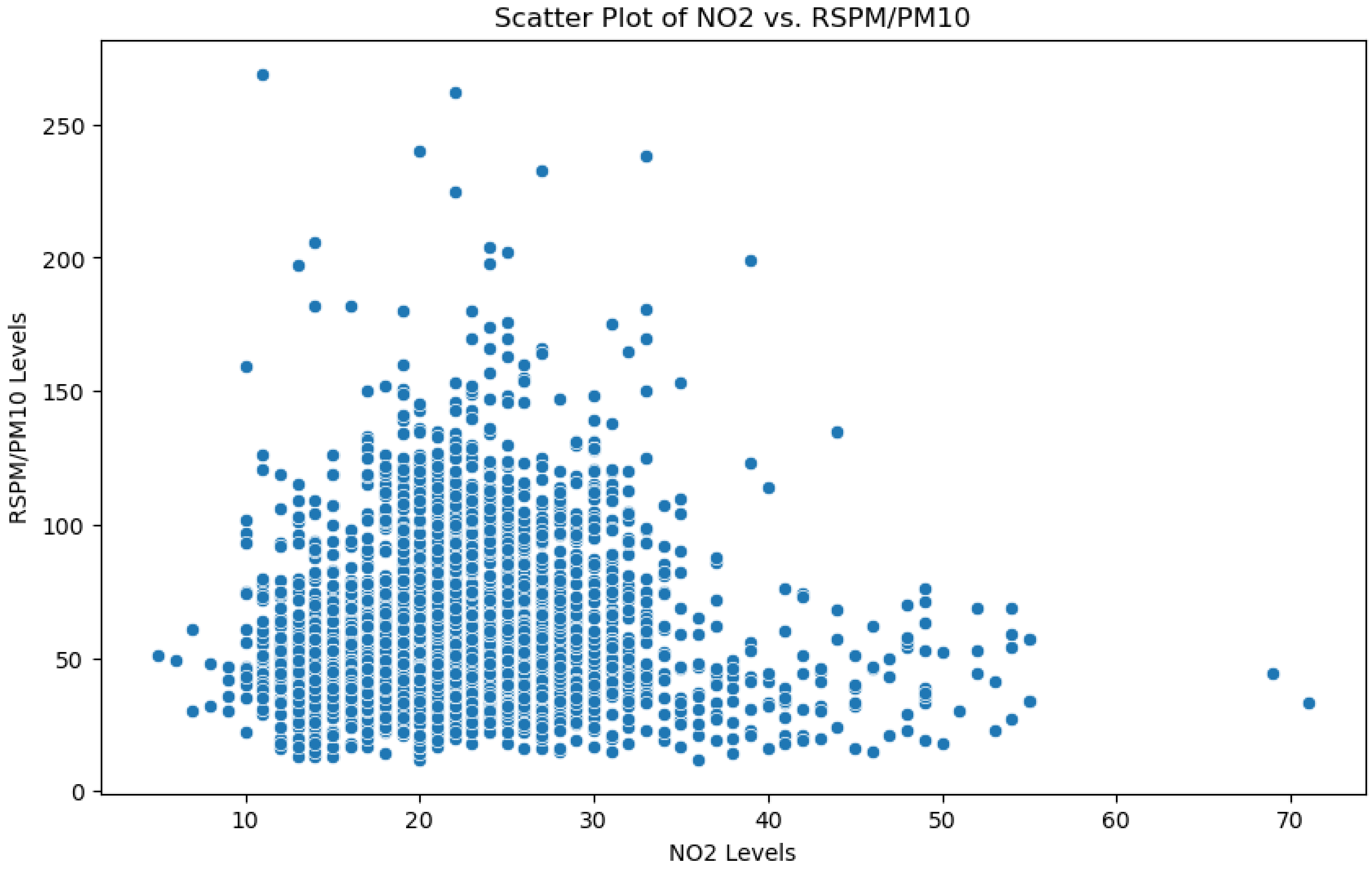
1.Histogram for SO2 levels

```
In [251]: plt.figure(figsize=(10, 6))
          plt.hist(dataset['SO2'], bins=20, color='blue', alpha=0.7)
          plt.title('Distribution of SO2 Levels')
          plt.xlabel('SO2 Levels')
          plt.ylabel('Frequency')
          plt.show()
```



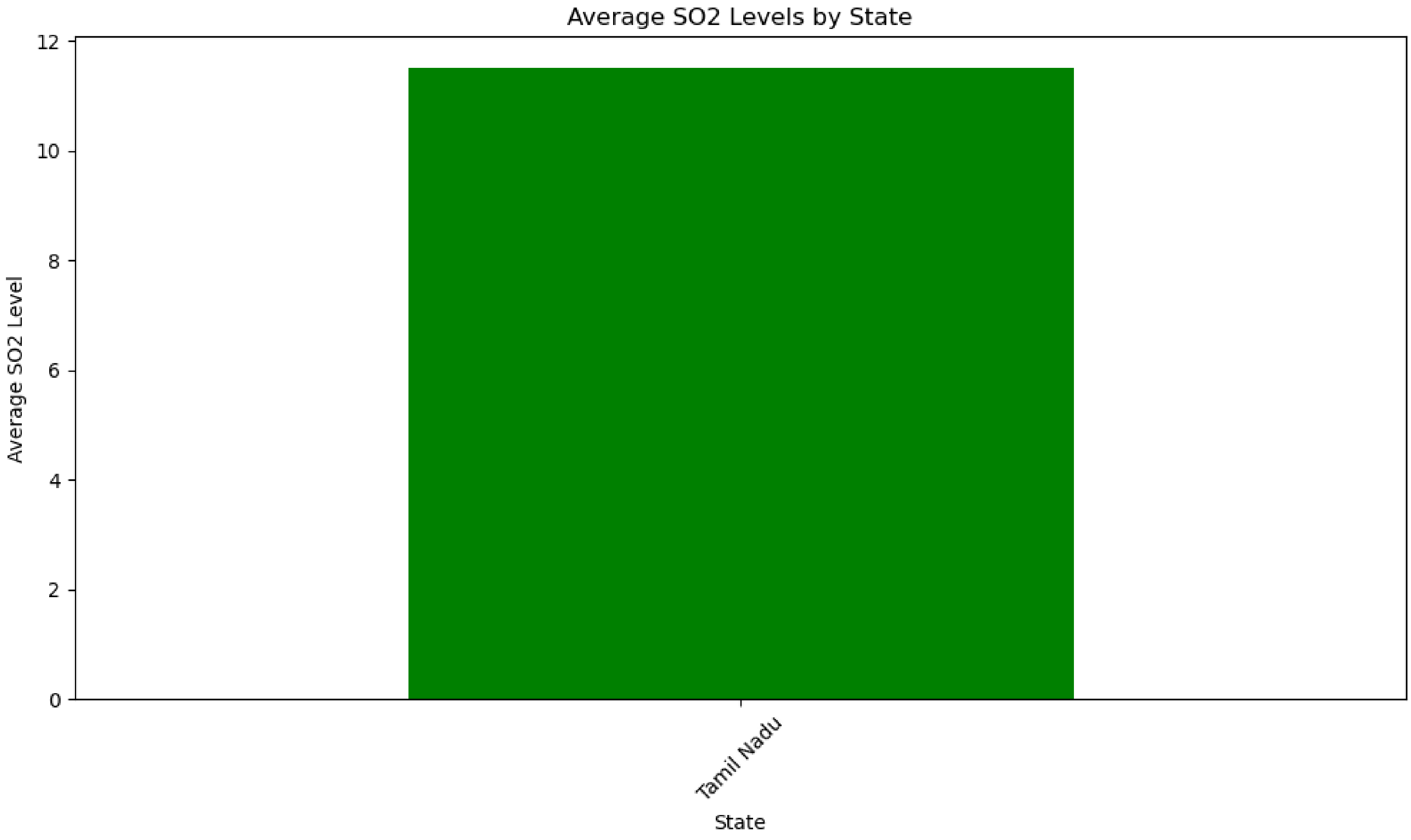
2.Scatter plot of NO2 vs. RSPM/PM10

```
In [252]: plt.figure(figsize=(10, 6))
          sns.scatterplot(x='NO2', y='RSPM/PM10', data=dataset)
          plt.title('Scatter Plot of NO2 vs. RSPM/PM10')
          plt.xlabel('NO2 Levels')
          plt.ylabel('RSPM/PM10 Levels')
          plt.show()
```



3.Bar chart for State-wise SO2 levels

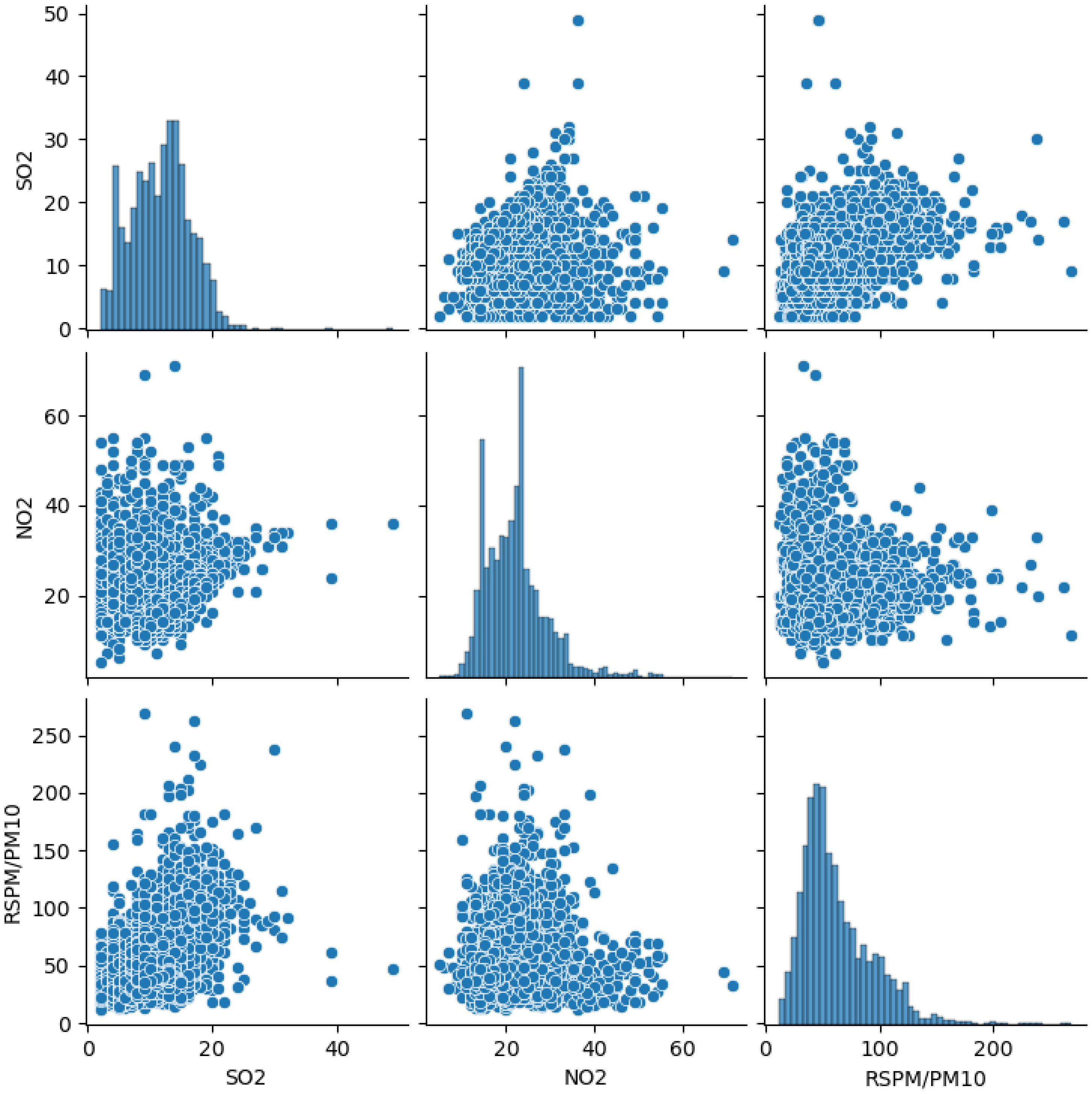
```
In [253]: statewise_so2 = dataset.groupby('State')['SO2'].mean().sort_values(ascending=False)
          plt.figure(figsize=(12, 6))
          statewise_so2.plot(kind='bar', color='green')
          plt.title('Average SO2 Levels by State')
          plt.xlabel('State')
          plt.ylabel('Average SO2 Level')
          plt.xticks(rotation=45)
          plt.show()
```



In [254]:
```
plt.figure(figsize=(12,8))
sns.pairplot(dataset)
```

Out[254]:
```
<seaborn.axisgrid.PairGrid at 0x2340fe67c40>

<Figure size 1200x800 with 0 Axes>
```

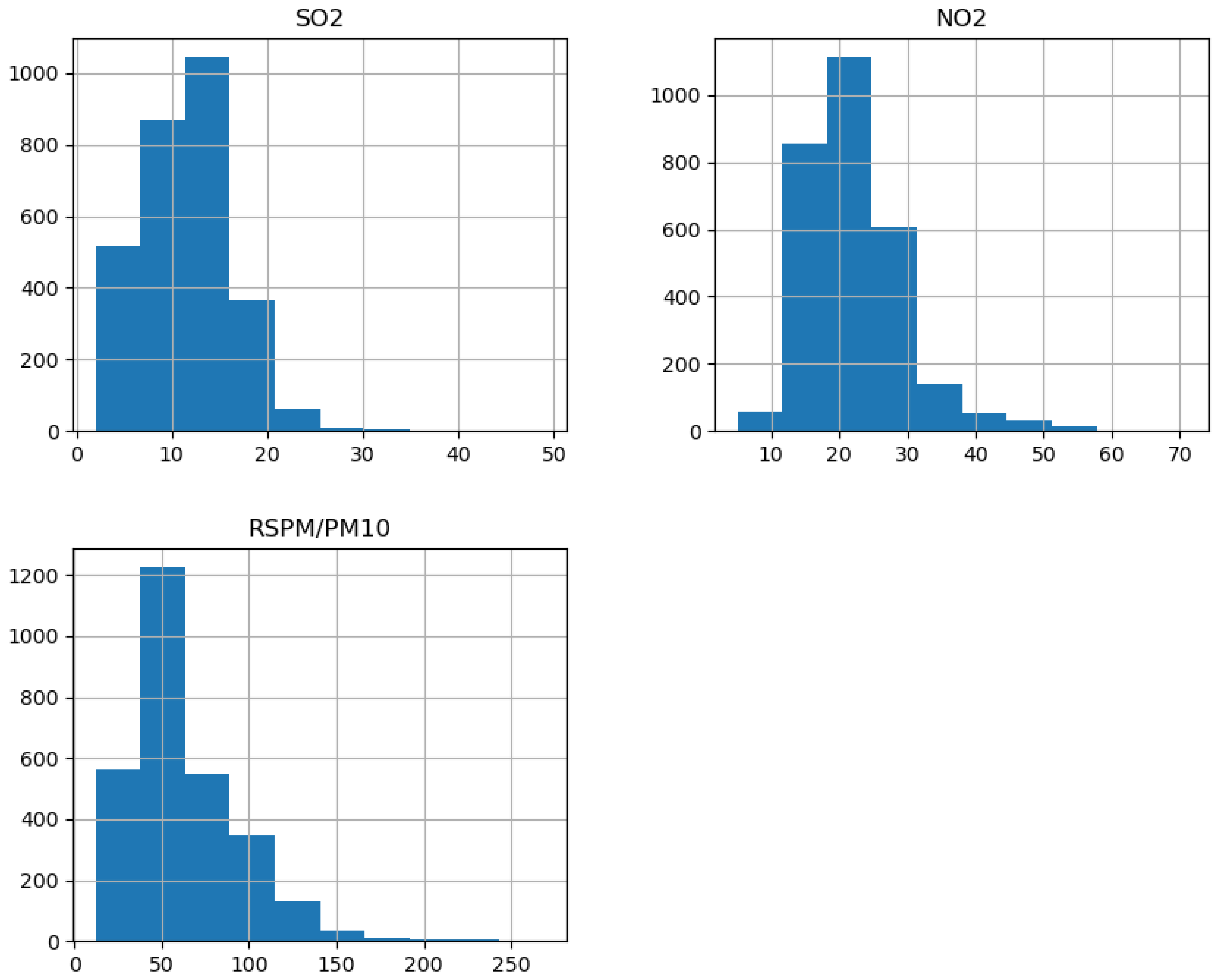


In [255]:

```
dataset.hist(figsize=(10,8))
```

Out[255]:

```
array([[<AxesSubplot:title={'center':'SO2'}>,
        <AxesSubplot:title={'center':'NO2'}>],
       [<AxesSubplot:title={'center':'RSPM/PM10'}>, <AxesSubplot:>]],
      dtype=object)
```



Visualising Correlation

In [256]:

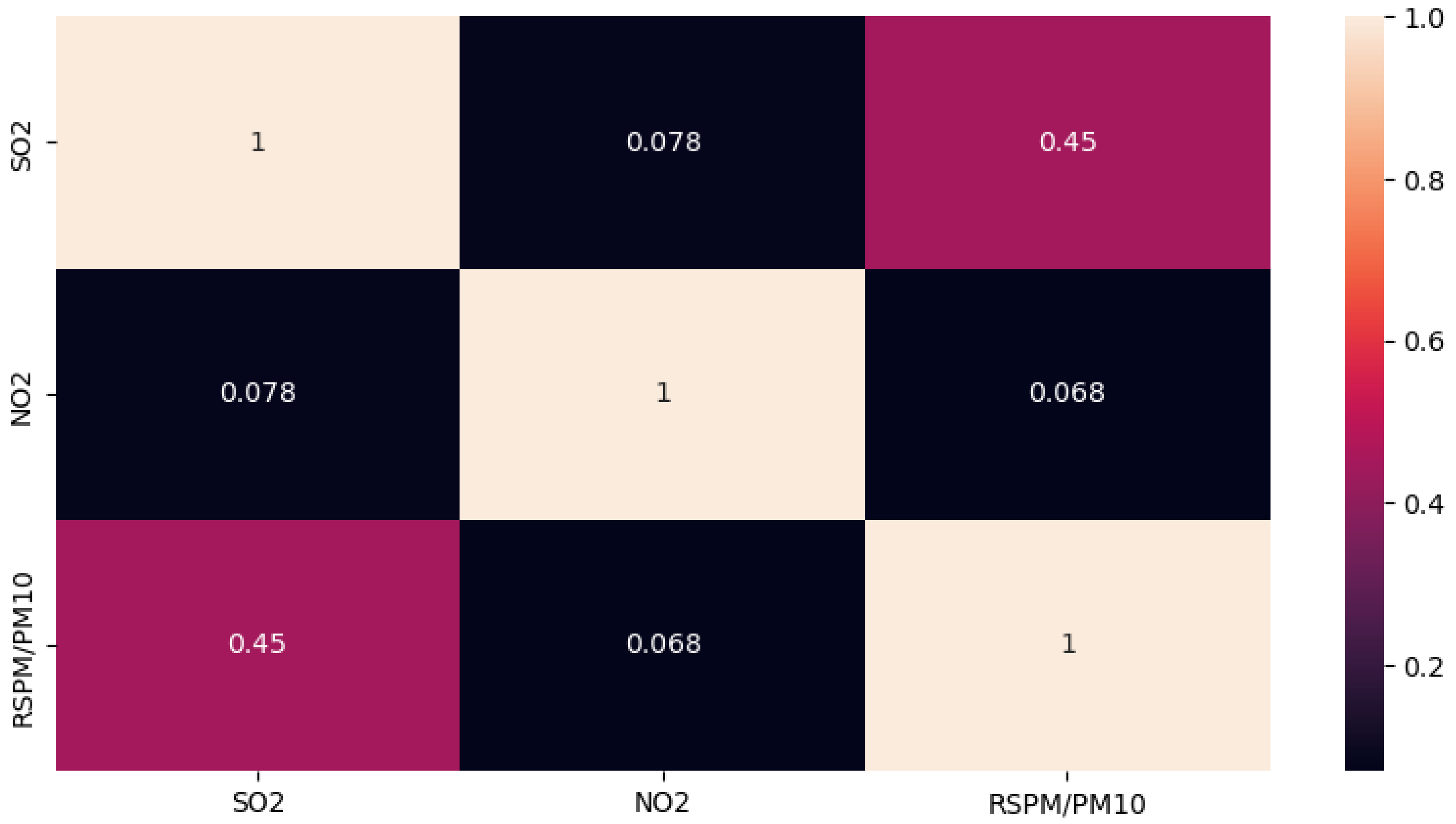
```
dataset.corr()
```

Out[256]:

	SO2	NO2	RSPM/PM10
SO2	1.000000	0.078246	0.445152
NO2	0.078246	1.000000	0.068277
RSPM/PM10	0.445152	0.068277	1.000000

In [257]:
```
plt.figure(figsize=(10,5))
sns.heatmap(dataset.corr(), annot=True)
```

Out[257]:
```
<AxesSubplot:>
```



In [ ]:
```
```