AIR QUALITY ANALYSIS AND PREDICTION IN TAMILNADU

PHASE 4

Air quality analysis

Calculate average SO2, NO2, and RSPM/PM10 levels across different monitoring stations, cities, or areas. Identify pollution trends and areas with high pollution levels.

Create visualizations

Create visualizations using data visualization libraries (e.g., Matplotlib, Seaborn).

STEPS

Step 1: Data Collection
Step 2: Data Preprocessing
Step 3: Calculate Averages
Step 4: Visualization
Step 5: Identify Pollution Trends
Step 6: Interpretation and Reporting

1.DATA COLLECTION

Collect air quality data from different monitoring stations, cities, or areas. You may obtain this data from government agencies, research organizations, or online databases.

2.DATA PREPROCESSING

Collect air quality data from different monitoring stations, cities, or areas. You may obtain this data from government agencies, research organizations, or online databases.

3.CALCULATE AVERAGES

Calculate the average levels of SO2, NO2, and RSPM/PM10 across the monitoring stations for each city or area. You can use Python and libraries such as Pandas for data manipulation.

4.VISUALIZATION

Create visualizations using data visualization libraries like Matplotlib or Seaborn to represent the calculated average pollutant levels.

5.IDENTIFY POLLUTION TRENDS

Analyze the visualizations to identify pollution trends. Look for cities or areas with consistently high levels of SO2, NO2, or RSPM/PM10. You can also create line charts to visualize trends over time, if your data includes timestamps.

6.INTERPRETATION AND REPORTING

Interpret the findings and create a report summarizing the pollution trends, areas with high pollution levels, and any insights gained from the visualizations.

IMPORTING PACKAGES

```
: import pandas as pd
```

```
: import numpy as np
```

```
: from sklearn import preprocessing
```

```
: import matplotlib.pyplot as plt
```

```
: from numpy import array
```

```
: from sklearn.preprocessing import MinMaxScaler
```

```
: from sklearn.metrics import mean_squared_error
```

LOAD THE DATAS

```
df = pd.read_csv('air_quality.csv')
print(df)
```

OUTPUT:

```
      No  year  month  day  hour  PM2.5  PM10   SO2   NO2     CO    O3  \
0      1  2013      3    1     0    4.0   4.0   4.0   7.0  300.0  77.0
1      2  2013      3    1     1    8.0   8.0   4.0   7.0  300.0  77.0
2      3  2013      3    1     2    7.0   7.0   5.0  10.0  300.0  73.0
3      4  2013      3    1     3    6.0   6.0  11.0  11.0  300.0  72.0
4      5  2013      3    1     4    3.0   3.0  12.0  12.0  300.0  72.0
```

VIEWING INFO ABOUT COLUMN

```
df.info()
```

VIEWING INFO ABOUT ROW

```
df.head()
```

OUTPUT FOR ROWS:

	No	year	month	day	hour	PM2.5	PM10	SO2	NO2	CO	O3	TEMP
0	1	2013	3	1	0	4.0	4.0	4.0	7.0	300.0	77.0	-0.7
1	2	2013	3	1	1	8.0	8.0	4.0	7.0	300.0	77.0	-1.1
2	3	2013	3	1	2	7.0	7.0	5.0	10.0	300.0	73.0	-1.1
3	4	2013	3	1	3	6.0	6.0	11.0	11.0	300.0	72.0	-1.4
4	5	2013	3	1	4	3.0	3.0	12.0	12.0	300.0	72.0	-2.0

OUTPUT FOR COLUMN:

```
RangeIndex: 35064 entries, 0 to 35063
Data columns (total 18 columns):
No          35064 non-null int64
year        35064 non-null int64
month       35064 non-null int64
day         35064 non-null int64
hour        35064 non-null int64
PM2.5       34139 non-null float64
PM10        34346 non-null float64
SO2         34129 non-null float64
NO2         34041 non-null float64
CO          33288 non-null float64
O3          33345 non-null float64
TEMP        35044 non-null float64
PRES        35044 non-null float64
DEWP        35044 non-null float64
RAIN        35044 non-null float64
wd          34983 non-null object
WSPM        35050 non-null float64
station     35064 non-null object
dtypes: float64(11), int64(5), object(2)
memory usage: 4.8+ MB
```

CHECKING NULL VALUES

```
df.null().sum()
```

OUTPUT

Out[7]:

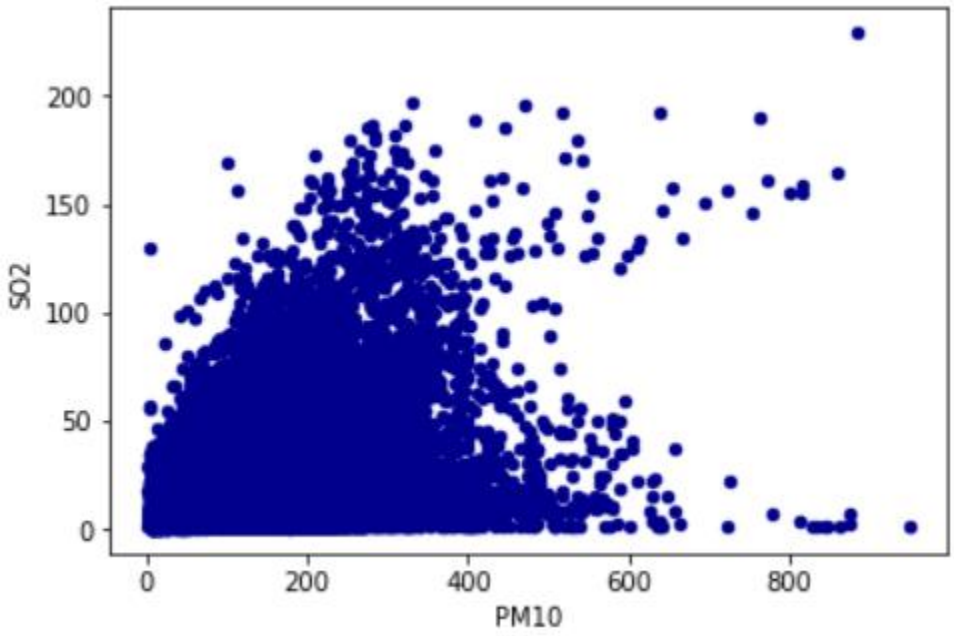
No	0
year	0
month	0
day	0
hour	0
PM2.5	925
PM10	718
SO2	935
NO2	1023
CO	1776
O3	1719
TEMP	20

DATA VISUALIZATION

FOR SO2:

```
df.plot.scatter(x='PM10', y='SO2', c='DarkBlue')
```

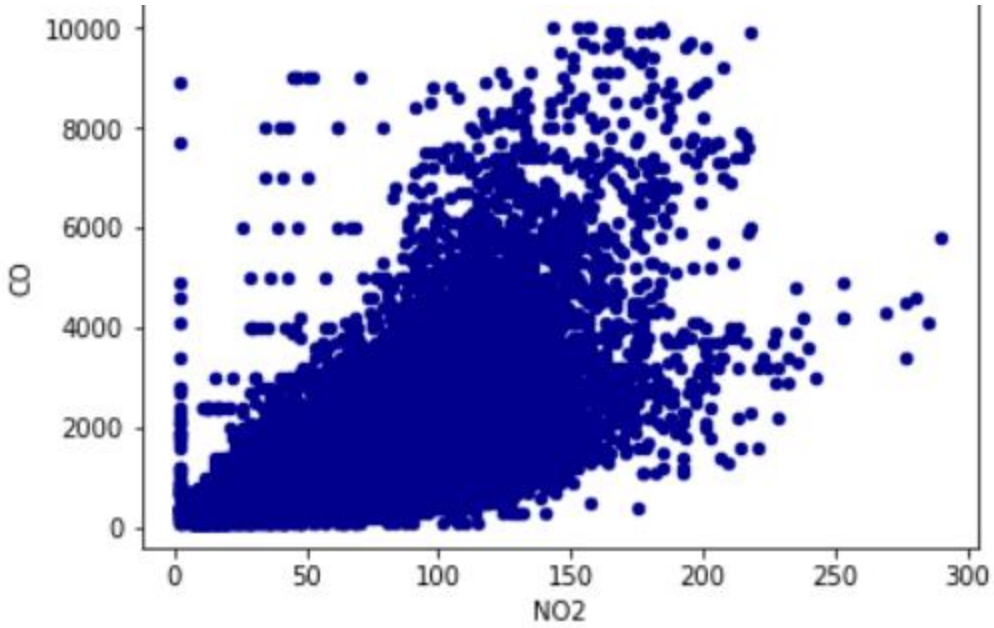
OUTPUT:



FOR NO2:

```
df.plot.scatter(x='SO2', y='NO2', c='DarkBlue')
```

OUTPUT:



CONCLUSION:

Thus the data visualization of SO2,NO2 is obtained.Values of air quality analysis is predicted.