AIR QUALITY ANALYSIS IN TAMILNADU - PHASE 5

Introduction:

Air Quality Monitoring Networks allow the measurement, operation and predictive analysis of the evolution of air pollution in different areas (urban areas, industrial areas, special nature conservation areas, etc.) Some stations are equipped with meteorological sensors and/or noise level meters to measure noise levels.

PROBLEM SOLUTION:

DATA COLLECTION AND MONITORING:

Enhance air quality monitoring infrastructure.

Collect and analyze data to identify pollution sources and hotspots.

EMISSION REDUCTION POLICIES:

Implement and enforce stricter emissions standards for industries and vehicles.

Promote the use of cleaner fuels and technologies.

PROMOTE PUBLIC TRANSPORTATION:

Invest in efficient and affordable public transportation systems to reduce the number of vehicles on the

road.

GREEN ENERGY SOURCES:

Encourage the adoption of renewable energy sources to reduce reliance on fossil fuels.

REFORESTATION AND GREEN SPACES:

Increase green spaces, parks, and tree planting to absorb pollutants and improve air quality.

AWARENESS AND EDUCATION:

Educate the public about the importance of reducing pollution and adopting sustainable practices.

REGULATE CONSTRUCTION:

Implement construction regulations to reduce dust and emissions from building activities.

Encourage sustainable urban planning and green building practices.

INDUSTRIAL COMPLIANCE:

Ensure industries comply with pollution control regulations.

Promote cleaner production processes.

COMMUNITY ENGAGEMENT:

Engage communities in air quality improvement efforts.

Encourage citizen initiatives like carpooling and reducing waste.

GOVERNMENT SUPPORT:

Allocate funds for air quality improvement projects.

Collaborate with other regions and countries to address cross-border pollution issues.

HEALTHCARE AND AWARENESS:

Improve healthcare infrastructure to handle pollution-related illnesses.

Raise awareness about the health risks of poor air quality.

RESEARCH AND INNOVATION:

Invest in research for innovative pollution control technologies.

Support startups and businesses working on air quality solutions.

DATA VISUALISATION:

It is used in representing model or graph.

The library used is data visualisation is matplot.

IDEAS AND INNOVATION :

1. Real-time :

Develop an air quality analyzer that provides real-time data on air pollution levels in different areas of Tamil Nadu. This can help residents and authorities to take timely actions to improve air quality.

2. Mobile Applications:

Create a mobile application that integrates with the air quality analyzer and provides users with personalized air quality alerts and recommendations for improving air quality in their immediate surroundings.

3. Indoor Air Quality:

Expand the capabilities of air quality analyzers to measure indoor air quality parameters such as volatile organic compounds (VOCs) and carbon dioxide levels. This can help individuals maintain healthy indoor environments.

4. IoT Integration:

Develop air quality analyzers that can connect to the Internet of Things (IoT) ecosystem. This allows for centralized monitoring and control of multiple analyzers in different locations, providing a comprehensive view of air quality across Tamil Nadu.

5. Data Visualization:

Enhance the data visualization capabilities of air quality analyzers, presenting pollution levels in an easily understandable format such as heat maps or color-coded graphs. This can help in identifying pollution hotspots and trends over time.

6. Sensor Miniaturization:

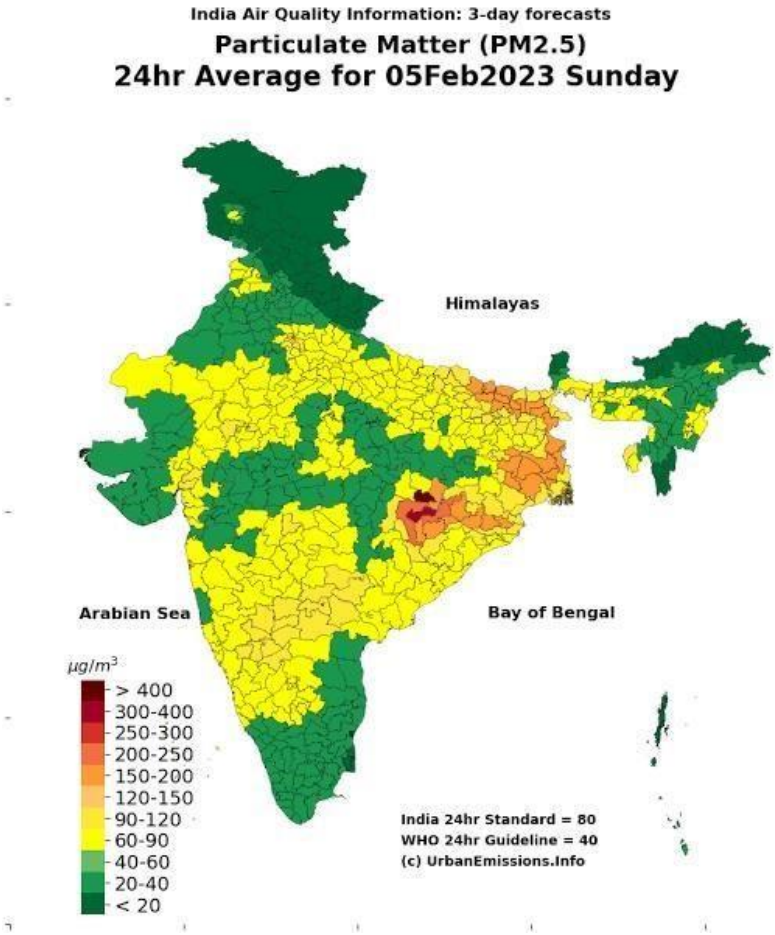
Explore miniaturization techniques to develop smaller, portable air quality analyzers that can be easily carried and used by individuals. This can encourage citizen involvement in monitoring air pollution levels.

7. Integration with Green Solutions:

Integrate air quality analyzers with other green solutions such as solar panels or electric vehicle charging stations. This creates an opportunity to promote sustainable practices while monitoring air quality.

8. Public Awareness Campaigns:

Develop innovative ways to raise public awareness about air pollution, utilizing the data collected by the analyzers. For example, interactive displays in public spaces that showcase real-time pollution levels and their impact on health.



9. Collaboration with Authorities:

Establish partnerships with local authorities, educational institutions, and research organizations to collect and analyze air quality data. This collaboration can lead to evidence-based policies and initiatives for improving air quality in Tamil Nadu.

10. Affordable Solutions:

Focus on developing cost-effective air quality analyzers that are accessible to a broader population, including low-income communities. This empowers individuals to actively monitor and address air pollution issues in their own communities.

LOADING AND PREPROCESSING METHODS:

- IMPORT LIBRARIES
- LOAD THE DATASET

- EXPLORE THE DATASET
- HANDLING MISSING DATA
- DATA CLEANING
- DATA TRANSFORMATION
- FEATURE ENGINEERING
- EXPLORATORY DATA ANALYSIS (EDA)
- SAVE PREPROCESSED DATASET

Dataset Link: *https://tn.data.gov.in/resource/location-wise-daily-ambientair-quality-tamil-nadu-year-2014*

```
In [2]: import pandas as pd
        import plotly.express as px
        import plotly.io as pio
        import plotly.graph_objects as go
        pio.templates.default = "plotly_white"

        data = pd.read_csv("D:\cpcb_dly_aq_tamil_nadu-2014.csv")
        print(data.head())

   stn code sampling date       state city/town/village/area  \
0        38      01-02-14  Tamil Nadu                Chennai
1        38      01-07-14  Tamil Nadu                Chennai
2        38      21-01-14  Tamil Nadu                Chennai
3        38      23-01-14  Tamil Nadu                Chennai
4        38      28-01-14  Tamil Nadu                Chennai

                    Location of Monitoring Station  \
0  Kathivakkam, Municipal Kalyana Mandapam, Chennai
1  Kathivakkam, Municipal Kalyana Mandapam, Chennai
2  kathivakkam, municipal kalyana mandapam, chennai
3  kathivakkam, municipal kalyana mandapam, chennai
4  kathivakkam, municipal kalyana mandapam, chennai

                                  Agency Type of Location   SO2   NO2  \
0  Tamilnadu State Pollution Control Board  Industrial Area  11.0  17.0
1  Tamilnadu State Pollution Control Board  Industrial Area  13.0  17.0
2  Tamilnadu State Pollution Control Board  Industrial Area  12.0  18.0
3  Tamilnadu State Pollution Control Board  Industrial Area  15.0  16.0
4  Tamilnadu State Pollution Control Board  Industrial Area  13.0  14.0

   RSPM/PM10  PM 2.5
0       55.0     NaN
1       45.0     NaN
2       50.0     NaN
3       46.0     NaN
4       42.0     NaN
```

DATASET :

```
In [3]: print(data.describe())

          Stn Code          SO2          NO2    RSPM/PM10  PM 2.5
count  2879.000000  2879.000000  2879.000000  2879.000000     0.0
mean    475.750261    11.515109    22.136158    62.511289     NaN
std     277.675577     5.071178     7.123029    31.393031     NaN
min      38.000000     2.000000     5.000000    12.000000     NaN
25%     238.000000     8.000000    17.000000    41.000000     NaN
50%     366.000000    12.000000    22.000000    55.000000     NaN
75%     764.000000    15.000000    25.000000    78.000000     NaN
max     773.000000    49.000000    71.000000   269.000000     NaN
```

HANDLING MISSING DATA:

Identify and handle missing data, which could involve removing rows with missing values or imputing missing values.

Check for missing values print(df.isnull().sum())

Handle missing values (example: impute with mean)

df['column_name'].fillna(df['column_name'].mean(), inplace=True)

```
# Check for missing values
print(data.isnull().sum())

# Handle missing values (example: impute with mean)
data['PM 2.5'].fillna(data['PM 2.5'].mean(), inplace=True)
```

```
Stn Code                        0
Sampling Date                   0
State                           0
City/Town/Village/Area          0
Location of Monitoring Station  0
Agency                          0
Type of Location                0
SO2                             0
NO2                             0
RSPM/PM10                       0
PM 2.5                       2879
dtype: int64
```

Data Cleaning:

Clean the data by addressing any data anomalies, inconsistencies, or outliers.

Data Transformation:

Depending on your project's requirements, you may need to transform the data. This could include converting date columns to datetime objects, encoding categorical variables, or scaling numerical features.

import matplotlib.pyplot as plt

from pandas.api.types import is_string_dtype, is_numeric_dtype

df = pd.read_csv("../input/marketing-data/marketing_data.csv") df.head()

```
import matplotlib.pyplot as plt
from pandas.api.types import is_string_dtype, is_numeric_dtype

df = pd.read_csv("D:\cpcb_dly_aq_tamil_nadu-2014.csv")
df.head()
```

	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	01-02-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	11.0	17.0	55.0	NaN
1	38	01-07-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	17.0	45.0	NaN
2	38	21-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	12.0	18.0	50.0	NaN
3	38	23-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	15.0	18.0	46.0	NaN
4	38	28-01-14	Tamil Nadu	Chennai	Kathivakkam, Municipal Kalyana Mandapam, Chennai	Tamilnadu State Pollution Control Board	Industrial Area	13.0	14.0	42.0	NaN

Feature Engineering:

Create new features or modify existing ones to improve your dataset's quality.

Exploratory Data Analysis (EDA):

Perform exploratory data analysis using visualizations (e.g., Matplotlib or Seaborn) to gain insights into your data.

Save Preprocessed Dataset:

Once you've completed preprocessing, save the cleaned and transformed dataset to a new file for future use.

df.to_csv('preprocessed_dataset.csv', index=False)

```
df.to_csv('cpcb_dly_aq_tamil_nadu-2014.csv', index=False)
```

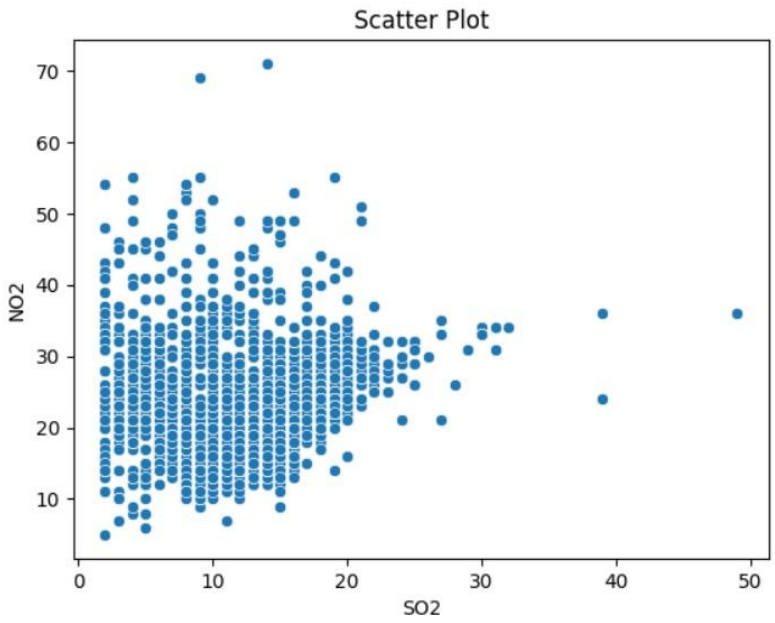
These steps provide a general guideline for loading and preprocessing a dataset. The specifics may vary depending on your dataset, project goals, and data quality

Program & Output:

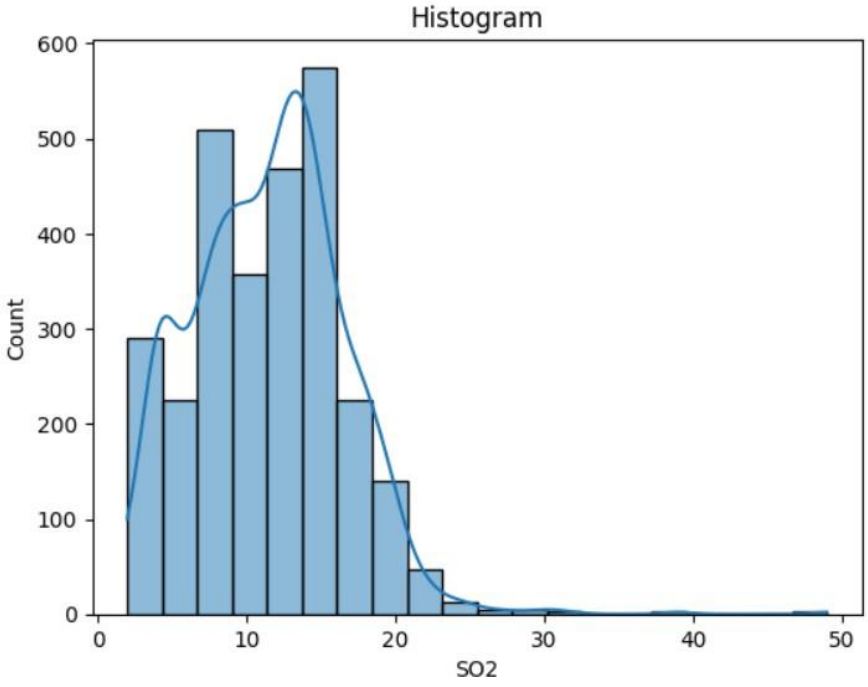
```
[1]  import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```
[2]  file_path = "/content/cpcb_dly_aq_tamil_nadu-2014.csv"
     df = pd.read_csv(file_path)
```

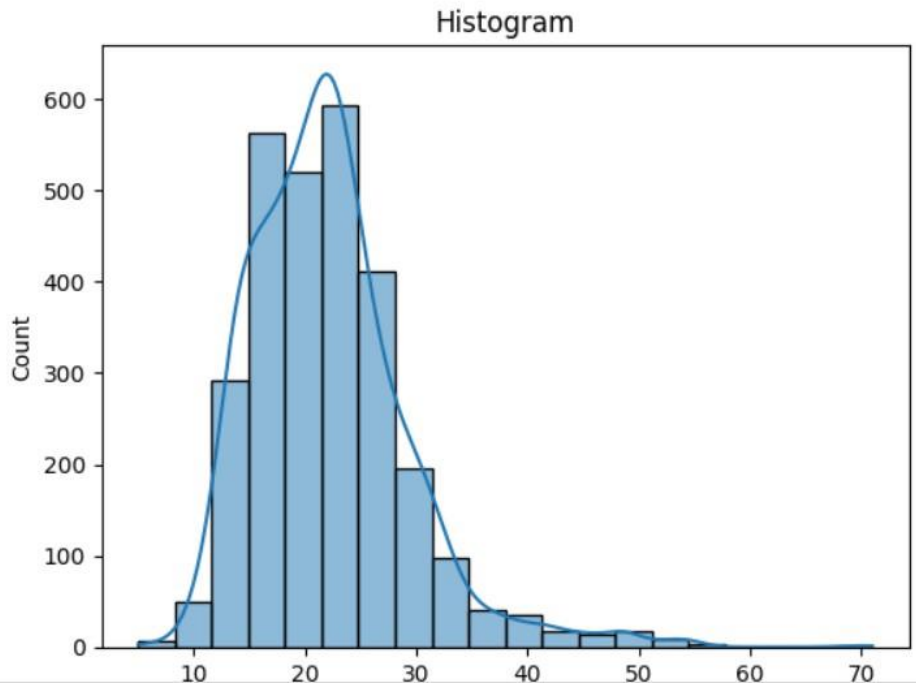
```
[3]  sns.scatterplot(data=df, x='SO2', y='NO2')
     plt.title('Scatter Plot')
     plt.show()
```



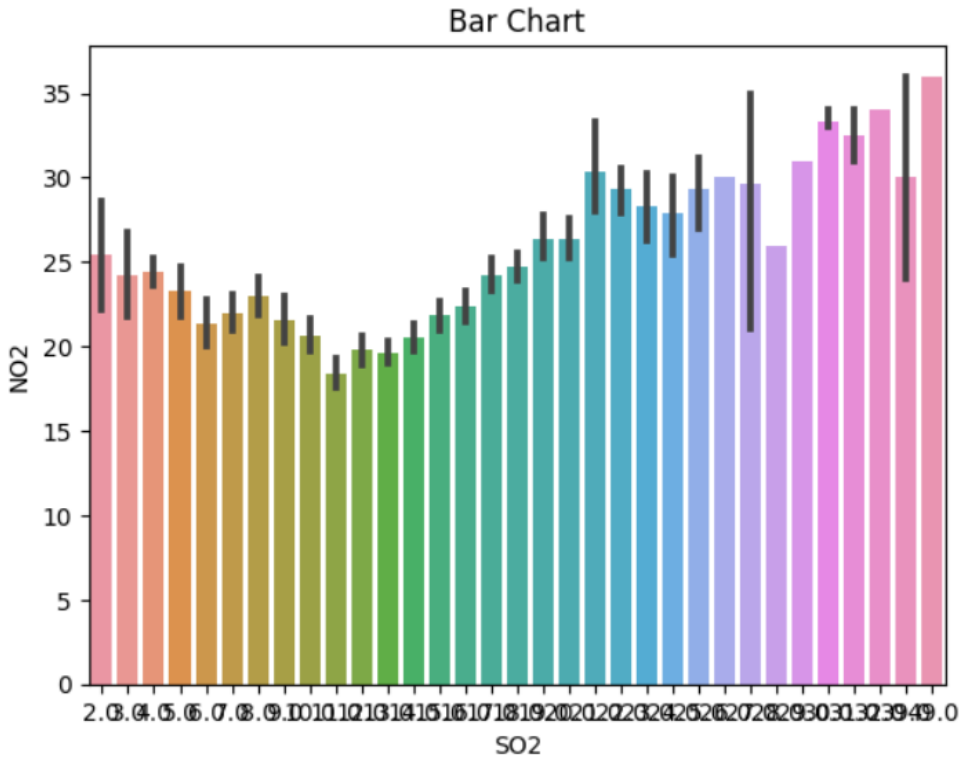

```
[4]  sns.histplot(data=df, x='SO2', bins=20, kde=True)
     plt.title('Histogram')
     plt.show()
```



```
[5]  sns.histplot(data=df, x='NO2', bins=20, kde=True)
     plt.title('Histogram')
     plt.show()
```



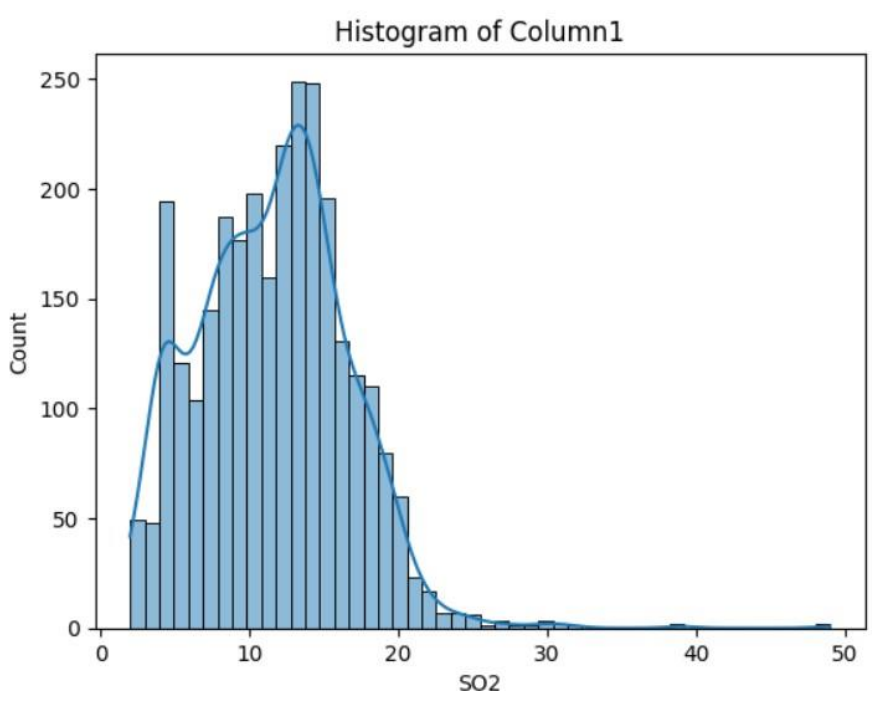

```
sns.barplot(data=df, x='SO2', y='NO2')
plt.title('Bar Chart')
plt.show()
```



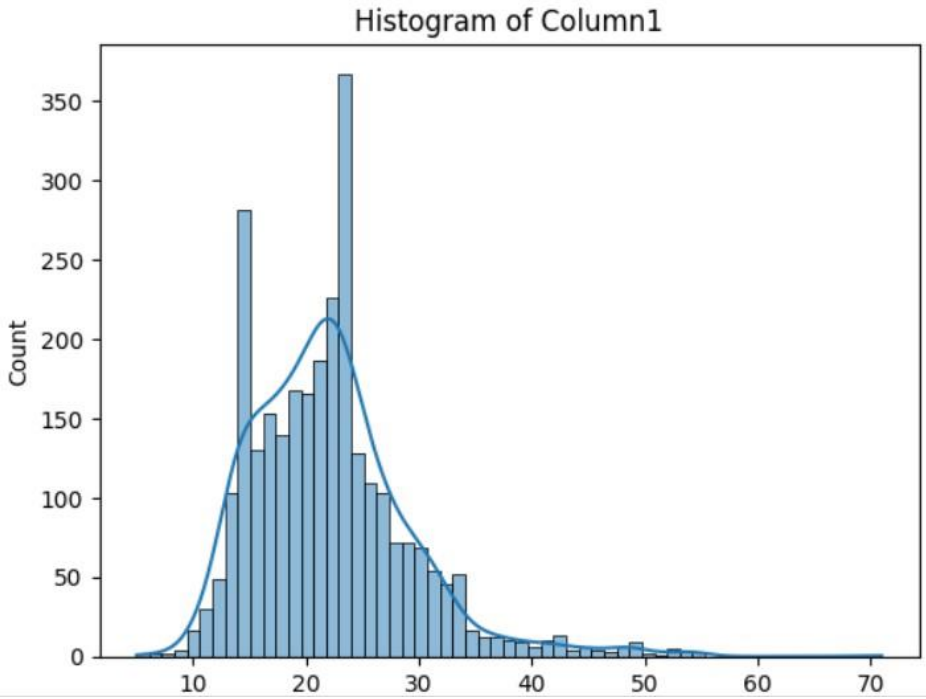
```
grouped = df.groupby(['State', 'City/Town/Village/Area', ])
averages = grouped[['SO2', 'NO2', 'RSPM/PM10']].mean()
averages = averages.reset_index()
print(averages)
```

```
        State City/Town/Village/Area        SO2        NO2  RSPM/PM10
0  Tamil Nadu                Chennai  13.014042  22.088442  58.998000
1  Tamil Nadu             Coimbatore   4.541096  25.325342  49.217241
2  Tamil Nadu              Cuddalore   8.965986  19.710884  61.881757
3  Tamil Nadu                Madurai  13.319728  25.768707  45.724490
4  Tamil Nadu                 Mettur   8.429268  23.185366  52.721951
5  Tamil Nadu                  Salem   8.114504  28.664122  62.954198
6  Tamil Nadu             Thoothukudi  12.989691  18.512027  83.458904
7  Tamil Nadu                 Trichy  15.293956  18.695055  85.054496
```

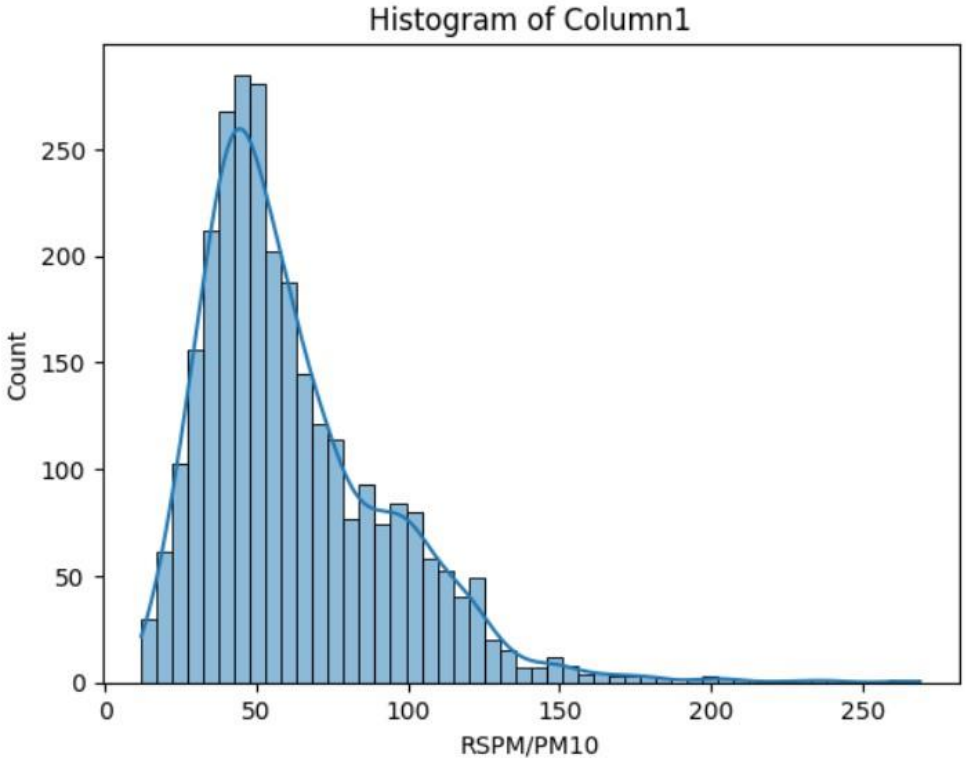
```
sns.histplot(data=df, x='SO2', kde=True)
plt.title('Histogram of Column1')
plt.show()
```



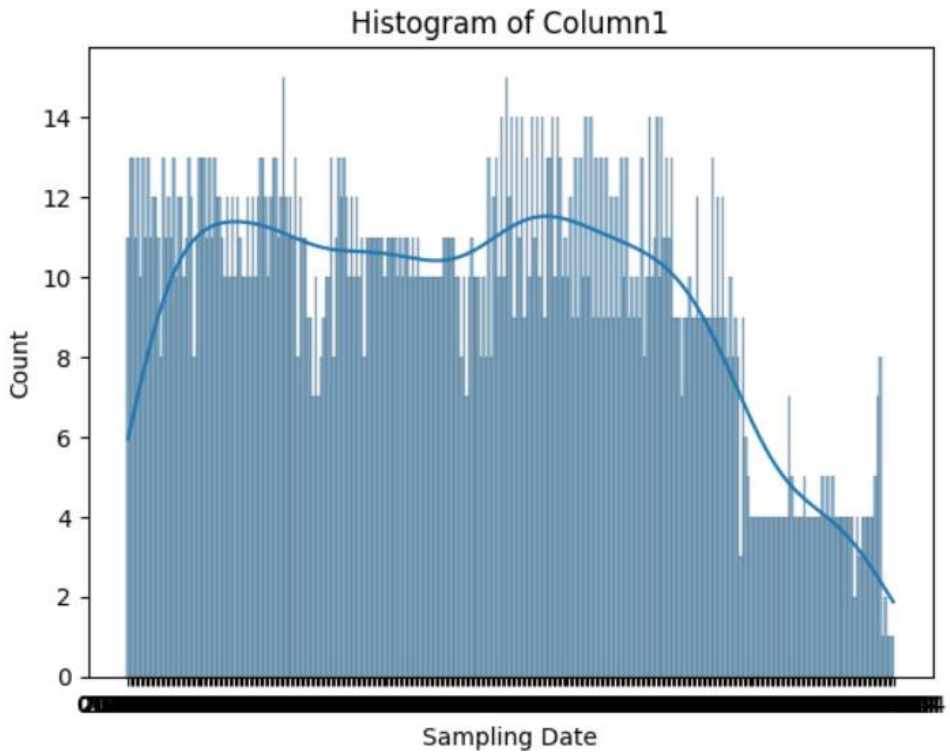
```
[10] sns.histplot(data=df, x='NO2', kde=True)
     plt.title('Histogram of Column1')
     plt.show()
```



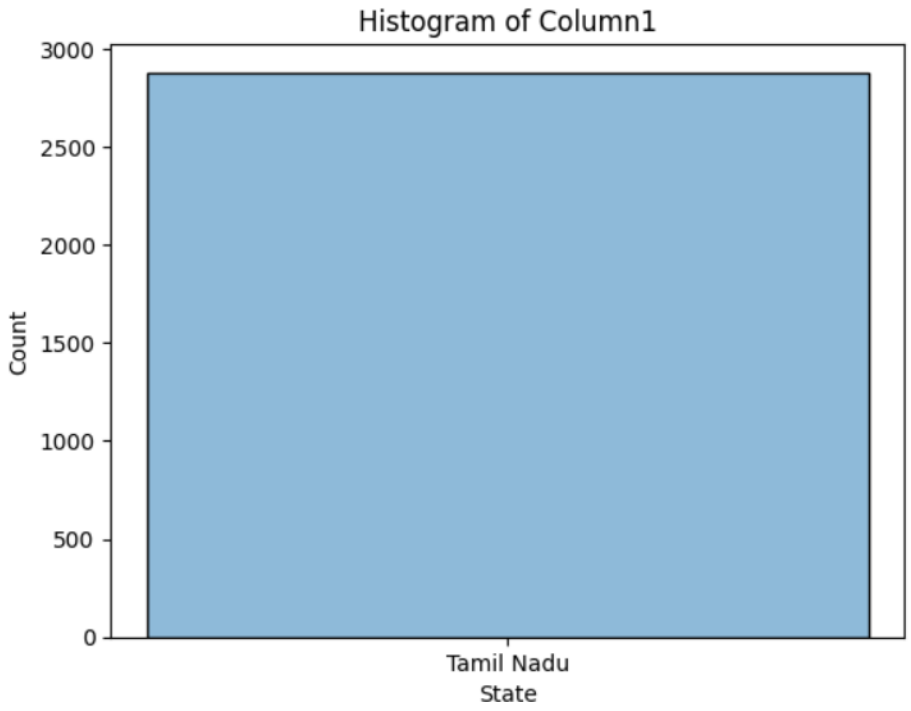

```
sns.histplot(data=df, x='RSPM/PM10', kde=True)
plt.title('Histogram of Column1')
plt.show()
```



```
sns.histplot(data=df, x='Sampling Date', kde=True)
plt.title('Histogram of Column1')
plt.show()
```

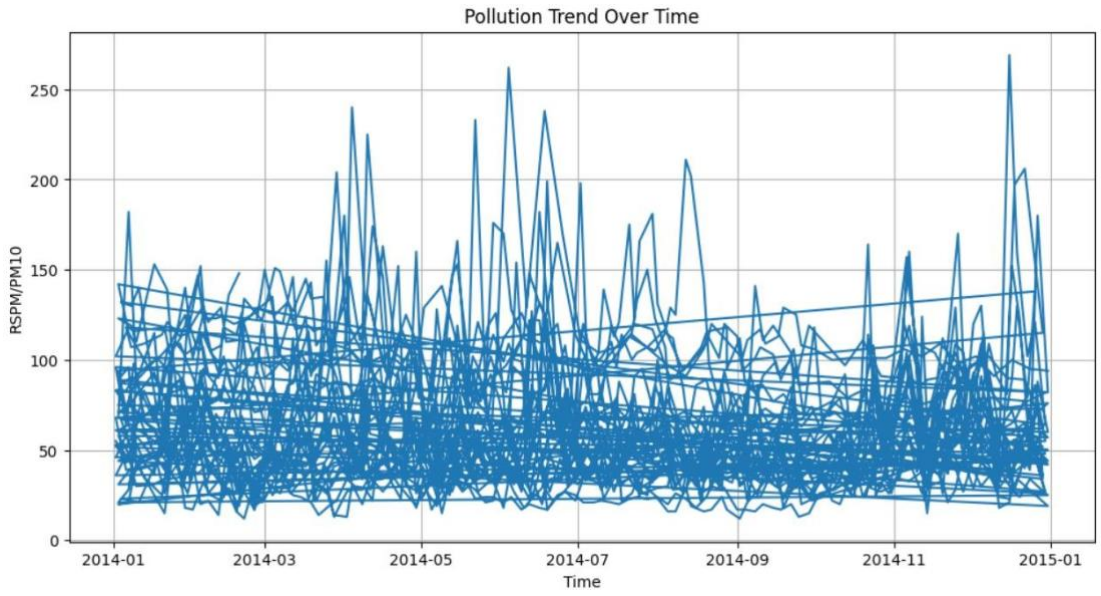



```
sns.histplot(data=df, x='State', kde=True)
plt.title('Histogram of Column1')
plt.show()
```



```
[15] df['Sampling Date'] = pd.to_datetime(df['Sampling Date'])
```

```
plt.figure(figsize=(12, 6))
plt.plot(df['Sampling Date'], df['RSPM/PM10'])
plt.title('Pollution Trend Over Time')
plt.xlabel('Time')
plt.ylabel('RSPM/PM10')
plt.grid()
plt.show()
```



In this we analysis the given sampling data's by comparing the values of RSPM, SO2, NO2 in the given area and also create a graph to represent the given data by using program.

CONCLUSION:

In conclusion, an DATA ANALYTICS-based air pollution monitoring system is a revolutionary solution that can provide accurate and real-time data about the air quality in a particular area.It can help identify the sources of pollution and take necessary measures to reduce it, protecting the environment and human health.