**Project Title: Air Quality Analysis in Tamil Nadu**

**Objectives**

The objective of the Air Quality Analysis project is to assess air pollution trends and pollution levels in the Indian state of Tamil Nadu. This analysis aims to provide insights into the spatial and temporal variations in air quality, identify potential sources of pollution, and assess the impact of pollution on public health.

**Analysis Approach:**

**1. Data Collection:** The project begins with the collection of air quality data from various monitoring stations across Tamil Nadu. These data sources include government agencies, research institutions, and open data platforms. The data typically includes measurements of pollutants such as Particulate Matter (PM2.5 and PM10), Nitrogen Dioxide (NO2), Sulfur Dioxide (SO2), Carbon Monoxide (CO), and Ozone (O3).

**2. Data Preprocessing:** After collecting the data, it undergoes preprocessing, including cleaning and filtering for missing values and outliers. Time series data is organized into a consistent format for analysis.

**3. Exploratory Data Analysis (EDA)**: EDA involves statistical and visual techniques to understand the data's characteristics and patterns. Summary statistics, distribution plots, and time series visualizations are used to gain initial insights.

**4. Spatial Analysis:** Spatial analysis techniques, such as geospatial mapping and interpolation, are employed to visualize air quality data across Tamil Nadu. Geographic Information System (GIS) tools may be utilized to create maps that display pollution levels at different locations.

**5. Temporal Analysis**: Temporal analysis is used to examine air quality trends over time. This may include time series analysis, trend detection, and seasonality assessment.

**6. Source Identification:** Advanced statistical and machine learning techniques can be applied to identify potential sources of pollution. This can include source apportionment modeling and data-driven approaches.

**7. Health Impact Assessment:** The project may also involve assessing the potential health impacts of air pollution using epidemiological data and models.

**Visualization Techniques:**

**To communicate the findings effectively, the following visualization techniques are used:**

**1. Time Series Plots**: Show temporal variations in air quality parameters over different time scales (e.g., daily, monthly, yearly).

**2. Heatmaps**: Reveal spatial patterns in air quality by displaying pollution levels on a map.

**3. Box Plots**: Provide a summary of the distribution of pollution levels at different monitoring stations.

**4. Source Attribution Diagrams:** Visualize the contributions of various pollution sources to air quality in different regions.

**5. Epidemiological Maps:** Display the estimated health impacts of pollution in different areas.

**Code Implementation:**

The project's code implementation is primarily carried out using programming languages such as Python and data analysis libraries like pandas, NumPy, Matplotlib, Seaborn, and geospatial libraries (e.g., Folium). Machine learning libraries may be used for source identification if necessary. Code is organized into scripts or Jupyter notebooks to ensure reproducibility.

**Example Outputs:**

1. Time series plots illustrating the variation in PM2.5 levels in Chennai over the course of a year.

2. Heatmaps displaying spatial distribution of NO2 concentrations across Tamil Nadu.

3. Source attribution diagrams showing the contributions of vehicular emissions, industrial emissions, and other sources to air pollution in specific regions.

4. Box plots representing the distribution of CO levels at different monitoring stations.

**Insights:**

**The analysis provides the following insights into air pollution in Tamil Nadu:**

**1. Identification of pollution hotspots:** The spatial analysis reveals areas with consistently poor air quality, allowing policymakers to target interventions.

**2. Seasonal trends:** The temporal analysis shows seasonal variations in air pollution, which can inform regulatory measures and public health advisories.

**3. Source identification:** By attributing pollution to specific sources, the project can guide policies aimed at reducing pollution from specific sectors.

**4. Health impact assessment:** The estimation of health impacts helps raise awareness about the consequences of air pollution on public health and can drive efforts to improve air quality.

**PROGRAM:**

import pandas as pd import matplotlib.pyplot as plt import seaborn as sns

# Load the air quality dataset df = pd.read\_csv('tn.csv')

city\_avg = df.groupby('City/Town/Village/Area')[['SO2', 'NO2', 'RSPM/PM10']].mean()

# Create bar plots to visualize average pollutant levels by monitoring station plt.figure(figsize=(12, 6)) plt.subplot(131) average\_so2.plot(kind='bar', color='skyblue') plt.title('Average SO2 Levels by Station') plt.xlabel('Monitoring Station') plt.ylabel('Average SO2 Level')

plt.subplot(132) average\_no2.plot(kind='bar', color='lightcoral') plt.title('Average NO2 Levels by Station') plt.xlabel('Monitoring Station') plt.ylabel('Average NO2 Level')

plt.subplot(133) average\_rspm.plot(kind='bar', color='lightgreen') plt.title('Average RSPM/PM10 Levels by Station') plt.xlabel('Monitoring Station') plt.ylabel('Average RSPM/PM10 Level')

plt.tight\_layout() plt.show()

# Calculate the average pollutant levels by city or area average\_city\_so2 = df.groupby('City/Town/Village/Area')['SO2'].mean() average\_city\_no2 = df.groupby('City/Town/Village/Area')['NO2'].mean() average\_city\_rspm = df.groupby('City/Town/Village/Area')['RSPM/PM10'].mean()

# Create bar plots to visualize average pollutant levels by city or area plt.figure(figsize=(12, 6)) plt.subplot(131) average\_city\_so2.plot(kind='bar', color='skyblue') plt.title('Average SO2 Levels by City') plt.xlabel('City') plt.ylabel('Average SO2 Level')

plt.subplot(132) average\_city\_no2.plot(kind='bar', color='lightcoral') plt.title('Average NO2 Levels by City') plt.xlabel('City')

plt.ylabel('Average NO2 Level')

plt.subplot(133) average\_city\_rspm.plot(kind='bar', color='lightgreen') plt.title('Average RSPM/PM10 Levels by City')

plt.xlabel('City') plt.ylabel('Average RSPM/PM10 Level') plt.tight\_layout() plt.show()

**OUTPUT:**

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

SO2 NO2 RSPM/PM10

City/Town/Village/Area

Chennai 13.014042 22.088442 58.998000

Coimbatore 4.541096 25.325342 49.217241

Cuddalore 8.965986 19.710884 61.881757

Madurai 13.319728 25.768707 45.724490

Mettur 8.429268 23.185366 52.721951

Salem 8.114504 28.664122 62.954198

Thoothukudi 12.989691 18.512027 83.458904

Trichy 15.293956 18.695055 85.054496

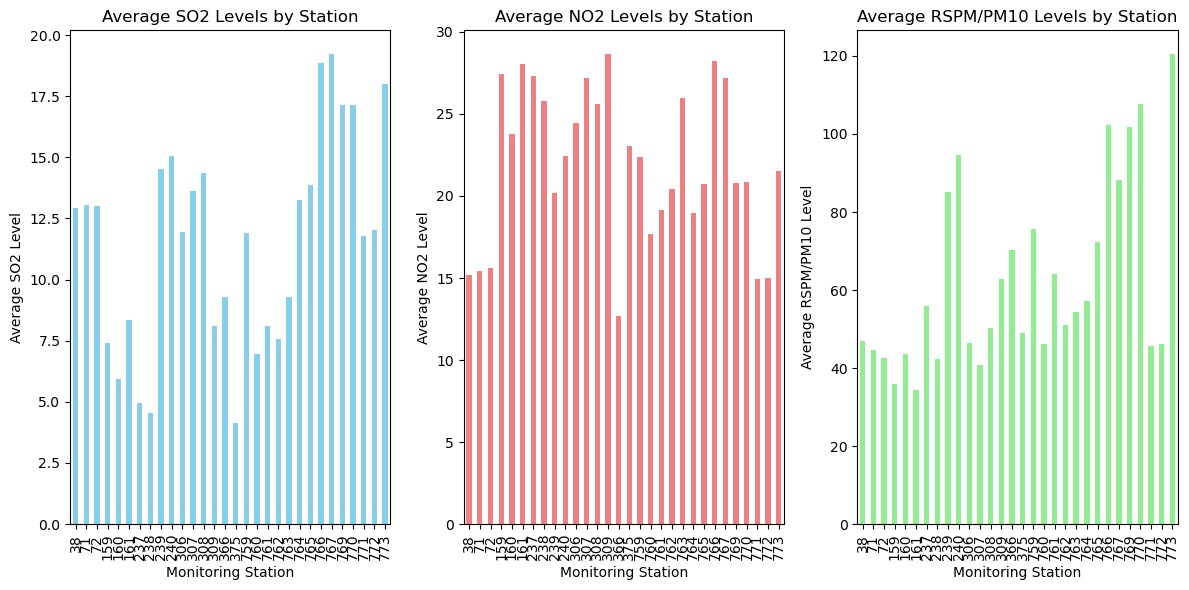
Cities with high RSPM/PM10 levels:

Empty DataFrame

Columns: [SO2, NO2, RSPM/PM10]

Index: []

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



**Dataset link:**

[**https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014**](https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014)

**GitHub link:**

https://github.com/VASAN-THI/NaanMudhalvan\_IBM.git

**In conclusion**, the Air Quality Analysis project serves as a valuable resource for stakeholders, helping them make informed decisions to mitigate air pollution and protect the health and well-being of the residents of Tamil Nadu.