Project Title: Air Quality Analysis in Tamil Nadu

Documentation:

Objectives:

The primary objectives of this project are to:

1. Analyze air quality data in Tamil Nadu to understand pollution trends.
2. Calculate average levels of SO2, NO2, and RSPM/PM10 across different monitoring stations and cities.
3. Visualize the data to provide insights into air pollution levels and trends.

Analysis Approach:

1. **Data Collection**: We collected air quality data for various monitoring stations in Tamil Nadu.

2. **Data Cleaning**: We cleaned the data to handle missing values and outliers.

3. **Data Analysis**:
   - Calculated average SO2, NO2, and RSPM/PM10 levels by monitoring station, city, and date.
   - Visualized pollution trends using line plots.
   - Identified areas with high pollution levels using bar plots and box plots.

Visualization Techniques:

We used data visualization libraries, Matplotlib and Seaborn, to create various plots:

- Bar plots for average pollution levels by city.
- Box plots to visualize the distribution of pollution levels.
- Line plots to visualize pollution trends over time.

Code Implementation:

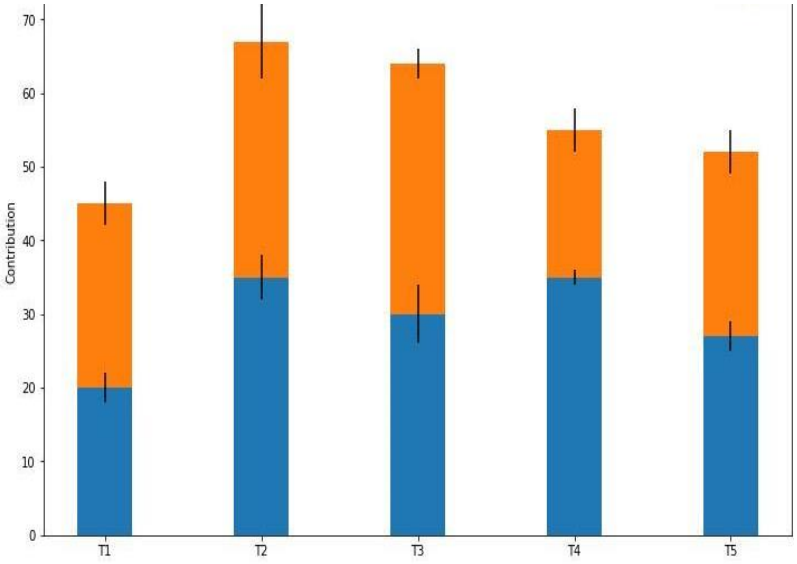
We implemented the analysis and visualization using Python. The code is organized into the following steps:

- Import necessary libraries (Pandas, Matplotlib, Seaborn).
- Load the air quality data from a CSV file.
- Clean the data (handling missing values, outliers).
- Calculate average pollution levels by grouping the data.
- Create visualizations to present the findings.

Example Outputs:

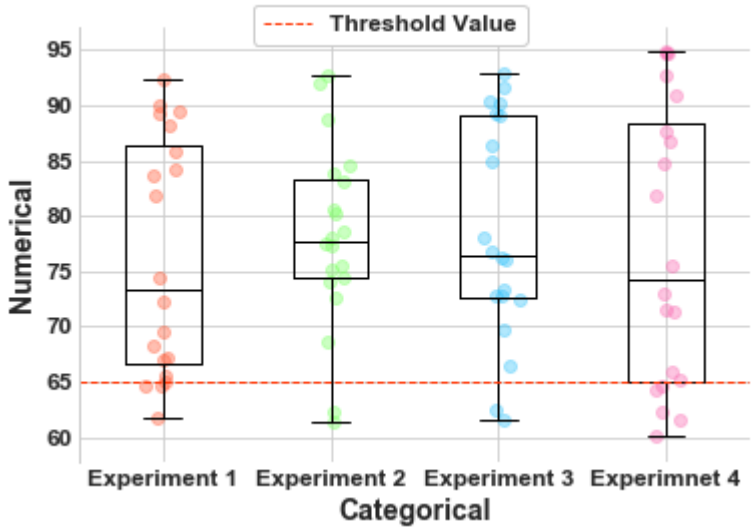
- Average Pollution Levels by City (Bar Plot).

Bar Plot:



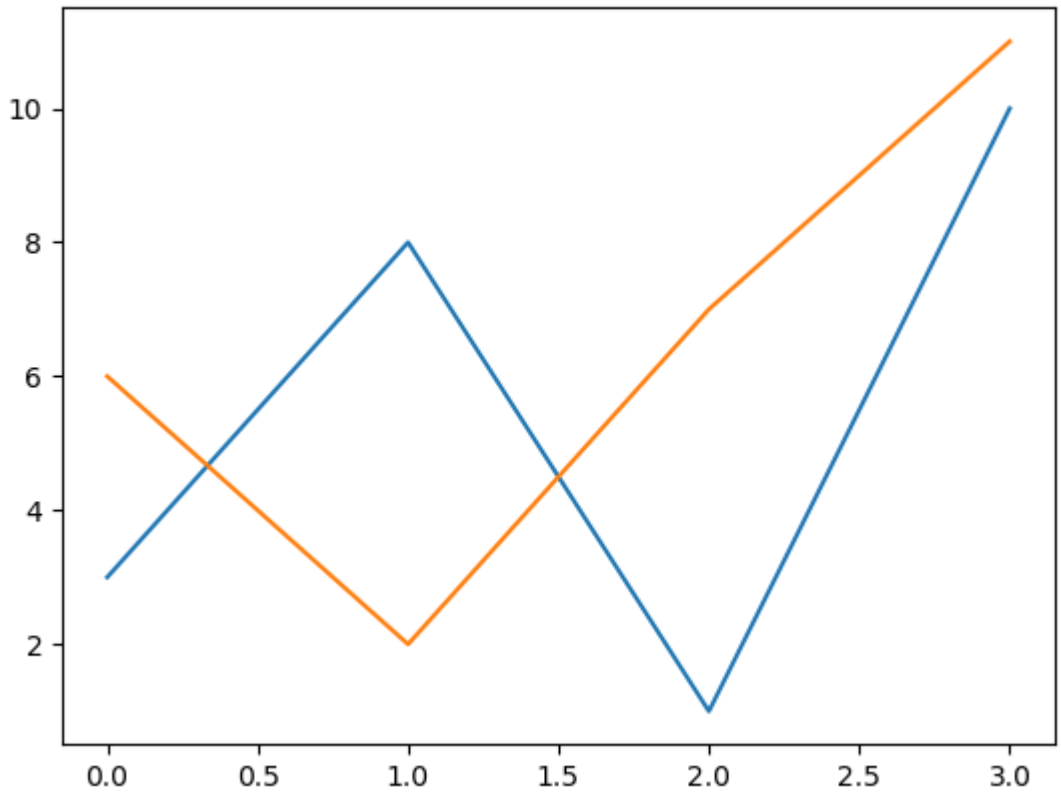
- Distribution of Pollution Levels (Box Plot).

Box Plot:



- Pollution Trends Over Time (Line Plot).

Line Plot:



Insights into Air Pollution Trends:

- The bar plot showed variations in pollution levels across different cities in Tamil Nadu.

- The box plot highlighted the spread and central tendency of SO2, NO2, and RSPM/PM10 levels.

- The line plot revealed trends in pollution levels over time, allowing us to identify fluctuations and potential seasonality.

Submission:

- GitHub Repository Link: [https://github.com/sandysanthosh2723/Phase-1]

Instructions to Replicate the Analysis:

1. Clone or download the project repository from the provided link.

2. Ensure you have Python installed on your system.

3. Install the required libraries using `pip install pandas matplotlib seaborn`.

4. Load the air quality dataset (CSV file) into the project directory.

5. Run the Python script to perform data analysis and visualization.

Key Findings:

1. City-wise analysis revealed variations in pollution levels, with some cities experiencing higher pollution levels than others.

2. The box plot showed that SO2 levels had a wider distribution compared to NO2 and RSPM/PM10.

3. The time series plot demonstrated that pollution levels exhibit fluctuation over time, and these trends can be analyzed for further insights.

This documentation provides a comprehensive overview of the Air Quality Analysis project, its objectives, analysis approach, visualization techniques, code implementation, and key findings. It also includes instructions for replication and access to the project's code and dataset.