Phase 3

Air Quality Analysis

Phase 3: Development Part 1

In this part we will:

- begin building your project by loading and preprocessing the dataset.
- Begin the analysis by loading and preprocessing the air quality dataset.
- Load the dataset using Python and data manipulation libraries (e.g., pandas)

Step 1: Download the Dataset:

- Access the provided link for Air Quality Analysis dataset.

Dataset Link: https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamilnadu-year-2014

- Download the dataset to local working directory or preferred location.

Step 2: Loading the Dataset:

Once you have the dataset downloaded, you can import and use the pandas library to load it into a Data Frame for further analysis. And also added with panda you can add some more data manipulation libraries to it.

Step 3: Exploratory Data Analysis (EDA):

EDA is a crucial step in understanding any dataset. For our "Air Quality Analysis" project, you can perform the following EDA tasks:

- Compute summary statistics to understand the distribution of air quality parameters.
- Create histograms, box plots, and scatter plots to visualize the distribution and relationships between variables.
- Check for missing data and decide on an appropriate strategy to handle it.
- Identify trends and patterns in air quality over time, and across locations within the dataset.

Step 4: Data Cleaning and Preprocessing

This step involves preparing the data for analysis. Tasks may include:

- Handling missing data by dropping, filling, or imputing values.
- Dealing with outliers if necessary.
- Formatting dates and times for time series analysis.
- Ensuring consistent data types.

- Handling any data quality issues identified during EDA.

Step 5: Preprocessing

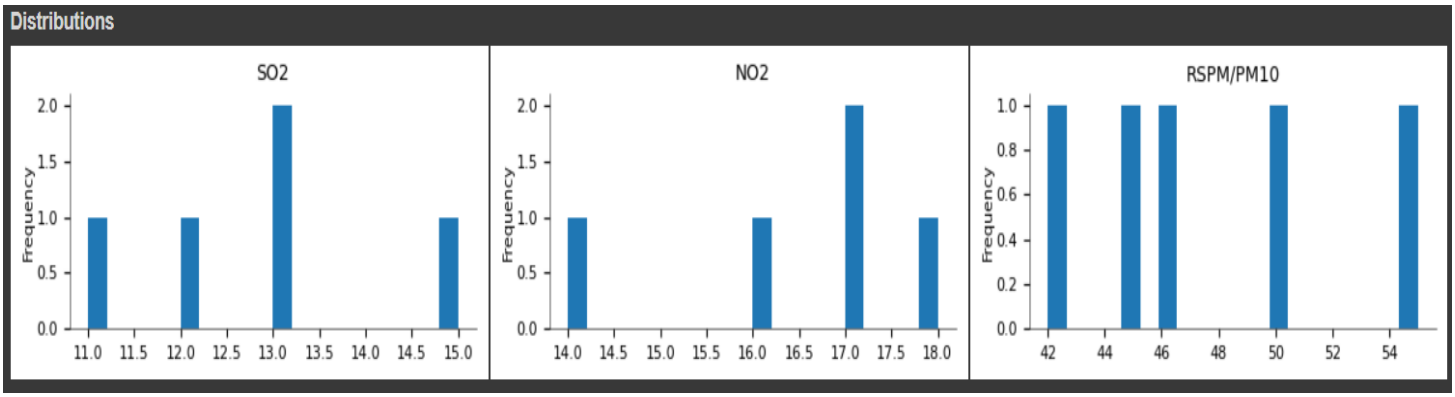
In this step, you can perform any additional preprocessing specific to your analysis objectives. For example, you may aggregate data to a daily or monthly level for trend analysis, or calculate averages across different monitoring stations.

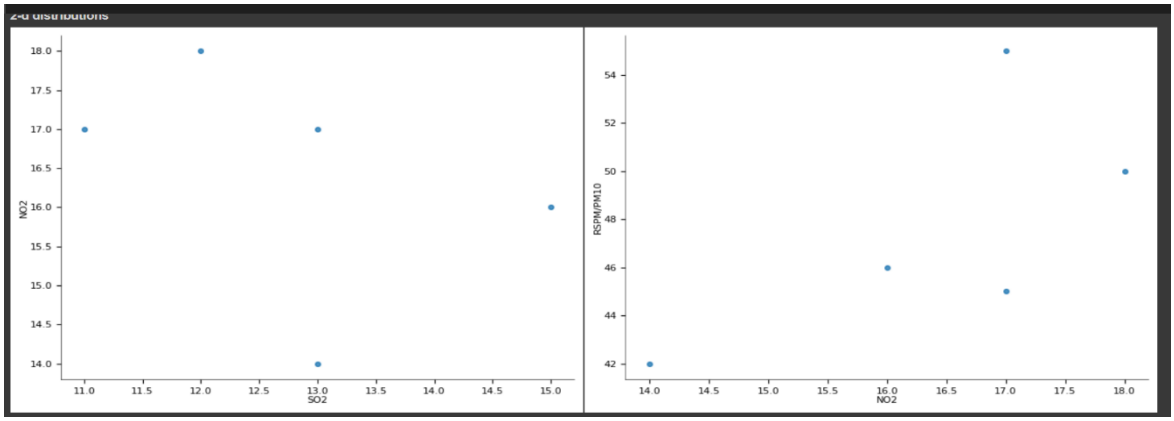
Step 6: Data Validation:

Before finalizing your analysis, validate the data to ensure its accuracy and reliability. Crosscheck data against known standards or external sources. Verify that your preprocessing and analysis steps have not introduced errors.

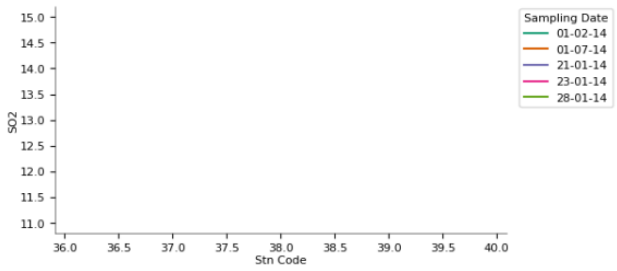
Step 7: Visualization

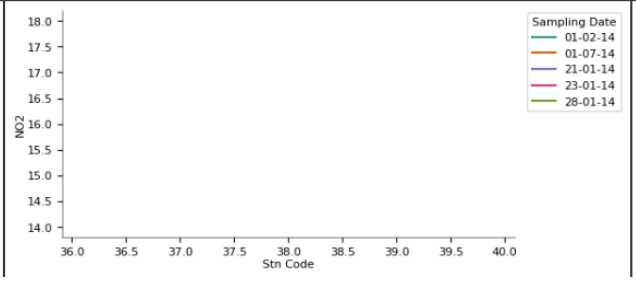
Visualizations are a key aspect of your analysis. Create various types of charts and plots to communicate your findings. For your "Air Quality Analysis," consider using line charts, bar charts, and geographic maps to visualize trends, comparisons, and spatial variations in air quality.

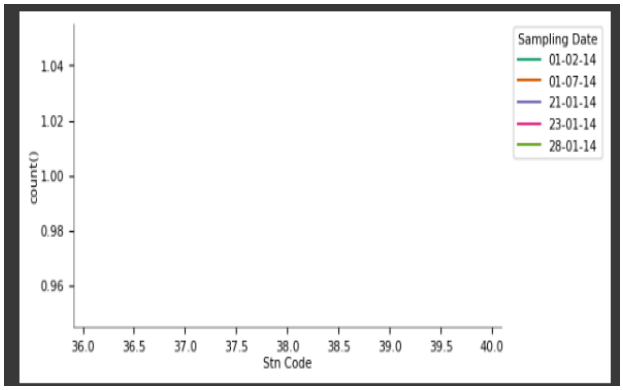


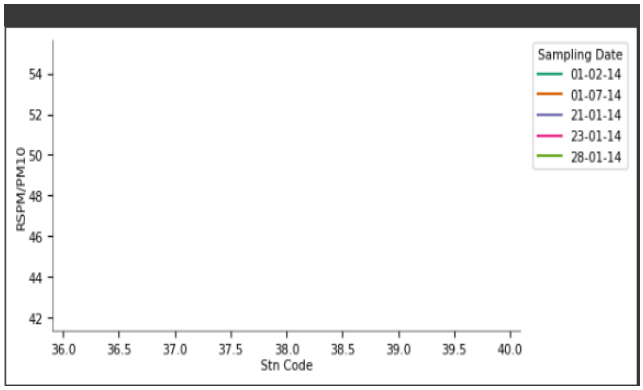


Time series

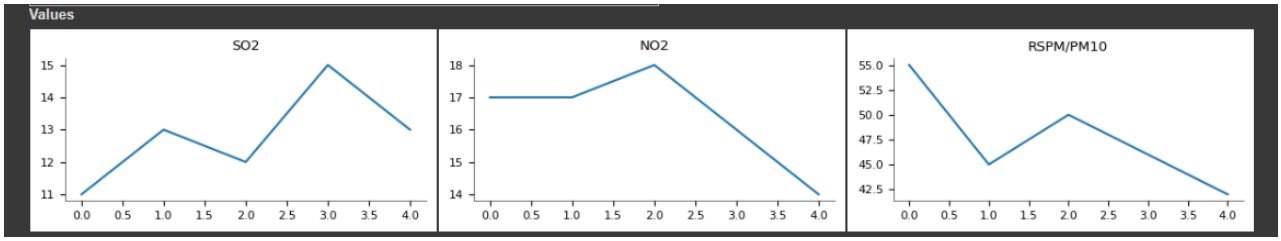


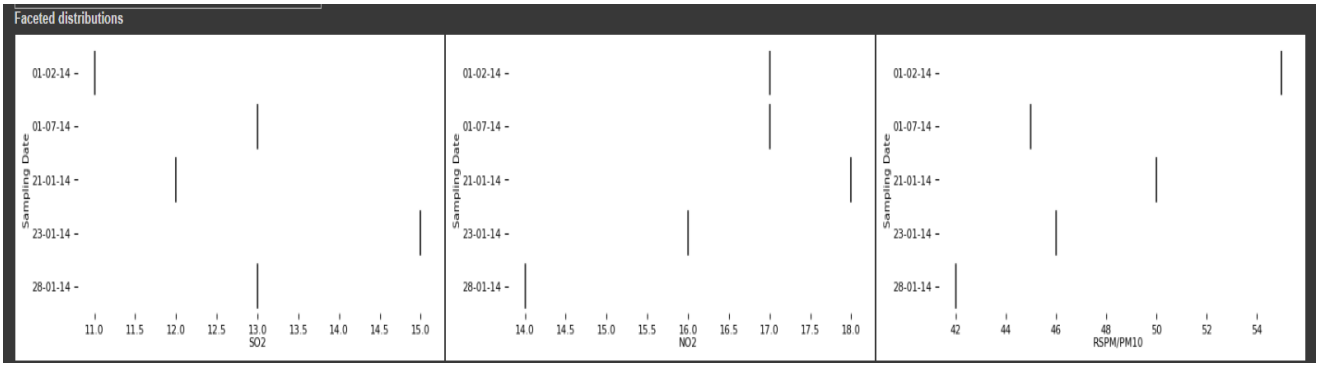




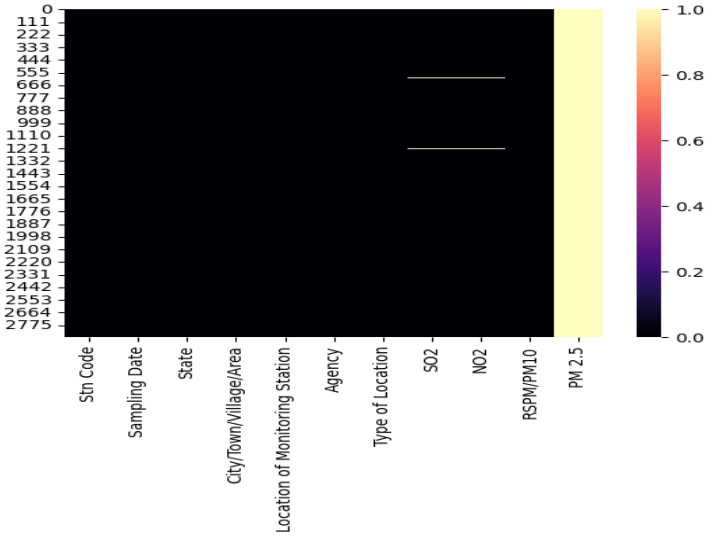


Values

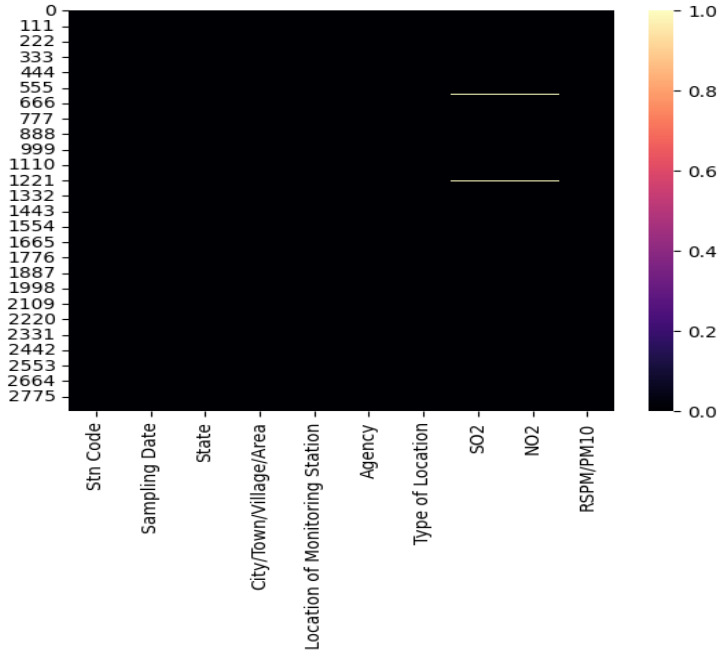




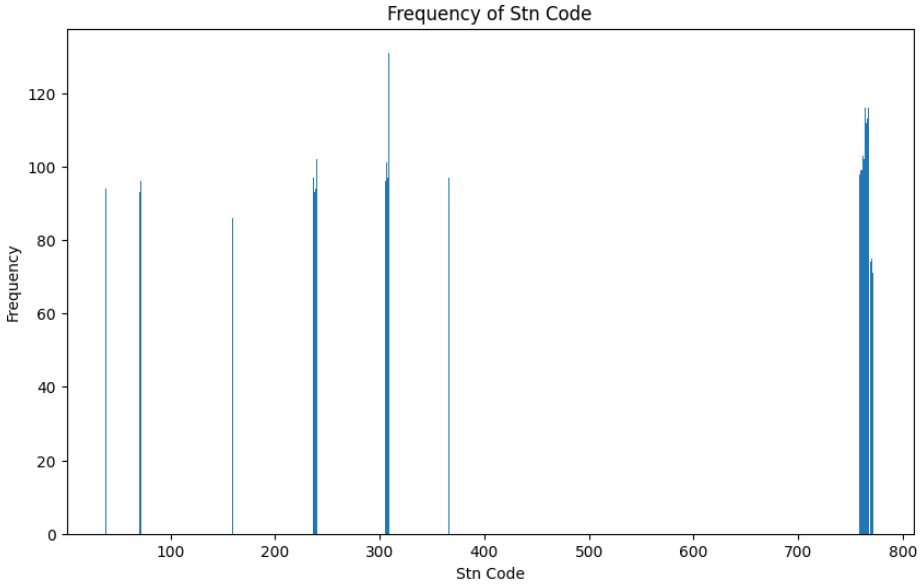
Null value detection using HeatMap:



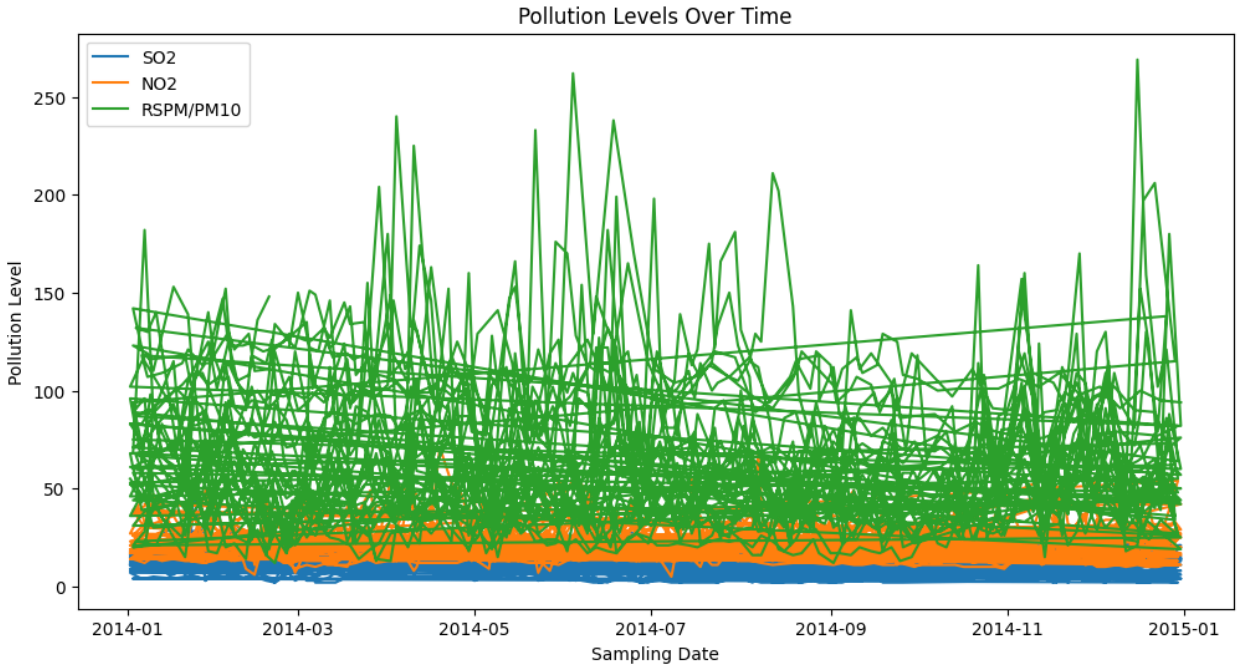
After removing null values:



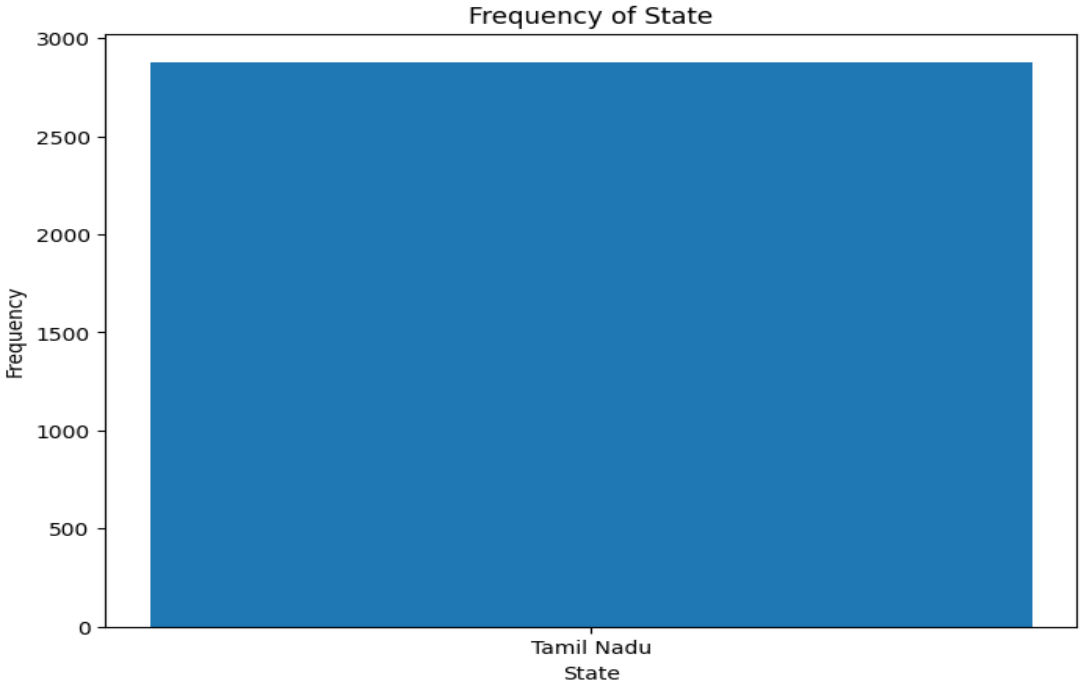
1.Bar Plot (Stn Code, State, Agency)



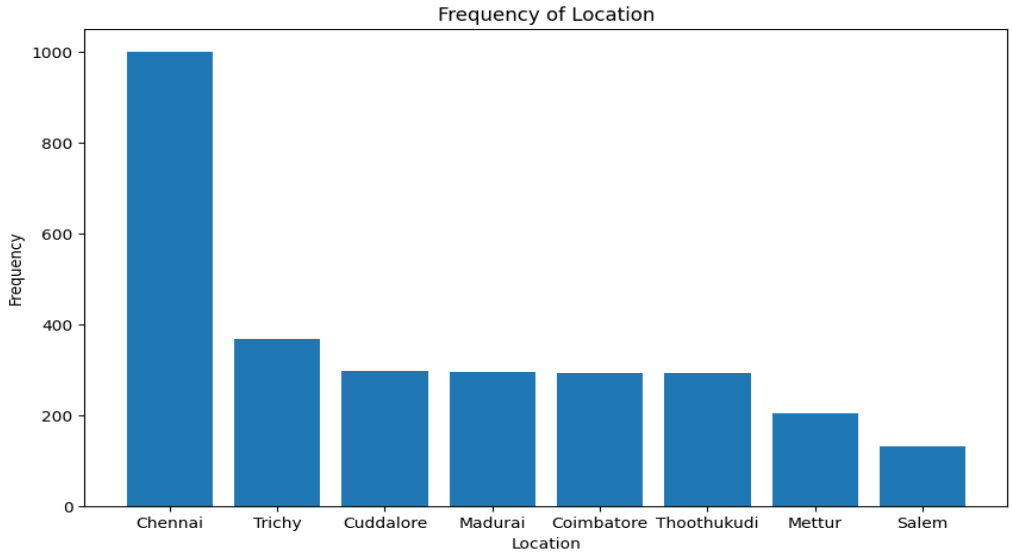
2. Time Series Plot (Sampling Date):



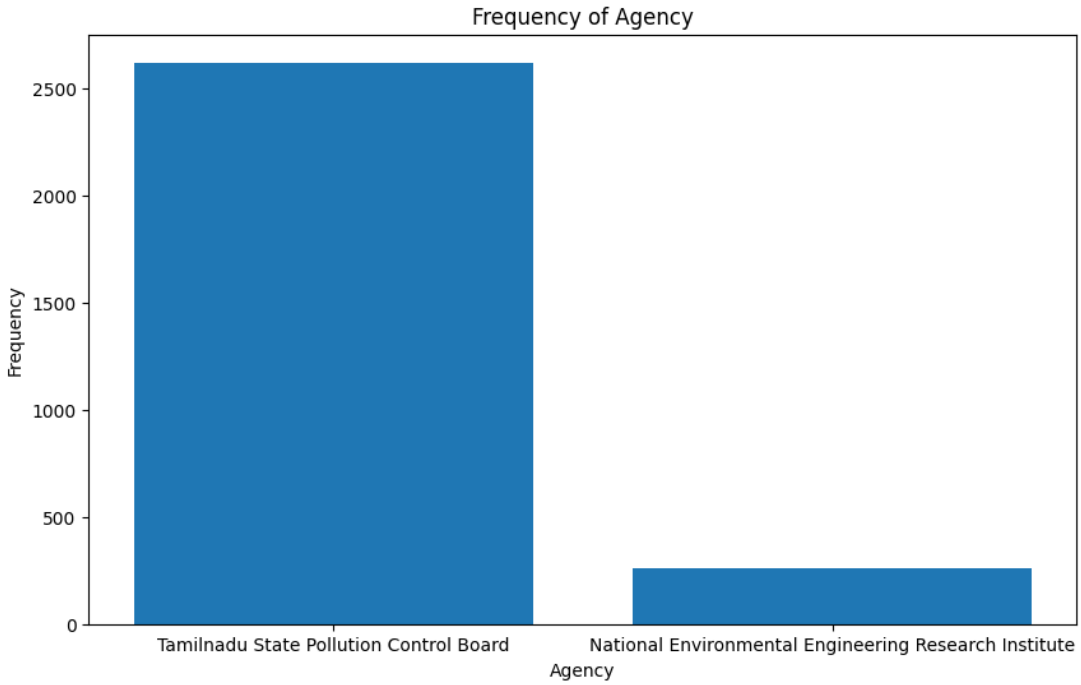
3. Stacked Bar Chart (Type of Location):



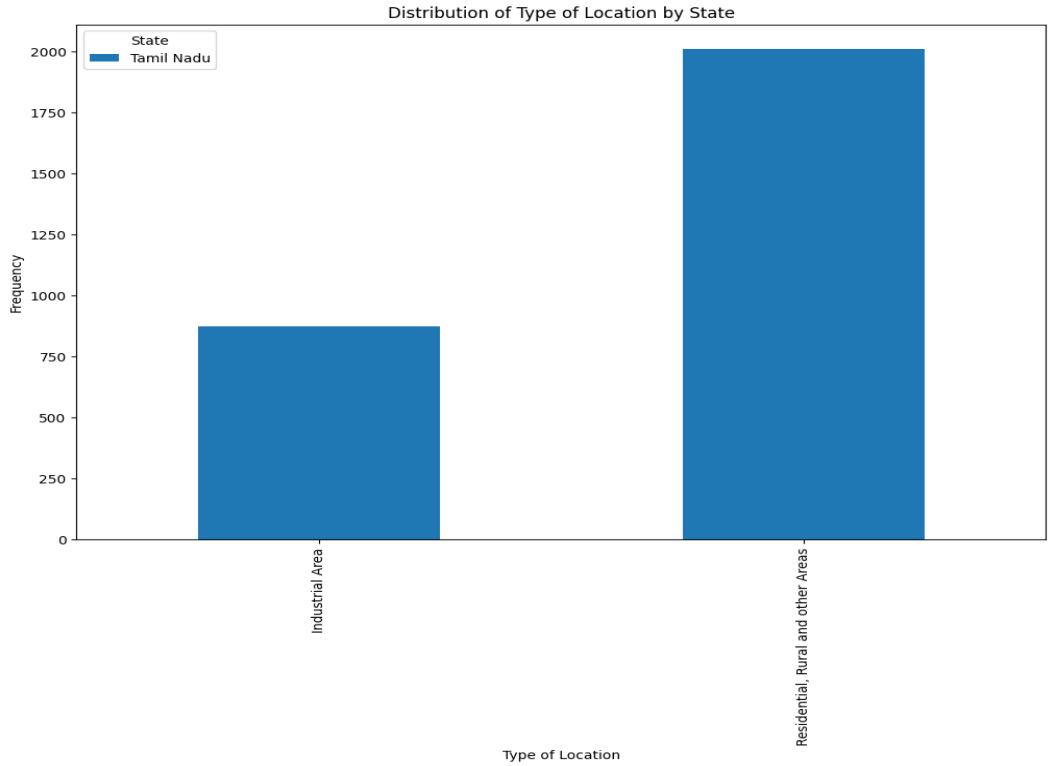
4. Bar Plot for 'City/Town/Village/Area:



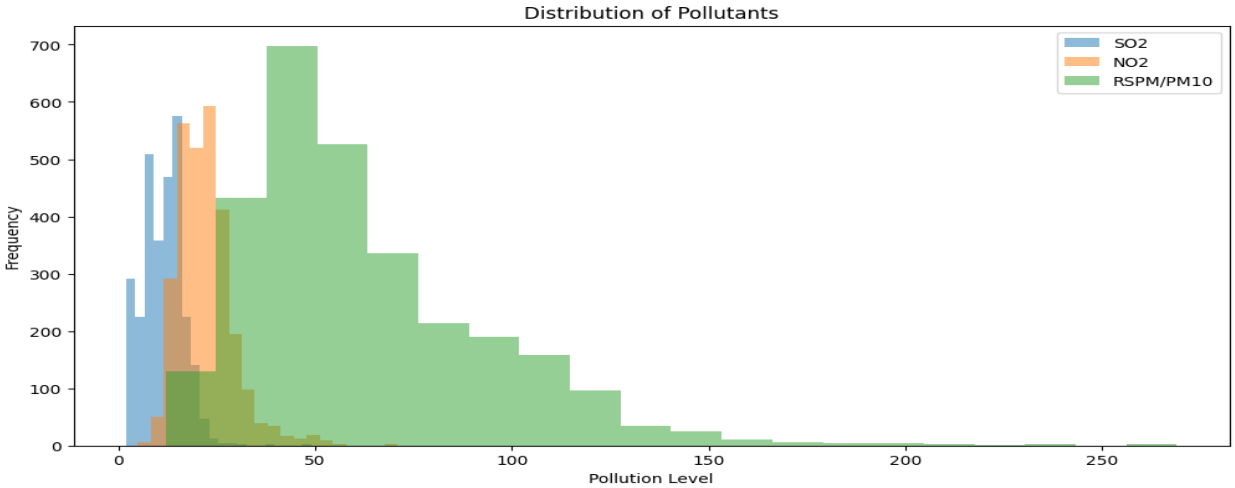
5. Bar Plot for 'Agency':



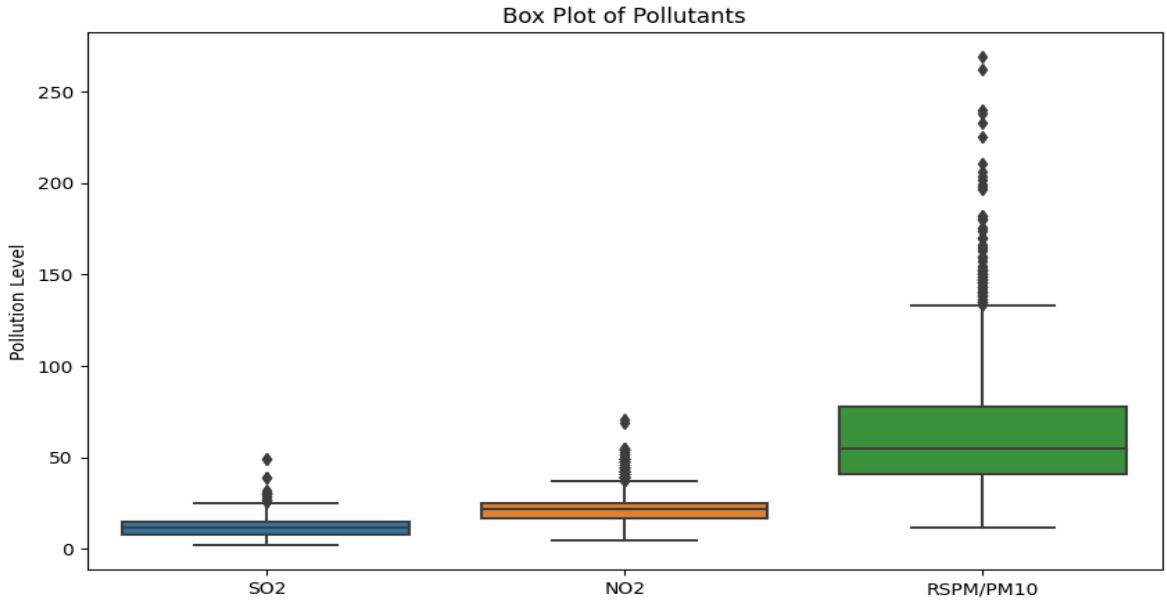
6. Stacked Bar Chart for 'Type of Location:



7. Histogram for 'SO2', 'NO2', 'RSPM/PM10':



8. Box Plot for 'SO2', 'NO2', 'RSPM/PM10':



Conclusion :

In this air quality analysis project, we initiated by downloading the dataset and subsequently loaded it into a pandas DataFrame and other data manipulation libraries for further analysis. During the exploratory data analysis (EDA) phase, we gained valuable insights into the dataset, including the distribution of air quality parameters, trends over time, missing data and potential data quality issues. With a defined analysis objective, our focus was to understand the dynamics of air pollutants, identify patterns, and uncover insights. Data cleaning and preprocessing steps were essential in ensuring data reliability, including handling missing values and formatting date columns. As a result, we transformed the dataset to facilitate meaningful analysis by loading the data and preprocessing it under phase 3 of 'Air Quality Analysis in Tamilnadu'.