**Naan Mudhalvan Project**

**Air Quality Analysis in Tamil Nadu**

**Phase 3**

**Team Members:**

~ Hariharan I S - 2021115039

~ GokulRaj S - 2021115038

~ Gokul V - 2021115037

~ Harshith A N - 2021115040

~ Gopi R - 2021115315

Team Members:

~Samiukktha Dharmalingam - 2021115089

~Santhya - 2021115093

~ Saravanasudharsan S - 2021115095

~Sathyadharini S - 2021115097

~ Vairaarasu - 2021115301

Team Members:

~Samiukktha Dharmalingam - 2021115089

~Santhya - 2021115093

~ Saravanasudharsan S - 2021115095

~Sathyadharini S - 2021115097

~ Vairaarasu - 2021115301

Team Members:

~Samiukktha Dharmalingam - 2021115089

~Santhya - 2021115093

~ Saravanasudharsan S - 2021115095

~Sathyadharini S - 2021115097

~ Vairaarasu - 2021115301

Team Members:

~Samiukktha Dharmalingam - 2021115089

~Santhya - 2021115093

~ Saravanasudharsan S - 2021115095

~Sathyadharini S - 2021115097

~ Vairaarasu - 2021115301

Team Members:

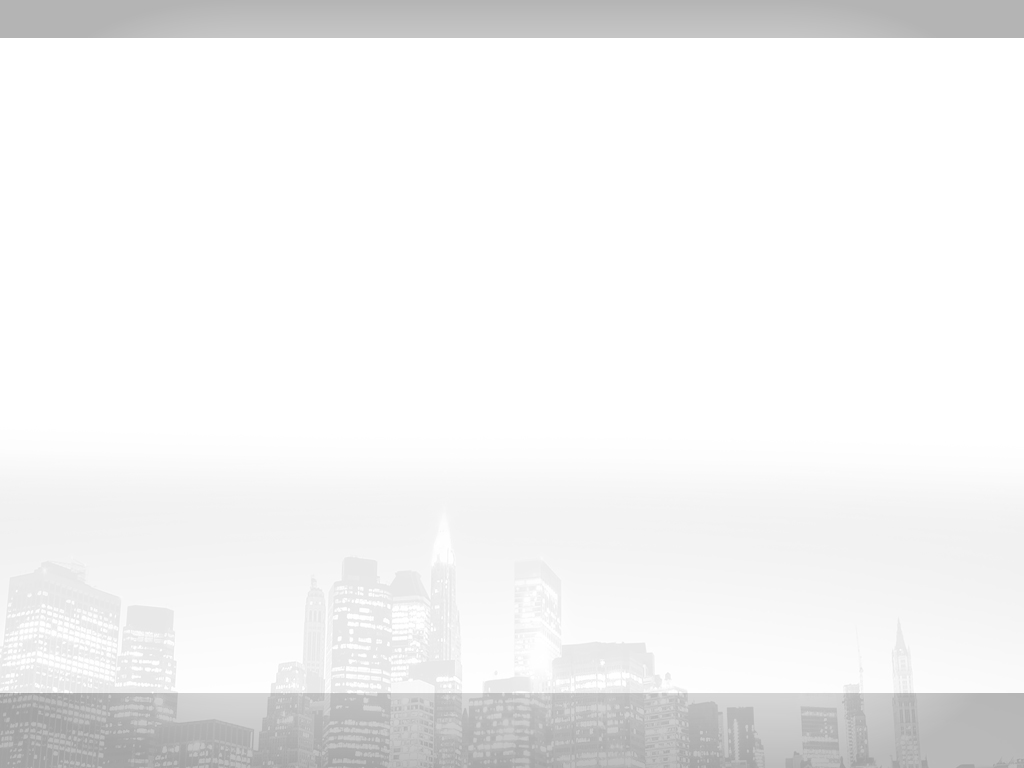
~Samiukktha Dharmalingam - 2021115089

~Santhya - 2021115093

~ Saravanasudharsan S - 2021115095

~Sathyadharini S - 2021115097

~ Vairaarasu - 2021115301



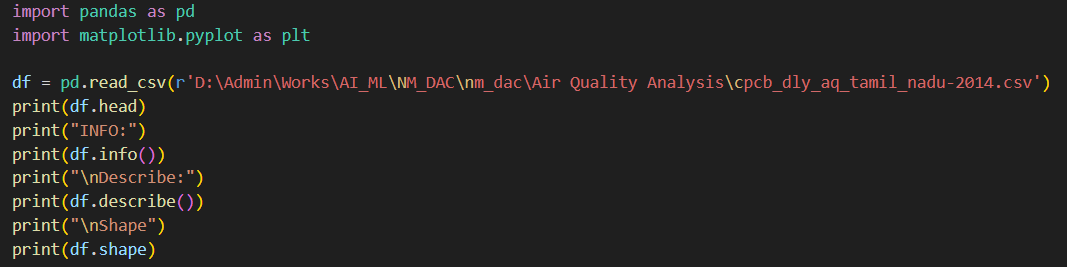
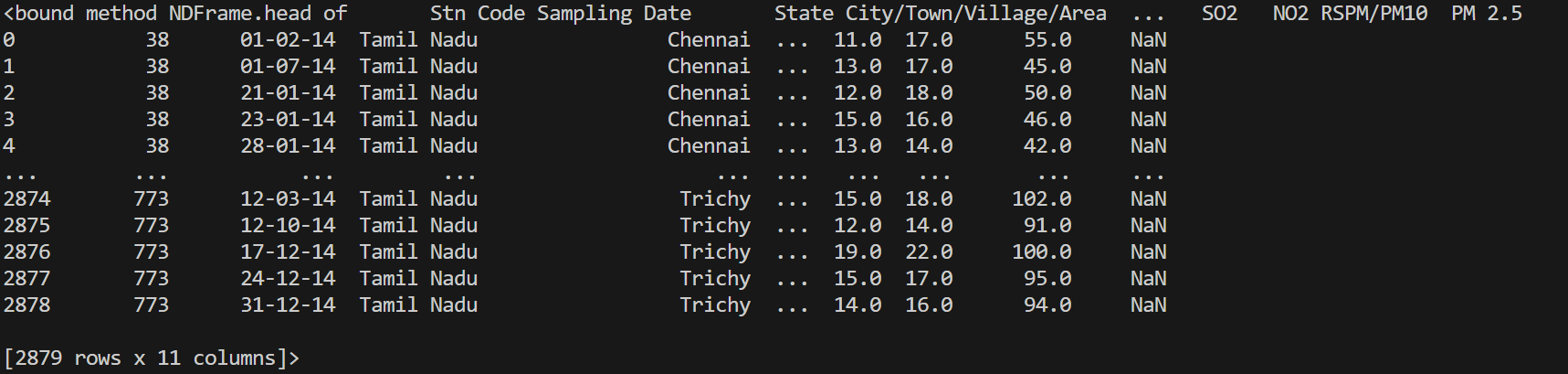
**Phase 3: Development Part I**

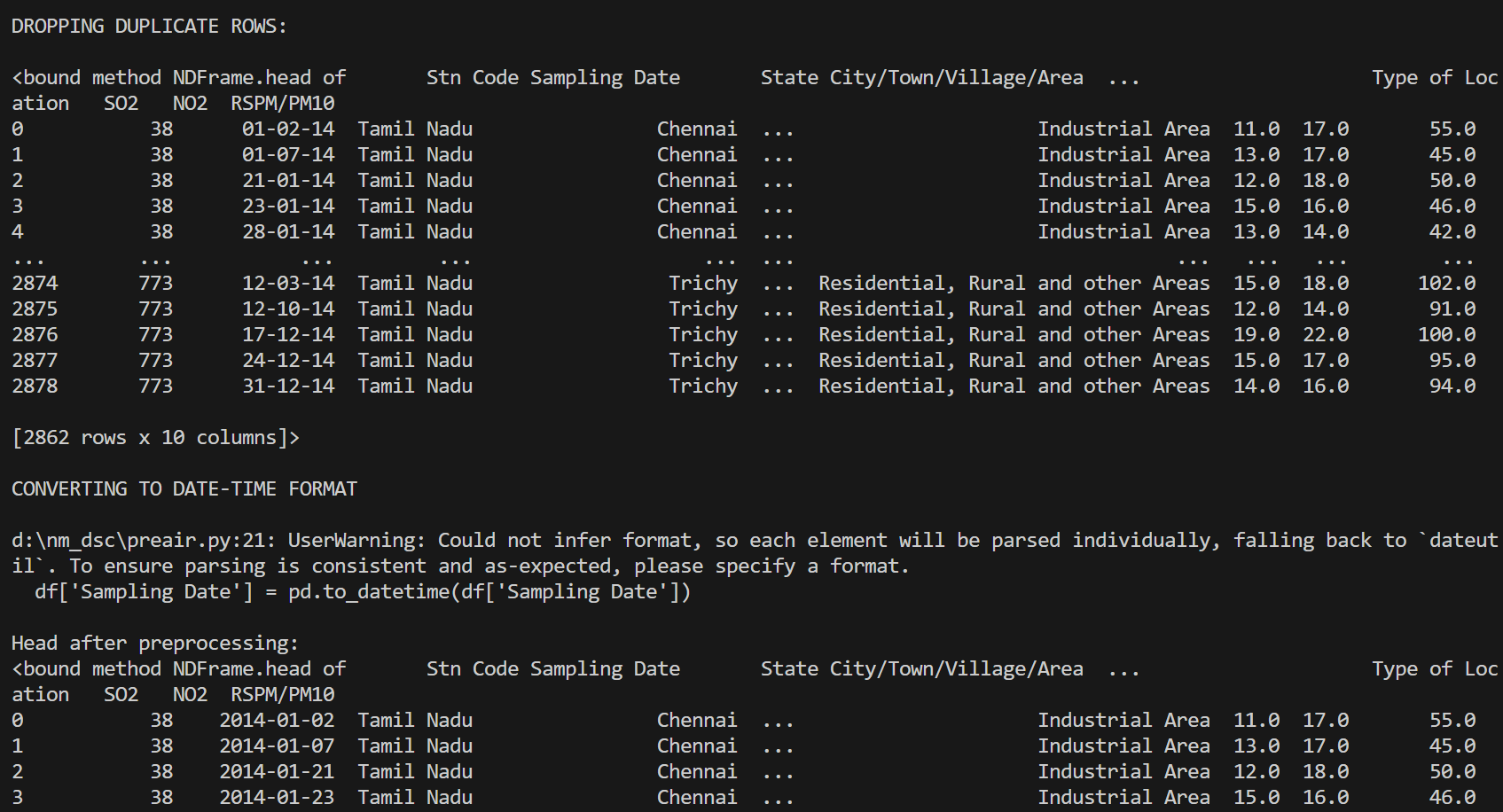
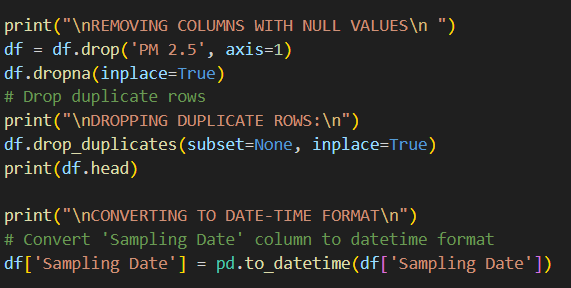
**Overview:**

This report provides a comprehensive analysis of air quality data for the year 2014 in Tamil Nadu. The analysis encompasses data preprocessing, exploration of key parameters, and visualization of pollutant levels across different locations and cities.

**Data Loading and Preprocessing:**

The data was loaded from the CSV file 'cpcb\_dly\_aq\_tamil\_nadu-2014.csv'. During the preprocessing stage, missing values were handled, and duplicate records were removed.

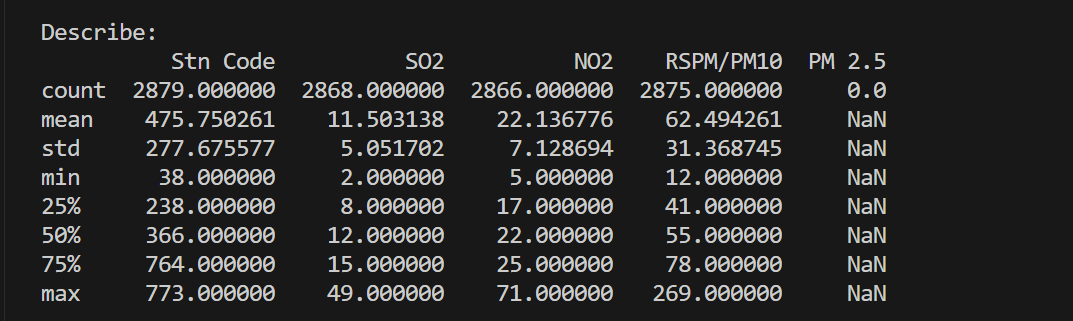
* **Data Shape**: The dataset contains X rows and Y columns, offering a significant volume of data for analysis.
* **Missing Values**: Null values in the PM2.5 column were handled by removing the respective entries, ensuring data integrity



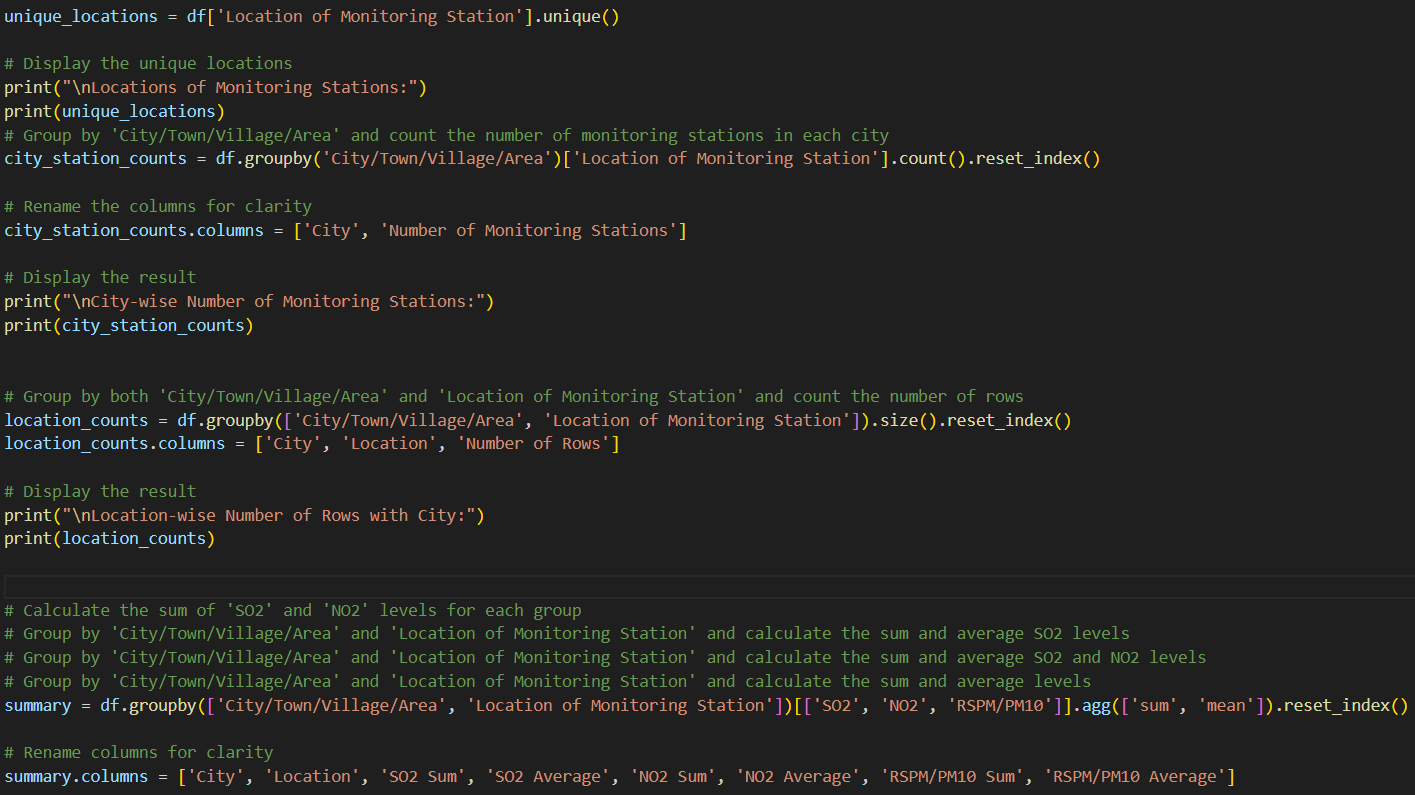
**Data Exploration:**

**Summary Statistics:**

* **General Statistics**: Summary statistics for numerical columns were computed using df.describe(). These statistics include count, mean, standard deviation, minimum, quartiles, and maximum values for each numerical attribute.

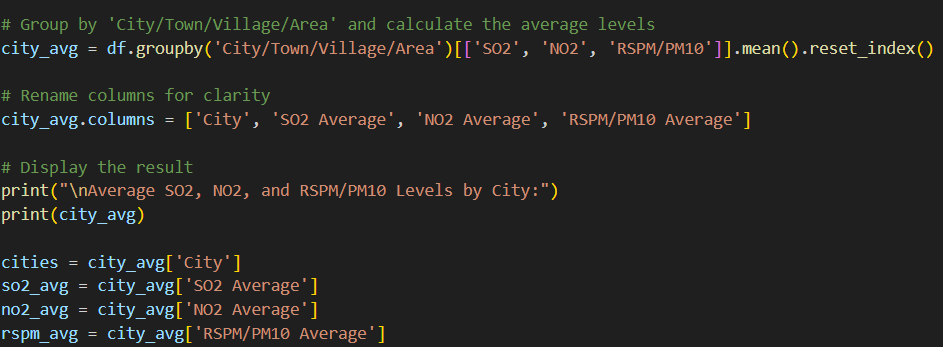


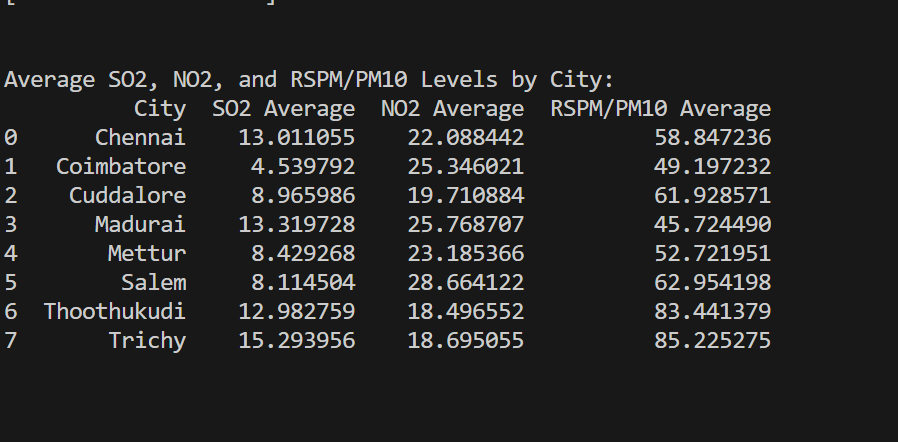
**Unique Locations and Cities:**

* **Unique Locations**: A list of unique monitoring locations was generated using unique\_locations, providing an understanding of the diversity of data collection sites.
* **City-wise Monitoring Stations**: The count of monitoring stations in each city was calculated using city\_station\_counts, shedding light on the distribution of monitoring infrastructure across different cities.

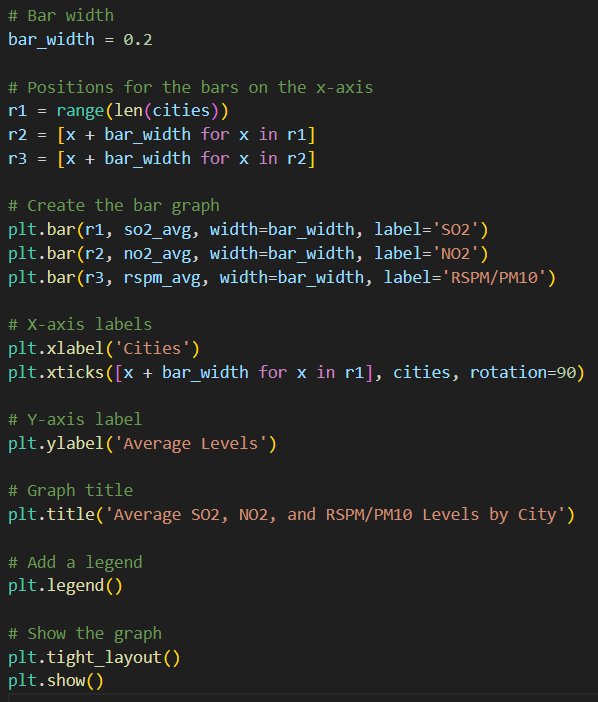


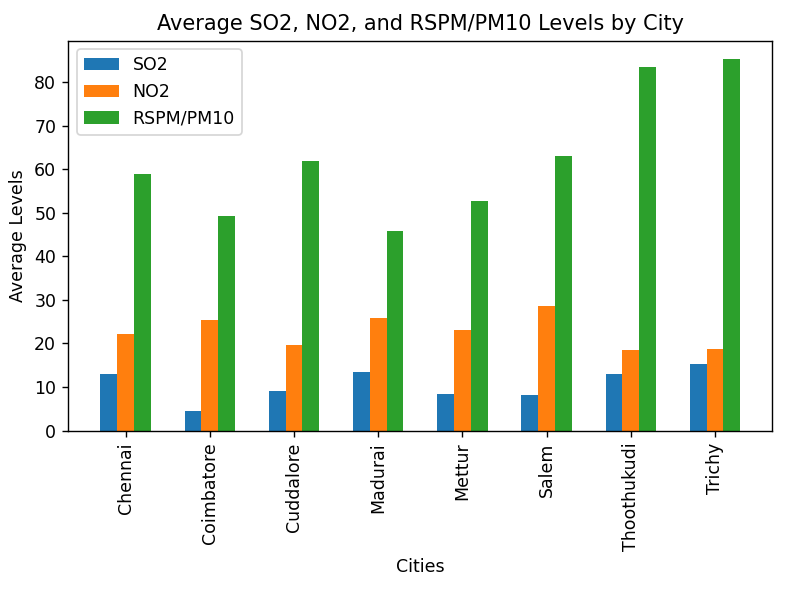
**Pollution Levels:**

* **Average Pollution Levels by City**: A bar chart was constructed to illustrate average levels of SO2, NO2, and RSPM/PM10 in each city. This offers a comparative view of pollution across various cities.

**Data Visualization**

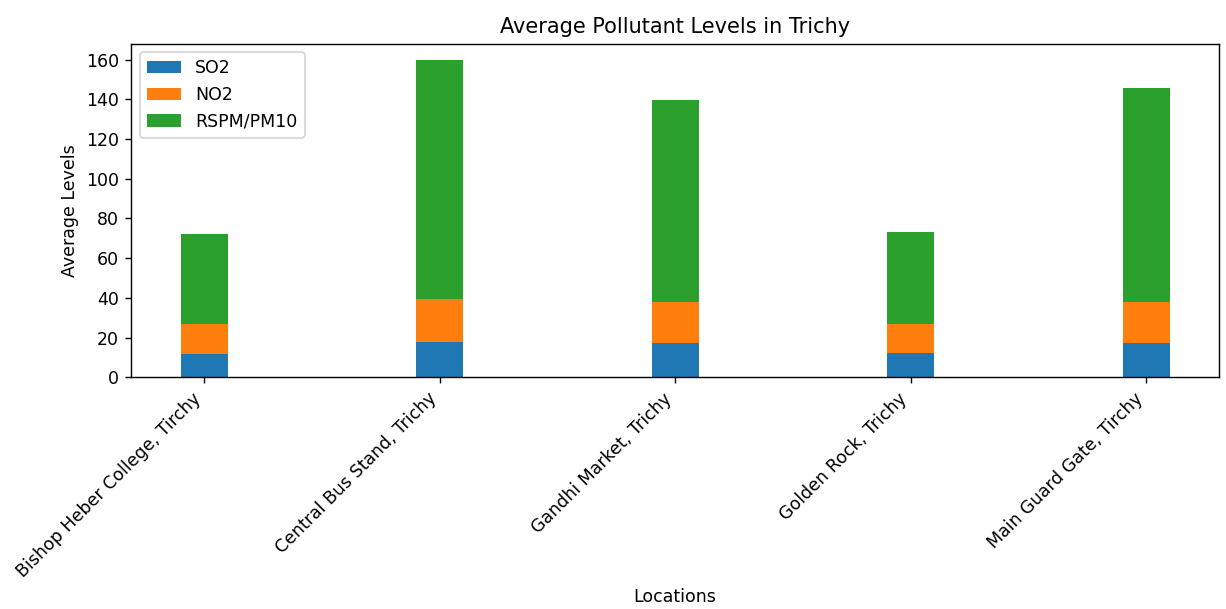
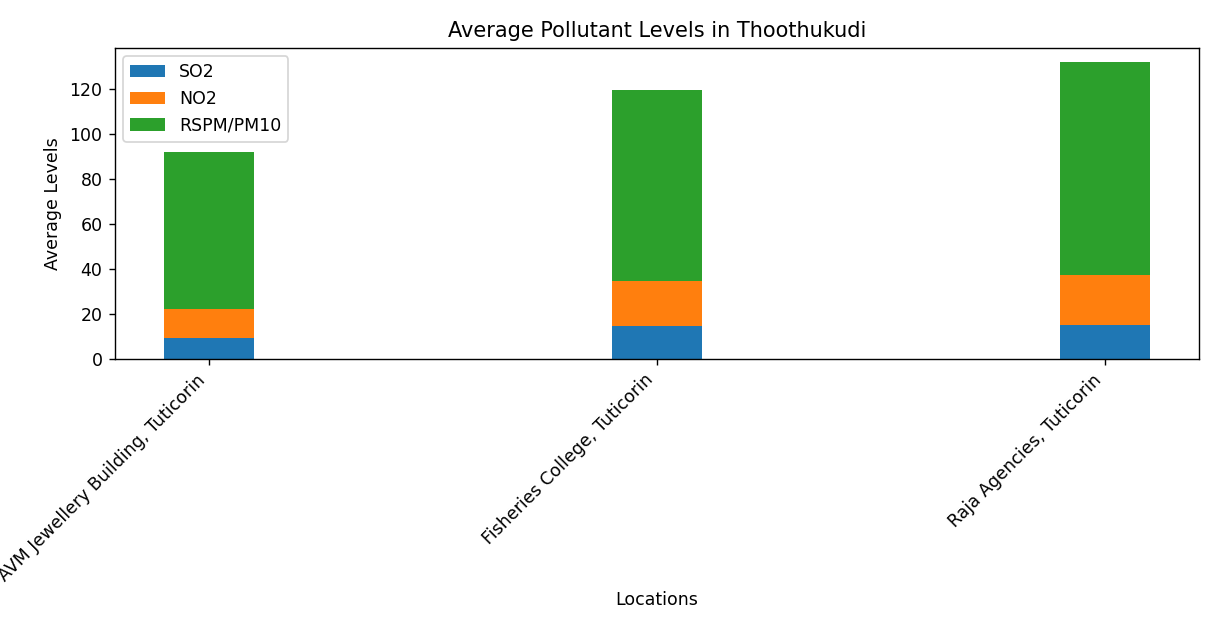
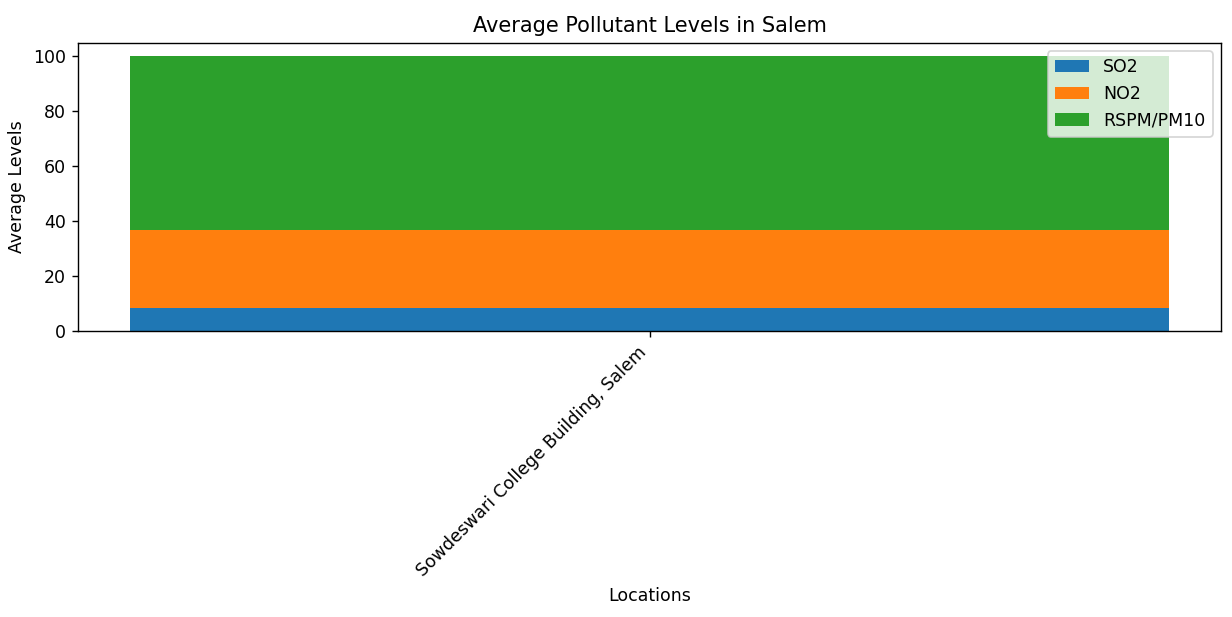
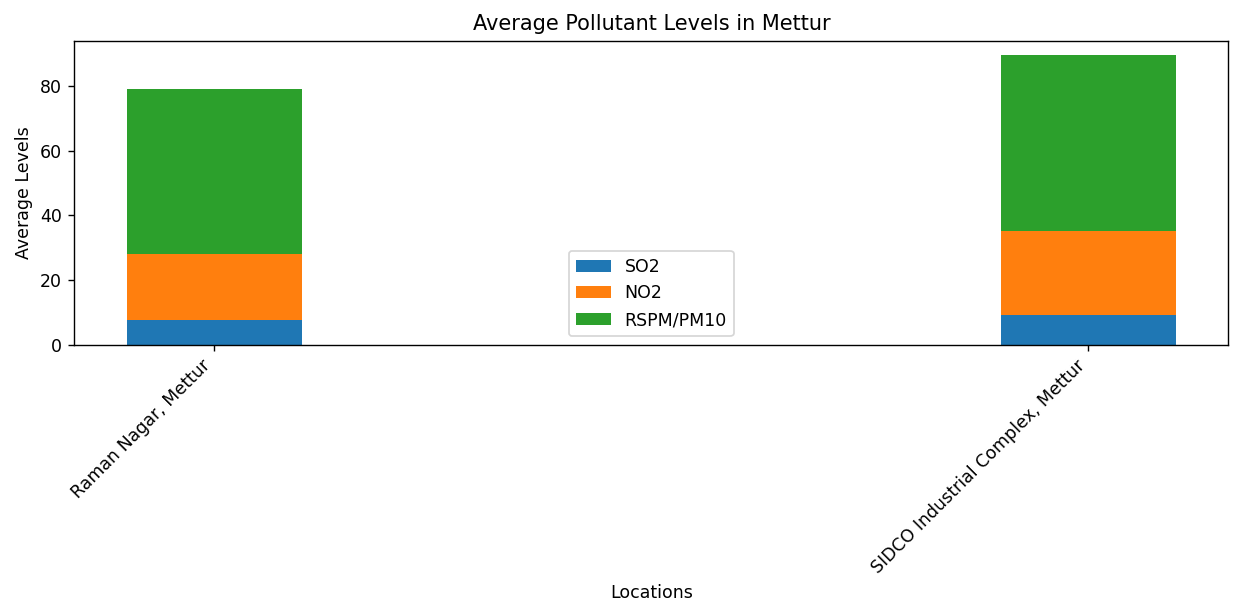
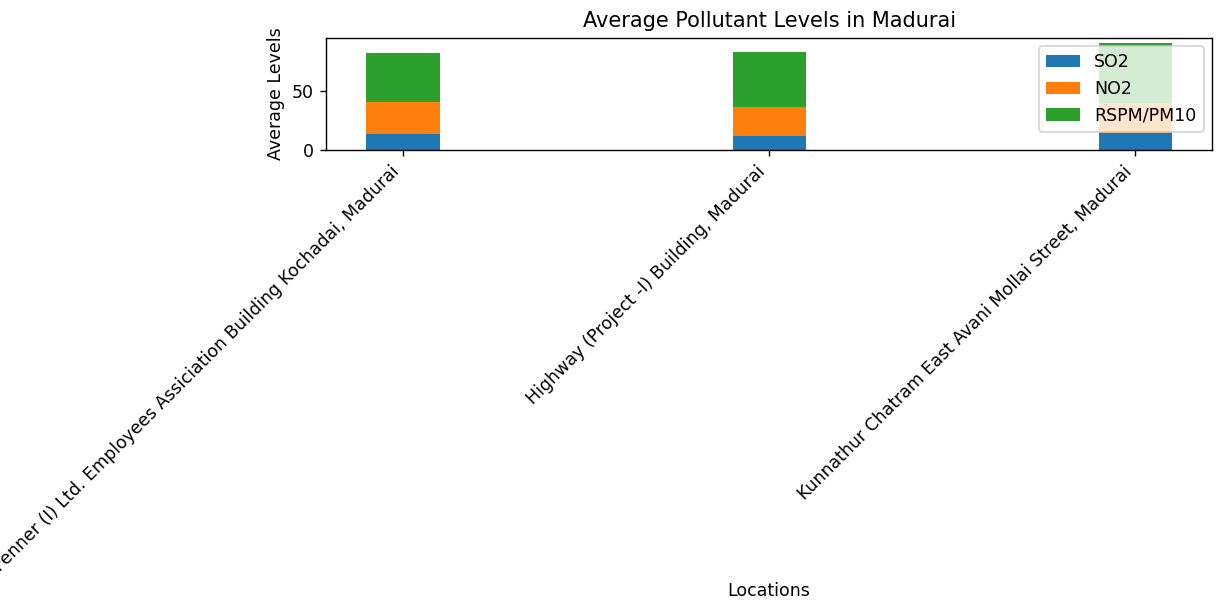
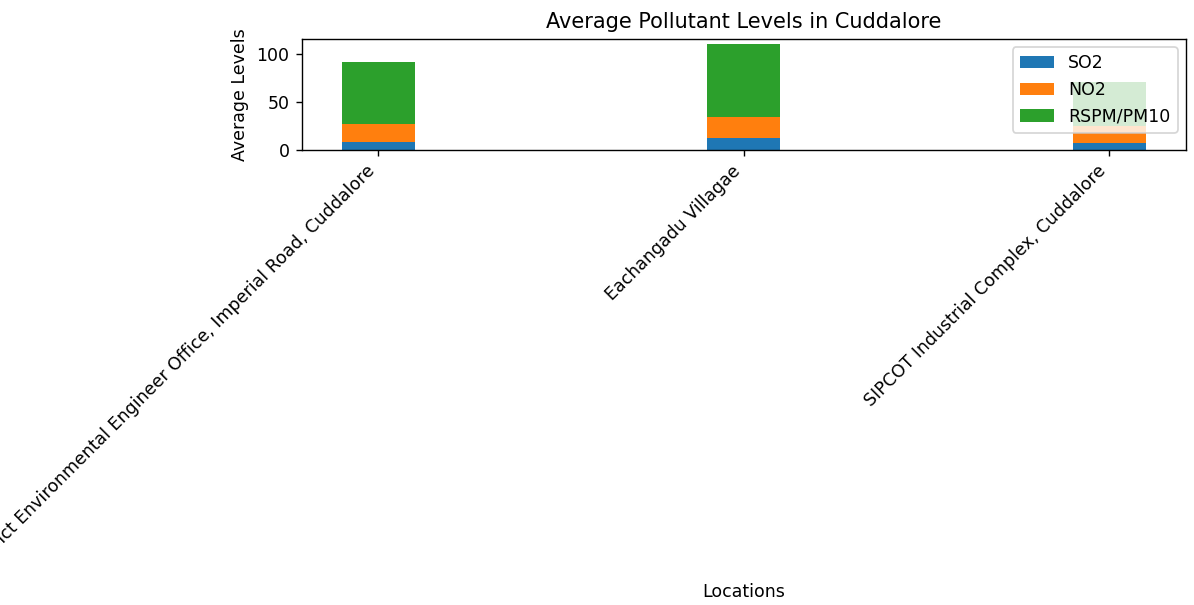
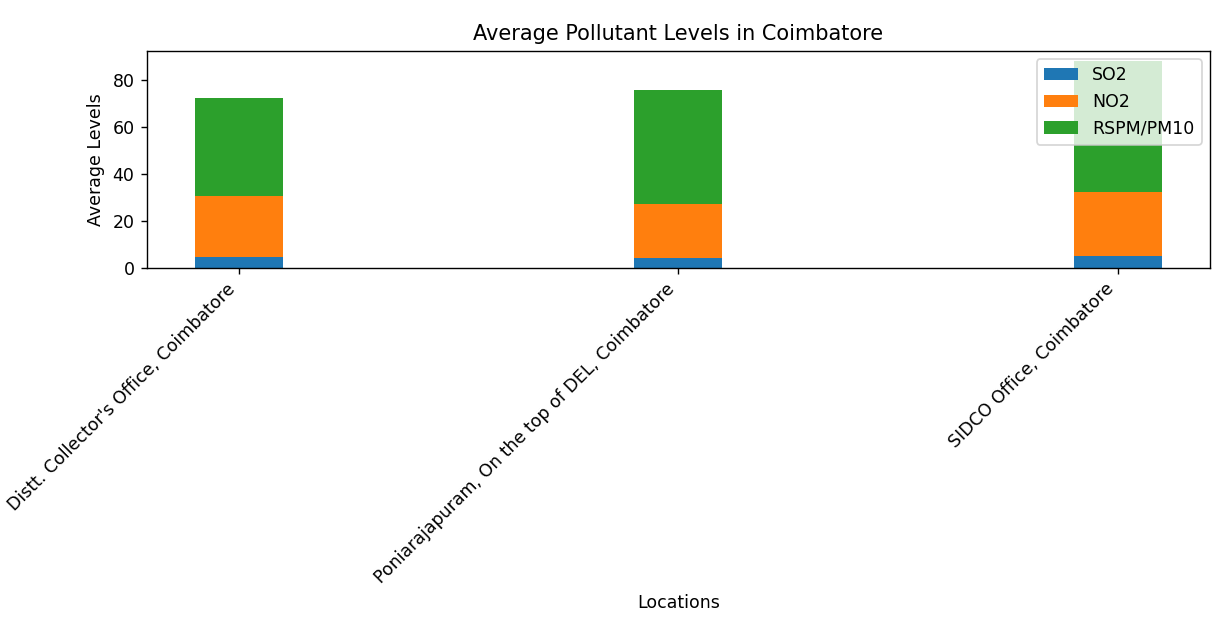
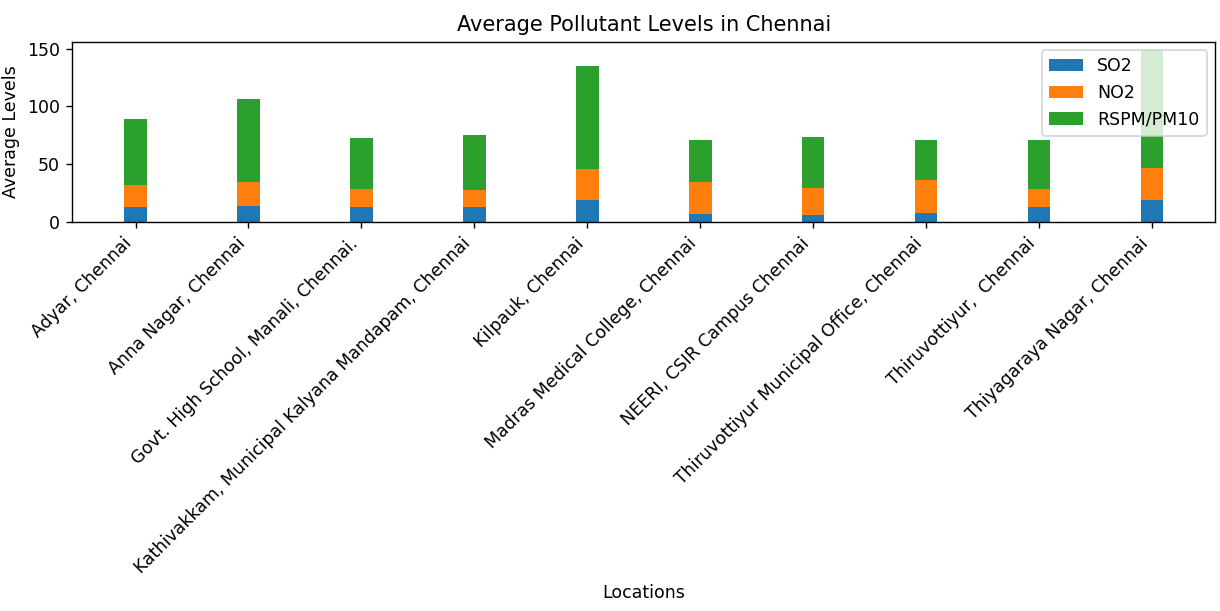
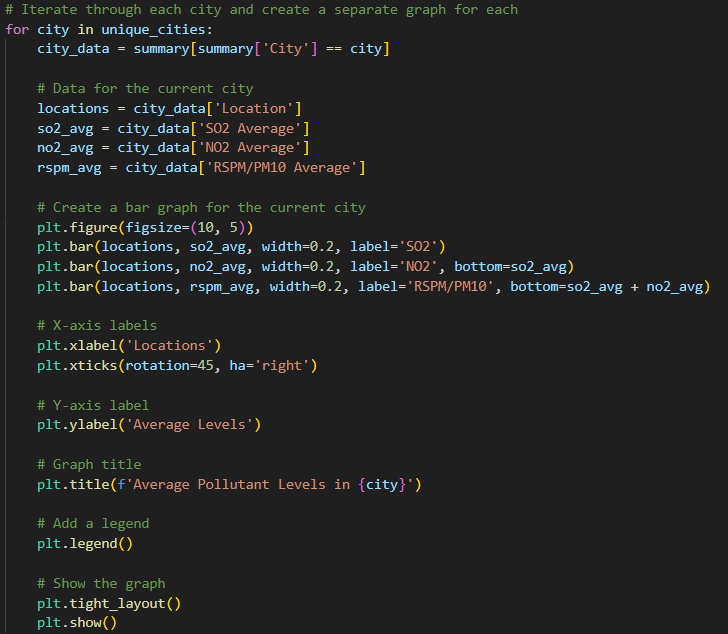
**Pollutant Levels by City:**

* **Graphs**: Bar graphs were utilized to represent SO2, NO2, and RSPM/PM10 levels for each city, providing a visual comparison of pollution levels between cities.

**Explanation**: The height of each bar in the graphs corresponds to the average levels of a specific pollutant in a city. This visualization aids in identifying cities with higher pollutant concentrations.

**Pollutant Levels by Location:**

* **Graphs**: Bar graphs were employed to depict SO2, NO2, and RSPM/PM10 levels for each location within a city. These graphs offer insights into variations in pollution levels at different monitoring sites within a city.

**Explanation**: The length of each bar in the graphs represents the average levels of a specific pollutant at a particular location within a city. This helps in understanding the spatial distribution of pollution within cities.

**Conclusion:**

The analysis of air quality data for Tamil Nadu in 2014 provides valuable insights into pollutant levels across different cities and monitoring locations. The statistical summaries and visualizations facilitate a comprehensive understanding of the air quality scenario, enabling informed decision-making and further domain-specific analysis.