**Development Part 1**

|  |  |
| --- | --- |
| **Date** | **17-10-2023** |
| **Team ID** | **499** |
| **Project Name** | **6112-AIR QUALITY ASSESSMENT IN TN** |

|  |  |
| --- | --- |
| 1 | Problem Statement |
| 2 | Data Pre-processing |
| 3 | Data Visualization |
| 4 | Conclusion |

**AIR QUALITY ASSESSMENT IN TN**

**Problem Statement**

Objective: Develop a comprehensive and localized air quality assessment system in Tennessee, leveraging advanced monitoring technologies and data analysis techniques.

Data: To conduct a comprehensive air quality assessment in Tennessee, critical data includes measurements of pollutants like PM2.5, PM10, NO2, SO2, CO, and O3 from monitoring stations. Meteorological data, encompassing temperature, humidity, wind patterns, and precipitation, is vital for understanding atmospheric conditions. Geographic information system (GIS) data aids in assessing how topography and land use influence air quality. Emission inventories offer insights into pollution sources, while population density and demographic data help identify vulnerable communities. Health records, historical data, and air quality modeling outputs provide additional context. Access to policy and regulatory information ensures evaluations align with existing guidelines, while remote sensing data offers a broader perspective on pollution patterns.

**Data Pre-Processing**

Data preprocessing is crucial before feeding the data into any machine learning model. It involves cleaning, transforming, and organizing the data to make it suitable for analysis. Here are some common data preprocessing steps you can perform for the provided dataset:

1. Handling Missing Values:

Identify and handle missing values. You can either remove rows with missing values or fill them using techniques like mean, median, or interpolation.

1. Handling Categorical Data:

If there are categorical variables, you might need to encode them (e.g., using one-hot encoding) so that they can be used in a machine learning model.

1. Feature Scaling:

Standardize or normalize numerical features to ensure they are on a similar scale. This can be important for models that are sensitive to the scale of the data.

1. Date Handling:

If the 'Date' column is not in a datetime format, convert it.

1. Feature Engineering :

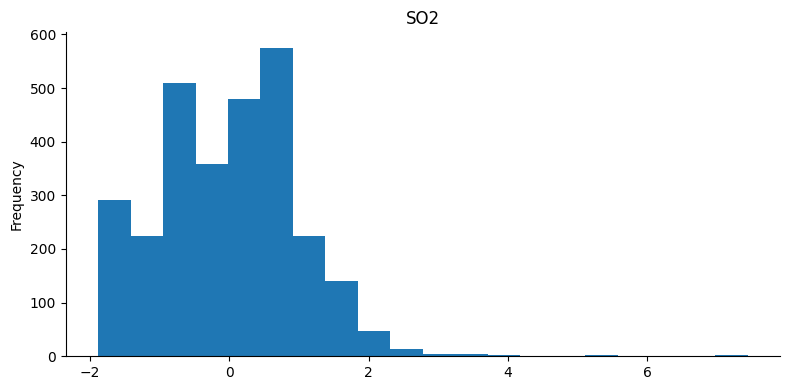
You might want to create new features based on existing ones. For instance, extracting day, month, and year from the date.

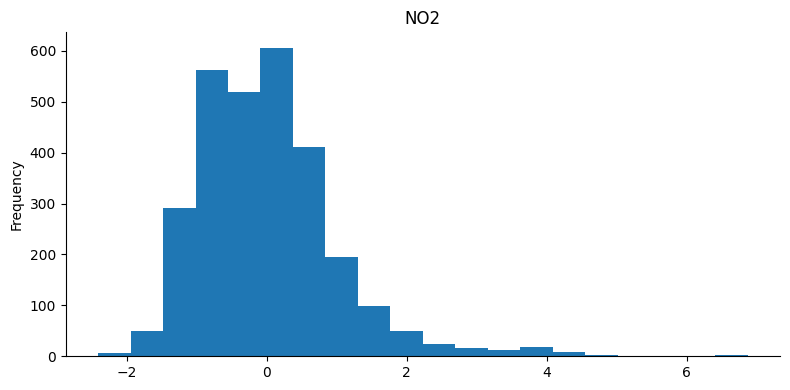
1. Remove Unnecessary Columns:

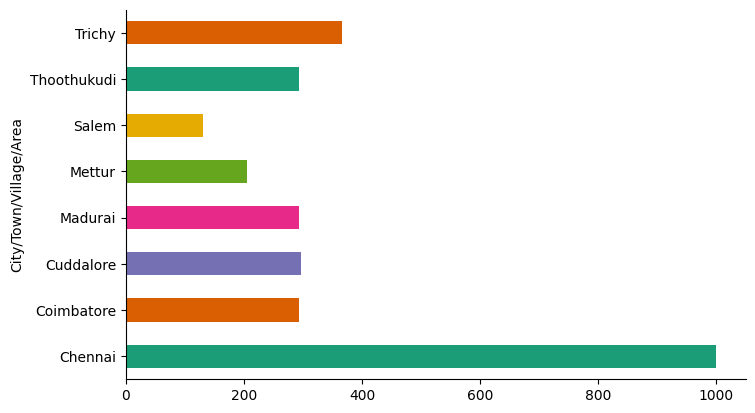
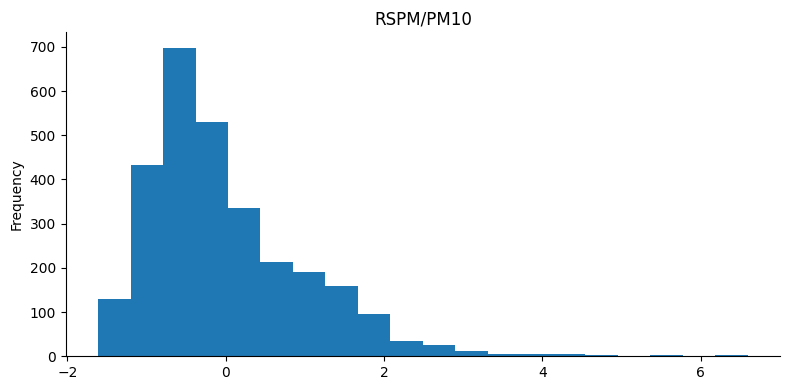
If certain columns don't contribute meaningfully to the analysis or modeling, consider dropping them.

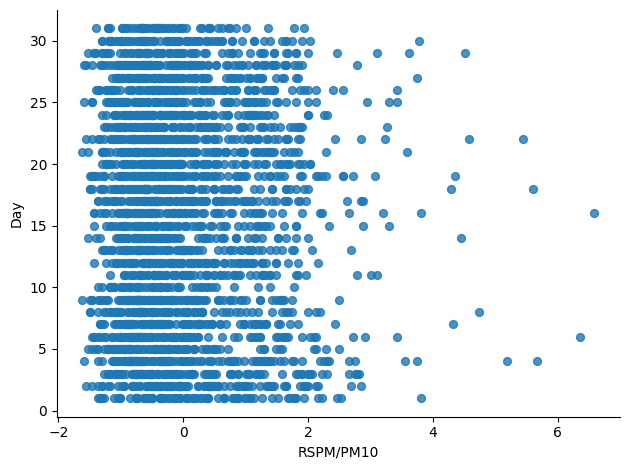
Remember, the specific preprocessing steps can vary depending on the nature of your analysis and the model you plan to use. Always understand your data and choose preprocessing steps accordingly.

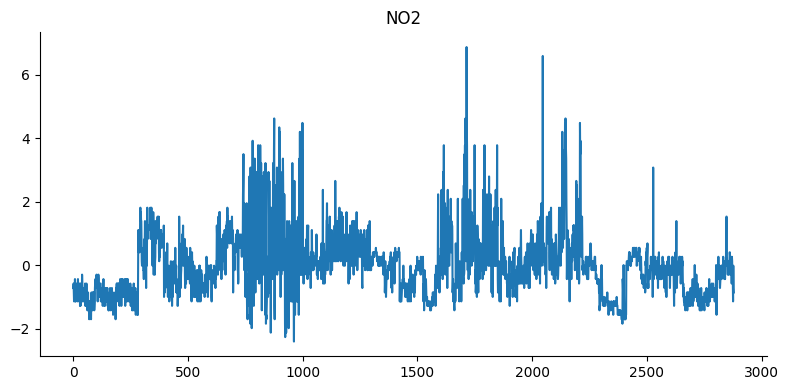
**Data Visualization**

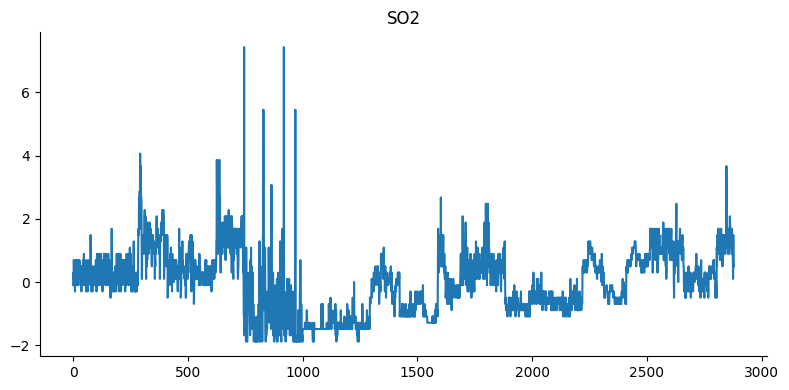


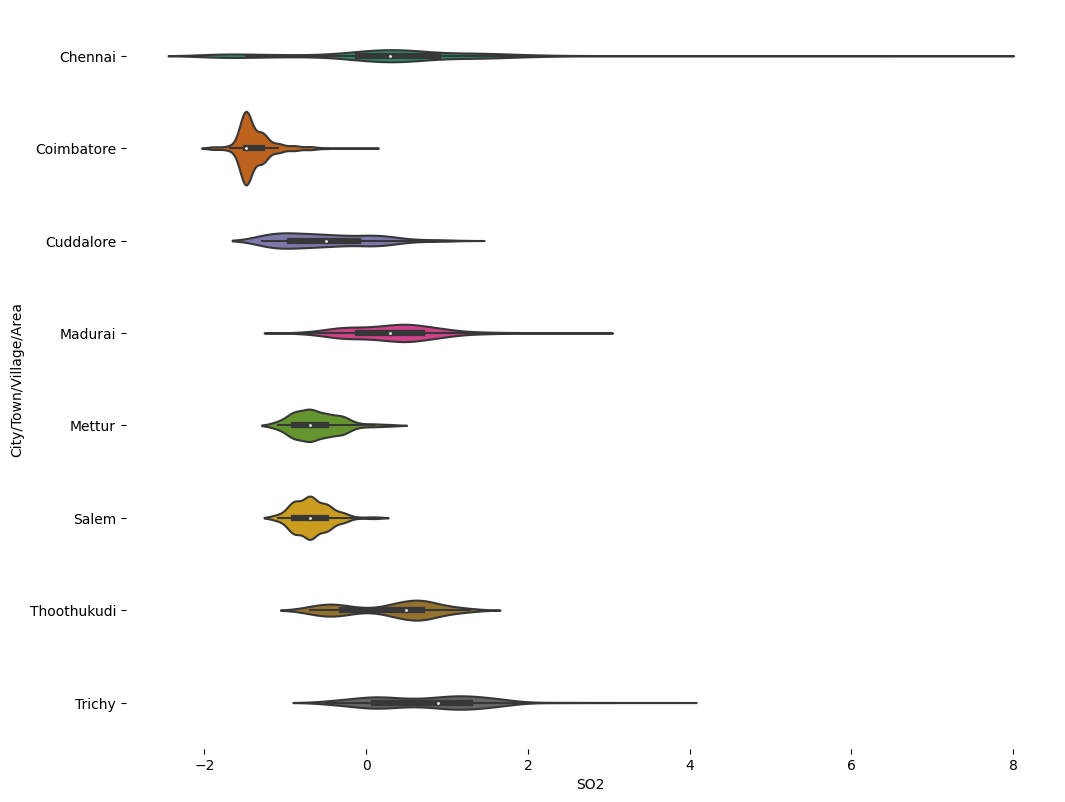


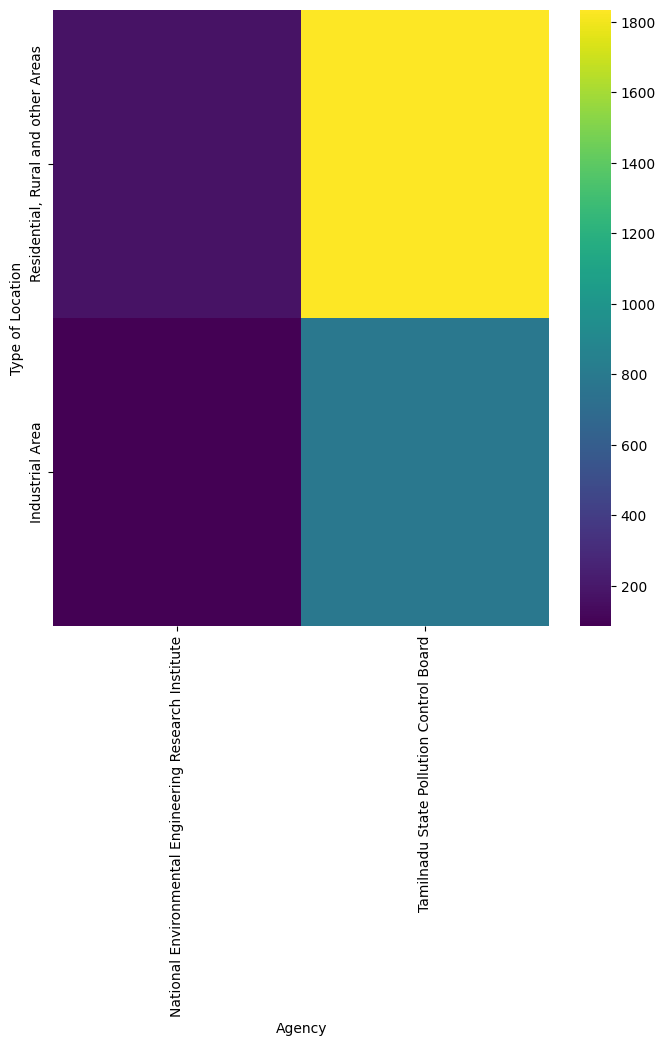










**Conclusion**

In conclusion, addressing air quality challenges in Tennessee demands a concerted and empathetic approach. The complexities, from limited monitoring infrastructure to diverse geographical influences, underscore the urgency of the issue. Through collaborative efforts, we can bridge gaps in data and technology, ensuring accurate assessments. Engaging communities and stakeholders, while prioritizing public health, will be pivotal in formulating effective policies. By empathizing with the concerns and experiences of affected individuals, we reaffirm our commitment to a cleaner, healthier Tennessee. Together, we can forge a path towards sustainable air quality improvements, safeguarding the well-being and vitality of our communities for generations to come.