Air Quality Analysis In Tamil Nadu 2014

Data Analytics with Cognos – Phase 5

Project Documentation

Development Part 2

AIR QUALITY ANALYSIS OBJECTIVE :

Assessment of Air Quality:

To monitor and assess the current air quality in various regions of Tamil Nadu, including urban and rural areas, to understand the levels of pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), and ozone (O3).

Identification of Pollution Sources:

Determine the major sources of air pollution in the region, which could include industrial emissions, vehicular traffic, construction activities, agricultural practices, and natural sources like dust and pollen.

Health Impact Assessment:

Evaluate the potential health risks associated with poor air quality, especially focusing on vulnerable populations, such as children, the elderly, and individuals with preexisting health conditions.

Compliance Monitoring:

Ensure that air quality standards and regulations set by the Central Pollution Control Board (CPCB) and the Tamil Nadu Pollution Control Board (TNPCB) are met and take necessary measures for compliance.

Data Collection and Analysis:

Collect and analyze air quality data from monitoring stations and satellite sources to provide a comprehensive view of air pollution trends and hotspots.

Public Awareness:

Educate the public about the importance of air quality, its impacts on health, and measures to reduce exposure to pollutants through awareness campaigns and outreach programs.

Policy Recommendations:

Provide data-driven recommendations for policy changes and initiatives to mitigate air pollution, such as stricter emissions standards, promoting public transportation, and reducing industrial emissions.

Technology Adoption:

Encourage the use of advanced technology, such as air quality monitoring devices and sensors, to enhance real-time data collection and public awareness.

Collaboration:

Collaborate with governmental agencies, non-governmental organizations, and research institutions to ensure a coordinated effort in addressing air quality issues.

Long-term Planning:

Develop strategies for long-term air quality improvement, which may involve urban planning, green infrastructure, and sustainable transportation solutions.

ANALYSIS APPROACH :

Data Collection and Monitoring:

* Establish a network of air quality monitoring stations across Tamilnadu, strategically located in urban, industrial, and rural areas.
* Collect data on key air pollutants such as PM2.5, PM10, NO2, SO2, CO, O3, and VOCs.
* Monitor meteorological parameters like temperature, humidity, wind speed, and wind direction to understand their impact on air quality.

Data Quality Assurance:

* Ensure data accuracy and reliability by calibrating and maintaining monitoring instruments regularly.
* Implement data quality control measures to identify and correct errors or inconsistencies in the collected data.

Data Integration:

* Integrate data from various monitoring stations into a centralized database for easy access and analysis.

Data Analysis:

* Conduct statistical analysis to identify trends and patterns in air quality data.
* Analyze temporal variations to detect daily, seasonal, and long-term trends.
* Use statistical tests to assess correlations between air pollutants and meteorological variables.

Spatial Analysis:

* Utilize Geographic Information System (GIS) tools to create spatial maps of air quality across Tamilnadu.
* Identify areas with high pollution levels and sources of pollution.

Emission Source Identification:

* Apply source apportionment techniques like chemical mass balance or receptor modeling to identify and quantify pollution sources contributing to poor air quality.

Health Impact Assessment:

* Evaluate the potential health impacts of poor air quality by considering factors like populatio densityandvulnerability.
* Estimate the number of health-related incidents attributed to air pollution.

Air Quality Index (AQI) Calculation:

* Calculate and report the AQI based on measured pollutant concentrations, providing an understandable indicator of air quality to the public.

Predictive Modeling:

* Develop predictive models using historical data to forecast future air quality conditions.
* These models can aid in issuing air quality alerts and advisories.

Data Visualization and Reporting:

* Create interactive data visualizations and dashboards to communicate findings to policymakers, stakeholders, and the public.
* Generate regular reports summarizing air quality trends and recommendations for improvement.

Policy Recommendations:

* Based on the analysis results, provide recommendations for policy interventions, such as emission controls, regulatory changes, or public awareness campaigns, to improve air quality.

Public Awareness and Education:

* Conduct public outreach and education programs to raise awareness about air quality issues and encourage responsible actions to reduce pollution.

Continuous Monitoring and Feedback Loop:

* Implement a system for continuous air quality monitoring and analysis to support real-time decision-making and the adaptation of policies as needed.

This comprehensive approach aims to provide a holistic understanding of air quality in Tamilnadu, facilitating informed decision-making and actions to mitigate pollution and protect public health and the environment.

VISUALIZATION TECHNIQUES :

Time Series Plots:

Line charts showing pollutant concentrations over time.

Heatmaps:

Visualize spatial variations in pollution levels across Tamil Nadu.

Bar Charts:

Illustrate pollutant levels by region or city.

Box Plots:

Depict the distribution of pollutant data, highlighting outliers.

Geospatial Maps:

Map pollution levels at different locations in Tamil Nadu.

Correlation Heatmaps:

Show the correlation between weather variables and pollution levels.

**Input 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)

CODE IMPLEMENTATION :

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import folium

from sklearn.preprocessing import LabelEncoder

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import classification\_report, confusion\_matrix, accuracy\_score

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score

df = pd.read\_csv('C:\\aqa.csv')

print(df.head())

df['SO2']=df['SO2'].fillna(0).astype('str').astype('float')

df['NO2']=df['NO2'].fillna(0).astype('str').astype('float')

df['RSPM/PM10']=df['RSPM/PM10'].fillna(0).astype('str').astype('float')

df['PM 2.5']=df['PM 2.5'].fillna(0).astype('str').astype('float')

df.drop(['Stn Code','Agency'],axis=1,inplace=True)

df=df.rename(index=str,columns={'Sampling Date':'year'})

print(df.info())

average\_so2 = df.groupby('City/Town/Village/Area')['SO2'].mean()

print (average\_so2)

average\_no2 = df.groupby('City/Town/Village/Area')['NO2'].mean()

print (average\_no2)

average\_rspm\_pm10 = df.groupby('City/Town/Village/Area')['RSPM/PM10'].mean()

print (average\_rspm\_pm10)

df['year'] = pd.to\_datetime(df['year'], format='%d-%m-%Y')

df.set\_index('year', inplace=True)

#Descriptive statistics

mean\_so2 = df['SO2'].mean()

median\_so2 = df['SO2'].median()

std\_dev\_so2 = df['SO2'].std()

print(f"Mean SO2 Level: {mean\_so2}")

print(f"Median SO2 Level: {median\_so2}")

print(f"Standard Deviation SO2 Level: {std\_dev\_so2}")

# Box Plot of rspm

plt.figure(figsize=(10, 6))

sns.boxplot(x='City/Town/Village/Area', y='RSPM/PM10', data=df)

plt.xlabel('City/Town/Village/Area')

plt.ylabel('RSPM/PM10 Levels')

plt.title('RSPM/PM10 Levels Across Cities (Box Plot)')

plt.xticks(rotation=45)

# Heatmap for rspm

pivot\_table = df.pivot\_table(index='City/Town/Village/Area', columns='year', values='RSPM/PM10', aggfunc='mean')

plt.figure(figsize=(10, 6))

sns.heatmap(pivot\_table, cmap='YlGnBu', annot=True)

plt.xlabel('year')

plt.ylabel('City/Town/Village/Area')

plt.title('Average RSPM/PM10 Levels by City and year')

#time series

plt.figure(figsize=(12, 6))

plt.subplot(3, 1, 1)

plt.plot(df['SO2'], label='SO2 Levels', color='blue')

plt.title('SO2 Levels Over Time')

plt.legend()

# Plot time series data for NO2

plt.subplot(3, 1, 2)

plt.plot(df['NO2'], label='NO2 Levels', color='green')

plt.title('NO2 Levels Over Time')

plt.legend()

# Plot time series data for RSPM/PM10

plt.subplot(3, 1, 3)

plt.plot(df['RSPM/PM10'], label='RSPM/PM10 Levels', color='red')

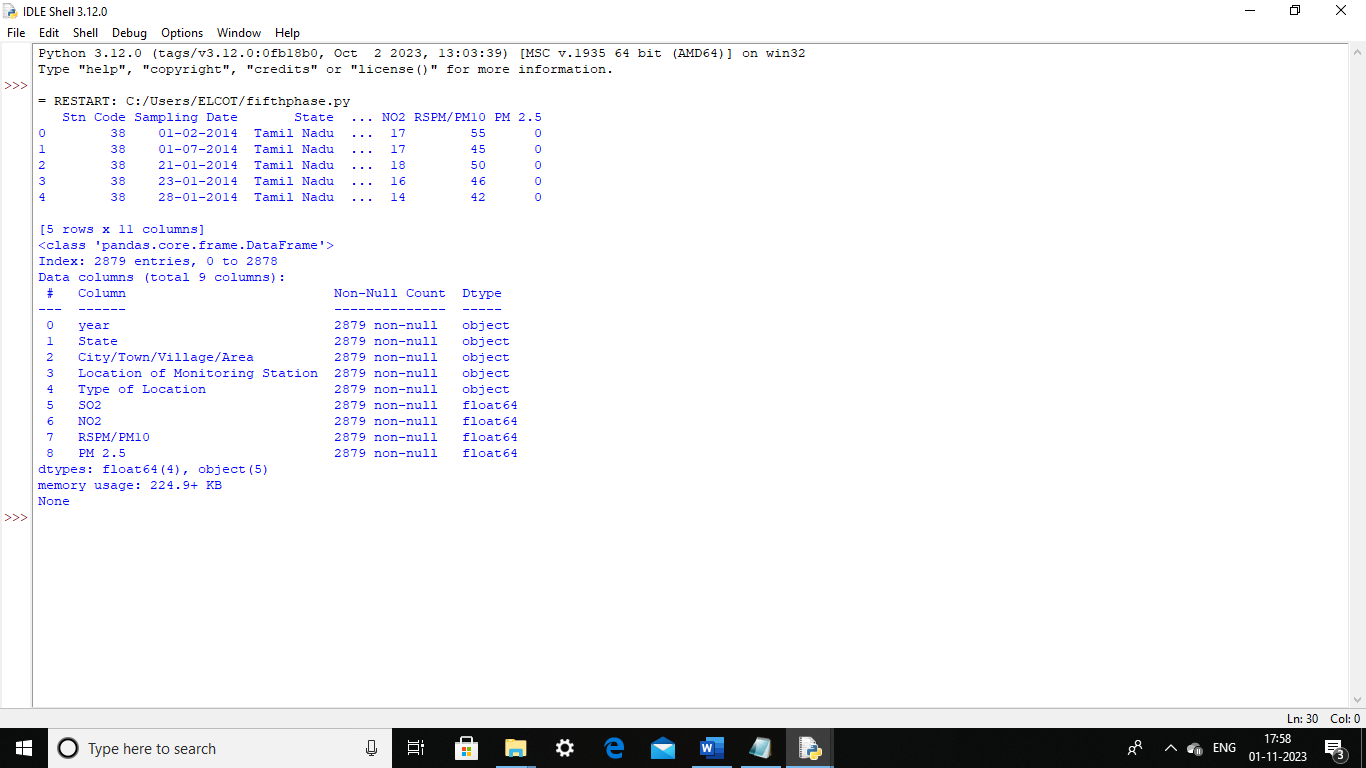
plt.title('RSPM/PM10 Levels Over Time')

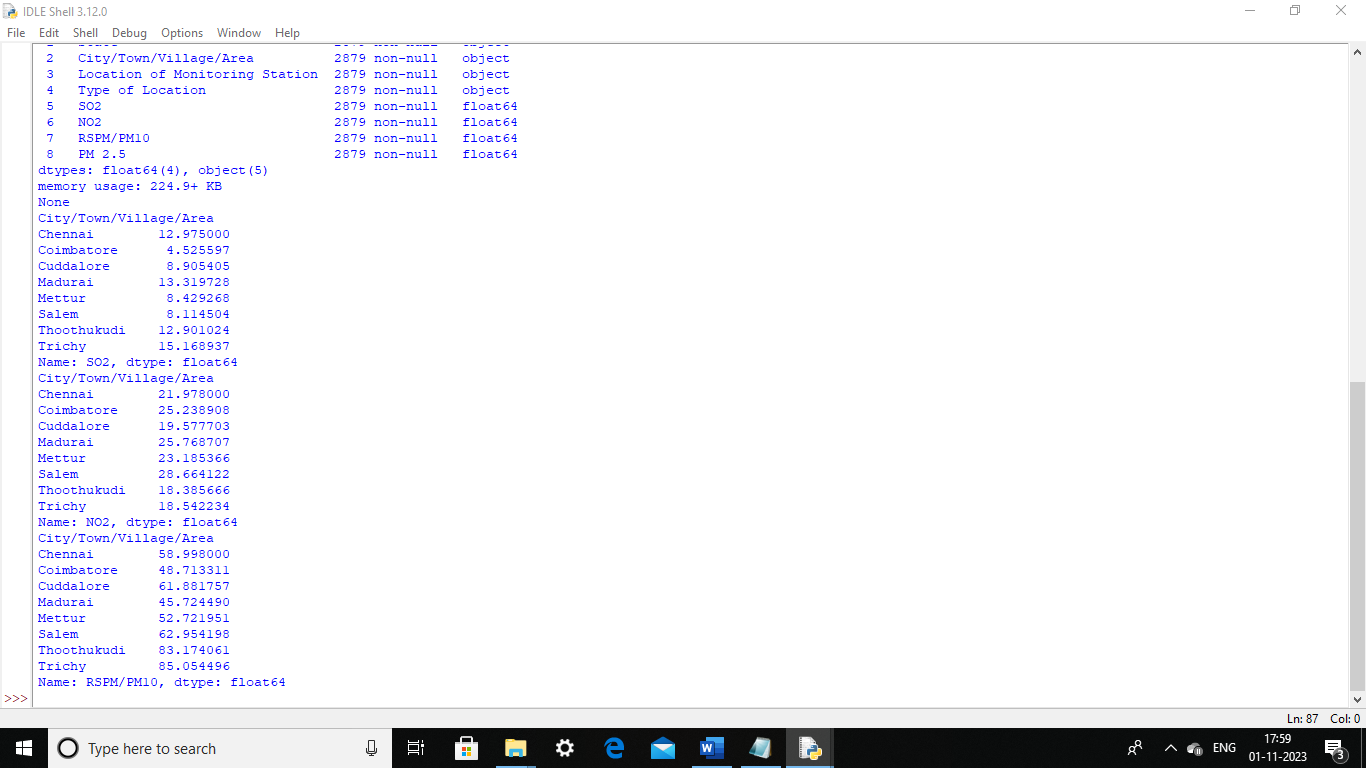
plt.legend()

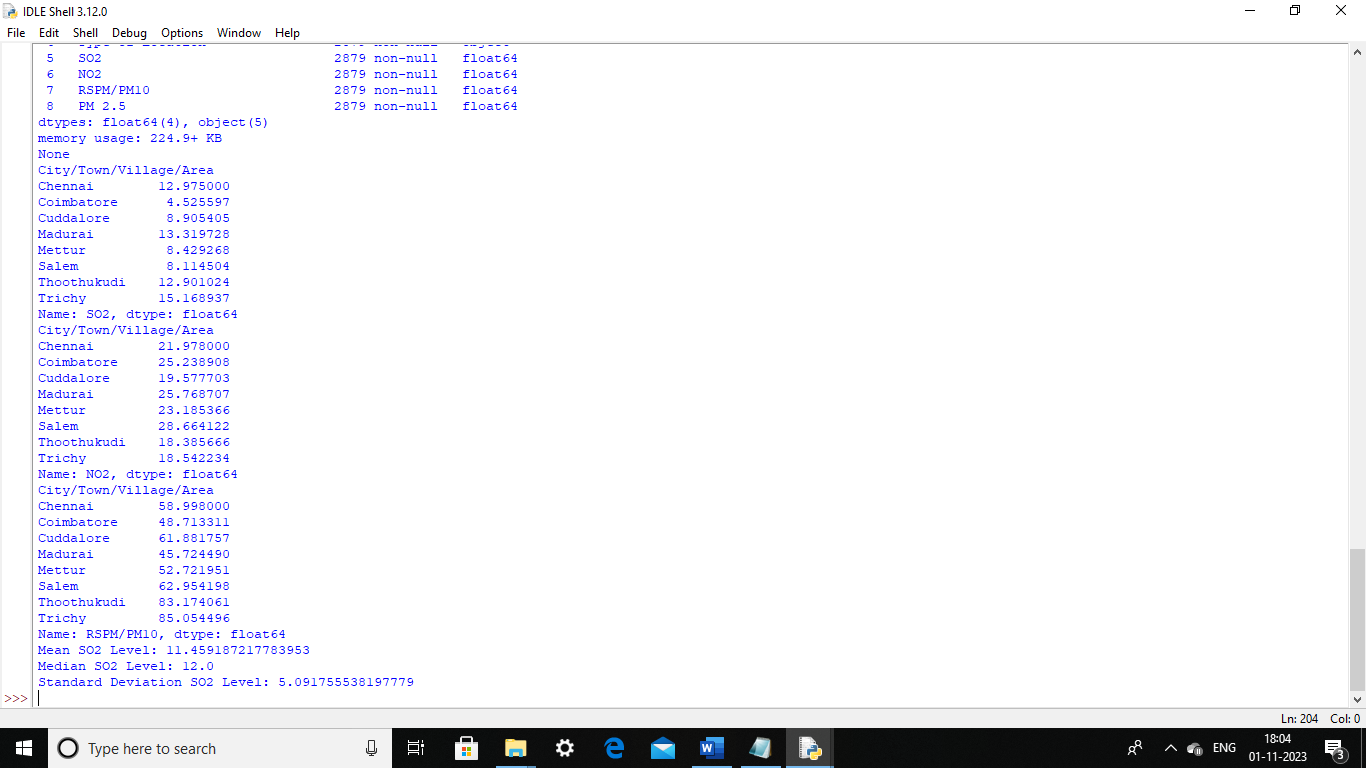
plt.tight\_layout()

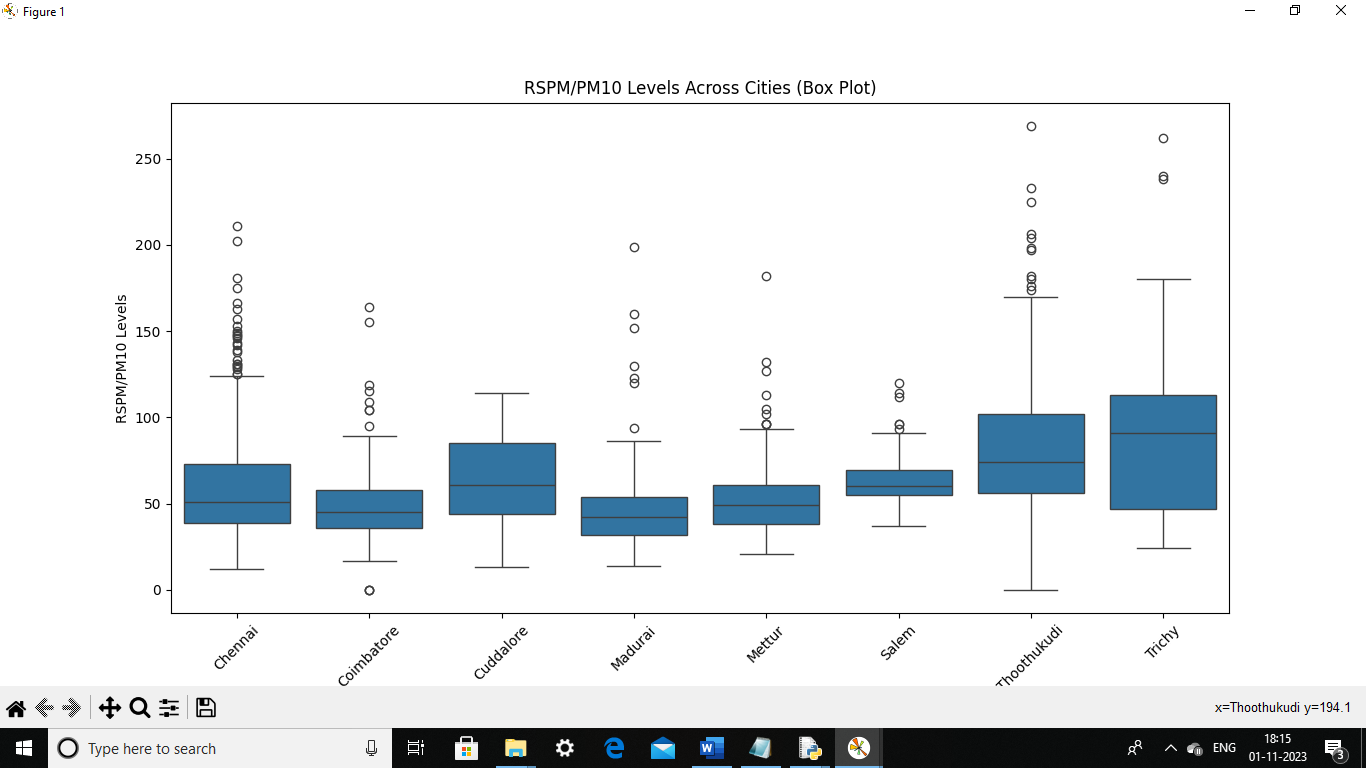
plt.show()

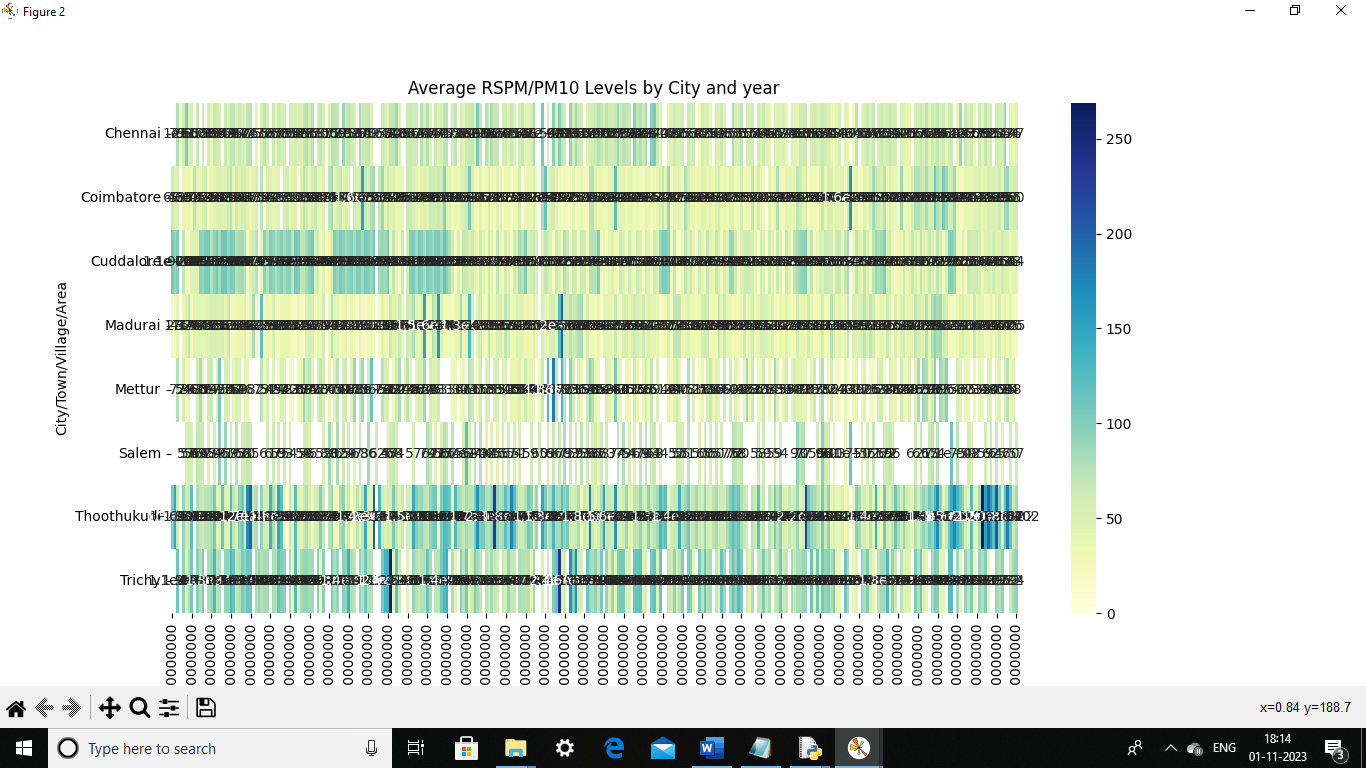
EXAMPLE OUTPUTS :

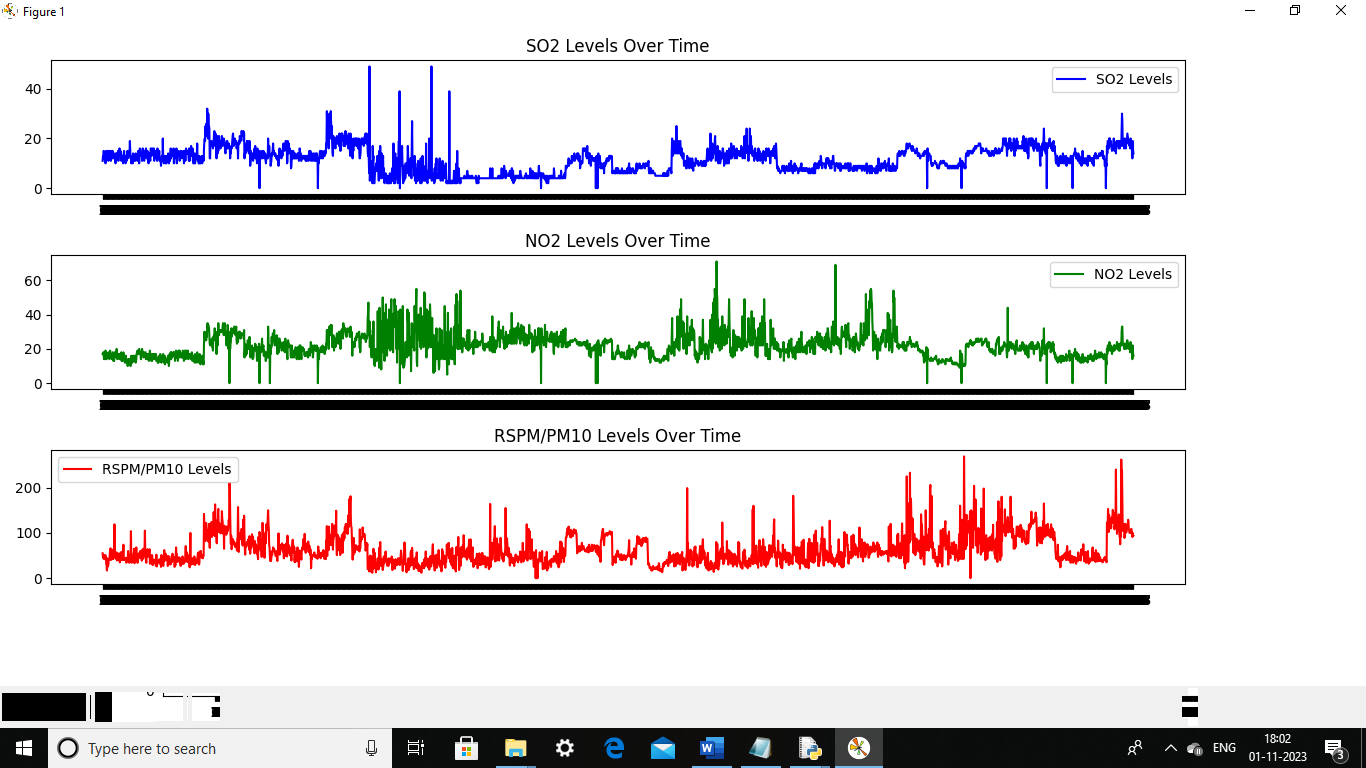






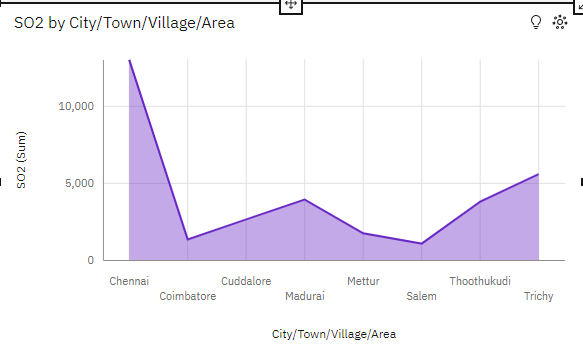






DATA VISUALIZATION WITH IBM COGNOS :

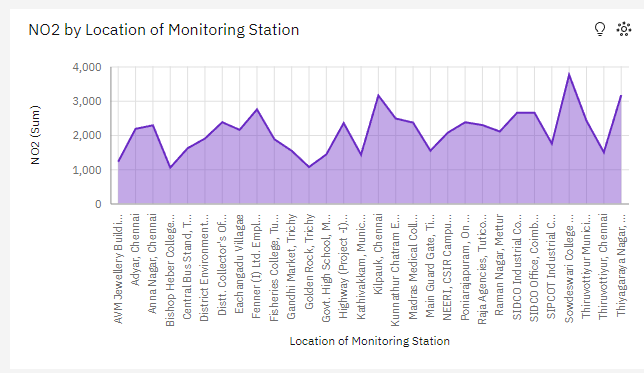
1.SO2 by City/Town/Village/Area:



Insights from the visualization:

* SO2 is unusually high when City/Town/Village/Area is Chennai.
* It is projected that by 2015-03-14, Thoothukudi will exceed Madurai in SO2 by 1.26.
* From 2014-01-23 to 2014-01-24, Chennai's SO2 increased by 163%.
* Over all values of City/Town/Village/Area, the sum of SO2 is almost 33 thousand.
* SO2 ranges from over a thousand, when City/Town/Village/Area is Salem, to nearly thirteen thousand, when City/Town/Village/Area is Chennai.

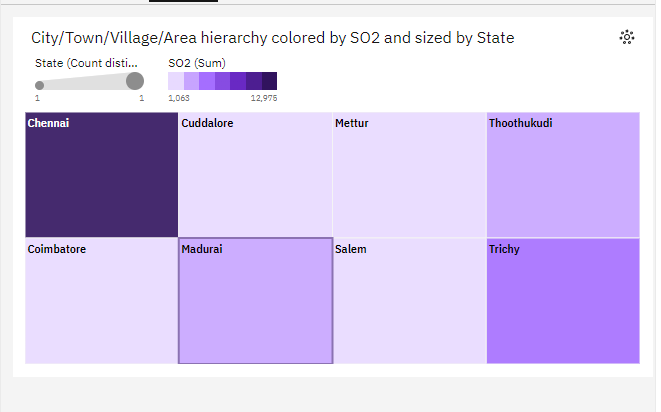
2.NO2 by Location of Monitoring Station :



Insights from the visualization:

* **NO2** is most unusual when the values of **Location of Monitoring Station** are **Sowdeswari College Building, Salem, Bishop Heber College, Tirchy, Golden Rock, Trichy, Thiyagaraya Nagar, Chennai, Kilpauk, Chennai** and more.
* From **2014-07-08** to **2014-07-10**, **SIDCO Industrial Complex, Mettur**'s **NO2** dropped by **65**%.
* Across all **location of monitoring stations**, the sum of **NO2** is **over 63 thousand**.
* **NO2** ranges from **over a thousand**, when **Location of Monitoring Station** is **Bishop Heber College, Tirchy**, to **nearly four thousand**, when **Location of Monitoring Station** is **Sowdeswari College Building, Salem.**

3. City/Town/Village/Area hierarchy colored by SO2 and sized by State :



Insights from the visualization:

* **City/Town/Village/Area** **Chennai** has the highest values of both **State** and **SO2**.
* It is projected that by **2015-03-14**, **Thoothukudi** will exceed **Madurai** in **SO2** by **1.26**.
* From **2014-01-23** to **2014-01-24**, **Chennai**'s **SO2** increased by **163**%.
* **Chennai** is the most frequently occurring category of **City/Town/Village/Area** with a count of **1000** items with **State** values (**34.7** % of the total).
* The total number of results for **State**, across all **City/Town/Village/Area**, is **nearly three thousand**.

CONCLUSION:

In conclusion, the air quality analysis in Tamil Nadu reveals a complex and concerning picture. While the state enjoys relatively better air quality compared to some other Indian regions, it still faces challenges such as seasonal variations in air pollution, industrial emissions, and vehicular pollution. Efforts to improve air quality must focus on stricter regulations, promoting clean energy sources, reducing vehicular emissions, and raising public awareness. Long-term monitoring and proactive measures are crucial to safeguard public health and the environment in Tamil Nadu.