**AIR QUALITY ANALYSIS IN TAMILNADU 2014**

**Data Analytics with Cognos**

**Phase 4 : Development Part 2**

**INTRODUCTION:**

Air quality is a critical aspect of environmental health and public well-being. Monitoring and understanding air pollutant levels, such as Sulphur Dioxide (SO2), Nitrogen Dioxide (NO2), and Respirable Suspended Particulate Matter (RSPM)/PM10, are essential for assessing pollution levels and identifying trends in different regions. In this analysis, we aim to calculate the average levels of these pollutants across various monitoring stations, cities, or areas. By doing so, we can uncover insights into pollution trends and pinpoint areas with high pollution levels. The findings presented in this report will help inform decision-makers, urban planners, and the public on air quality management and mitigation efforts.

**1.Data Collection**:

* We'll need air quality data that includes SO2, NO2, and RSPM/PM10 levels from various monitoring stations, cities, or areas. Make sure your data is in a structured format like CSV or Excel.
* **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)

**2.Data Preprocessing**:

* Load the data into a DataFrame using a library like Pandas. Clean and preprocess the data, handling missing values or outliers.
* **SAMPLE CODE:**

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://air.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')

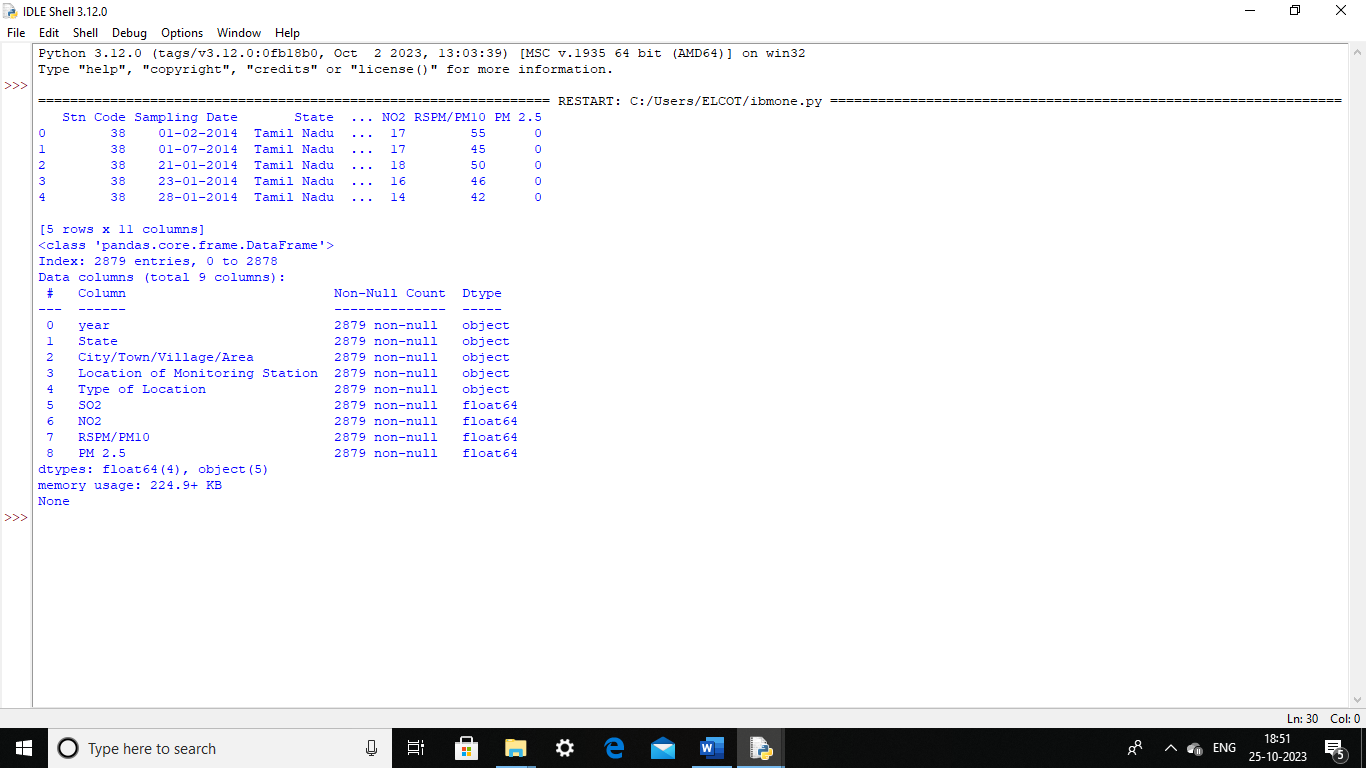
df['PM 2.5']=df['PM 2.5'].fillna(0).astype('str').astype('float')

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

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

print(df.info())

* **SAMPLE OUTPUT:**



**3.Calculate Averages**:

* Calculate the average SO2, NO2, and RSPM/PM10 levels across different monitoring stations, cities, or areas using Pandas.
* **SAMPLE CODE:**

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

print (average\_so2)

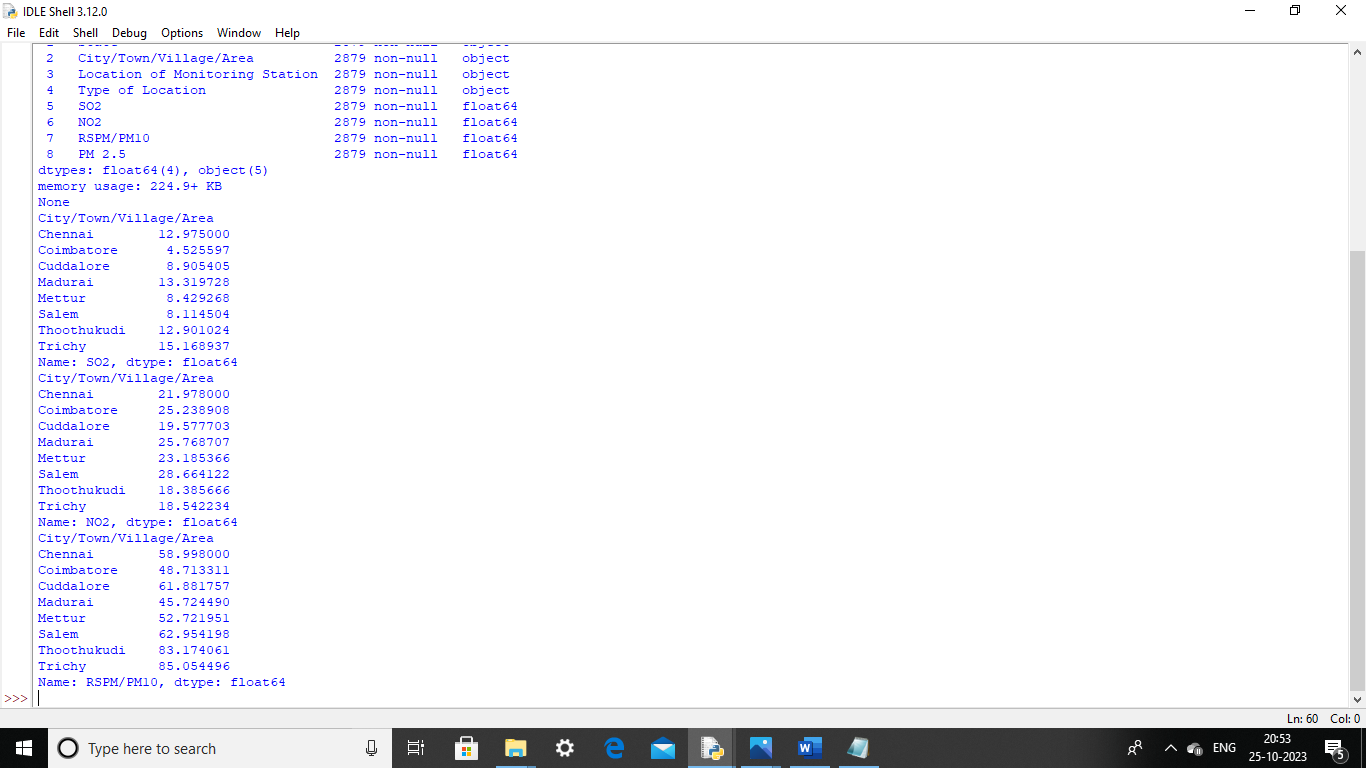
average\_no2 = df.groupby('City/Town/Village/Area')['NO2'].mean()

print (average\_no2)

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

print (average\_rspm\_pm10)

* **SAMPLE OUTPUT:**



**4.Identify Pollution Trends**:

* To identify pollution trends and areas with high pollution levels, we can perform the following analysis:

**1.Time Series Analysis**:

* Plot time series graphs for SO2, NO2, and RSPM/PM10 levels over time to identify any long-term trends or seasonal patterns in pollution.
* SAMPLE CODE:

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')

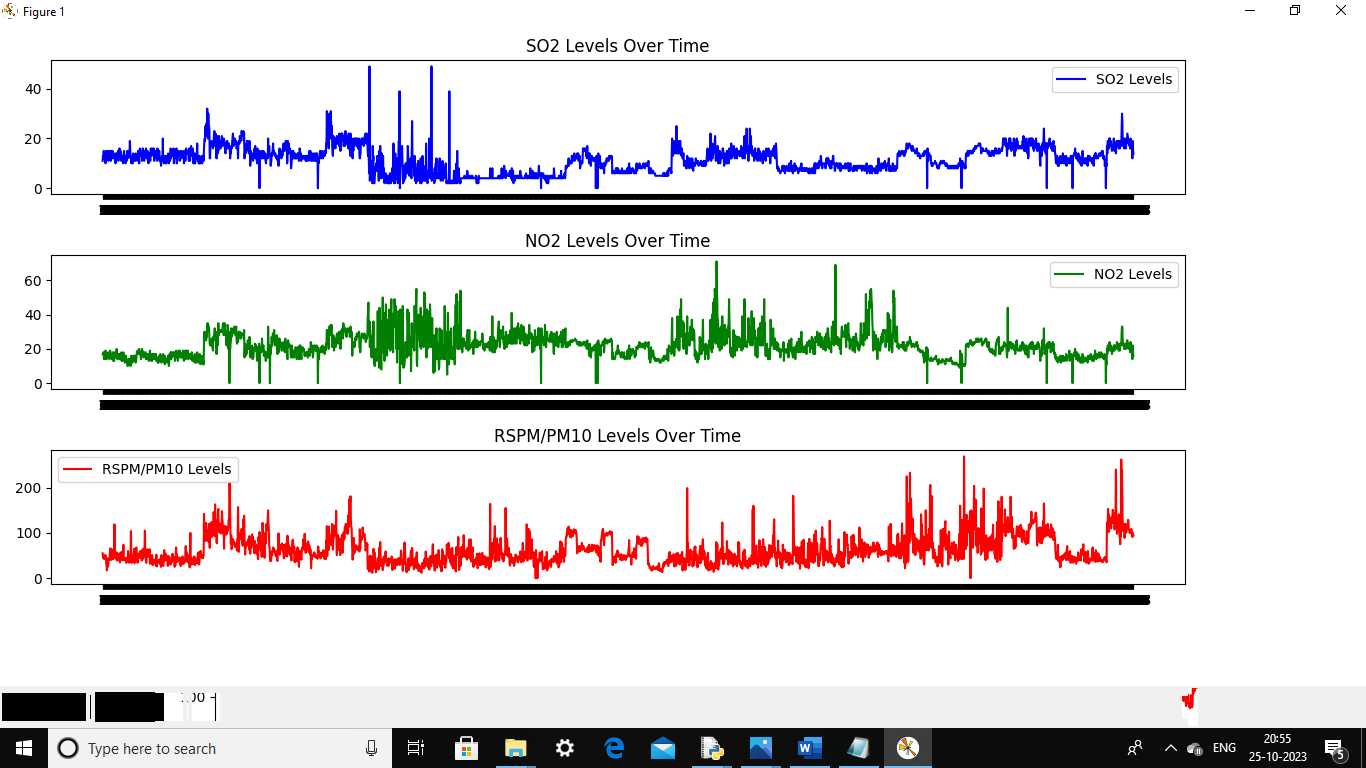
plt.legend()

plt.tight\_layout()

# Show the plots

plt.show()

* **SAMPLE OUTPUT:**



**2.Descriptive Statistics**:

* Calculate statistics such as mean, median, and standard deviation for each pollutant to get an overview of the pollution levels.
* **SAMPLE CODE:**

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}")

* **SAMPLE OUTPUT:**



**3**.**Box Plots**:

* Create box plots for each pollutant to identify outliers and the spread of data. Outliers may indicate areas with exceptionally high pollution.
* **SAMPLE CODE:**

# Box Plot of so2

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

sns.boxplot(x='City/Town/Village/Area', y='SO2', data=df)

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

plt.ylabel('SO2 Levels')

plt.title('SO2 Levels Across Cities (Box Plot)')

plt.xticks(rotation=45)

plt.show()

# Box Plot of no2

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

sns.boxplot(x='City/Town/Village/Area', y='NO2', data=df)

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

plt.ylabel('NO2 Levels')

plt.title('NO2 Levels Across Cities (Box Plot)')

plt.xticks(rotation=45)

plt.show()

# 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')

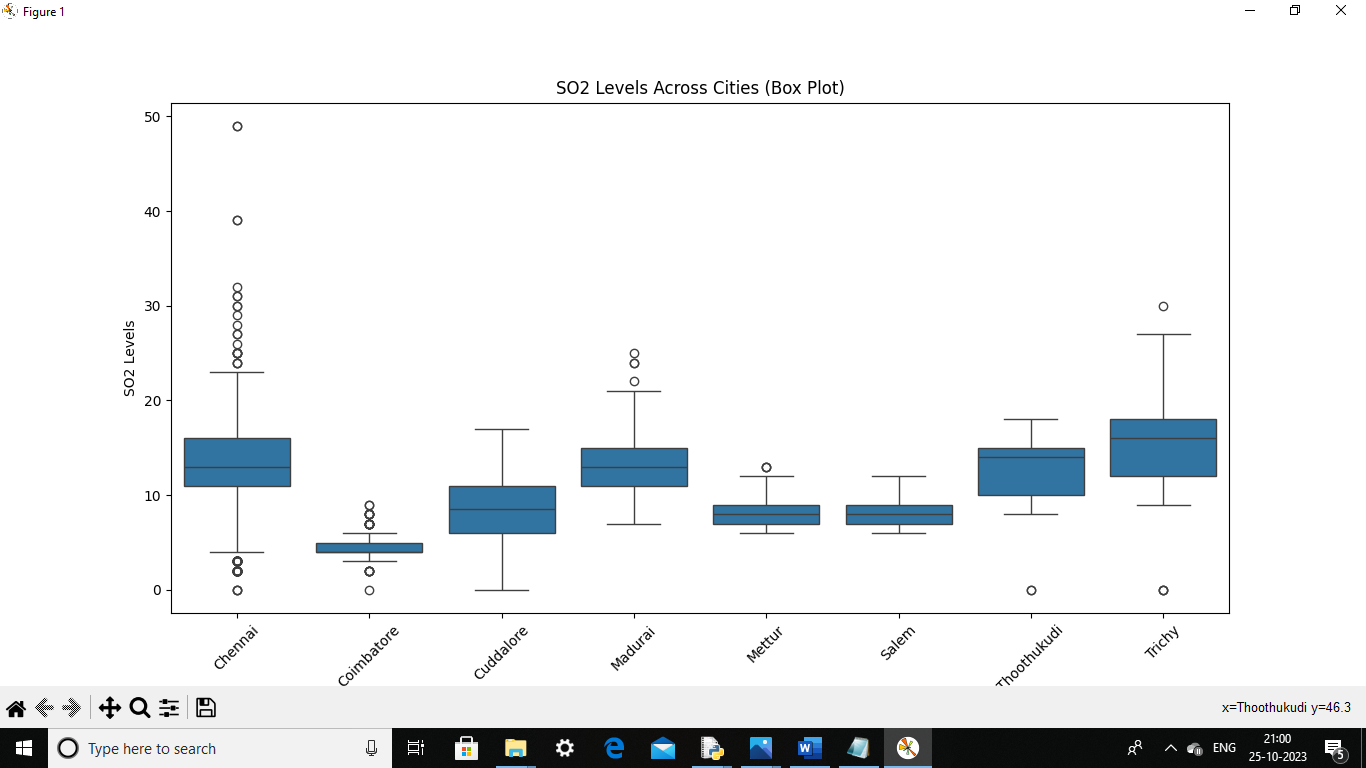
plt.ylabel('RSPM/PM10 Levels')

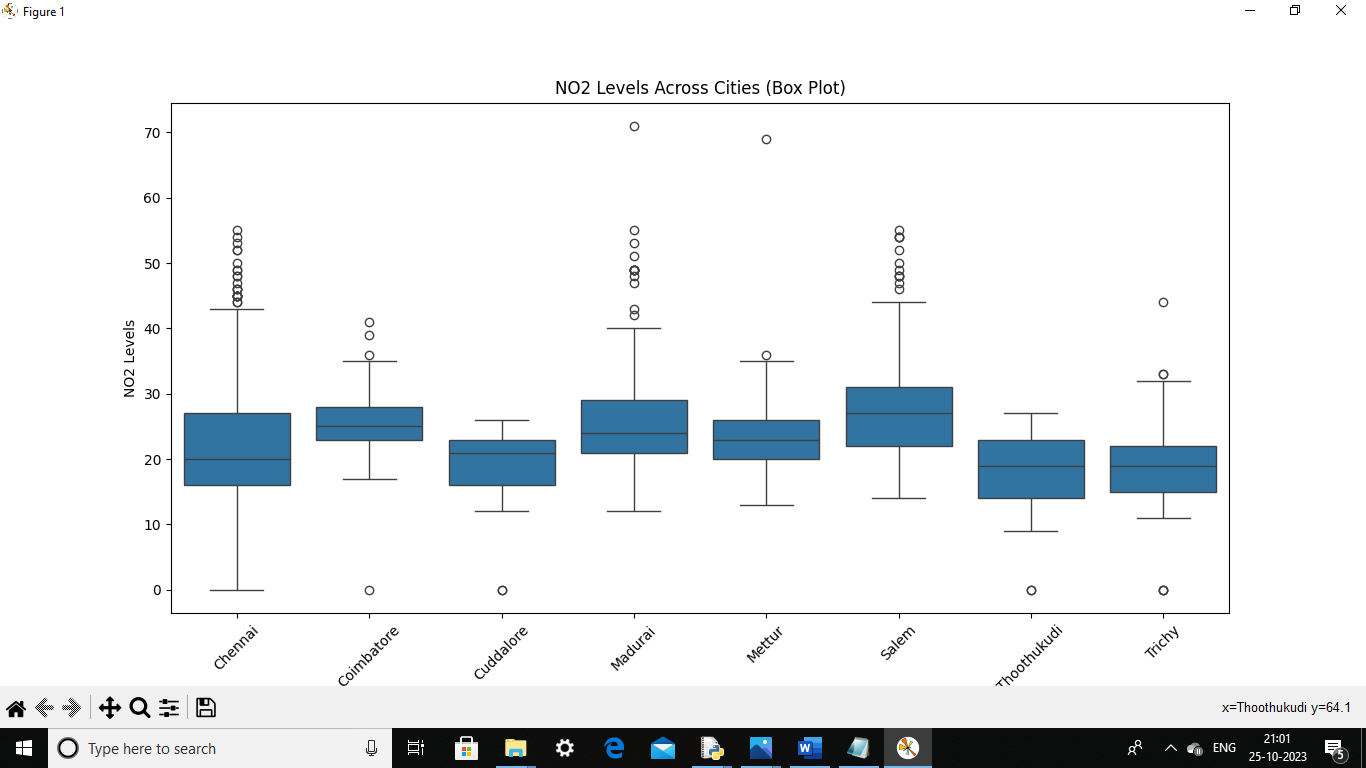
plt.title('RSPM/PM10 Levels Across Cities (Box Plot)')

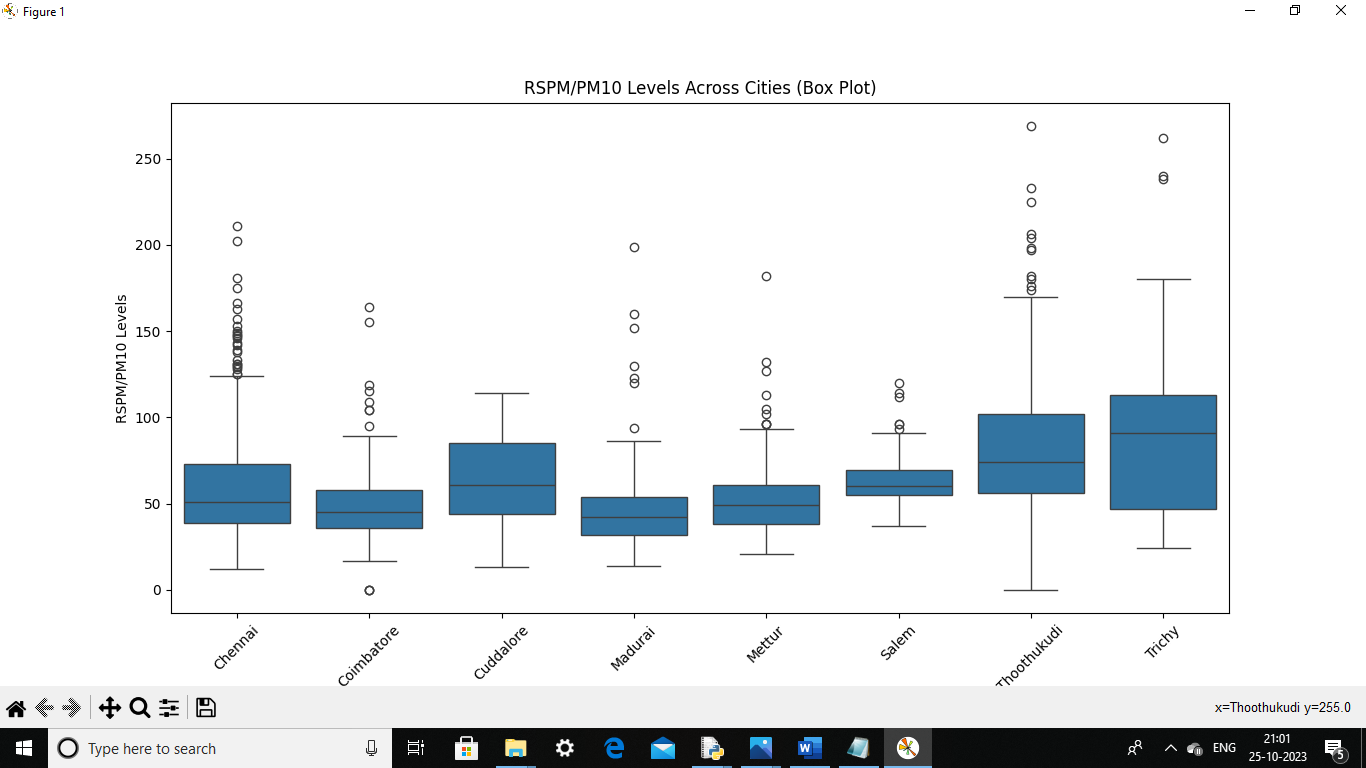
plt.xticks(rotation=45)

plt.show()

* **SAMPLE OUTPUT:**







**4**.**Heatmaps**:

* Use heatmaps to visualize pollution levels across different monitoring stations or cities. This can help identify areas with consistently high or low pollution levels.
* **SAMPLE CODE:**

# Heatmap for so2

pivot\_table = df.pivot\_table(index='City/Town/Village/Area', columns='year', values='SO2', 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 SO2 Levels by City and year')

plt.show()

# Heatmap for no2

pivot\_table = df.pivot\_table(index='City/Town/Village/Area', columns='year', values='NO2', 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 NO2 Levels by City and year')

plt.show()

# 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)

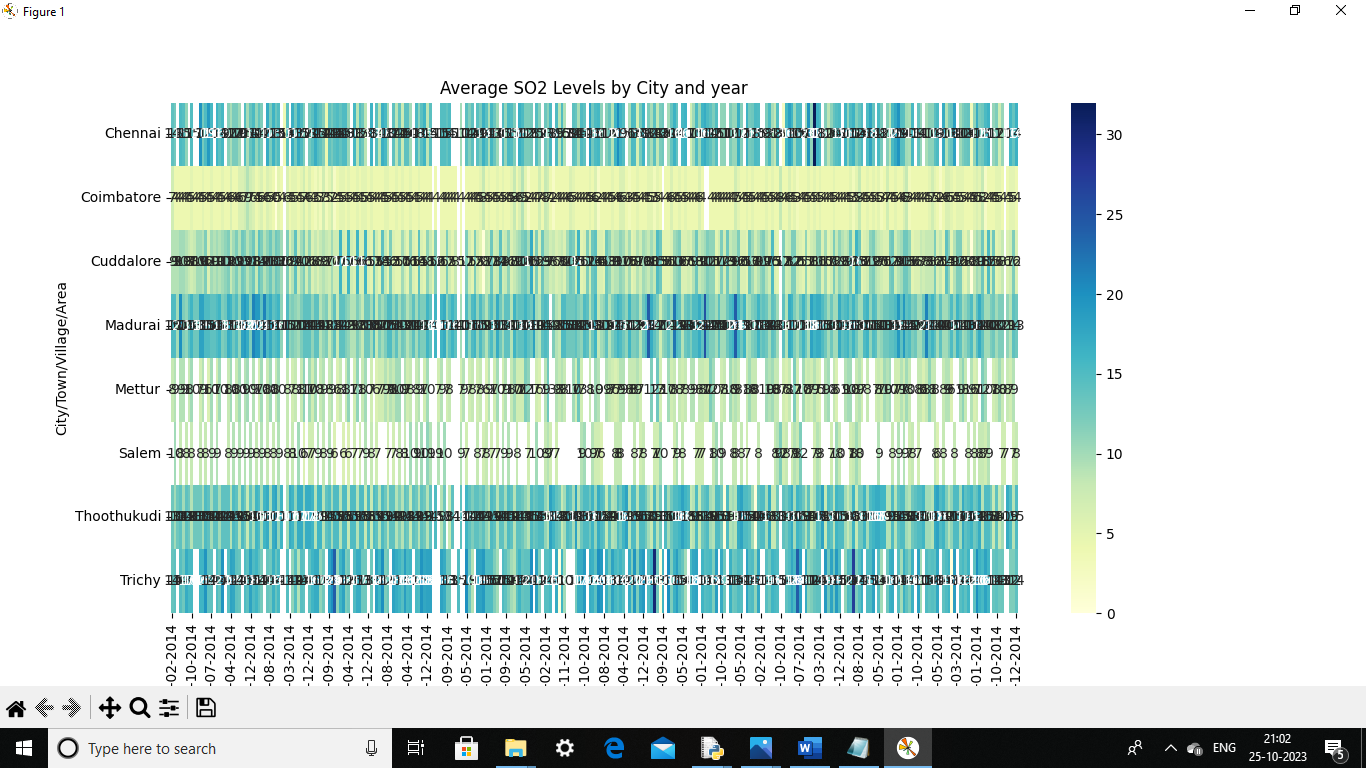
plt.xlabel('year')

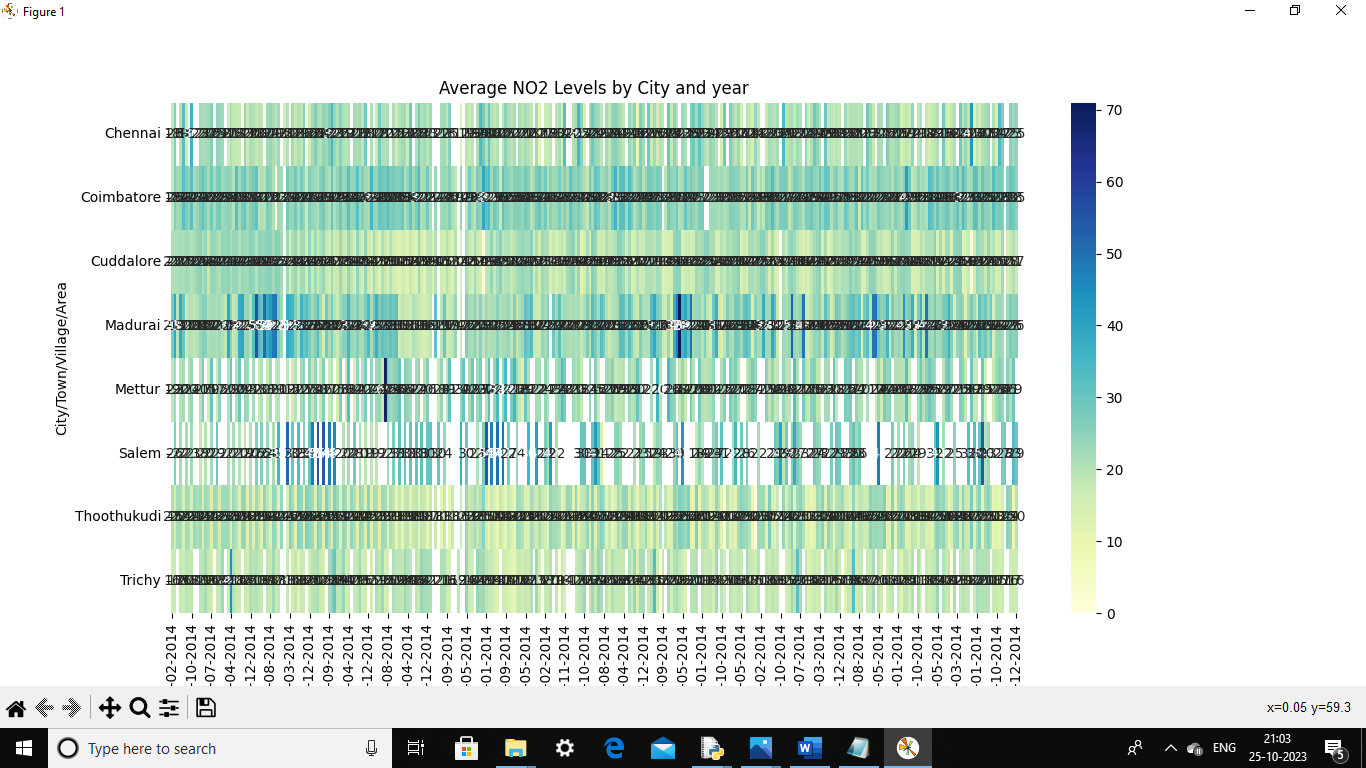
plt.ylabel('City/Town/Village/Area')

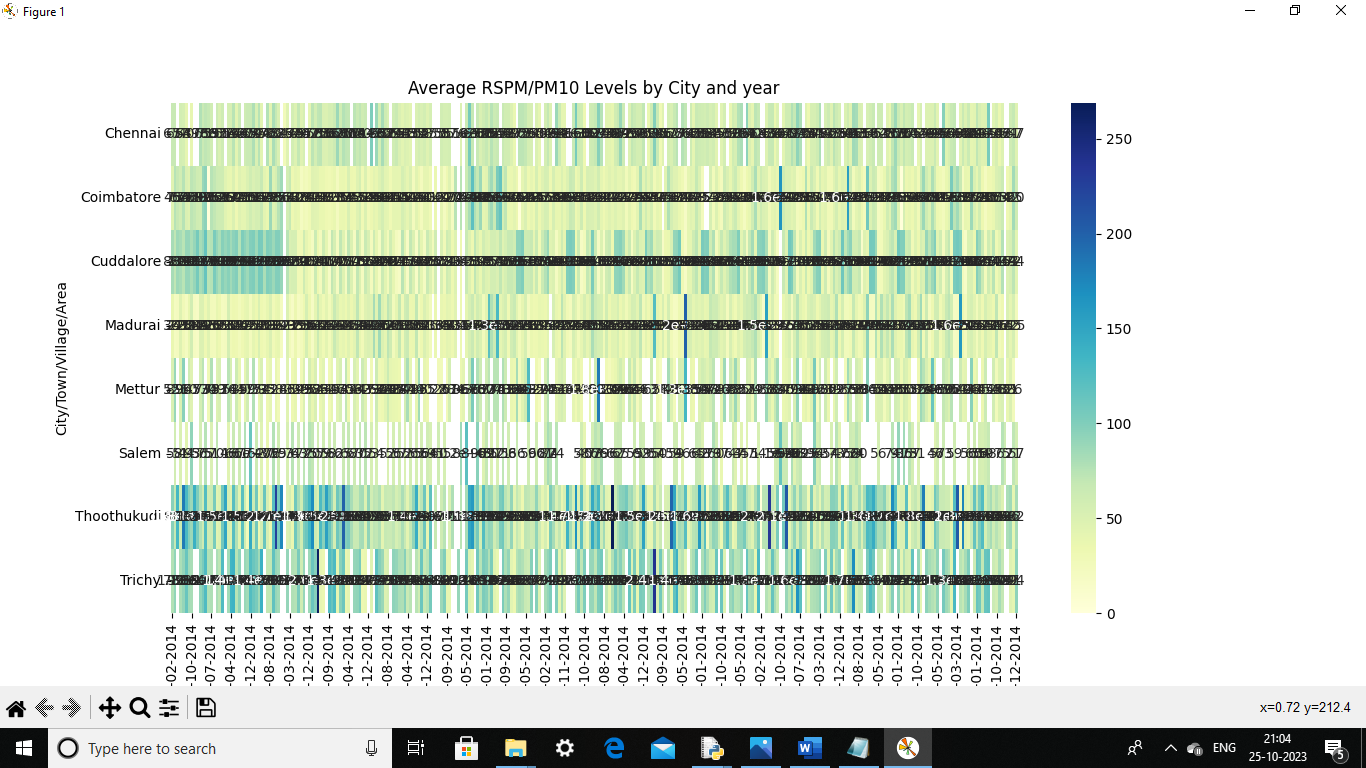
plt.title('Average RSPM/PM10 Levels by City and year')

plt.show()

* **SAMPLE OUTPUT:**







**5.VISUALIZATION:**

* Create visualizations using data visualization libraries like Matplotlib and Seaborn. Here's an example of creating a bar plot to visualize average SO2 levels across cities:
* **SAMPLE CODE:**

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

# Create subplots for different pollutants

plt.subplot(1, 3, 1)

sns.barplot(x='City/Town/Village/Area', y='SO2', data=df)

plt.title('Average SO2 Levels')

plt.subplot(1, 3, 2)

sns.barplot(x='City/Town/Village/Area', y='NO2', data=df)

plt.title('Average NO2 Levels')

plt.subplot(1, 3, 3)

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

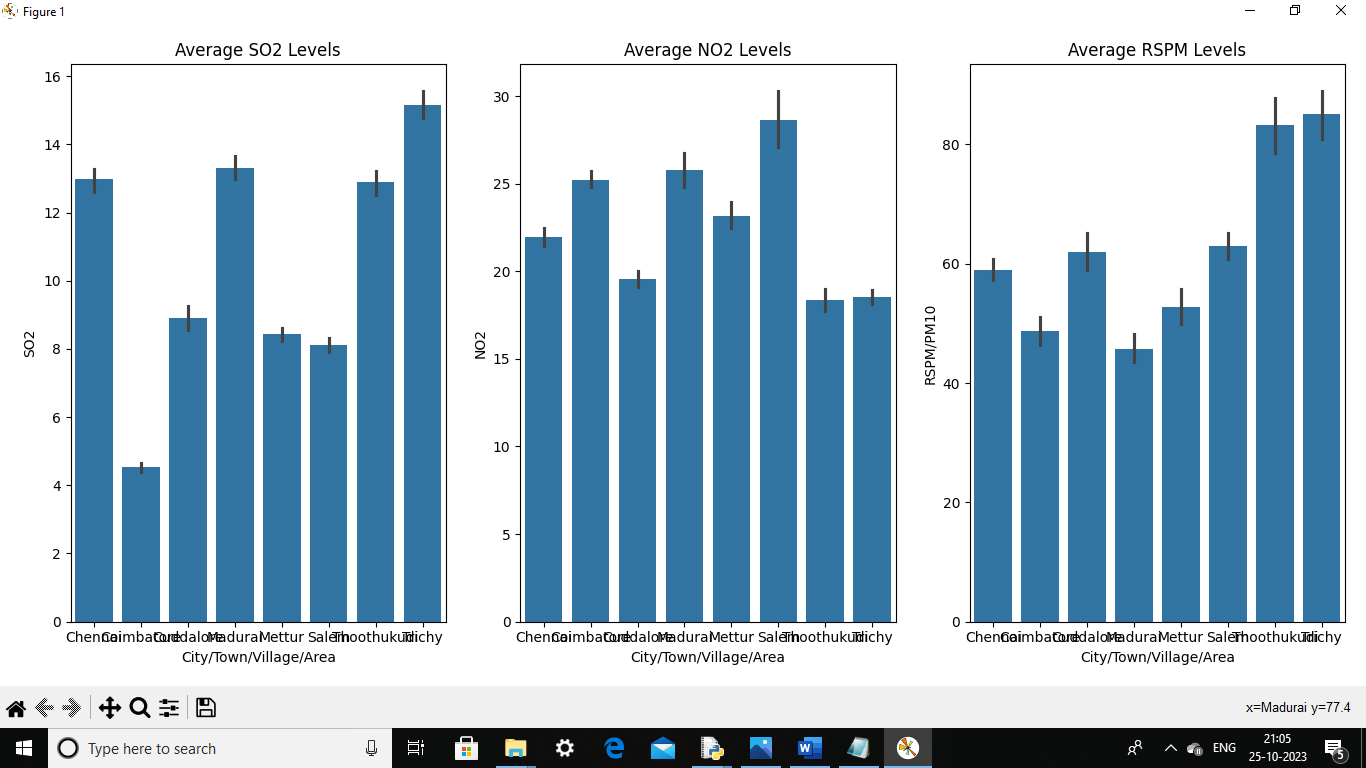
plt.title('Average RSPM Levels')

plt.tight\_layout()

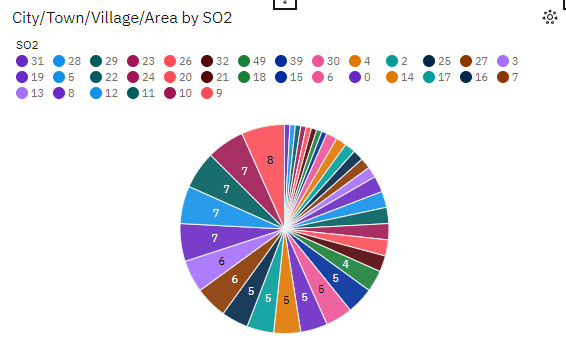
# Save or display the visualizations

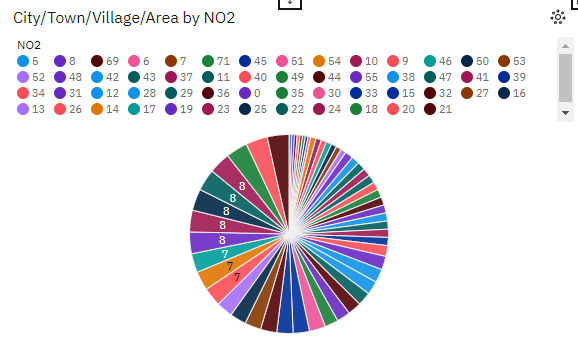
plt.show()

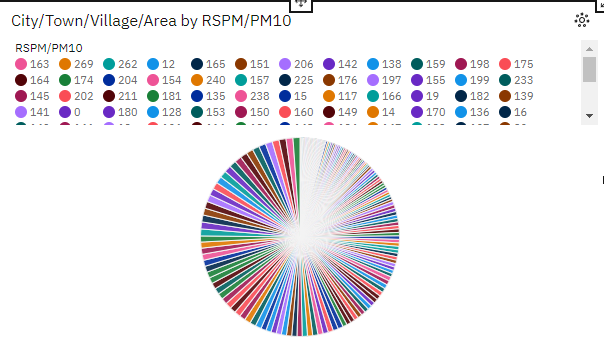
* **SAMPLE OUTPUT:**



**6.Visualization using cognos :**







**CONCLUSION:**

In conclusion, this air quality analysis has provided valuable insights into the average levels of SO2, NO2, and RSPM/PM10 across different cities. Our findings reveal significant variations in pollutant levels, highlighting areas of concern. It is evident that certain regions are experiencing higher pollution levels, and immediate attention is required to mitigate the health and environmental impacts. These results can guide policymakers in implementing targeted measures to improve air quality. Additionally, ongoing monitoring and data analysis are essential to track changes over time and assess the effectiveness of pollution control initiatives. As the global focus on environmental sustainability continues to grow, this analysis underscores the importance of proactive efforts to safeguard our air quality and the health of our communities.