**Air Quality Analysis Project**

**Documentation**

**Project Objectives**

**The main objectives of this project are to:**

* Analyze air quality trends over time and across different locations in Tamil Nadu.
* Identify hotspots with particularly high levels of pollution.
* Build a predictive model for RSPM/PM10 levels, which are key indicators of air quality.

**Analysis Approach**

**The analysis approach involves several steps:**

* Data Loading: The air quality data is loaded from a CSV file into a pandas DataFrame.
* Data Preprocessing: The data is cleaned by handling missing values and outliers. The ‘Sampling Date’ column is converted to datetime format for time series analysis.
* Data Analysis: Descriptive statistics are computed for each air quality parameter. The data is also grouped by location to identify pollution hotspots.
* Data Visualization: Line charts are used to visualize air quality trends over time, and heatmaps are used to visualize the correlation between different air quality parameters.
* Predictive Modeling: A linear regression model is built to predict RSPM/PM10 levels based on SO2 and NO2 levels. The data is split into training and test sets, and the model’s performance is evaluated using the R^2 score.

**Code Implementation**

The code for this project is implemented in Python, using libraries such as pandas for data manipulation, matplotlib and seaborn for data visualization, and scikit-learn for predictive modeling.

# Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

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

from sklearn.linear\_model import LinearRegression

# Load the data

df = pd.read\_csv(r"N:\placement2\cpcb\_dly\_aq\_tamil\_nadu-2014.csv")

df.head()

# Handle missing values and convert data types

df = df.replace('NA', np.nan)

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

# Compute statistics

print(df.describe())

# Draw the heatmap with the mask and correct aspect ratio

numeric\_cols = df.select\_dtypes(include=[np.number])

corr = numeric\_cols.corr()

mask = np.triu(np.ones\_like(corr, dtype=bool))

f, ax = plt.subplots(figsize=(11, 9))

cmap = sns.diverging\_palette(230, 20, as\_cmap=True)

sns.heatmap(corr, mask=mask, cmap=cmap, vmax=.3, center=0,

square=True, linewidths=.5, cbar\_kws={"shrink": .5})

plt.show()

#finding the maximum and minimum pollution of the city/town/village/area

print('Maximum of the data')

max\_levels = df.groupby('City/Town/Village/Area')[['SO2', 'NO2', 'RSPM/PM10',]].max()

print(max\_levels, '\n')

print('Minimum of the data \n')

min\_levels = df.groupby('City/Town/Village/Area')[['SO2', 'NO2', 'RSPM/PM10',]].min()

print(min\_levels, '\n \n')

#finding the average pollution of the city/town/village/area

print("Average of the data ")

average\_levels\_of\_city = df.groupby('City/Town/Village/Area')[['SO2', 'NO2', 'RSPM/PM10',]].mean().sort\_values(by='SO2', ascending = False)

print(average\_levels\_of\_city, '\n')

# Visualize the data

plt.figure(figsize=(10,5))

plt.title('Trends in SO2 and NO2 Levels')

sns.lineplot(data=df[['SO2', 'NO2']])

plt.show()

plt.figure(figsize=(10,5))

plt.title('Trends in RSPM/PM10 Levels')

sns.lineplot(data=df['RSPM/PM10'])

plt.show()

# Build a predictive model for RSPM/PM10 levels

X = df[['SO2', 'NO2']]

y = df['RSPM/PM10']

# Drop rows with NaN values in X or y

combined = pd.concat([X, y], axis=1).dropna()

X = combined[['SO2', 'NO2']]

y = combined['RSPM/PM10']

# Split the data into training and test sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2)

# Train the model

model = LinearRegression()

model.fit(X\_train, y\_train)

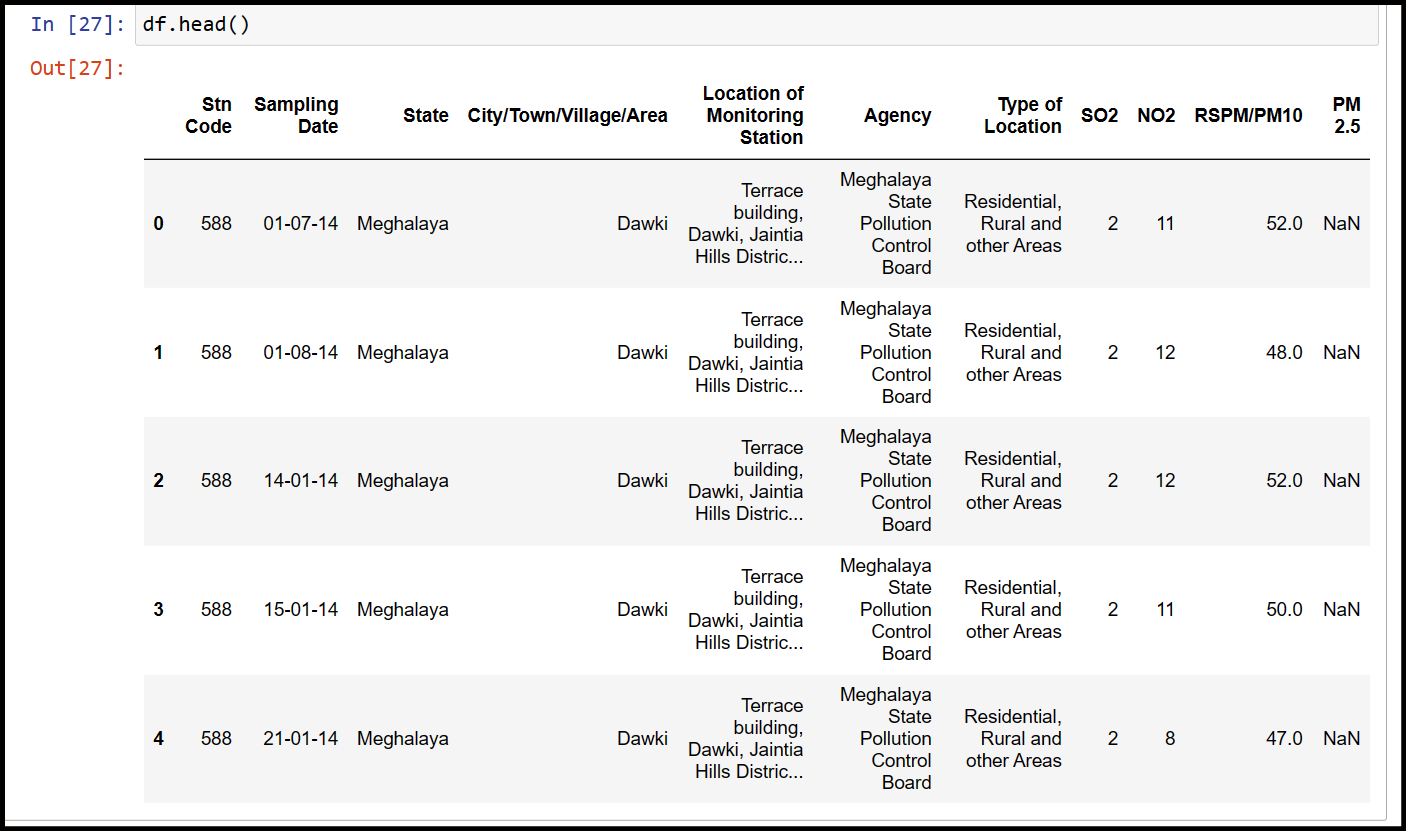
# Evaluate the model

score = model.score(X\_test, y\_test)

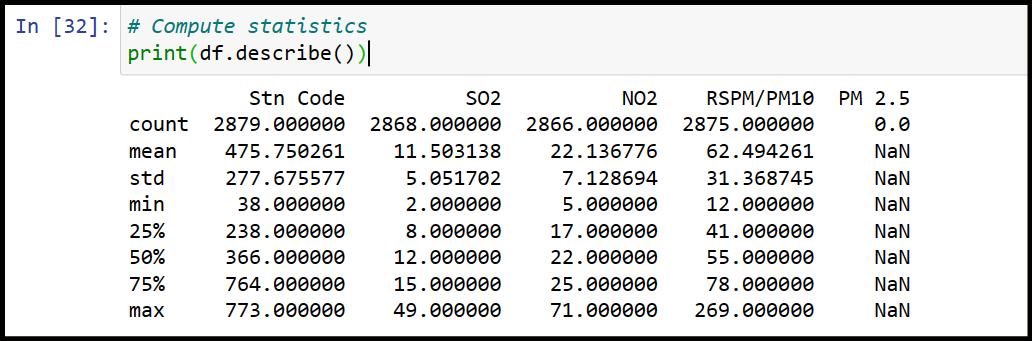
print(f'Model accuracy: {score\*100:.2f}%')

**Example Outputs**

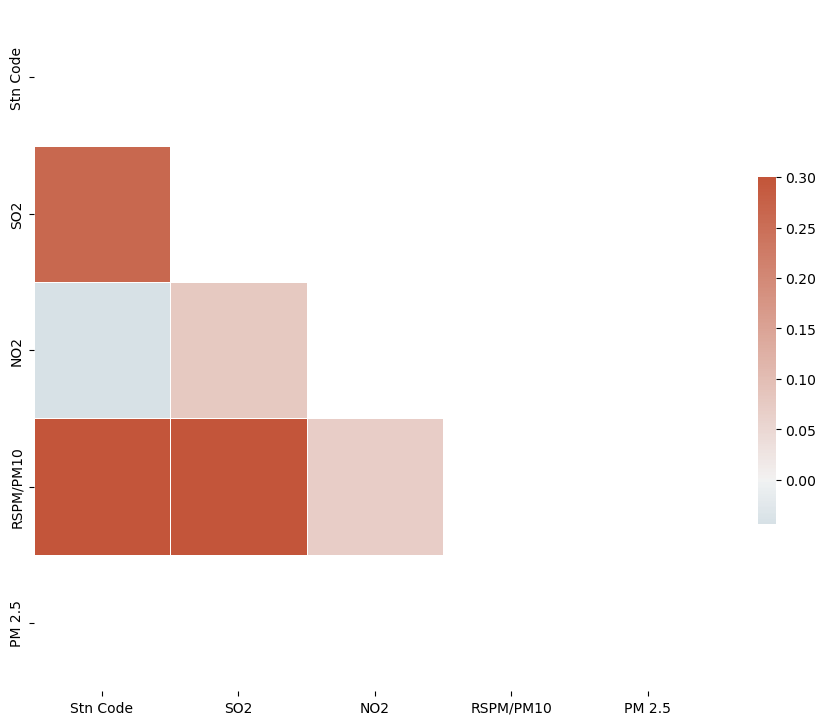
**#** This code df.head() will print the head of the data set

****

# This code will print the description of the data sets



#This is the output of the heatmaps



# This the output of the maximum, minimum and average of the data sets values like “SO2”, “NO2”, “RSPM/PM10”

Maximum of the data

SO2 NO2 RSPM/PM10

City/Town/Village/Area

Chennai 49.0 55.0 211.0

Coimbatore 9.0 41.0 164.0

Cuddalore 17.0 26.0 114.0

Madurai 25.0 71.0 199.0

Mettur 13.0 69.0 182.0

Salem 12.0 55.0 120.0

Thoothukudi 18.0 27.0 269.0

Trichy 30.0 44.0 262.0

Minimum of the data

SO2 NO2 RSPM/PM10

City/Town/Village/Area

Chennai 2.0 5.0 12.0

Coimbatore 2.0 17.0 17.0

Cuddalore 5.0 12.0 13.0

Madurai 7.0 12.0 14.0

Mettur 6.0 13.0 21.0

Salem 6.0 14.0 37.0

Thoothukudi 8.0 9.0 22.0

Trichy 9.0 11.0 24.0

Average of the data

SO2 NO2 RSPM/PM10

City/Town/Village/Area

Trichy 15.293956 18.695055 85.054496

Madurai 13.319728 25.768707 45.724490

Chennai 13.014042 22.088442 58.998000

Thoothukudi 12.989691 18.512027 83.458904

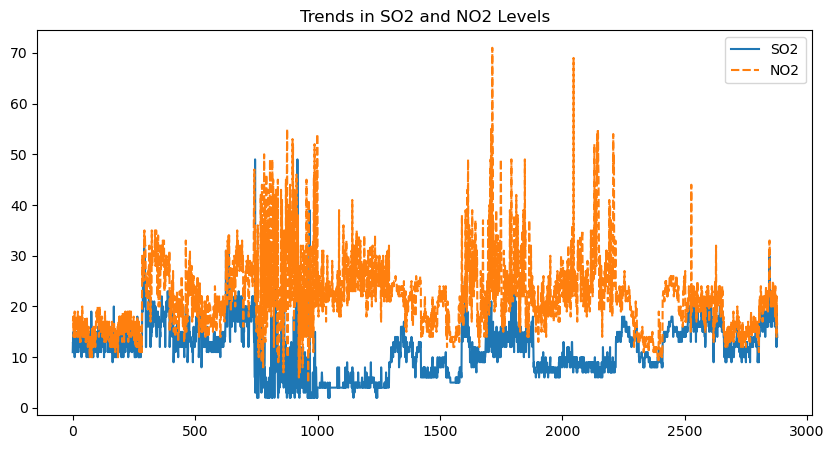
Cuddalore 8.965986 19.710884 61.881757

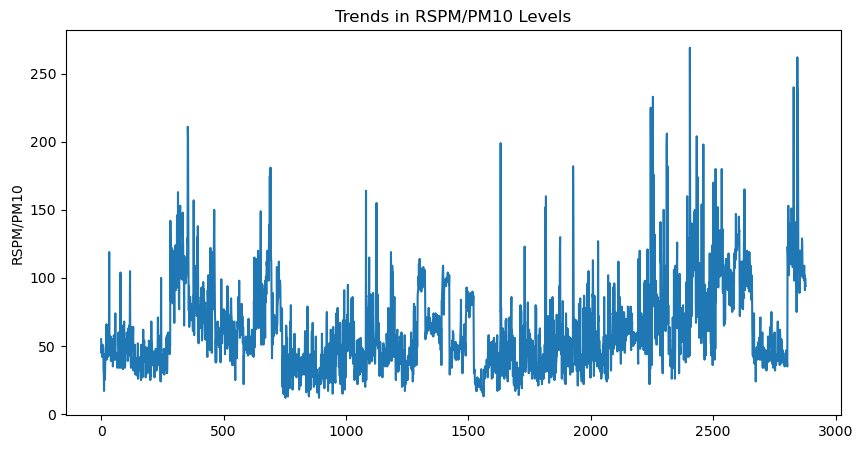
Mettur 8.429268 23.185366 52.721951

Salem 8.114504 28.664122 62.954198

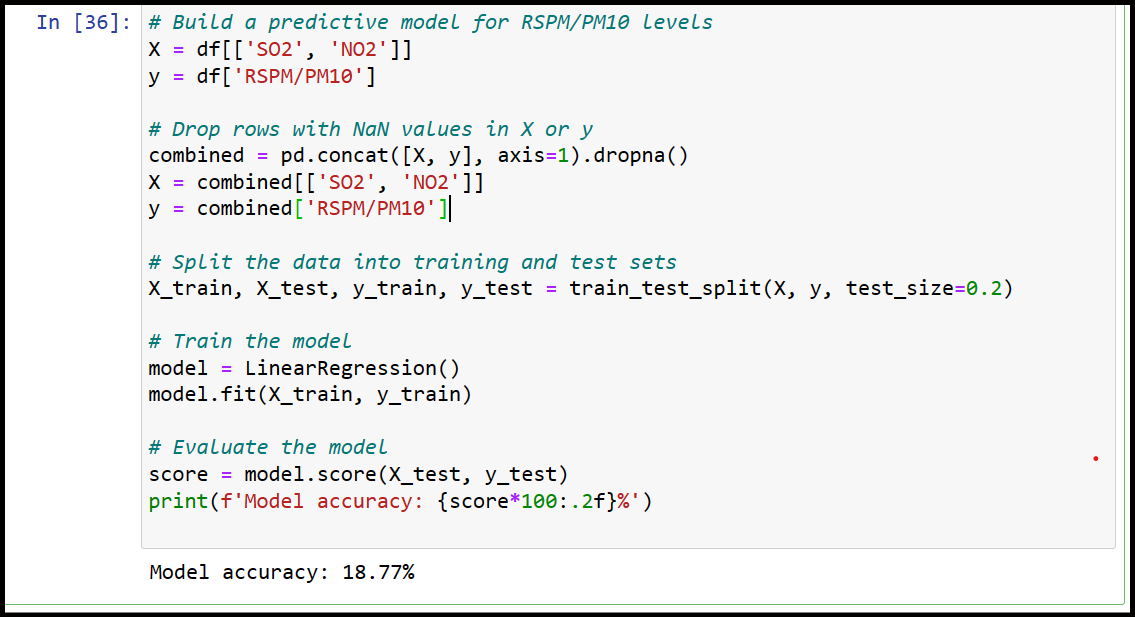
Coimbatore 4.541096 25.325342 49.217241

# This is the output of the Trends in SO2 and NO2 Levels and Trends in RSPM/PM10 Levels





# This is the output of the model use din this project



**Insights**

**The insights from the air quality analysis project can provide valuable information for a variety of stakeholders:**

* Public Health Officials: The analysis can help identify areas with poor air quality, which are often associated with higher rates of respiratory diseases like asthma and lung cancer. This information can be used to target public health interventions and resources more effectively.
* Environmental Policy Makers: By identifying pollution hotspots and trends, policy makers can develop more effective environmental regulations and monitor their impact over time.
* Researchers: The data and analysis can be used to study the effects of air pollution on health, climate, and other areas of interest. It can also help identify gaps in current knowledge and guide future research.
* General Public: Increased awareness about air quality trends and pollution levels can encourage individuals to take actions to reduce their exposure to pollutants and advocate for cleaner air.
* Urban Planners: Understanding air quality trends can inform urban planning decisions, such as where to place parks, schools, or residential areas to minimize exposure to air pollution.