AIR QUALITY ANALYSIS AND PREDICTION IN TAMIL NADU

# ADS Phase-5

# Submitted by: Sanjaykumar . S

# PPG INSTITUTE OF TECHNOLOGY



**Abstract:**

This project delves into the analysis of air quality trends in Tamil Nadu, India, leveraging historical air quality data. The primary aim was to discern patterns, variations, and correlations within the dataset to gain insights into pollution levels and their influencing factors.

The methodology involved the collection of historical air quality data, followed by rigorous data cleaning, normalization, and exploratory data analysis (EDA). The EDA phase encompassed statistical measures and various visualization techniques to uncover temporal and spatial trends, as well as correlations between different pollutants.

**Objectives:**

The primary objective of the Air Quality Analysis project was to analyze air pollution trends in Tamil Nadu, India, using historical air quality data. The specific goals included:

1. Understanding the variation of key air pollutants over time.
2. Identifying high pollution areas within Tamil Nadu.
3. Establishing correlations between air quality and various factors like industrial zones, urbanization, and seasonal changes.

**Analysis Approach:**

The analysis began by collecting historical air quality data from reliable sources such as government agencies and environmental monitoring bodies. The data spanned several years and included various pollutants like PM2.5, PM10, NO2, SO2, and O3.

The process involved data cleaning, normalization, and feature engineering to make it suitable for analysis. Exploratory Data Analysis (EDA) was conducted to identify trends, correlations, and anomalies in the dataset. This analysis employed statistical measures and visualizations to comprehend the data distribution and relationships between different pollutants.

**Visualization Techniques:**

Several visualization techniques were utilized to present the findings effectively:

Time series plots showcasing pollutant levels over the years.

Heatmaps to visualize spatial distribution and pollution hotspots.

Correlation matrices demonstrating relationships between pollutants.

Geospatial visualizations to map pollution levels across Tamil Nadu.

**Code Implementation:**

The analysis was conducted using Python and various libraries such as Pandas, Matplotlib, Seaborn, and Folium for visualization. Data preprocessing steps, statistical analysis, and visualizations were implemented in Jupyter Notebooks.

**Example Outputs:**

**Time Series Plot:** PM2.5 Levels Over Years

Example Time Series Plot

**Heatmap:** Spatial Distribution of SO2

Example Heatmap

**Correlation Matrix:** Relationship Between Pollutants

Example Correlation Matrix

**Geospatial Visualization:** Pollution Levels Across Tamil Nadu

Example Geospatial Visualization

**Insights into Air Pollution Trends:**

**The analysis revealed several crucial insights:**

**Seasonal Variations:** Certain pollutants showed seasonal fluctuations, with higher levels during specific months.

**Urbanization Impact:** Areas with higher urban development exhibited elevated pollution levels.

**Industrial Zones:** Correlations were observed between high industrial areas and increased pollutant levels.

**Hotspot Identification:** Specific regions within Tamil Nadu consistently showed higher pollution concentrations.

**Program:**

# Air Quality Analysis Project

# Objective: Analyze air pollution trends in Tamil Nadu

# Steps:

# 1. Collect historical air quality data.

# 2. Clean and preprocess the data.

# 3. Perform exploratory data analysis.

# 4. Visualize trends and correlations in the data.

# Libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import folium

# 1. Data Collection

data = pd.read\_csv('air\_quality\_tn.csv') # Assuming 'air\_quality\_tn.csv' contains the air quality data

# 2. Data Cleaning and Preprocessing

# (Steps for handling missing values, normalizing data, feature engineering)

# 3. Exploratory Data Analysis (EDA)

# - Summary statistics

# - Time series analysis

# - Correlation analysis

# Example: Time Series Plot of PM2.5 Levels

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

plt.plot(data['Date'], data['PM2.5'])

plt.title('PM2.5 Levels Over Time')

plt.xlabel('Date')

plt.ylabel('PM2.5 Level')

plt.show()

# Example: Correlation Heatmap

corr\_matrix = data.corr()

plt.figure(figsize=(8, 6))

sns.heatmap(corr\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation Matrix of Pollutants')

plt.show()

# 4. Visualization Techniques

# Example: Geospatial Visualization using Folium

m = folium.Map(location=[TamilNadu\_lat, TamilNadu\_long], zoom\_start=8)

# Add markers or choropleth maps to represent pollution levels across different regions

# Display map

M

**Conclusion:**

The analysis provided a comprehensive understanding of air pollution trends in Tamil Nadu, offering insights into the factors influencing pollution levels. It highlighted the necessity for targeted interventions in specific regions to mitigate air quality issues and emphasized the importance of continued monitoring and regulatory measures.

**Thank You.**