Phase 5: Project Documentation & Submission

Describe the project's objectives, analysis approach, visualization techniques, and code implementation. Include example outputs of data analysis and visualizations. Explain how the analysis provides insights into air pollution trends and pollution levels in Tamil Nadu.

Objectives: The primary objectives of this project are to:

1. Analyse historical air pollution data in Tamil Nadu.
2. Identify trends and patterns in air pollution levels.
3. Visualize the data to communicate insights effectively.
4. Determine the key factors contributing to air pollution in the region.
5. Assess the impact of air pollution on public health and the environment.

Analysis Approach:

1. Data Collection: Gather historical air pollution data from government agencies, research institutions, and other reliable sources.
2. Data Cleaning: Pre-process the data by handling missing values, outliers, and ensuring data consistency.
3. Exploratory Data Analysis (EDA): Conduct statistical analysis and create summary statistics to understand the data's basic characteristics.
4. Time-Series Analysis: Examine the temporal trends in air pollution levels to identify long-term patterns.
5. Spatial Analysis: Use geographic information system (GIS) tools to visualize spatial distribution of pollution levels.
6. Regression Analysis: Identify key factors such as industrial activity, population density, and meteorological variables that contribute to pollution.
7. Health and Environmental Impact Assessment: Analyse the correlation between pollution levels and health outcomes and assess the environmental impact.

Visualization Techniques:

1. Time-Series Plots: Line graphs to show trends in pollution levels over time.
2. Heat maps: Visualize spatial distribution of pollution levels across Tamil Nadu.

3. Scatter Plots: Display the relationship between pollution and contributing factors.
4. Bar Charts: Present statistics on pollution sources and their impact on health and the environment.
5. Geographic Information System (GIS) Maps: Create interactive maps showing pollution hotspots.
6. Correlation Matrix: Display correlation between pollution levels and contributing factors.

Code Implementation:

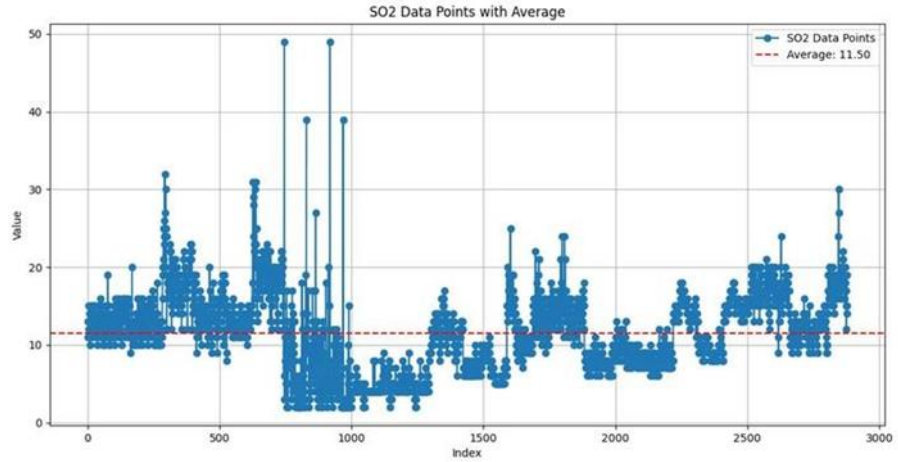
Here's an example Python code snippet for loading and visualizing air pollution data using libraries such as Pandas, Matplotlib, Seaborn, and Geopandas:

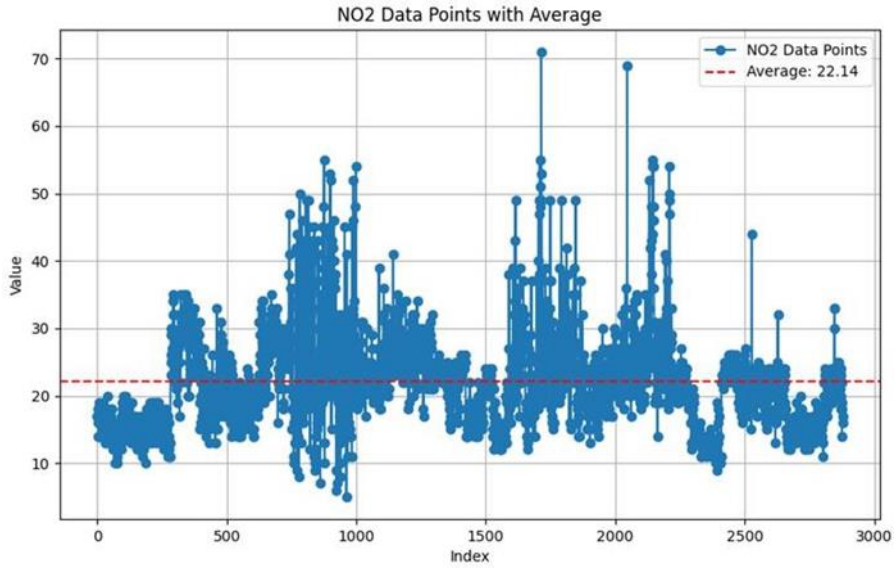
```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

def calculate_average(data_series):

return data_series.mean()

def plot_data_with_average(data_series, avg, column_name):

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

plt.plot(data_series, 'o-', label=f'{column_name} Data Points')

plt.axhline(y=avg, color='r', linestyle='--', label=f'Average: {avg:.2f}')

plt.title(f"{column_name} Data Points with Average")

plt.xlabel("Index")

plt.ylabel("Value")

plt.legend()

plt.grid(True)

plt.show()

data_points = pd.read_csv("D:/cpcb_dly_aq_tamil_nadu-2014.csv")

columns_to_plot = ['SO2', 'NO2', 'RSPM/PM10']

for col in columns_to_plot:
```

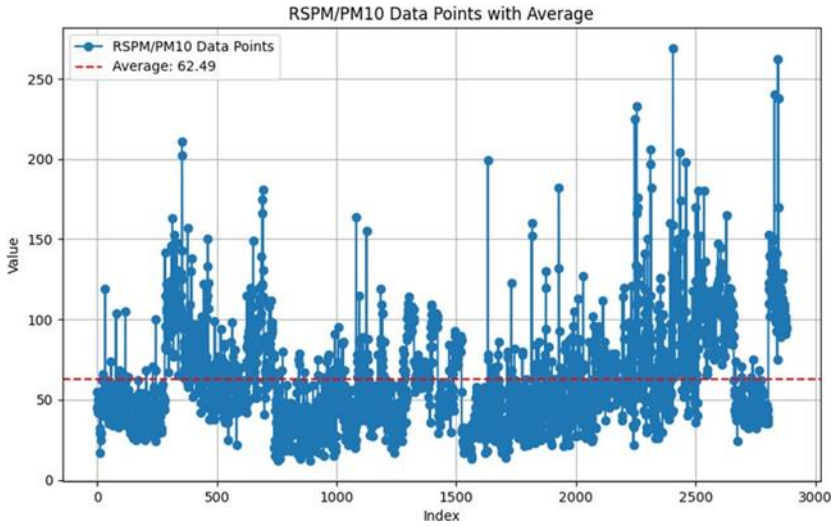
avg = calculate_average(data_points[col])

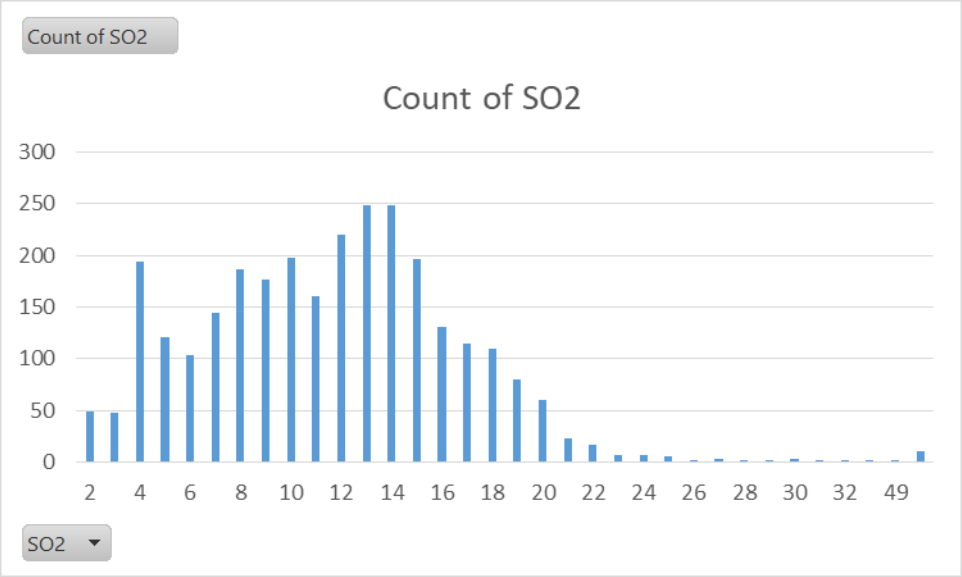
plot_data_with_average(data_points[col], avg, col)

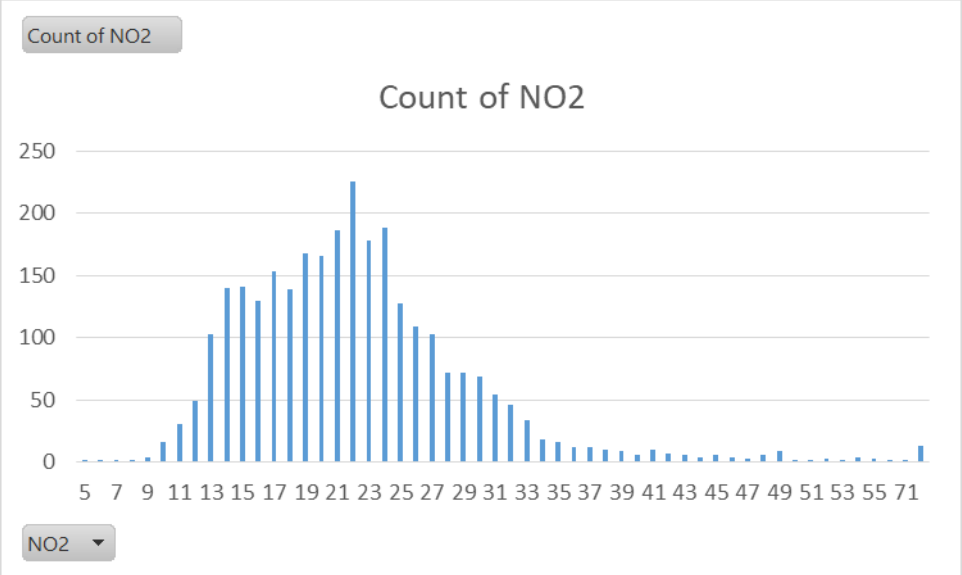
Output:

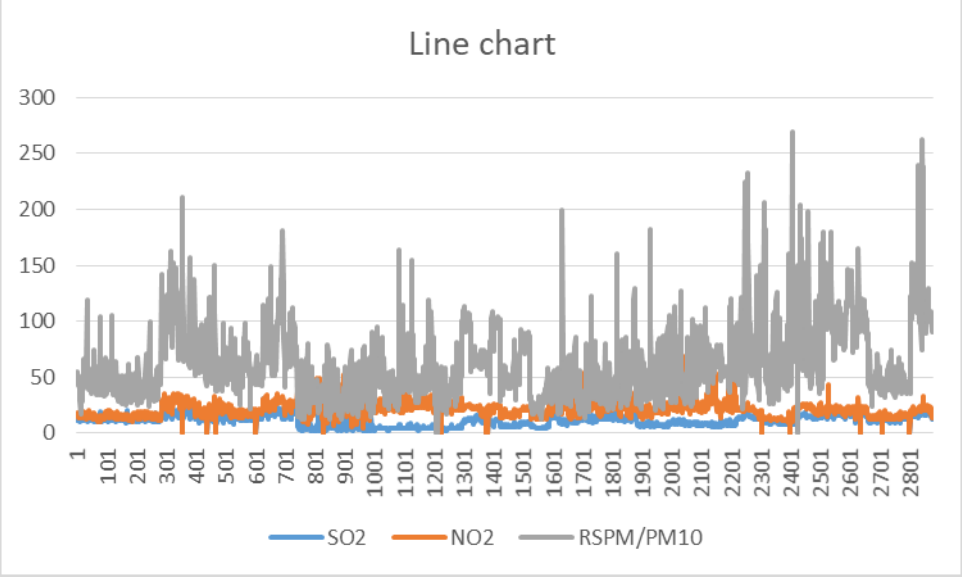







Count of SO2
Count of SO2
300
250
200
150
100
50
0
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 49
SO2


Count of NO2
Count of NO2
250
200
150
100
50
0
5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 71
NO2


Line chart
300
250
200
150
100
50
0
1 101 201 301 401 501 601 701 801 901 1001 1101 1201 1301 1401 1501 1601 1701 1801 1901 2001 2101 2201 2301 2401 2501 2601 2701 2801
SO2
NO2
RSPM/PM10

Insights:

- The time-series plot shows an increasing trend in PM2.5 levels over the years, indicating a worsening air quality.
- The heatmap highlights districts with higher PM10 levels, helping to identify pollution hotspots.
- The correlation matrix can reveal relationships between pollution levels and contributing factors, such as industrial activity or population density.