

Phase 5

Air Quality Analysis Project

Project Objectives

The main objectives of this project are:

1. **Data Collection and Preprocessing:** We obtained the air quality dataset for Tamil Nadu in 2014 from the official data portal. We loaded the dataset using Python's `pandas` library and performed data preprocessing tasks, such as handling missing values, data cleaning, and column renaming.
2. **Data Exploration:** We conducted exploratory data analysis (EDA) to gain an initial understanding of the dataset. This involved summary statistics, data distribution, and the identification of outliers.
3. **Trend Analysis:** We examined the time series data to identify trends in air quality over the year, including variations in pollutant levels in different seasons.
4. **Location-wise Analysis:** We analyzed air quality data on a location-by-location basis to compare and contrast air quality across different areas in Tamil Nadu.
5. **Visualization:** To make the analysis more accessible, we used Python libraries such as `matplotlib` and `seaborn` for creating visualizations. We generated line charts, scatter plots, and heatmaps to illustrate trends and variations in air quality.

Code Implementation

Below are simplified example code snippets used for data analysis and visualization:

```
# Python code for data analysis and visualization
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the preprocessed dataset
df = pd.read_csv("air_quality_data.csv")

# Example 1: Time Series Analysis
```

```
# Calculate and plot the monthly average of a specific pollutant
monthly_avg_pm25 = df.groupby('Month')['PM2.5'].mean()
monthly_avg_pm25.plot(kind='line')
plt.title('Monthly Average PM2.5 Levels')
plt.xlabel('Month')
plt.ylabel('PM2.5 Level')
plt.show()

# Example 2: Location-wise Analysis
# Create a heatmap to visualize the air quality across different locations
location_air_quality = df.pivot_table(values='PM10', index='Location', columns='Month', aggfunc='mean')
sns.heatmap(location_air_quality, cmap='YlGnBu')
plt.title('Air Quality Across Locations')
plt.show()
```

Replicating the Analysis:

1. **Data Collection:** Obtain the air quality dataset for Tamil Nadu in 2014 from the official data portal, or use the provided dataset link:
<https://tn.data.gov.in/resource/location-wise-daily-ambient-air-quality-tamil-nadu-year-2014>.
2. **Data Preprocessing:** Load the dataset using Python's `pandas` library and perform data preprocessing tasks, such as handling missing values, data cleaning, and column renaming. Ensure you have the necessary libraries installed.

```
import pandas as pd
df = pd.read_csv("air_quality_data.csv") # Load the dataset
# Perform data preprocessing tasks here (e.g., handling missing values).
```

3. **Data Exploration:** Conduct exploratory data analysis (EDA) to gain an initial understanding of the dataset, including summary statistics, data distribution, and identification of outliers.

```
# Explore the data using various pandas functions.
```

4. **Trend Analysis:** Examine the time series data to identify trends in air quality over the year, including variations in pollutant levels in different seasons.

```
# Analyze the data for time series trends.
```

5. **Location-wise Analysis:** Analyze air quality data on a location-by-location basis to compare and contrast air quality across different areas in Tamil Nadu.

```
# Perform location-wise analysis.
```

6. **Visualization:** Create visualizations to illustrate trends and variations in air quality using Python libraries such as `matplotlib` and `seaborn`.

```
import matplotlib.pyplot as plt
import seaborn as sns

# Generate visualizations (e.g., line charts, scatter plots, and heatmaps).
```

Key Findings:

Here's a summary of the key findings from the air quality analysis:

- **Monthly PM2.5 Trend Analysis:** The analysis revealed that air quality deteriorated in certain months, suggesting potential pollution sources or weather-related factors affecting air quality.
- **Location-wise Analysis:** Variations in air quality across different regions of Tamil Nadu were observed. Some areas consistently exhibited better air quality than others, indicating potential hotspots of pollution.
- **Heatmap Visualizations:** Heatmaps helped identify locations with consistent air quality issues, which can be valuable for policymakers and the public in addressing air quality-related challenges.

By following the provided instructions, you can replicate the air quality analysis and obtain similar findings to gain insights into air pollution trends and pollution levels in Tamil Nadu for the year 2014.