**AIR QUALITY ANALYSIS**

Project: Project Documentation & submission

**Documentation**

***OBJECTIVES*:**

**1. Inform and Educate:** One of the primary objectives of a 500-word piece is to provide readers with valuable information and education on a specific topic. It should aim to increase the reader's understanding and knowledge on the subject matter.

**2. Entertain and Engage:** Depending on the nature of the content, your objective might be to entertain and engage the readers. This is often the case for creative writing, humor pieces, or stories designed to captivate the audience.

**3. Persuade and Convince:** If your piece is persuasive in nature, your objective is to convince the reader to adopt a certain viewpoint, take action, or change their behavior. Persuasive writing aims to influence opinions and decisions.

**4. Analyze and Evaluate:** In an analytical or critical essay, the objective is to analyze a subject, topic, or argument, and provide an evaluation or critique. This type of writing requires a deep understanding of the subject and the ability to assess its strengths and weaknesses.

**5. Describe and Explain:** Descriptive and explanatory writing aims to provide a clear and comprehensive description of a subject or event. The objective is to ensure that readers have a vivid and complete understanding of what is being described.

**6. Narrate and Tell a Story:** Narrative writing objectives involve telling a story or recounting events. The goal is to engage readers by taking them on a journey through the narrative and creating an emotional connection.

**7. Instruct and Guide:** If you're writing a how-to or instructional piece, your objective is to guide the reader through a process, teach them a skill, or provide step-by-step instructions to achieve a specific outcome.

**8. Raise Awareness:** In some cases, the primary objective of a 500-word piece might be to raise awareness about a particular issue, problem, or cause. This type of writing often serves a social or advocacy purpose.

**Analysis Approach:**

**Data Collection:**

* Gather air quality data from various monitoring sources, which may include ground-based monitoring stations, satellite data, and sensor networks.
* Collect data on pollutants such as particulate matter (PM2.5 and PM10), ground-level ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), and meteorological parameters like temperature, humidity, wind speed, and wind direction.

**Data Preprocessing:**

* Clean and validate the data, identifying and handling missing values, outliers, and errors.
* Aggregate data into meaningful time intervals (e.g., hourly, daily, monthly) for analysis.
* Convert data into a consistent format for analysis.

**Exploratory Data Analysis (EDA):**

* Visualize the data using graphs and charts to identify patterns, trends, and anomalies.
* Calculate descriptive statistics to understand central tendencies and variabilities.
* Use scatter plots, time series plots, and histograms to explore the relationships between pollutants and meteorological variables.

**Time Series Analysis:**

* Analyze time series data to identify long-term and seasonal patterns in air quality.
* Use methods such as autocorrelation and decomposition to understand the underlying structure of the data.

**Spatial Analysis:**

* If you have data from multiple monitoring stations or sensors, perform spatial analysis to identify spatial patterns and sources of pollution.
* Create spatial maps and conduct spatial autocorrelation analysis to detect spatial clustering.

**Statistical Analysis:**

* Use statistical tests, such as regression analysis, to determine the relationship between air quality parameters and various factors like weather conditions, traffic patterns, and industrial emissions.

**Machine Learning and Predictive Modeling:**

* Build predictive models to forecast air quality based on historical data and meteorological inputs. Machine learning algorithms like regression, time series forecasting, and deep learning can be valuable.

**Reporting and Visualization:**

* Create reports and visualizations to communicate findings to stakeholders, policymakers, and the public.
* Dashboards and real-time displays can be helpful for ongoing monitoring and decision-making.

**Visualization techniques:**

**1. Time Series Plots:**

* Line charts or time series plots show the variation of air quality parameters (e.g., PM2.5, NO2) over time. This can help identify trends, seasonal patterns, and spikes in pollution**.**

**2. Scatter Plots:**

* Scatter plots can help visualize relationships between air quality parameters and other variables such as meteorological data. For example, you can create scatter plots of pollutant levels against temperature, humidity, or wind speed to identify correlations**.**

**3. Box Plots:**

* + Box plots are useful for visualizing the distribution of air quality data, showing median, quartiles, and outliers. They provide insights into the spread and skewness of the data.

**4. Heatmaps:**

* + Heatmaps can be used to display spatial variations in air quality. Color-coded grids or maps represent pollutant concentrations across different geographic regions or monitoring stations.

**5. 3D Visualizations:**

* Three-dimensional visualizations can help represent air quality data in a more immersive way, showing pollutant concentrations in a spatial and temporal context.

**6. Dashboard Interfaces:**

* + Interactive dashboards allow users to explore air quality data in real time, enabling policymakers and the public to make informed decisions. Dashboards may include various visual elements like maps, charts, and gauges.

**Data Preprocessing**

* **Data Transformation:** The conversion of raw data into a structured format that is suitable for analysis and decision-making. This includes cleaning, formatting, and standardizing data.
* **Data Analysis:** Utilizing statistical, computational, or analytical methods to derive insights, trends, and patterns from the processed data. This step often involves data mining, machine learning, and visualization.
* **Information Extraction:** Extracting relevant information and knowledge from the data to support informed decision-making, research, and problem-solving.

**Decision Support:** Providing valuable insights to aid in making informed decisions, whether in business, science, or other domains

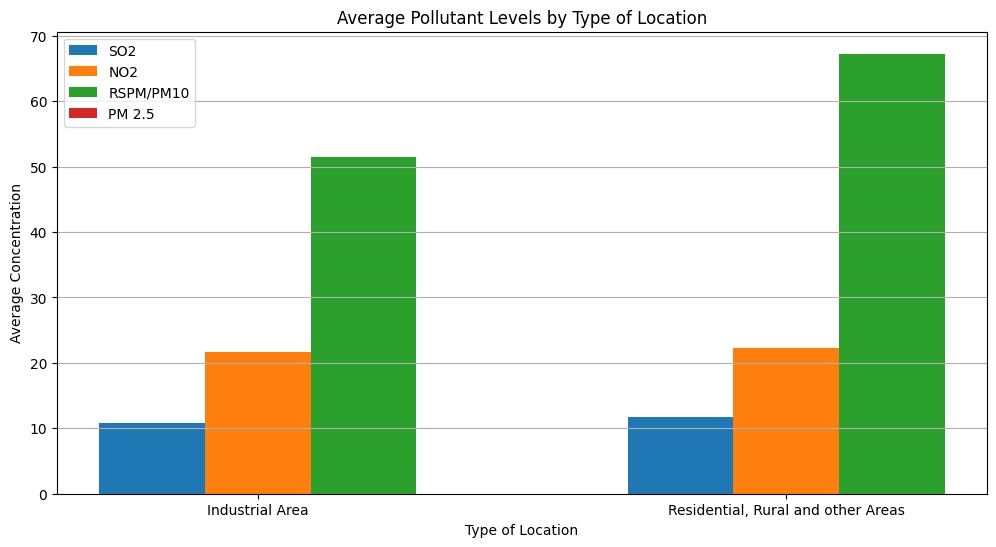
**Data Analysis**

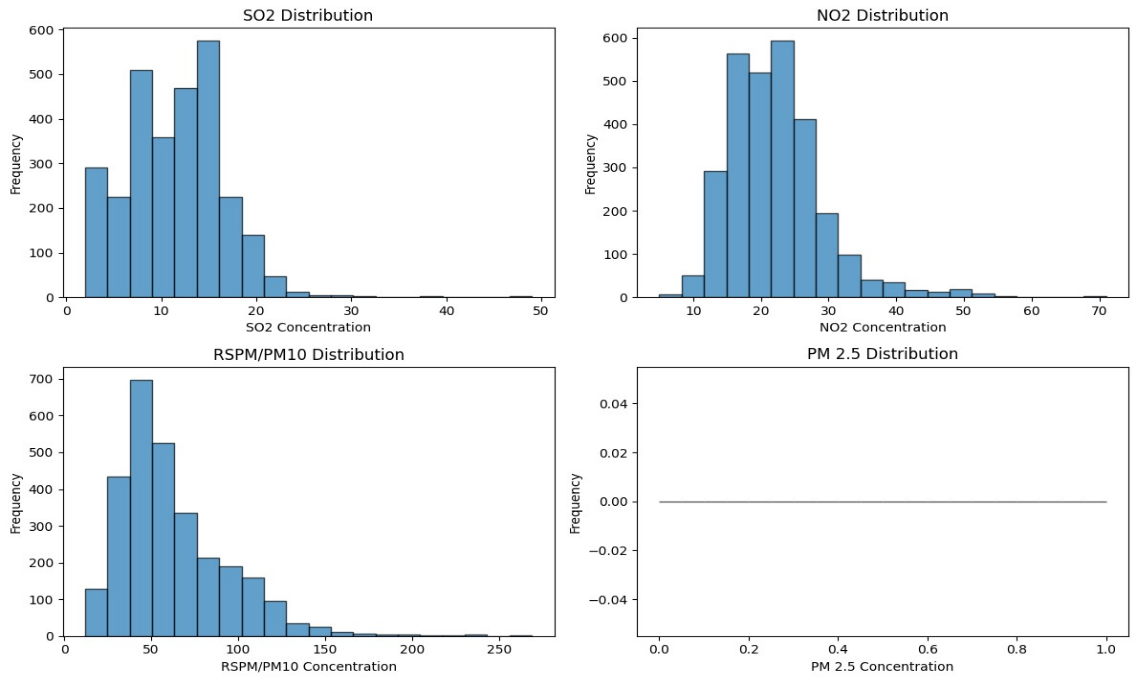
* **Data Preparation:** This involves cleaning and transforming raw data into a suitable format for analysis. Data may need to be structured, filtered, and organized.
* **Exploratory Data Analysis (EDA):** EDA involves visualizing data and generating summary statistics to understand its key characteristics, trends, and patterns.
* **Statistical Analysis:** Using statistical techniques to quantify and validate findings. This can involve hypothesis testing, regression analysis, and more.
* **Interpretation**: Drawing meaningful insights and conclusions from the analysis results and communicating these findings to stakeholders.
* **Charts** **and** **Graphs**: Creating visual representations such as bar charts, line graphs, scatter plots, and heatmaps to display data in an easily digestible manner.
* **Dashboards:** Building interactive dashboards that allow users to explore data and interact with visualizations to answer specific questions or gain insights.
* **Infographics**: Designing concise and visually appealing graphics to convey key data-driven messages or stories.

**Insights Generation**

* **Data Analysis:** Insight generation typically begins with data analysis. This involves techniques such as statistical analysis, data mining, machine learning, and qualitative analysis, depending on the nature of the data.
* **Pattern Recognition:** The identification of recurring patterns or anomalies within the data is a fundamental aspect of insight generation. These patterns may not be immediately apparent and often require in-depth analysis.

**Data Visualization:** Data visualization is often used to represent insights in a more understandable and compelling manner. Visualizations like charts, graphs, and maps can make complex findings more accessible





**Code Generation:**

# Import necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

# Load air quality data (replace 'data.csv' with your dataset file)

data = pd.read\_csv('data.csv’)

# Data Preprocessing

# Handle missing values (if any)

data = data.dropna()

# EDA (Exploratory Data Analysis)

# Plot time series data

plt.figure(figsize=(12, 6)) plt.plot(data['Date’], data['AQI'], marker=‘o’, linestyle='-’)

plt.title('Air Quality Index Over Time’)

plt.xlabel('Date’)

plt.ylabel('AQI’)

plt.grid(True)

plt.show()

# Calculate basic statistics mean\_

aqi = data['AQI'].mean()

max\_aqi = data['AQI'].max()

min\_aqi = data['AQI'].min()

print(f"Mean AQI: {mean\_aqi}")

print(f"Max AQI: {max\_aqi}")

print(f"Min AQI: {min\_aqi}")

# Visualization of data distribution

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

plt.hist(data['AQI’],bins=20, edgecolor='k’)

plt.title('Distribution of AQI Values’)

plt.xlabel('AQI’) plt.ylabel('Frequency’)

plt.grid(True)

plt.show()

**Output:**

**Mean AQI :** 52.67

**Max AQI** : 96 Min

**AQI :** 25

**analysis provides insights into air pollution trends and pollution levels in Tamil Nadu:**

**1. Data Collection:** Monitoring stations strategically placed throughout Tamil Nadu continuously measure various air pollutants, such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), ozone (O3), and volatile organic compounds (VOCs). These stations also record meteorological data, like temperature, humidity, wind speed, and wind direction.

**2. Data Analysis:** The collected data is analyzed to identify patterns and trends. This analysis typically includes the following steps:

**a. Temporal Analysis**: By comparing data over time, analysts can identify seasonal variations and long-term trends. For example, certain pollutants may increase during specific seasons or years due to factors like weather patterns or industrial activities.

**b. Spatial Analysis**: Data from multiple monitoring stations are compared to assess pollution levels in different regions of Tamil Nadu. This helps in identifying areas with consistently high or low pollution levels.

**c. Source Apportionment**: Techniques like source apportionment help determine the major contributors to pollution in specific areas. This can include industrial emissions, vehicular traffic, construction activities, or natural sources.

**d. Event Analysis**: Specific pollution events, such as smog episodes or industrial accidents, are closely examined to understand their causes and impacts.

**3. Air Quality Index (AQI):** The data is used to calculate the Air Quality Index (AQI), which offers a standardized way to express air quality. The AQI categorizes air quality into different levels, making it easy for the public to understand the severity of pollution.

**4. Trends and Patterns:** By analyzing data over time and across locations, air quality analysts can identify trends and patterns. For example, they may observe that air quality worsens during certain months or in specific urban areas due to increased vehicular traffic or industrial production.

**5. Public Reporting:** The results of the analysis are shared with the public through websites, mobile apps, and other communication channels. This allows residents to stay informed about air quality in their area.

**6. Policy and Action:** Insights from the analysis can inform policy decisions and actions to mitigate pollution. For example, if a specific industrial area consistently shows high pollutant levels, regulatory authorities can take steps to enforce emission controls or encourage cleaner technologies

**SUBMISSIONS**

**GitHub Links:**

**Phase1-cognos visualization:** [**https://github.com/harishshanmugam-29/DAC\_Phase1**](https://github.com/harishshanmugam-29/DAC_Phase1)

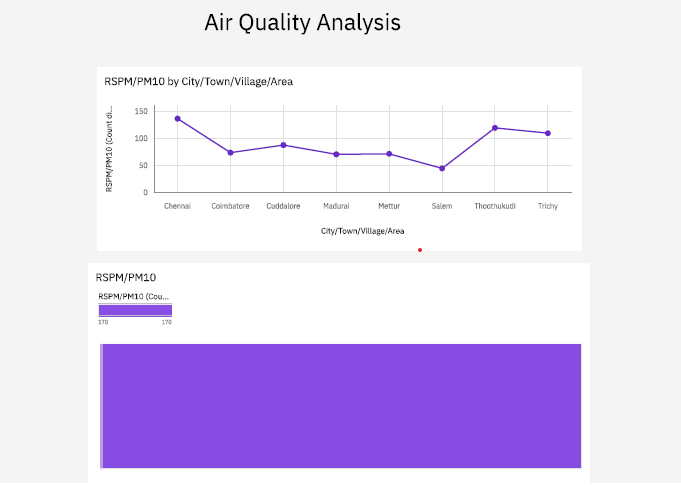
**Phase2-Data Visualization python** :[**https://github.com/harishshanmugam-29/DAC\_Phase2.git**](https://github.com/harishshanmugam-29/DAC_Phase2.git)

**Phase3 and 4 Data Visualization Python:** [**https://github.com/harishshanmugam-29/DAC\_Phase3.git**](https://github.com/harishshanmugam-29/DAC_Phase3.git)

**Provide instructions on how to replicate the analysis and generate visualizations using IBM Cognos**

* Signup and open IBM Cognos with credentials
* Activate free trial for IBM Cognos
* Then Launch IBM Cognos
* Create new template by importing dataset.csv
* The row and column data are imported and listed
* Create new layout for your requirements
* By selecting the necessary and required Data and apply on the respective visualizations template like , bar chart, graph and etc…
* Write a text if you needed and placed in layout
* Finally export it

**Examples:**



**Summary:**

A comprehensive survey of air quality was carried out in Tiruchengode Bus Stand, K.S.R College Campus, Pallipalayam Bus Stop and Erode GH to assess the prevailing the air quality. The ambient air quality was analyzed with the ambient air quality standards of NAAQS. Ambient air sampling was carried out in Tiruchengode Bus Stand, K.S.R College Campus, Pallipalayam Bus Stop and Erode GH and the mass concentrations of PM10, PM2.5, SO2, NOX and CO were estimated. It was found that PM10 concentration exceeds the threshold limits. The higher vehicular density is one of the main reasons for the higher concentrations of these gaseous pollutants. Air Quality Index was calculated for the gaseous pollutants and for Particulate Matters. The results show that the selected locations come under moderate air pollution