**AIR QUALITY ANALYSIS**

**TEAM MEMBER**

**311521205004: ABUBACKER SIDDIQUE S**

**Phase 2 Submission Document**

**Project:**Air Quality Analysis using data analytics

A diagram of a computer and a machine

Description automatically generated with medium confidence

**INTRODUCTION:**

Air quality is a major public health and environmental concern. According to the World Health Organization, air pollution is responsible for about 7 million deaths each year. Air pollution can cause a variety of health problems, including asthma, heart disease, stroke, and cancer. It can also damage ecosystems and reduce agricultural yields.

Air quality analysis is important for understanding the sources and distribution of air pollution, assessing the health risks posed by air pollution, and developing strategies for reducing air pollution. Data analytics can play a vital role in air quality analysis.

**AIR QUALITY ANALYSIS PROJECT FOR A GIVEN DATASET**

This project will use data analytics to analyze a given air quality dataset. The dataset may contain data on a variety of air pollutants, such as particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. The dataset may also contain data on meteorological factors, such as temperature, wind speed, and wind direction.

The goal of the project is to gain insights into the air quality data using data analytics techniques. The project may focus on the following tasks:

**DATA CLEANING AND PREPROCESSING**:

The first step is to clean and preprocess the data. This may involve removing outliers, filling in missing values, and converting the data into a consistent format.

**EXPLORATORY DATA ANALYSIS**:

Exploratory data analysis is used to identify patterns and trends in the data. This can be done using a variety of visualization and statistical techniques.

**FEATURE ENGINEERING:**

Feature engineering is the process of creating new features from existing features. This can be done to improve the performance of machine learning models.

**MACHINE LEARNING:**

Machine learning can be used to develop models that can predict air quality levels or identify the sources of air pollution.

The following figure illustrates the methodology involved in the air quality analysis project:

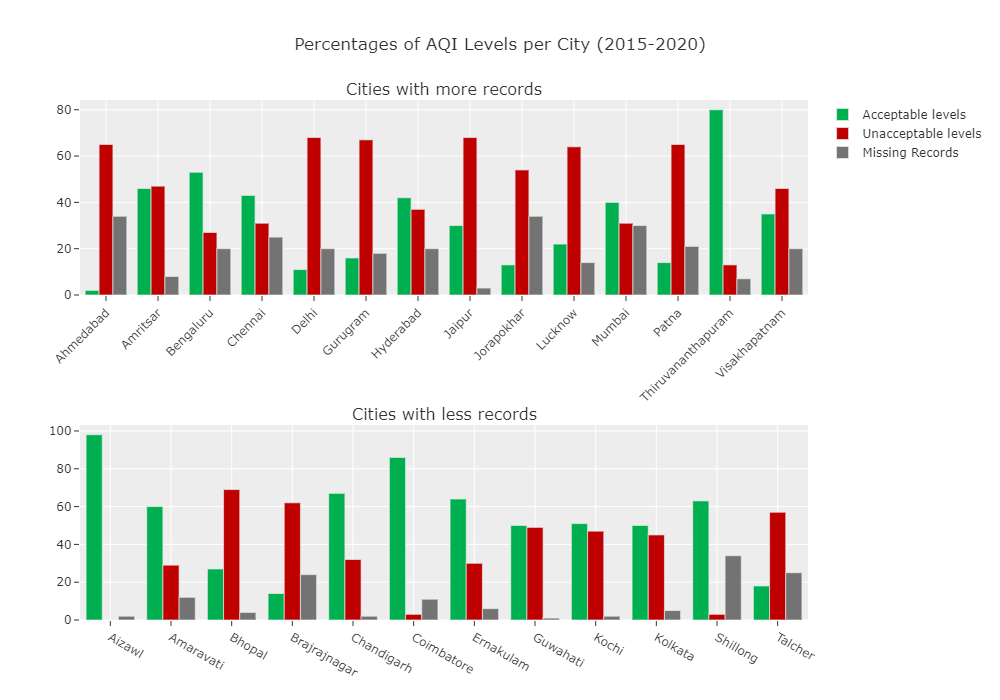
A diagram of a scientific process

Description automatically generated with medium confidence

**BENEFITS OF DATA ANALYTICS FOR AIR QUALITY ANALYSIS**

Data analytics can provide a number of benefits for air quality analysis, including:

**IDENTIFYING PATTERNS AND TRENDS:**

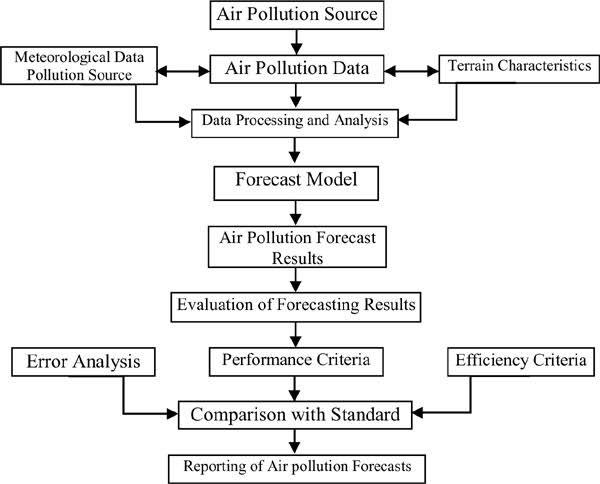
Data analytics can be used to identify patterns and trends in air quality data that would be difficult to see with the naked eye. This information can be used to develop a better understanding of the factors that contribute to air pollution and to predict future air quality conditions.

**QUANTIFYING THE HEALTH AND ENVIRONMENTAL IMPACTS OF AIR POLLUTION:**

Data analytics can be used to quantify the health and environmental impacts of air pollution. This information can be used to make informed decisions about air pollution control policies and regulations.

**DEVELOPING AND EVALUATING AIR POLLUTION CONTROL STRATEGIES:**

Data analytics can be used to develop and evaluate air pollution control strategies. For example, data analytics can be used to identify the most cost-effective ways to reduce air pollution from different sources.

**STEPS INVOLVED IN AIR QUALITY ANALYSIS:**

**PROGRAM:**

import pandas as pd

import matplotlib.pyplot as plt

# Load the air quality dataset from a CSV file

data = pd.read\_csv('air\_quality\_dataset.csv')

# Display basic information about the dataset

print("Dataset Info:")

print(data.info())

# Summary statistics

print("\nSummary Statistics:")

print(data.describe())

# Plotting time series data (e.g., PM2.5 levels over time)

plt.figure(figsize=(12, 6))

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

plt.xlabel('Date')

plt.ylabel('PM2.5 Levels')

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

plt.legend()

plt.grid(True)

plt.show()

# Correlation between variables

correlation\_matrix = data.corr()

print("\nCorrelation Matrix:")

print(correlation\_matrix)

# Visualize correlations using a heatmap

import seaborn as sns

plt.figure(figsize=(10, 8))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f")

plt.title('Correlation Heatmap')

plt.show()

**OUTPUT:**

PM2.5 Levels Over Time:

A line graph will be displayed showing the trend of PM2.5 levels over time using the "Date" and "PM2.5" columns from your dataset.

Correlation Matrix:

A table will be printed showing the correlation coefficients between different numerical columns in your dataset. This indicates how strongly variables are related to each other.

Correlation Heatmap:

A heatmap will be displayed, visually representing the correlation matrix. Darker colors represent stronger correlations, while lighter colors indicate weaker or negative correlations.

**CONCLUSION:**

Data analytics can play a vital role in air quality analysis. By using data analytics to analyze air quality data, we can gain insights into the sources and distribution of air pollution, assess the health risks posed by air pollution, and develop strategies for reducing air pollution.