**Sustainable Development Goal (SDG) Problem Definition Document**

**SDG Goal: SDG 13 - Climate Action**

**Problem Definition:**

**Title:** Monitoring and Reducing Air Pollution in Urban Areas

**Introduction:** Air pollution is a significant environmental issue that affects the health and well-being of urban populations. High levels of pollutants like CO2, NO2, and particulate matter (PM10 and PM2.5) can lead to respiratory and cardiovascular diseases, reducing the quality of life and increasing healthcare costs. Monitoring air pollution is essential for developing effective strategies to mitigate its impact and ensure sustainable urban development.

**Specific Problem:** The specific problem addressed in this project is the need for comprehensive monitoring and analysis of air pollution levels in urban areas. This includes identifying pollution hotspots, understanding the correlation between weather conditions and pollution levels, and evaluating the effectiveness of pollution control measures.

**Objectives:**

1. **Data Collection:**
   * Gather data from various air quality monitoring stations in the city.
   * Collect measurements for key pollutants such as CO2, NO2, PM10, and PM2.5.
   * Obtain weather data including temperature, humidity, and wind speed.
2. **Database Design and Implementation:**
   * Design and implement a relational database to store air quality and weather data.
   * Ensure the database can handle large volumes of data and supports efficient querying.
3. **Data Retrieval and Analysis:**
   * Write SQL queries to retrieve relevant data for analysis.
   * Perform descriptive statistical analysis to summarize pollutant levels.
   * Conduct time series analysis to identify trends and seasonal patterns.
   * Analyze the correlation between weather conditions and pollutant levels.
   * Compare pollutant levels across different monitoring stations.
4. **Data Visualization and Reporting:**
   * Use Microsoft Excel to create visualizations and dashboards for data analysis.
   * Generate reports summarizing key findings and insights.
   * Provide recommendations for reducing air pollution based on the analysis.

**Methodology:**

1. **Data Collection:**
   * Use CSV files (stations.csv, pollutants.csv, weather.csv, and measurements.csv) to collect and organize data.
   * Ensure data is clean, accurate, and well-structured for analysis.
2. **Database Design:**
   * Create tables in MySQL to store data:
     + Stations table for air quality monitoring stations.
     + Pollutants table for different types of pollutants.
     + Weather table for weather conditions.
     + Measurements table for pollutant measurements.
3. **SQL Queries for Data Retrieval and Analysis:**
   * Retrieve all measurements for a specific pollutant (e.g., CO2) from a specific station.
   * Calculate summary statistics (mean, median, min, max) for pollutant levels.
   * Analyze pollutant levels over time to identify trends.
   * Examine the correlation between weather conditions and pollutant levels.
4. **Data Visualization and Reporting:**
   * Import CSV files into Excel and use pivot tables, charts, and conditional formatting for analysis.
   * Create dashboards to summarize and visualize data insights.
   * Compile findings into a comprehensive report with recommendations for air pollution control.

**Expected Outcomes:**

1. **Comprehensive Air Quality Database:**
   * A well-structured MySQL database containing air quality and weather data for urban areas.
2. **Insights and Trends:**
   * Identification of pollution hotspots and trends over time.
   * Understanding the impact of weather conditions on pollutant levels.
3. **Data-Driven Recommendations:**
   * Evidence-based recommendations for policymakers to reduce air pollution.
   * Strategies for improving air quality monitoring and reporting.
4. **Enhanced Data Visualization:**
   * Interactive Excel dashboards and visualizations to facilitate data analysis and decision-making.

**Conclusion:**

By leveraging data collection, database design, SQL queries, and Excel visualizations, this project aims to provide a comprehensive analysis of air pollution in urban areas. The insights gained will inform strategies to mitigate air pollution, contributing to the achievement of SDG 13: Climate Action. This project not only addresses a critical environmental issue but also empowers stakeholders with the information needed to make informed decisions for a sustainable future.

**Appendix:**

**Sample SQL Queries:**

1. **Retrieve All Measurements for a Specific Pollutant (e.g., CO2) from a Specific Station:**

SELECT MeasurementDate, AVG(MeasurementValue) AS AverageCO2Level

FROM Measurements

WHERE PollutantID = (SELECT PollutantID FROM Pollutants WHERE PollutantName = 'CO2')

GROUP BY MeasurementDate

ORDER BY MeasurementDate;

1. **Calculate Summary Statistics for CO2 Levels:**

SELECT

AVG(MeasurementValue) AS AverageValue,

MEDIAN(MeasurementValue) AS MedianValue,

MIN(MeasurementValue) AS MinValue,

MAX(MeasurementValue) AS MaxValue

FROM Measurements

WHERE PollutantID = (SELECT PollutantID FROM Pollutants WHERE PollutantName = 'CO2');

1. **Analyze CO2 Levels Over Time:**

SELECT MeasurementDate, AVG(MeasurementValue) AS AverageCO2Level

FROM Measurements

WHERE PollutantID = (SELECT PollutantID FROM Pollutants WHERE PollutantName = 'CO2')

GROUP BY MeasurementDate

**ORDER BY MeasurementDate;**

1. **Correlation Between CO2 Levels and Temperature:**

SELECT

CORR(m.MeasurementValue, w.Temperature) AS CO2\_Temperature\_Correlation

FROM Measurements m

JOIN Weather w ON m.MeasurementDate = w.Date

WHERE m.PollutantID = (SELECT PollutantID FROM Pollutants WHERE PollutantName = 'CO2');