# **Project Manager: Motamedi, ali**

# **Air Quality Dataset Analysis**

### **Problem/Opportunity/Research Question(s):**

Air quality has a significant impact on human health, the environment, and the economy. The fluctuating levels of pollutants such as Carbon Monoxide (CO), Nitrogen Oxides (NOx), and Benzene (C6H6) in urban areas pose a challenge for environmental policies and public health. This project addresses the opportunity to utilize a comprehensive dataset on air quality to analyze the patterns, correlations, and impacts of various pollutants. The primary research question focuses on understanding how these pollutants vary over time and their relationship with environmental conditions such as temperature, humidity, and atmospheric pressure.

### **Dataset:**

The dataset encompasses a range of air quality indicators recorded hourly, including:

* **Date & Time**: Timestamps for each observation.
* **CO(GT)**: Concentration of Carbon Monoxide in mg/m^3.
* **PT08.S1(CO)**: Tin oxide sensor response for CO.
* **NMHC(GT)**: Concentration of Non-Methane Hydrocarbons in microg/m^3.
* **C6H6(GT)**: Concentration of Benzene in microg/m^3.
* **PT08.S2(NMHC)**: Titania sensor response for NMHC.
* **NOx (GT)**: Concentration of Nitrogen Oxides in ppb.
* **PT08.S3(NOx)**: Tungsten oxide sensor response for NOx.
* **NO2(GT)**: Concentration of Nitrogen Dioxide in microg/m^3.
* **PT08.S4(NO2)**: Tungsten oxide sensor response for NO2.
* **PT08.S5(O3)**: Indium oxide sensor response for Ozone.
* **T**: Temperature in °C.
* **RH**: Relative Humidity (%).
* **AH**: Absolute Humidity.

### **Goal:**

The goal of this project is to develop a comprehensive analysis of air quality trends and pollutant behaviour over time, leading to actionable insights for environmental policy-making and public health advisories. This analysis will be conducted by applying statistical and machine learning techniques to identify patterns, correlations, and causations within the dataset. The final deliverable will be a detailed report including data visualizations, findings, and recommendations, expected to be completed within next three months.

### **Objectives:**

* **Data Cleaning and Preparation**: Standardize the dataset format, handle missing values, and remove outliers to prepare the data for analysis.
* **Exploratory Data Analysis (EDA)**: Conduct a thorough EDA to understand the distribution of pollutants, seasonal variations, and preliminary correlations between variables.
* **Trend Analysis**: Identify long-term trends in air quality data and the impact of environmental factors using time series analysis.
* **Predictive Modelling**: Develop predictive models to forecast air quality indices based on historical data and environmental parameters.
* **Report Generation and Recommendations**: Compile findings into a comprehensive report detailing analysis results, insights, and actionable recommendations for stakeholders.

### **Success Criteria:**

The project will be considered successful if it:

* Project will be completed within 4 months
* Provides clear, actionable insights for reducing pollutant levels in urban areas and effects of external factors.
* Achieves an accuracy of 80% or higher in predictive modelling for future air quality indices.
* Receives manager approval on the comprehensiveness and applicability of the final report.

### **Assumptions, Risks, Obstacles:**

* **Assumptions**: The dataset is accurate, timely, and representative of a wide range of urban environments.
* **Risks**: Potential data gaps or inaccuracies could affect analysis outcomes. Environmental factors not included in the dataset may influence pollutant levels.
* **Obstacles**: Limited access to computing resources for large-scale data analysis and model training. Delays in project milestones due to unforeseen challenges in data preprocessing or model complexity.

This documentation outlines a structured approach to analysing an air quality dataset, addressing critical environmental and public health issues through data-driven insights.