

Project 4: Outdoor Air Quality Forecasting

Project Overview

Outdoor air quality is heavily influenced by **weather conditions**, with factors like **temperature**, **humidity**, and **wind speed** playing a significant role in pollution levels. In this project, you will develop a tool that forecasts **air quality** based on weather data for a large city (population > 500,000). By analyzing how different weather patterns influence pollution, you will create a simple model that predicts the **best** and **worst** times of day for air quality.

Objective

You will gather **forecasted weather** (temperature, humidity, wind speed) and **forecasted air quality (AQI)** data for a **large city** (population > 500,000) of your choosing and develop a **predictive model** that identifies the times of day when air quality is likely to be the worst or best. By exploring the **relationship** between weather conditions and pollution levels, you will discover how air quality fluctuates over a forecast period (e.g., the next few days) and share insights that can help the public make informed decisions.

Major Questions for Your Proposal

1. City Choice & Data Approach

- Which large city (population > 500,000) are you planning to focus on, and why?
- Do you expect the city's climate or daily weather variation to significantly impact air quality?
- What relationship do you expect between various weather conditions and air quality?

2. Predictive Model Setup

- How do you intend to use the forecasted weather data (temperature, humidity, wind speed) and air quality data (AQI) to develop a predictive model?
- Will you use a **simple regression**, a **correlation-based** approach, or something else to predict future air quality?
- How will you handle any anomalies or missing data?

3. Visualization Strategy

- How do you plan to **visualize** the relationship between the data you gather for the forecasted air quality and weather conditions?
- What figures will you use to explore the data - which approach will provide the greatest insights before you start developing your predictive model?
- Which figures will you use to present your findings?

Project Definitions

Weather Parameters

1. **Temperature:** Warmer temperatures can worsen air quality by increasing the rate of chemical reactions that form pollutants.

2. **Humidity:** High humidity can trap pollutants closer to the ground.
3. **Wind Speed:** Strong winds can help disperse pollutants, potentially improving air quality.

Pollution Levels

- **Air Quality Index (AQI):** A standardized measure of how polluted the air is, indicating potential health impacts. You will forecast AQI values using your chosen model.

Data Sources

- **Forecast API:** [OpenWeatherMap 5-day Forecast](#) Provides forecasted data on temperature, humidity, and wind speed.
- **Air Pollution API:** [OpenWeatherMap Air Pollution](#) Offers real-time and/or forecasted AQI data for the same city.
- You will need to **sign up** for an API key to authenticate your requests.

Project Requirements

1. Data Collection

- Select **one large city** (population > 500,000) to focus on. Base your choice on your responses to **Major Question 1**
- Use the **Forecast API** to fetch weather data (temperature, humidity, wind speed) for the next few days in the chosen city.
- Use the **Air Pollution API** to fetch AQI data over the same time period.

2. Visualization

- As you outlined in **Major Question 3**, create visuals to explore the relationship between weather parameters and air quality.
- Try using annotations, highlights, or simply denote **peak pollution** periods (worst AQI) and **lowest pollution** periods (best AQI).
- Ensure your visualizations clearly communicate the existence (or absence) of a **relationship** between weather variables and air quality over time.

3. Analysis

- Follow your plan from **Major Question 2** to create a regression or other predictive approach that correlates weather conditions with AQI.
- Use the outcome from that analysis to identify the **times of day** when air quality is likely to be the **worst** or **best**, based on the weather forecast.
- Use your model with synthetic data to predict air quality for different weather conditions. Do these predictions make sense? Are there limitations to your model i.e. are there certain weather conditions that your model cannot predict or certain conditions that invalidate your model?
- Consider data quality issues: handle missing, anomalous, or inconsistent data in a systematic way.

4. Data File

- Save your synthetic weather data with the predicted air quality data to a **csv** or **json** file.