

Michael Christopher – s224830467

CAPRI – Correlation of Air Pollution and Respiratory Illness

What have been done?

1. Extended Data Cleaning & Preprocessing

Continued refining preprocessing pipelines by improving handling of missing values and ensuring pollutant concentration consistency across all datasets.

So what? This strengthens data integrity, ensuring the modelling phase is built on reliable and valid inputs.

2. AQI Calculation Enhancement

Applied and validated the EPA AQI formula across the cleaned datasets, verifying that the maximum pollutant score is captured for each interval.

So what? This provides a robust and standardised measure of air quality that can be used consistently in predictions and health data integration.

3. Machine Learning Modelling Progress

Ran multiple AQI prediction models (Linear Regression, Random Forest, Gradient Boosting), comparing initial performance results.

So what? This identifies promising approaches for forecasting air quality trends and sets the stage for fine-tuning models to improve accuracy.

4. API Integration Testing

Continued testing the WAQI API to ensure stable and repeatable real-time data calls.

So what? This prepares the project for potential real-time data ingestion and interactive features in later stages.

Next actions?

1. Finalise model comparison and tune hyperparameters for improved predictive performance.
2. Begin exploratory correlation analysis with AIHW health datasets.
3. Develop prototype visualisations (pollutant trends, AQI over time, regional comparisons).
4. Explore integration of real-time AQI API data into predictive workflows.