­

|  |
| --- |
| A city with many tall buildings  Description automatically generated |
| Urban Air Quality  Predictive Modeling and Analysis of Air Quality in Mongolia |
| |  |  |  | | --- | --- | --- | | Nomin Lkhagvasukh | 10/27/24 | Strategic Thinking | |

A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Nomin Lkhagvasukh |
| *Student Number* | 2024461 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date* |  |
| *Date of Submission* | Sunday, 27th October 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Introduction 2](#_Toc180838595)

[Objectives and Problem Definition 4](#_Toc180838596)

[Project Scope 4](#_Toc180838597)

[Explore Data: 4](#_Toc180838598)

[Data Preprocessing: 5](#_Toc180838599)

[Model Development: 5](#_Toc180838600)

[Model Evaluation: 5](#_Toc180838601)

[Document Processes: 5](#_Toc180838602)

[Boundaries 5](#_Toc180838603)

[Data Sources 5](#_Toc180838604)

[Timeline 6](#_Toc180838605)

[Ethical Considerations 6](#_Toc180838606)

[References 6](#_Toc180838607)

# Introduction

In many cities, air pollution is becoming a big problem that affects people’s health. When the air is polluted, people can have breathing issues and other health problems. This project is about understanding what causes bad air quality, analysing data to see the patterns, and predicting future air quality. By doing this, we can give advice to city planners and leaders on ways to keep the air cleaner. The project is important because clean air is essential for people’s health and the environment.

A screenshot of a computer

Description automatically generated

Figure 1. Global Risk Factors for Death in 2021, Source: UNICEF, 2024)

For both the entire world's population and children under five, air pollution is now the second-leading cause of death.

Many pollutants are important contributors to human illness. These include Particulate Matter (PM), which are particles with varying but extremely small diameters that enter the respiratory system through inhalation and can lead to cancer, reproductive and cardiovascular disorders, and malfunctions of the central nervous system and reproductive system.

A graph of numbers and a number of people

Description automatically generated with medium confidence

Figure 2.Why is India's pollution much worse than China's, Source: BBC, 2019

According to studies by Greenpeace and IQ AirVisual, a Swiss organization that collects air-quality data worldwide, 22 of the 30 most polluted cities in the world are in India. This shows that Mongolia's capital (Ulaanbaatar) is more polluted than China's Beijing.

According to the IQAir AirVisual air quality index (AQI), Ulaanbaatar, Mongolia, has the worst air quality in the world on October 23, 2020. Since October 11, the city's average air quality index readings have ranged from 100 to 200, with a current level of 203.

A screenshot of a web page

Description automatically generated

Dangerously high air quality levels can have a direct impact on human health. Ulaanbaatar, Mongolia air pollution levels have registered as dangerous on the Air Quality Index (AQI). (IQAir, Ulaanbaatar residents face worst air quality in the world, 2020)

A table of health and medical information

Description automatically generated with medium confidence

Out of all the most polluted cities in the world, Mongolia ranked third with an annual average of 62 μg/m³ for PM2.5 values in 2019. As this amount continues to increase, air pollution is one of Mongolia's most significant concerns.

# Objectives and Problem Definition

**Objectives (Goals):**

1. **Analyse Past Air Quality Data:** Look at old data about air pollution to understand when and why air quality becomes bad. This includes pollutants (like PM2.5, PM10 and CO) and what times of year or days they are worse.
2. **Create a Model to Predict Air Quality:** Build a simple program (a model) that can predict when air quality might be bad in the future. This will help cities warn people before pollution gets worse.
3. **Study the Effect of Weather and Traffic on Air Quality:** Look at how things like weather (temperature, humidity) and traffic patterns make pollution better or worse.
4. **Suggest Ideas to Improve Air Quality:** Based on what we learn, provide some suggestions for how city leaders can reduce air pollution, such as controlling traffic during certain hours.

**Problem Definition:**

Air pollution in cities is a growing issue. It’s caused by many factors, like vehicle traffic and weather conditions, and it can lead to health issues. The problem this project focuses on is: “How can we predict when air pollution will become bad and help cities take action to reduce it?” By finding answers, we can help cities improve air quality and protect people’s health.

# Project Scope

The purpose of this study is meant to inform decision makers on the causes of air pollution in Ulaanbaatar, the harm that PM causes to human health, and the available strategies for reducing its effects. This report intends to highlight the nature and scale of air pollution and associated health problems and outline some possible interventions to address this issue.

* Collect data on air quality, weather, and traffic. This data will come from online sources like government air quality reports.
* Build a prediction model to estimate air quality for a long periods.
* Study trends and patterns in the data to understand what causes bad air quality.

### Explore Data:

In order find out the primary features of the data, identify trends, and test predictions, I will analyze the data using Exploratory Data Analysis (EDA) and visualization methods.

**Exploratory Data Analysis (EDA)**

**Visualize Pollution Levels:** Use time series plots to observe trends in pollutant levels over days, weeks, or years.

**Correlation Analysis:** Check for relationships between pollutants and weather conditions (how wind affects PM2.5 levels).

**Spatial Analysis:** If data is available across multiple locations, use mapping techniques to visualize air quality across regions

### Data Preprocessing:

Data Preprocessing: To make sure missing values are handled, errors are solved, and formatting problems are fixed, the quality of the data will be checked. For the model to have a strong foundation, this stage is essential.

### Model Development:

Explain Prediction Finding out if the model will forecast particular pollutant levels (regression) or air quality categories (classification) is the goal.

Choose Features: Make use of relevant data, including past pollutant levels, weather (temperature, humidity), and time-based variables (season, time of day) that affect air quality.

Engineering Features: To improve model performance, add more helpful features such as pollutant averages, percentages, or time-based statistics.

Select the Model:

For Regression: For predicting pollutant concentrations, use models such as Linear Regression, Random Forest, or Gradient Boosting.

For classification, apply models such as Random Forest Classifiers for AQI categories, Decision Trees, or Logistic Regression.

### Model Evaluation:

We check the model on unknown information during the model evaluation step to determine its accuracy and predictability. We improve the model by changing hyperparameters, experimenting with different techniques, or including additional features if the results show problems. Through this iterative process, the model's accuracy, dependability, and optimization to accurately anticipate air quality are ensured, supporting quick health screenings.

### Document Processes:

Every step and process will be recorded and documented during the project's execution, both in the machine learning model's Jupyter notebook and in reports. These will be stored in a GitHub folder with version control.

## Boundaries

Developing new air quality sensors or systems in the actual world is not part of this project.

While suggestions based on findings will be provided, a comprehensive plan for changing government policy will not be included.

# Data Sources

Data from trustworthy sources is necessary to collect pollution levels, the climate, and other relevant environmental parameters in order to create an air quality prediction model. The following popular data sources can offer thorough datasets on air quality:

**Environmental Protection Agency (EPA):** Updated hourly or daily, the EPA provides comprehensive air quality statistics for the United States, including pollutant levels (PM2.5, CO, NO2, O3).

**Data includes air quality indexes (AQI):** Thorough monitoring station data are among the data, which are useful for figuring out local pollution patterns.

**OpenAQ:** Compiles current air quality information from reliable government and academic sources in numerous nations.

This dataset facilitates a more comprehensive geographic study by incorporating pollution data from several international cities.

**The World Air Quality Index (WAQI)** is a valuable tool for tracking air pollution at the city level globally since it offers real-time updates on global air quality data.

# Timeline

Based on what I currently know about the amount of time available throughout semesters one and two, I have projected a high-level timeline:

|  |  |
| --- | --- |
| First Semester: | |
| **Weeks 1–4:** | *Review information and understand the project topic.* |
| **Weeks 5–8:** | *Collect and clean the data so it’s ready for analysis.* |
| **Weeks 9–12:** | *Study the data to find patterns in air pollution, Start building the prediction model.* |
| **Second Semester:** | |
| **Weeks 1–6:** | *Start building the prediction model.* |
| **Weeks 7–10:** | *Study how weather and traffic affect pollution.* |
| **Weeks 11–14:** | *Build a dashboard to show results clearly, finish up the project.* |

Table 1. Timeline

# Ethical Considerations

As this initiative is committed to all ethical standards, data will be used in a way that respects people's privacy and health. For instance, any air quality alerts or forecasts will be communicated in a straightforward and calm manner to assist, not to frighten, people.

# References