

National College of Ireland

Project Proposal

An exploratory data analysis on air pollution

02/11/21

BSHCEDA4

Data Analytics

2021/2022

Sean Burke

X17132118

X17132118@student.ncirl.ie

Contents

[1.0 Objectives 2](#_Toc83590881)

[2.0 Background 2](#_Toc83590882)

[3.0 State of the Art 2](#_Toc83590883)

[4.0 Data 3](#_Toc83590884)

[5.0 Methodology & Analysis 3](#_Toc83590885)

[6.0 Technical Details 3](#_Toc83590886)

[7.0 Project Plan 4](#_Toc83590887)

# Objectives

(Max half Page)

What does this project set out to achieve?

The objective of this project is to apply what I have learned in my data analysis modules to a topic that Interests me. I have chosen the topic of air pollution.

I plan to gather data from an API from <https://docs.openaq.org> , This API holds the world air quality data that also contains many datasets within countries so I will be able to provide a more refined overview within each country.

I’m also planning to source another dataset via web scraping. Currently the data set I’m looking to use is the most polluted countries 2020 <https://www.iqair.com/world-most-polluted-countries> .

My aim is to be able to provide values and graphs that report my findings comparing different countries and cities. Find out the “Why” on the high or low values of air pollution in different parts of the world.

# Background

(Max half Page)

Why did you choose to undertake this project? How will you meet the objectives set out in Section 1.0?

I chose to undertake this project because, air pollution has become an increasing problem over the years and has many negative impacts.

I plan to use programming languages like R and Python to take in the data sets that I have chosen (either via reading in from csv files, accessing an API, or Web scraping). From there I will clean the data to show that specific data that I plan to use (not every part of the dataset will be used). Once I have my cleaned the data, I can compare the data against other cleaned data with the aid of graphs and then give a detailed report on my findings

# State of the Art

(Max half page)

What similar analysis has been carried out by others already? What makes your work stand out? How does it differ from similar work of others?

The first academic Article looked at was: “The Spatial analysis of air pollution and mortality in Los Angeles” <http://scientificintegrityinstitute.org/Jerrett110105.pdf> , this study was done to determine how a high particle mass 2.5 could result in a higher risk of people in that area getting ischemic heart disease of lung cancer.

The second academic article I looked at was: “Air Pollution: The effects of air pollution” <http://repositorio.cenpat-conicet.gob.ar:8081/xmlui/bitstream/handle/123456789/492/airPullotion.pdf?sequence=1&isAllowed=y> , this article shows the different negative effects that air pollution causes, e.g. vegetation, Indoor air quality, Ozone etc.

The third academic article I looked at was : “Unprecedented Temporary Reduction in Global Air Pollution Associated with COVID-19 forced confinement: A Continental and City Scale Analysis” <https://www.mdpi.com/2072-4292/12/15/2420/pdf> , This study shows how a global lockdown shown huge signs of improvement in air quality such as a reduction of 49% for NO2 emission in London. Showing how changes for a relatively small amount of time can make a huge difference.

The premise of my project was to get data from all over the world, not just the countries, but the cities in those countries. I want to find out the reasons why certain countries are so heavily polluted. Could there be small changes made to apply a big impact or are these small changes even feasible in these countries (for example poverty, culture etc.)

This differs from the articles I have provided because they are more based around the premise of the effects of air pollution and a report on how the COVID-19 pandemic has forced countries into better air quality.

# Data

(Max 1 page)

What data is required for this work? How will you access or compile the necessary data?

The data I will need for this project is the fine particle matter values (usually pm2.5, pm10) of each of the countries/ cities I want to gather data from. I’m going to gather this data by downloading the .csv files and using R to get comfortable with the data as I’m quite comfortable with using R for projects, once I know the data is usable in R, I would like to attempt using the api to gather the data. I would preferably use python for most of this project as I would like to use this project as a learning experience to learn a new language like python. This data will be used for the current year dataset.

The second dataset needed will come from the iqair website, specifically the table data that shows the most polluted countries from 2020, and this data will be obtained via web scraping. Web scraping is the process of extracting data from a website, the data is collected and exported into a chosen format, an example of this is table data.

# Methodology & Analysis

(Max 1 page)

What methodology will you follow and why? What approach will you take to analysis? How will you break down your project work into project tasks, activities, and milestones?

I plan to us an Agile approach to this project. The agile methodology provides guidance on how to choose methods and procedures within the project.

I plan to use sprint cycles starting from December to make sure that the project stays on track, and If I’m falling behind or the project needs to be adjusted, it can be updated for the next sprint cycle.

At the start of each month I will assess what stage I’m currently on in the sprint cycle and then adjust specific topics that I had aimed to be done at certain dates.

# Technical Details

(Max 1 page)

Outline any technical development to be carried out as part of this project. What are the important algorithms or approaches under consideration for this work?

There is

# Project Plan

(Max 2 pages)

Project plan with detailed steps and timelines. This project plan should provide as much detail as possible for now and will be revised with more detail with the mid point documentation.