|  |  |
| --- | --- |
| Real Time air pollution,Comparing continents and evaluating performance | Abstract  By reviewing live data of pollution indicators it will be shown which are the best and worst performers by continent. The data obtained can be expanded to include historical data to analyse long term performance. The snapshot provided can give an insight into where improvements could be targeted.  John Truong  Wei Ke (William)  Callum Linnegan  Karissa Malseed  James Rydlewski  Data Analytics Bootcamp: Project 1 |

# Introduction

Air Pollution is produced from a host of different sources and is closely linked to population size and heavy industries in a locality. This report is focussing on consideringa cross section of cities from different continents to make a comparison of the pollution data. From this it will be possible to glean which region is the greatest contributor to global air pollutants.

The data obtained is real time snap shot of air pollution in different cities this will be taken as a typical output for each of the cities and regions considered. For future iterations of this project historical data can be used which would show how air pollution has increased or decreased with respect to time.

# Data Presentation

The joy of data analytics is the story that can be presented illustrating the findings of the research performed. The data gathered for our analysis is real time air pollution data. The graphics elected to show this data are:

* Pie chart showing the concentration of pollutants in each regions worst performing city
* Bar chart showing the air quality index for each city within a region.
* Top 10 worst performing cities by region for each pollutant category
* Box and whisker showing the range of performance for each region for each pollutant type
* Scatter plot of pollutants vs location to see what correlation there is.
* Scatter plot of pollutants vs AQI to see what contributes most to AQI
* Google map to visualise the location and intensity of Air pollution.

# Background

This project is based on API calls for air quality data from the World Air Quality Index[[1]](#footnote-1).

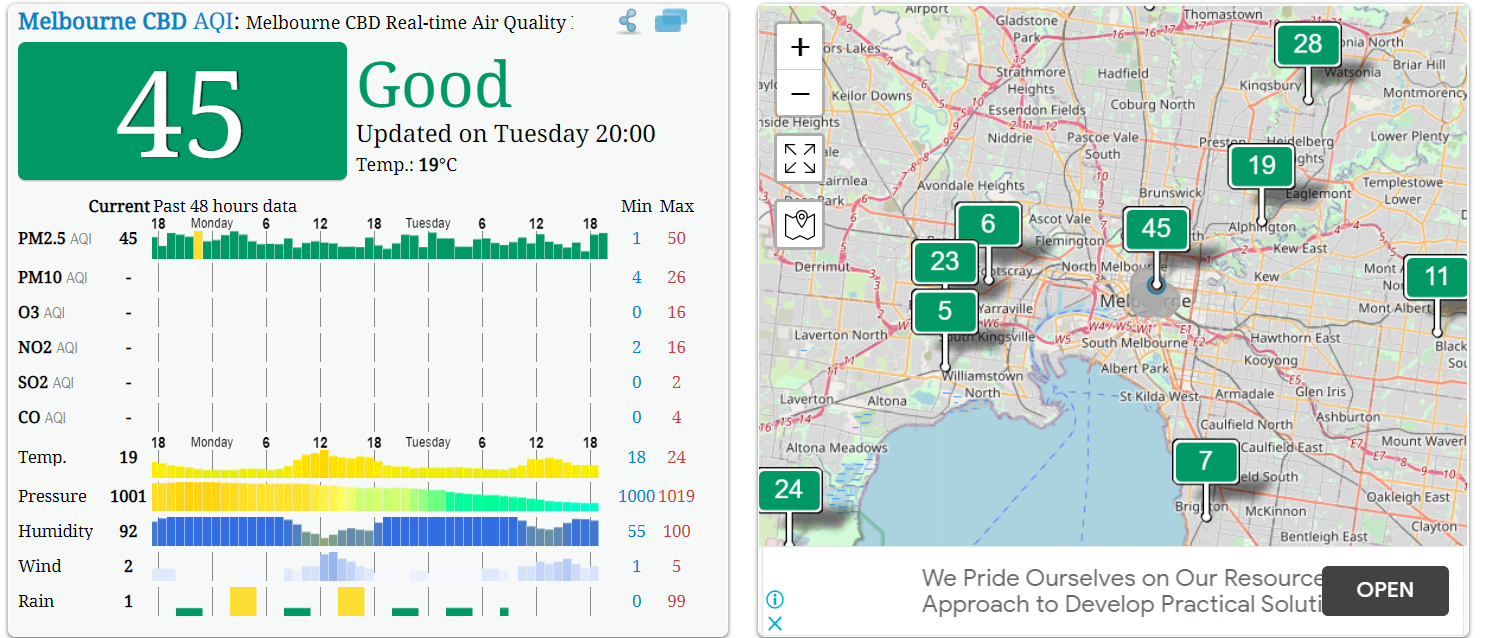


Figure 1 Sample of Data that can be obtained from WAQI

Figure 1 illustrates the sample data that will be explored in this analysis. Our objective is to obtain the following parameters for a number of cities across each continent.

|  |  |
| --- | --- |
| Parameter | Definition |
| PM 2.5 | PM 2.5 is particulate matter that is less than 2.5µm in diameter. These particles are small enough to be inhaled deeply which can have significant health implications. Often in regions where there is a high PM 2.5 count people will experience difficulty breathing. These particles are also problematic as they can be absorbed into your blood stream which can lead to further health complications of particular concern is cancer. |
| PM 10 | PM 10 is particulate matter that is less than 10µm in diameter. Although the particle size is greater than PM 2.5 it is still problematic. Typically found in smoke and smog in terms of air pollution but can also occur as suspended solids during the manufacturing process. An example would be rock dust whilst quarrying. In the context of air pollution. PM pollution causes significant health concerns. A particularly vicious episode occurred in London during the 1950’s which led to the death of over 4000 people. |
| O3 (Ozone) | Whilst Ozone occurring in the upper atmosphere has a beneficial effect for humanity as it shields the earth from harmful solar rays, at ground level it has the propensity to damage health. It is caused by the reaction of Nitrates and Volatile Organic Compounds (VOC’s) in the presence of sunlight. It is the main pollutant in “smog” and causes breathing difficulties for many. In addition to harming human health it can damage the wellbeing of plant life. Photosynthesis can be reduced slowing the plants growth and potentially killing affected species. As this happens there is less biodiversity which has a knock on effect to other entities dependent on the affected plant. |
| NO2 | Motor Vehicle exhaust gasses and heavy industry are responsible for the emission of NO. In the presence of air this reacts to form NO2. The effects are breathing difficulties especially for the elderly and children. Significant quantities of aerosol NO2 can lead to acid rain which damages other elements of the eco system. Acid rain can further damage plant and marine life. |
| SO2 | SO2 pollution is heavily caused by the burning of fossil fuels. Similar to NO2 It leads to acid rain and harms the respiratory tract and eyes. |
| CO | CO is caused by the incomplete combustion of fuels. This occurs when not enough oxygen is present during the combustion process. The major issue of CO pollution is the tendency for it to form a blanket at surface level. This again causes breathing difficulties, and potentially death. It also affects photosynthesis in plants further damaging the environment |
| Air Quality Index (AQI) | AQI takes the level of pollutants present and combines this data with the prevailing weather conditions to index the overall air quality for a location. The data obtained from this can be used to see which pollutants contribute to AQI performance, and by how much. |

# Methodology

The process used to collate and analyse global pollution data is as follows:

* Generating lists of world cities divided by continent
* API calls from World Air Quality Index Site for each city.
* Storing pollution data for each city within each continent.
* Building data frames for each continents data removing null values.
* Generating rudimentary graphs of the AQI and Pollutant data.
* Storing all data frames as .csv for further analysis.
* Visualising the data on a map.

An .ipynb and .config file to store api keys were created in addition to the above. Upon creating the file dependencies were imported. The dependencies used are there to help obtain and manipulate the data for world pollution.

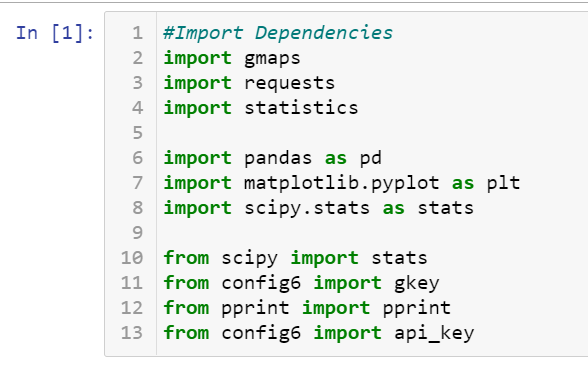


Figure 2 Dependencies Imported for further analysis

## World Cities by Continent

To perform this research the most populated cities by continent. The data for this has been taken from an open source website, which has then been copied into the.ipynb environment



Figure 3 North American Cities list, This process has been repeated for each continent granting 6 lists featuring the most populated cities in each continent

## API Calls

The city list data was used in combination with API keys for each parameter that is needed to be observed. To discover where the pertinent information was being stored in the API call a trial run for a single city was run to check how the data was being called.

|  |  |
| --- | --- |
|  |  |
| Figure 4 Process showing how data was found in the API | |

## Storing pollution data

The Pollution Data that was obtained was found using this method, figure 5 illustrates the code used for this. For each pollutant an exception clause was added to ensure that the code did not stall when null value were found. The excerpt featured shows one such call which had the data appended to a list.

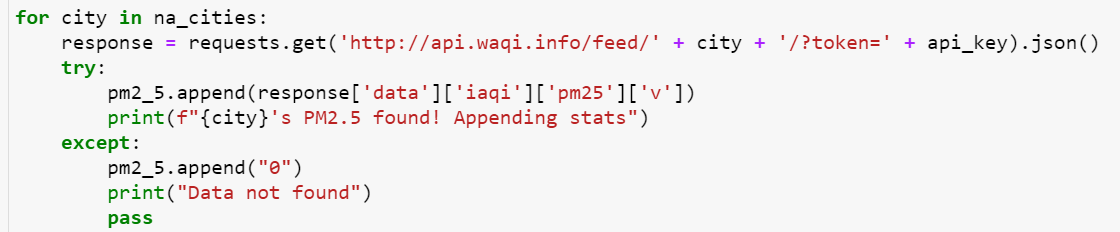


Figure 5 Calling and Appending pollution data

## Data frame creation

The list data for each pollutant was combined into a single data frame for the first step of visualisation. This single data frame is the backbone for the visualisations that will be displayed of the pollution data.

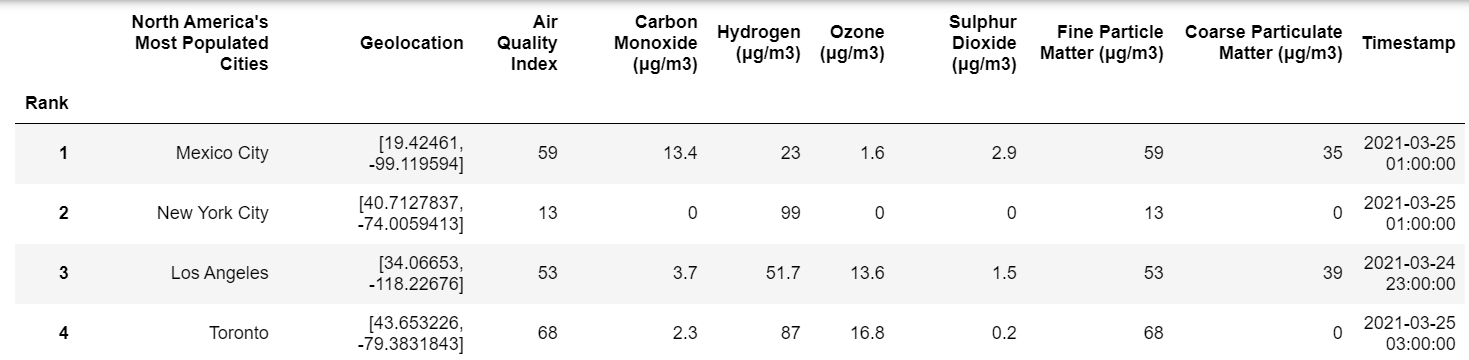


Figure 6 North American Pollution Dataframe

## Cursory Graphs

For the purpose of quickly visualising the data for each continent Bar and Pie charts have been used to show how each city is tracking with air quality and what proportions of pollutants are present.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  | Figure 7 AQI of N. Americas most populated cities |  |

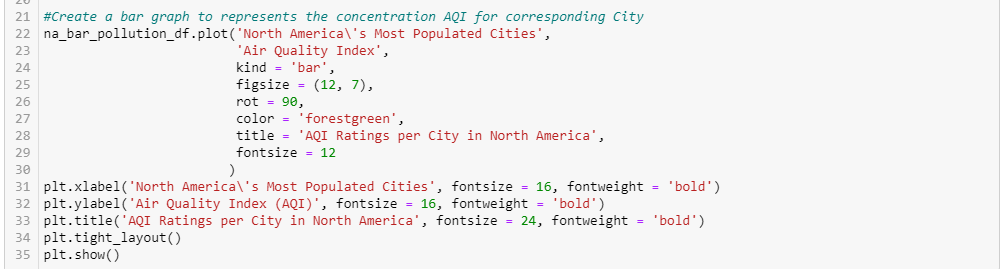


Figure 8 Bar Plot Code

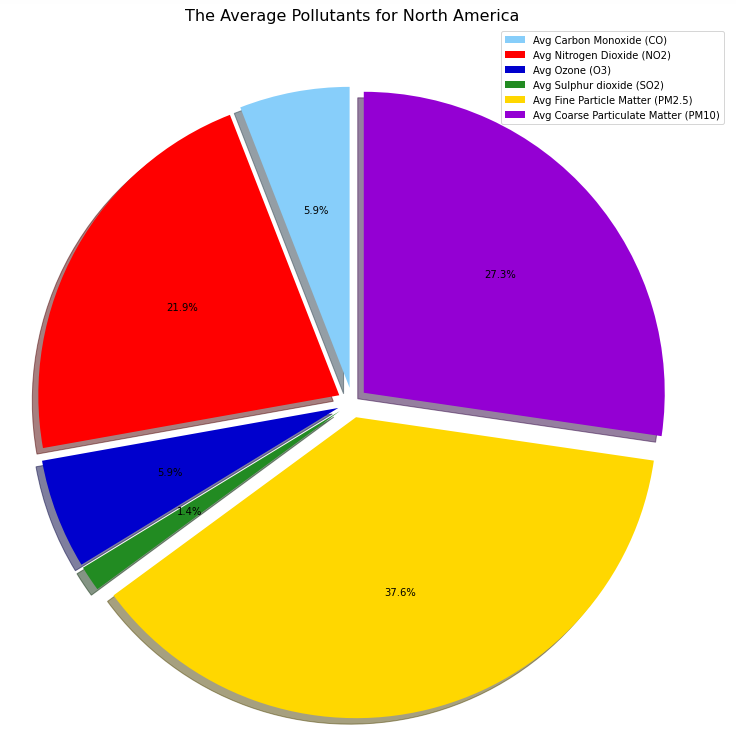


Figure 9 Pie chart of average N.American Pollutants



Figure 10 Pie Chart Code

## Saving csv files

By saving each data frame as a CSV further analysis can be performed by importing the data into a new .ipynb file.



Figure 11 Saving Data Frames as .csv files

## Map Visualisation

TO complete the general visualisation to show how polluted each city is a map showing the location of the city and its AQI is used. This can be performed on an individual city basis and also for all cities that have been considered. Shown below is the visualisation and code for a single city.

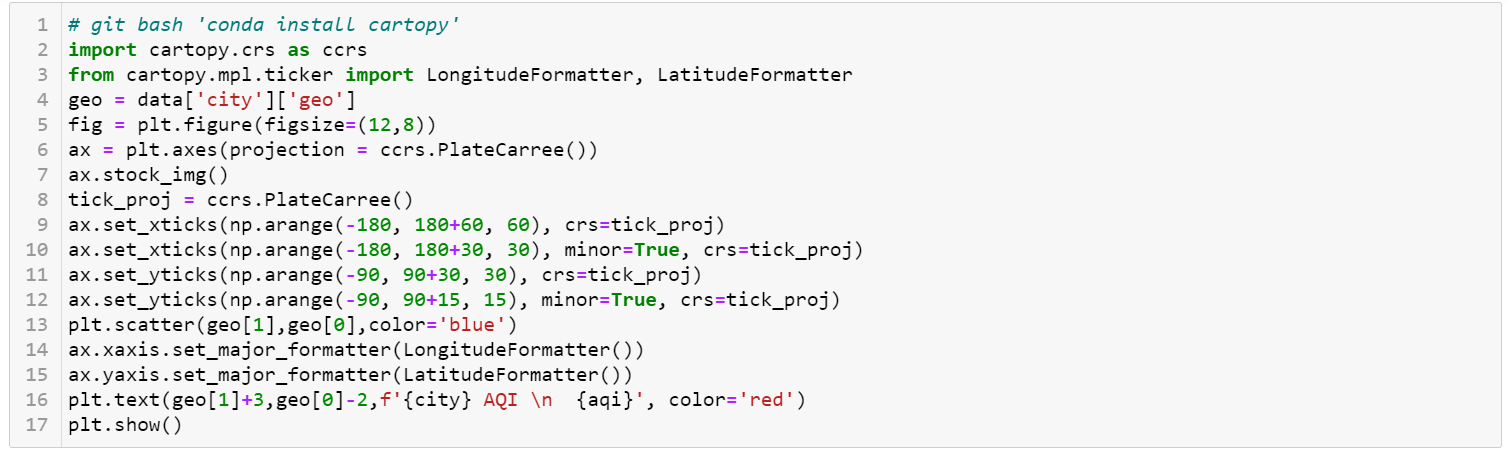


Figure 12 Map printing code

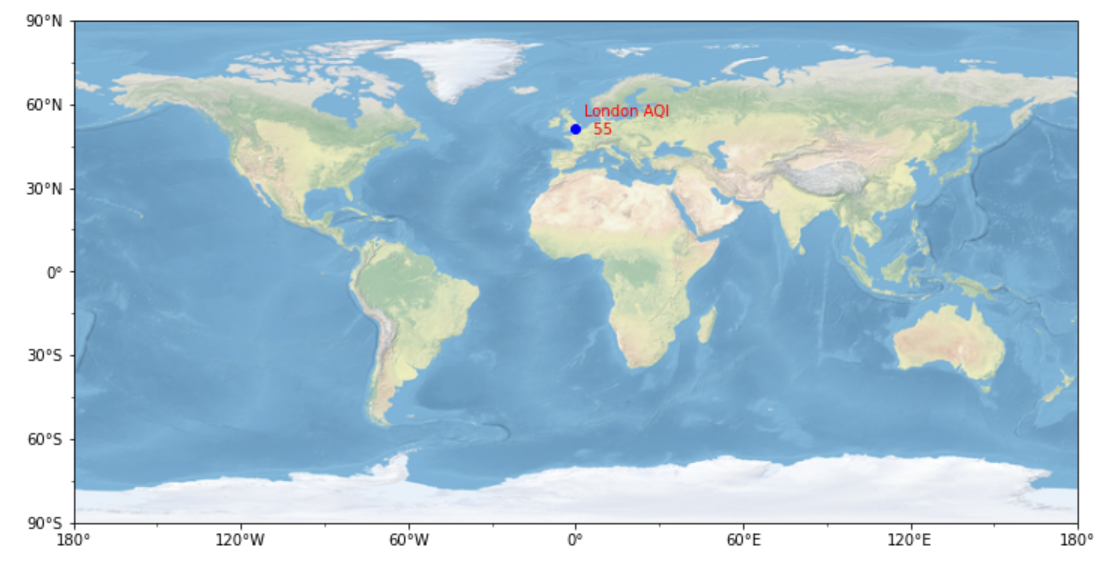


Figure 13 Single city and AQI visualised on map

# Research Questions

The purpose of this research is to gain an understanding of how pollution varies by continent. The questions that are posed are a small section of what can be asked of the data obtained. Further questions will be suggested upon conclusion of this report which would be worthy of further analysis

* Does AQI perfomance correlate to continent?
* How do the extremities of pollution differ based on location?
* If pollution is location dependent then what type of pollutant is the most prevalent?
* Which pollutant corelates most closely to the AQI ?
* If there a correlation between pollutant and location, can predictions on pollution be made?

# Data Analyses

1. Website data obtained from is waqi.info. Wonderful resource with interactive map featuring air quality data from around the world. [↑](#footnote-ref-1)