

# Project proposal

Viraj Rathod  
u1269659

- Data you plan to use

The data I plan to use is the real-time air quality. I believe this dataset will help me drive the story of air quality around the globe and help me derive some conclusions about the ranking of quality in different regions.

The URL for the same is:

<https://console.cloud.google.com/bigquery?project=hazel-torus-281022&folder=&organizationId=&p=bigquery-public-data&d=openaq&page=dataset>

- Techniques you intend to use

The techniques I intend to use will be around the advanced SQL concepts learned during the coursework. The window analytical functions, clauses, temporary tables using the WITH AS clause. The aggregations would be quite handy since I am planning to compare different parts of the world for the air quality.

I will also use the date functions to manipulate the dates given for comparing the moving average of air quality over a particular season/month. The arithmetic functions for calculating the statistics using various formulae to uncover certain conclusions. Also, grouping the data for using functions over a window of data. At the end, I will give a report of the air quality for different parts of the world, best and worst countries by the air quality.

- Questions you intend to ask

- How the air quality is different in separate zones or parts of the world?
- Where is the worst air quality from?
- Which country has the greatest number of bad performing cities in terms of air qualities?
- Calculate the moving average for the worse and best cities in air quality indices.

- Visualizations you intend to create

- Average air quality over Fall, Spring, Summer for best and worst cities around the globe.
- Top cities for the top country and bottom most country.
- Worst cities from a particular part of the world performing bad and best.
- Lastly, I will develop a story-telling dashboard. I have not decided a specific tool yet, but I will go ahead with Tableau or/and Google Data Studio.