Final Project

1. **Explain the technical details of the dataset that you chose to investigate.**

The dataset is related to the air quality which depends on the pollutant levels in the city. In the US, EPA (Environmental Protection Agency) measures the air quality index (AQI) through various procedures and sensory process. The various pollutant values used to measure the AQI are:

Row	Pollutant
1	Co (Carbon monoxide)
2	NO_2 (Nitrogen Dioxide)
3	O_3 (Ozone)
4	PM_{25} (Fine particulate matter)
5	SO_2 (Sulfur dioxide)
6	PM_{10} (Particulates)
7	BC (Ballistic Coefficients)

The data includes air quality measurements from 5490 locations in 47 countries.
Update Frequency: **Weekly**

Measurement units a. ug/m3: micro gram/cubic meter b. ppm: Parts Per Million

There are 0 null values across the dataset, I checked using where and IS NULL conditions.

The basic statistics summary is as follows:

```
SELECT
            MAX(latitude) as `max_latitude`,
            MIN(latitude) as `min_latitude`,
            MAX(longitude) as `max_longitude`,
            MIN(longitude) as `min_longitude`,
            MAX(averaged_over_in_hours) AS max_average,
            MAX(value) as max_value_pollutant,
            MIN(value) as min_value_pollutant,
            AVG(value) as avg_value_pollutant
            FROM `bigquery-public-data.openaq.global_air_quality`
            WHERE country = 'US'
```

Processing location: US

Run | Save query | Save view | Schedule query | More

Query results | SAVE RESULTS | EXPLORE DATA

Query complete (0.3 sec elapsed, 735.9 KB processed)

Job information | Results | JSON | Execution details

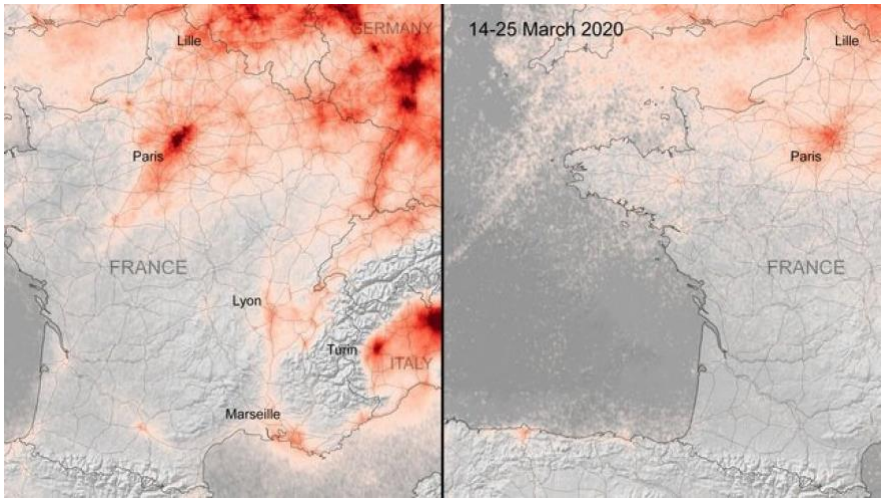
Row	max_latitude	min_latitude	max_longitude	min_longitude	max_average	max_value_pollutant	min_value_pollutant	avg_value_pollutant
1	64.84593	18.009556	-66.141655	-161.767	24.0	410.19	-1111.0	3.560479049153913

2. **Explain the importance of the dataset from a policy or business perspective.**

Most countries and their environmental regulators operate AQI, all of them with some differences. AQI may be calculated and presented at a local scale such as a City, or at State or Country levels. Their bands will differ from region to region, depending on their local conditions. Some AQI are based purely on concentration range, while some are based on the health impacts at given concentrations.

The businesses and product companies target the high air quality index cities and offer beneficial options such as an air purifier and it helps the public as well as businesses.
One such policy was implemented in IN, even odd, meaning that the number plates having odd number as their last digit can only take their vehicles out during odd number of the weekday. This was implemented in **Delhi** because of the rising air quality index which became unhealthy for certain health groups.
Reference: https://en.wikipedia.org/wiki/Odd–even_rationing

IAQ Index	Air Quality	Impact (long-term exposure)	Suggested action
0 – 50	Excellent	Pure air; best for well-being	No measures needed
51 – 100	Good	No irritation or impact on well-being	No measures needed
101 – 150	Lightly polluted	Reduction of well-being possible	Ventilation suggested
151 – 200	Moderately polluted	More significant irritation possible	Increase ventilation with clean air
201 – 250[9]	Heavily polluted	Exposition might lead to effects like headache depending on type of VOCs	optimize ventilation
251 – 350	Severely polluted	More severe health issue possible if harmful VOC present	Contamination should be identified if level is reached even w/o presence of people; maximize ventilation & reduce attendance
> 351	Extremely polluted	Headaches, additional neurotoxic effects possible	Contamination needs to be identified; avoid presence in room and maximize ventilation

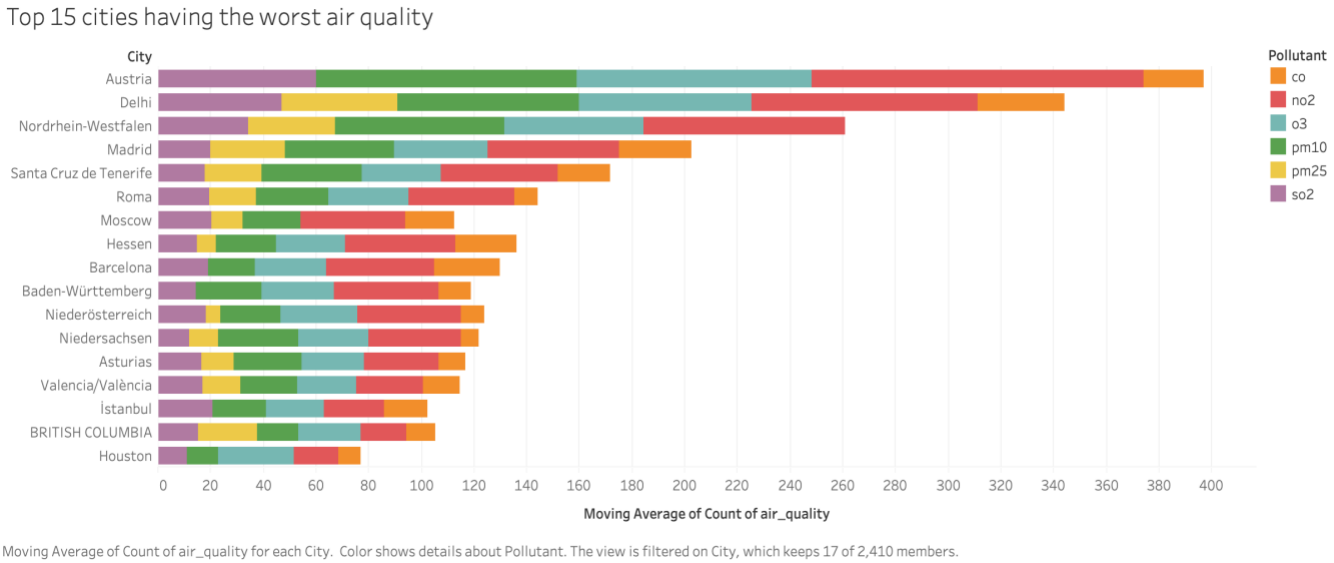


Picture of the country France, before and after the COVID-19 hit the world.

3. Explain the questions that you asked, and the importance of this question.

- Which 10 locations have had the worst air quality this month (as measured by high PM10)?
  - ➔ The top 10 locations with the highest concentrations of PM10.

    This tells you the much needed information about the top 10 countries where air quality is the worst and the measures which can be taken in order to curb the quality which is healthy for its population.
    The visualization is show below.



- Which city in the US has improved its air quality (PM2.5) the most since 2010?
  - ➔ This query joins the EPA Historical Air Quality dataset with OpenAQ to retrieve the cities in the US with the most improved PM2.5 concentrations since 2010

- What is the current global picture of PM10 concentration?
  - ➔ PM10 (small particulate matter of 10 microns or less in diameter) is known to cause disease and cancer, contributing to an estimated 3 million premature deaths worldwide per year in 2012. This query reveals PM10 concentrations by location.

- Which country has the greatest number of bad performing cities in terms of air qualities?
  - ➔ Certainly because of the population boom, India is the worst performing country in terms of air quality.

4. Explain your conclusions

- There are 47 countries and when you plot the values of the AQI over a range of several years, you see that the quality has degraded in every part of the world.
- The factors being, traffic increase, population boom, globalization and industrialization.
- The social aspect: understanding the challenges of air pollution, how it is measured, where it's more problematic, how it's affected by temperature ...
- The data science aspect: how to use Big Query to get the most out of our database and how to play with a query when you want to do a lot of different things.

5. **Provide your visualizations**

   https://github.com/virajrathod/IS6850

6. **Methodologies**

   - How did you analyze the data?
     ➔ I used Big Query and Tableau calculations for the custom SQL queries/
   - Provide all queries and code.
     https://github.com/virajrathod/IS6850

   - Explain your visualization methodology.
     ➔ I used tableau for my visualizations.