Global Air Quality

Team 2: M. Borunda, D. Brown, K. Childers, J. Cisneros, and R. Yusuff

Southern Methodist University

Data Science Bootcamp (09/2023)

Abstract

Project 3 encompasses a data visualization and data engineering track for students to showcase their skillset. Team 2 has found an open-source Kaggle dataset to manipulate and answer Project 3 requirements. A story will be revealed by using visualizations and skills learned throughout the course to design a robust database.

Global Air Quality

According to the World Health Organization (WHO), air pollution is one of the greatest environmental risks to health that affects people from all different backgrounds. Specifically, outdoor air pollution in both cities and rural areas were estimated to have caused 4.2 million premature deaths worldwide per year in 2019; this mortality is due to exposure to fine particulate matter (PM2.5), which causes cardiovascular and respiratory disease, and cancers. In addition, air pollution is considered the second highest risk factor for noncommunicable diseases and examining it is key to protecting public health. Most sources of outdoor air pollution are well beyond the control of individuals, and this demands concerted action by local, national, and regional policymakers working in sectors like energy, transport, waste management, urban planning, and agriculture to develop initiatives and counteract the negative effects of air pollution.

# Data Visualization/Engineering

The dataset was transformed into visualizations by first cleaning it and creating a data frame that could be used for queries. Python was utilized to drop null values reflected in the raw data as shown in the figure below. The column titled “Country” only had 16,393 values and the other columns reflected 16,695. There was a total of 302 rows of data that did not have a Country associated with it and 2 null values. These values were dropped from the data set to have more consistency.

A screenshot of a computer

Description automatically generated

**Figure 1: Image of df.info on dataset**

The dataset did have some other quality issues to include it containing some incorrect latitude and longitude points that do not correspond to the country or city they are supposed to be identifying. In addition, the last update for this dataset was performed 9 months ago.

The team then used mySQL to filter the data with queries. Those filters were then transposed to a JavaScript file that allows users to interface with the dashboard by choosing a country or city to see its AQI values. The HTML and JavaScript files created by Team 2 can be found in the eda and app folders in their project GitHub. A link was also provided ([kokimber.pythonanywhere.com)](http://kokimber.pythonanywhere.com/) to show a successful broadcast of data to the web interface. The data was broken out by these data points for the user to view it by Country or City. The data visualizations were created in response to these three questions:

1. Which Countries have good and bad air quality ratings?
2. How do big cities compare with more rural areas?
3. What are the main pollutants affecting air quality?

A screenshot of a computer

Description automatically generated

**Figure 2: Image of query for countries with the worst AQI**

A screenshot of a computer

Description automatically generated

**Figure 3: Image of query for countries with the best AQI**

A table with numbers and names

Description automatically generated

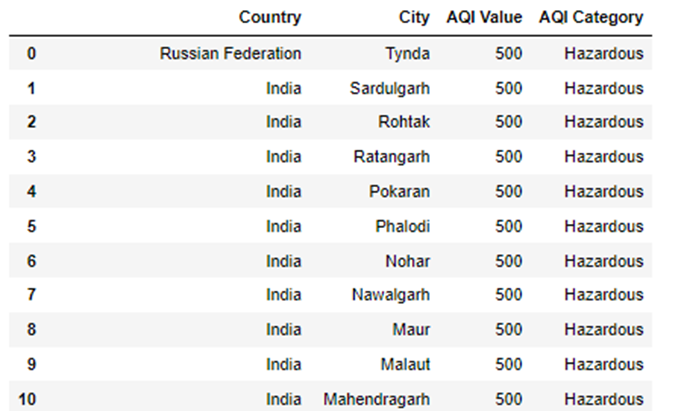
**Figure 4: Second image of query result for countries with good AQIs**

**Findings**

After exploring the data, Team 2 had six findings to highlight. Various locations around the world were picked to analyze for a global spread of AQI values. Before proceeding with the findings, it is important to know that AQI values range from 0 to 500 with 0 falling under the “Good” Air Quality Category and 500 falling under the “Hazardous” Category (Figure 7 shows the range of values for AQI).

Finding 1 was that India is one of the countries with the worst ratings for air quality. Nine of the top 10 cities in the world with the worst ratings are in India. In addition, 40.7% (Figure 6) of its cities fall under the “Unhealthy” category. This makes sense considering multiple cities in India have AQI values of 500 which means they have “Hazardous” air quality conditions. Despite these findings, there are a handful of cities with “Good” ratings with AQI values under 50 such as Port Blair (an island far from the mainland), Chengam, and Hindupur. Below are some visual representations of Finding 1:

***Top 10 Cities in the World with worst AQI values***



**Figure 5: Cities with the worst AQI**

A pie chart with different colored circles

Description automatically generated

**Figure 6: AQI Categories of India**

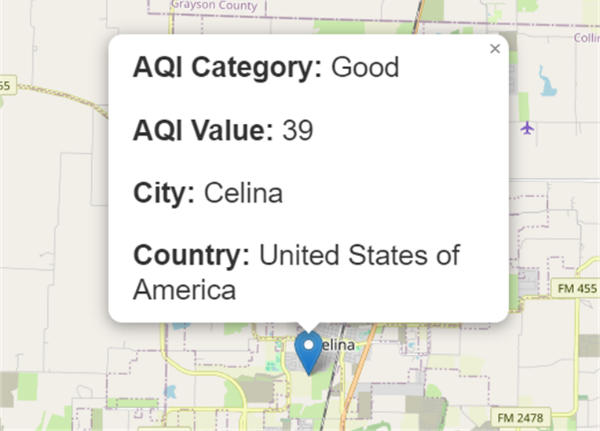
A speedometer with text on it

Description automatically generated

**Figure 7: AQI Categories**

Finding 2 explored SMU’s location in Dallas, TX. Team 2 found downtown Dallas as having a moderate Air Quality Index value of 70. Interestingly a city 50 miles north of Dallas called Celina, TX had a much lower AQI value of 39. This makes it fall under the “Good” air quality category. This supports the research question on how AQI values compare between rural areas and cities. The specific values of these two locations show how AQI values are higher in bigger cities compared to more rural areas.

A screenshot of a map

Description automatically generated 

**Figure 8: AQI of each Texas city location**

Finding 3 revolved around Argentina. It found that 84.7% of cities in Argentina fall under the “Good” air quality category. The city with the worst AQI value rating ended up being Villeta, which has a rating of 90 (moderate air quality). The issue of limited datapoints on cities per country would affect any conclusions made on whether Argentina has an overall average AQI value of “Good”. More research should be done on more cities to give an accurate depiction of the country’s average status.

A green and red circle with a red circle with numbers

Description automatically generated

**Figure 9: AQI Categories of Argentina**

A blue and green bar graph

Description automatically generated

**Figure 10: Average AQI of cities in Argentina**

Finding 4 focused on Iceland and it has “Good” air quality ratings. This conclusion cannot be taken as 100% accurate since there was only a single rating that was reported. More research should be completed on multiple cities throughout the country to reach the conclusion that Iceland has an overall “Good” air quality.

A screenshot of a computer screen

Description automatically generated

**Figure 11: AQI Categories and values of Iceland**

Finding 5 was completed on Chihuahua City (Misha’s hometown) which has a much better air quality compared to Mexico City. Again, this proves how bigger cities experience worse air quality compared to rural areas. Team 2 found a pattern overall that countries located away from industrialized ones (Pacific and other isolated islands) as well as rural areas that were not near big cities tended to have lower AQIs.

A screenshot of a map

Description automatically generated A screenshot of a map

Description automatically generated

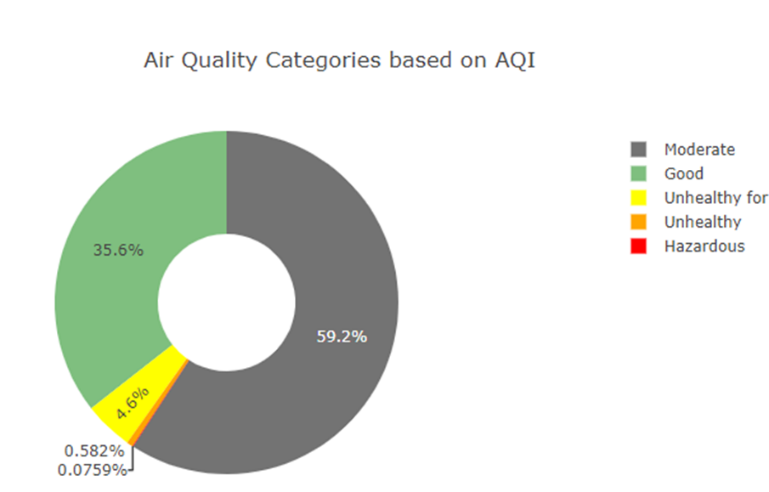
**Figure 12: AQI values of cities in Mexico**

Finding 6 explored the United States and discovered that based on this dataset, the main pollutant affecting air quality was fine particulate matter (PM 2.5) with 60.6% and it is followed by Ozone with 29%. On average, the United States has “Good” to “Moderate” air quality. Overall, 94.8% of the ratings fall under the above two categories.

A pie chart with text on it

Description automatically generated

**Figure 13: Concentration of Pollutants in The United States**



**Figure 14: AQI Categories of The United States**

**Dashboard**

The Dashboard that was created to have an approachable way for people to explore Air Quality across the world and it includes 4 different tabs. The first tab is the Home page with useful information about Air Quality. The second tab is the actual Dashboard with an interactive map, and two donut charts that are adjusted and modified accordingly to a location filter. The third tab includes information about Team 2, and lastly the fourth tab includes links to works cited. The colors chosen by the AQI Category donut chart range from red to green to visually represent “Good” AQI values (in green) and “Hazardous” AQI values (in red). For the pollutant percentage donut chart, very distinct colors were chosen to clearly differentiate each unique pollutant. The use of these distinct colors ensured that the user would not interpret these values with being relational to the other diagrams on the dashboard. For the Bar Chart, blue was the color of choice since it is usually the color associated with air.

A screenshot of a computer

Description automatically generated

A map with blue pins

Description automatically generated

**Figure 15: Screenshot of the Dashboard**

**Limitations and Bias**

This study was performed with some limitations due to the data available. One limitation was that not all air quality pollutants are included in this data set such as Sulfur Dioxide (SO2). There was also no date and/or time information included therefore it is unsure if the readings were all coming from the same point in time or over a span of different dates. Finally, an additional limitation was that some Cities have multiple entries while others only have one. A more accurate analysis could be made if all locations had the same count of readings done on the same date.

**Future Implementation**

Naturally this study is intriguing due to its global effect and further research should be done to determine any global trends as well as forecast and future implications from air pollution. Some questions to consider are:

1. Can AQI readings fluctuate during the same day?
2. Does temperature influence AQI values?
3. What age range is more sensitive to bad Air Quality?
4. Which Country has been the most successful in improving their overall air quality?
5. What industries are being more proactive in implementing solutions?

In a broader aspect, Team 2 could delve into health data to see which ailments are more sensitive to a bad AQI or use these measurements to determine if a country or industry is impacting an AQI value for the good or bad. This could support conclusions on if initiatives to save the environment are positive or negative.

**Conclusion**

Team 2’s web interface provides an opportunity for individuals to explore their air quality in an easy-to-read dashboard. It also brings awareness to the importance of maintaining clean air and the importance to tracking air quality. Overall, a dashboard was created to provide three views (Concentration Per Pollutant, Air Quality Categories based on AQI, and Average AQI Value) by filtering per Country.

References

Ambient (outdoor) air pollution (December 2022). World Health Organization.

https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Air Quality Index, Ozone Alerts & PM Alerts, and Health Advisories. Oklahoma Environmental Quality.

https://www.deq.ok.gov/air-quality-division/ambient-monitoring/aqi-alerts-advisories/

Earth Day 2020: A Guide for All Ages (2020). <https://digitalprojects.davidson.edu/earthday2020/air-pollution/>

Ramachandran, A. World Air Quality Index by City and Coordinates (CC BY-NC-SA 4.0).Kaggle. <https://www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates/data>

World Economic Forum (Sep 2020). Can We Put a Price on Clean Air? Yes – And Here It Is.

<https://www.weforum.org/agenda/2020/09/we-can-put-a-price-on-clean-air/>

Terminology

**Air Quality Index (AQI)**: Index is used for reporting daily air quality. It tells you how clean or polluted the air is in a region.

**Particulate matter (PM2.5)**: Fine Particulates such as sulfates, nitrates, ammonia, sodium chloride, black carbon, mineral dust and water.

**Carbon monoxide (CO):** Toxic gas produced by the incomplete combustion of carbonaceous fuels.

**Ozone (O3)**: Ozone at ground level – not to be confused with the ozone layer in the upper atmosphere – is one of the major constituents of photochemical smog and it is formed through the reaction with gases in the presence of sunlight.

**Nitrogen dioxide (NO2)**: NO2 is a gas that is commonly released from the combustion of fuels in the transportation and industrial sectors.