Term Project 1 Report

Wangui Gachugu, Zain Farooq, Mintesnot Yimer, Jade Loaëc

Evolution of Air Quality in the USA

The objective of Term Project 1 is to conduct an analysis of air quality. Our group chose to focus specifically on the United States, examining the evolution of air quality over time and the distribution of polluted areas across the country.

Technologies + Tools Used

- The following python libraries: numpy, pandas, seaborn, matplotlib
- Power BI
- Canva (for the poster)
- Google Docs (as a collaborative tool)

Methodology:

Once the topic was assigned, we took time to explore the selected databases and considered various ways to analyze air quality.

The analysis aimed to assess changes in air quality (AQI) across U.S. states from 2013 to 2023. The methodology included data preprocessing, aggregation, percentage change calculation, slope calculation, state selection, and visualization, enabling a comparison of states with improving or deteriorating air quality trends over the decade.

What was Implemented

1. Change in median AQI recorded for the five states with the largest increases and decreases in AQI

2. Frequency polygon to track the evolution of the states with the highest and lowest median AQI from 2013 to 2023

3. Geographic distribution of maximum recorded AQI in 2023

How it was Implemented

1. To get the change in median AQI recorded, we used the following process:
    a. Aggregating AQI data by state and year to calculate median values.
    b. Extracting AQI data for 2013 and 2023, then merging them for comparison.
    c. Calculating the AQI change by subtracting 2013 values from 2023 values.
    d. Sorting states by AQI change to identify the top 5 most improved and worsened states.
    e. Visualizing the data with a horizontal bar plot to highlight these changes.

2. The frequency polygon implementation was as follows:
    a. Data from 2013-2023 was imported and cleaned.
    b. Median AQI values were calculated for each state.
    c. Percentage changes in AQI were determined.
    d. Linear regression showed AQI trends by state.
    e. The top and bottom 5 states with extreme changes were identified.
    f. Frequency polygons highlighted the decade's air quality shifts.

3. Geographic distribution of maximum recorded AQI in 2023 was calculated using Power BI:
    a. We imported the raw air quality data into Power BI, making sure every coordinate and data point matched our required format for accurate geographic placement.
    b. We filtered through the dataset to extract 2023 records, removing any data gaps and confirming all coordinates would map correctly to their locations.
    c. We constructed the map visual in Power BI, connecting our latitude and longitude points while setting maximum AQI as our main measurement value.
    d. We built the color scheme to show AQI intensity levels, adding detailed information panels that would appear when users hover over specific cities.
    e. We divided our AQI values into distinct groups based on EPA guidelines, applying specific colors to help users understand different air quality ranges.
    f. We polished the final visualization with clear headings, detailed legends, and interactive features that would help users explore the data effectively.

Conclusion

There was a general increase in air quality in many states from 2013 to 2023.

Challenges faced + how we managed them

We came up with many ideas, but implementing them required finding datasets that had data input in specific formats, which was difficult to find.

Although the United States has a wealth of public data, finding datasets with precise information on air quality, wildfires, road traffic, and socioeconomic factors—factors we wanted to connect with our main topic—proved difficult.

The global dataset we found (World Air Quality) had the following problems:

- The coverage was not comprehensive and only captured a small fraction of cities in some countries

- Omission of data known to exist, but which could not yet be accessed due to language issues or limited
- Accessibility
- The measures were localized (city average are an average of ground measurement stations)

We then decided to approach the problem differently by basing our analysis on the two datasets we had found, rather than searching for datasets to match our initial analysis ideas. This allowed us to move forward and adjust our topic accordingly.

Future Work on the Project

1. Expanded Analysis Scope:
The next phase could explore how factors like income levels affect air quality in different areas. We could study how heavy traffic contributes to poor air quality and examine whether wildfires make air quality worse in certain seasons of the year.

2. Data Enhancement:
Future work should include more cities to get a better picture of each state's air quality. We need to collect data from more monitoring stations and translate valuable information from non-English sources. This would help fill in missing data for areas we couldn't study before.

3. Advanced Analytics:
We could build tools to predict future air quality trends and use artificial intelligence to spot patterns in pollution levels. This would help us identify high-risk areas early and understand how new environmental rules affect air quality.

4. Policy Impact Assessment:
Future research could show whether clean air laws actually work and how different states' environmental rules affect their air quality. This would help calculate the financial impact of both good and bad air quality on communities.

5. Community Impact:
The next step would be to study how air quality affects people's health in different neighborhoods. We could examine whether some communities face more air pollution than others

and look at how aware people are about their local air quality.

References

World Air Quality - OpenAQ

Annual AQI by County