Assessing the Impact of Air Quality on Climate Variability Across the United States

Introduction

Understanding air quality trends is crucial for identifying environmental challenges and implementing necessary public health measures and other precautions. This report analyses Air Quality Index (AQI) trends in the United States of America and any correlation between the change in AQI data with respect to the climate variability factors. The analysis explores how AQI values fluctuate over time (from 2013 to 2022) and identifies regions in and around the USA that are most affected by air quality issues.

Background of the Project

Fluctuations in the air quality has a major impact on public health and environmental sustainability. Monitoring and analysing the Air Qualityby Index (AQI) provide critical insights into pollution levels, sources, and their impacts on populations. In the United States, understanding AQI trends is essential to identify environmental challenges. Implementing policies to improve air quality, and mitigate health risks is a crucial factor. The Environmental Protection Agency (EPA) measures AQI, considering pollutants like particulate matter (PM2.5 and PM10), ozone, carbon monoxide, sulphur dioxide, and nitrogen dioxide. Tracking these pollutants helps to assess compliance with air quality standards and evaluate health advisories.

AQI trends vary widely across the country due to differences in industrial activities, urbanization, transportation, and natural events like wildfires and dust storms. Additionally, climate variability factors, such as temperature, humidity, and wind patterns, influence pollutant diffusion and concentration. Research shows that climate change may intensify air pollution, affecting human health and ecosystems (Fiore et al., 2015; Jacob & Winner, 2009).

The ambitious scope of this project involves:

1. **Nationwide Analysis:** Investigating AQI trends across the entire United States, comparing states, territories, and counties to identify areas with the highest pollution levels.
2. **Climate Correlation:** Exploring the relationship between AQI changes and climate variability factors to understand how shifts in weather patterns impact air quality.
3. **Policy Implications:** Analysing the effectiveness of current regulations and proposing recommendations for better air quality management.
4. **Health and Environmental Impact:** Assessing the potential risks associated with poor air quality and its implications for public health and environmental protection.

By conducting a comprehensive analysis, this project aims to contribute to a deeper understanding of AQI trends, climate-related impacts on air quality, and how policy measures can improve environmental and public health outcomes in the United States.

References:

- Fiore, A. M., Naik, V., & Leibensperger, E. M. (2015). Air Quality and Climate Connections. *Journal of the Air & Waste Management Association*, 65(6), 645-685. Full Link
- Jacob, D. J., & Winner, D. A. (2009). Effect of Climate Change on Air Quality. *Atmospheric Environment*, 43(1), 51-63. Full Link

Research Questions

1. **Trends over time**

- How has the air quality changed over the years for specific states?
- Are there any noticeable trends in air quality anomalies over the years?

2. **Geographical Analysis**

- Which states or counties have consistently experienced the best or worst air quality?
- In which regions are unhealthy air quality days most common?

3. **Health and Safety Insights**

- Correlation between AQI levels and the occurrence of very unhealthy or hazardous days.
- Identify areas with frequent 'Unhealthy for Sensitive Groups' days to target health advisories.

Methodology

Data Sources

While I had to collect a large number of datasets based on yearly reports from NOAA and AQS links below, I had to use Python to preprocess the data using the ETL processing procedure through Jupyter Notebook initially. For this I had to adjust the NOAA datasets according to their size and content, remove any irregularities, adjust the values to remove any null or uncertain information and then finally combine them into a single data source for the NOAA dataset. Once the dataset has been adjusted accordingly to match the content in the AQI dataset, I had to merge them, either using the common columns such as ‘Year’ and ‘State’ using the Notebooks or using the Tableau data source.

https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/county/mappingLinks to an external site.

https://aqs.epa.gov/aqsweb/airdata/download_files.htmlLinks to an external site.

Data Preprocessing

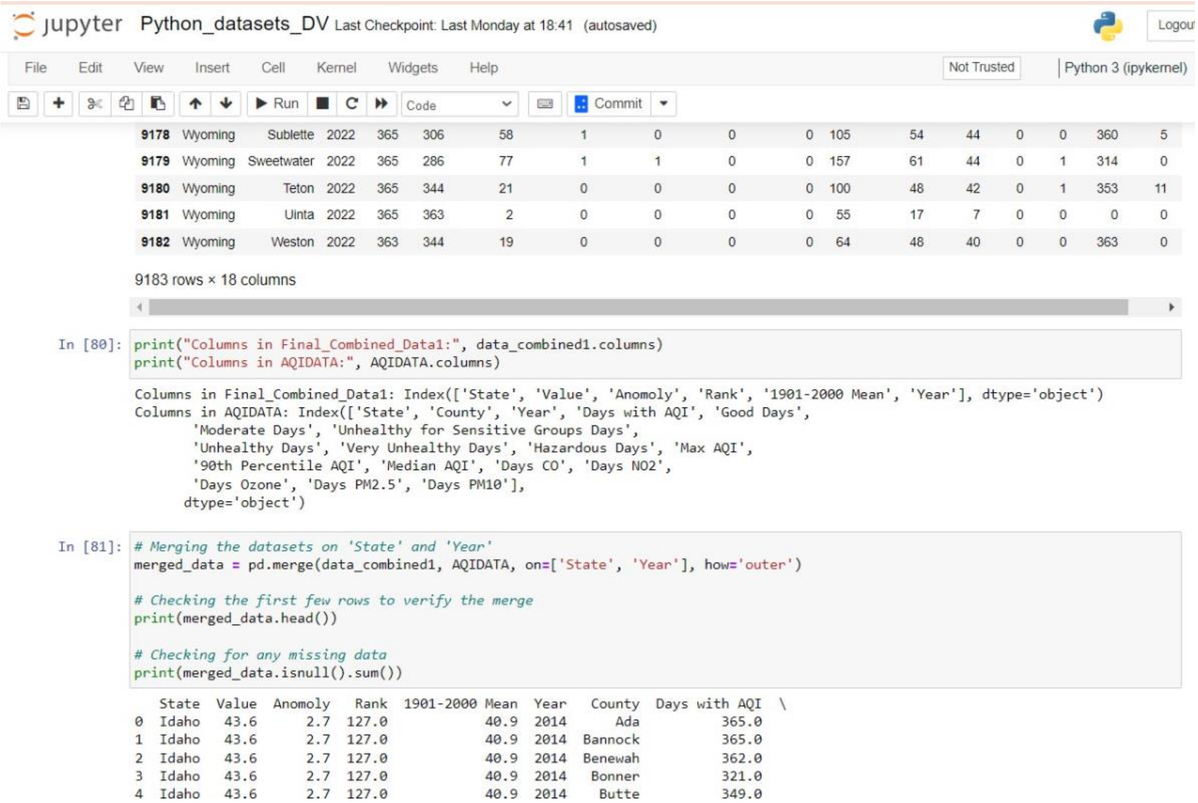
To ensure data consistency and quality, the following steps were taken:

- **Combining Datasets:** The **final_combined_data1** and **Annual_AQI** datasets were merged based on common columns such as **Year** and **State**.
- **Cleaning Data:** Missing values were handled by either replacing them with appropriate default values or removing incomplete records.
- **Normalizing Data:** All numerical columns were normalized to ensure consistency in the analysis.

Please see the complete description of how this is achieved below:

```
In [1]: import pandas as pd

In [10]: file_path = 'E:/Downloads/data.csv'
         column_headers = ['ID', 'Name', 'State', 'Value', 'Anomaly (1901-2000 base period)', 'Rank', '1901-2000 Mean']
         data = pd.read_csv(file_path, names=column_headers, skiprows=5)

In [15]: years = [2017, 2018, 2019, 2020, 2021, 2022]
         total_rows = len(data)

         # Calculating how many rows each year should cover
         rows_per_year = total_rows // len(years)
         remaining_rows = total_rows % len(years)

         # Generating a list of years where each year repeats according to the calculated distribution
         year_assignments = []
         for year in years:
             year_assignments.extend([year] * rows_per_year)

         # For any remaining rows, extending it with the last year
         year_assignments.extend([years[-1]] * remaining_rows)

         # Assigning the list of years to a new column in the DataFrame
         data['Year'] = year_assignments

         # Verifying the distribution
         print(data[['ID', 'Name', 'State', 'Year']].head())
         print(data[['ID', 'Name', 'State', 'Year']].tail())

                               ID  Name State  Year
         AL-009 Blount County    Alabama  62.6     2  2017
         AL-011 Bullock County   Alabama  65.6   1.5  2017
         AL-013 Butler County    Alabama  66.3   1.8  2017
         AL-015 Calhoun County   Alabama  62.9     2  2017
         AL-017 Chambers County  Alabama    63   1.5  2017
                                   ID  Name State  Year
```

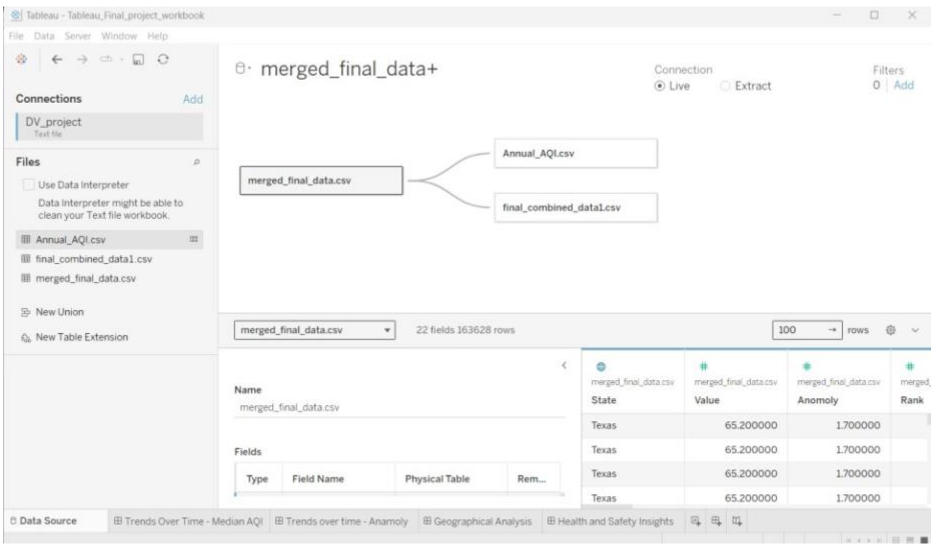
1. Loaded the dataset data.csv files containing different years of data containing Climatic information and then had to process them to have a common ‘Year’ column. I had to do this for each year from 2012 to 2022.



2. Later I had to combine the datasets and then preprocess the datasets for any irregularities. Then I saved them and combined them as some data sources that I can later combine as a single source.

Challenges faced:

The data source has too many problems while merging properly and each issue has to be solved by more processing of the data using Python. The final merged data file contains all the merged information which I later had to connect to this source with the AQI and the Climate data.



The primary dataset used for this analysis is the Annual AQI data, which provides detailed information on AQI levels for various regions. Additional datasets were sourced to include information on specific anomalies and environmental factors that could affect air quality.

Analysis:

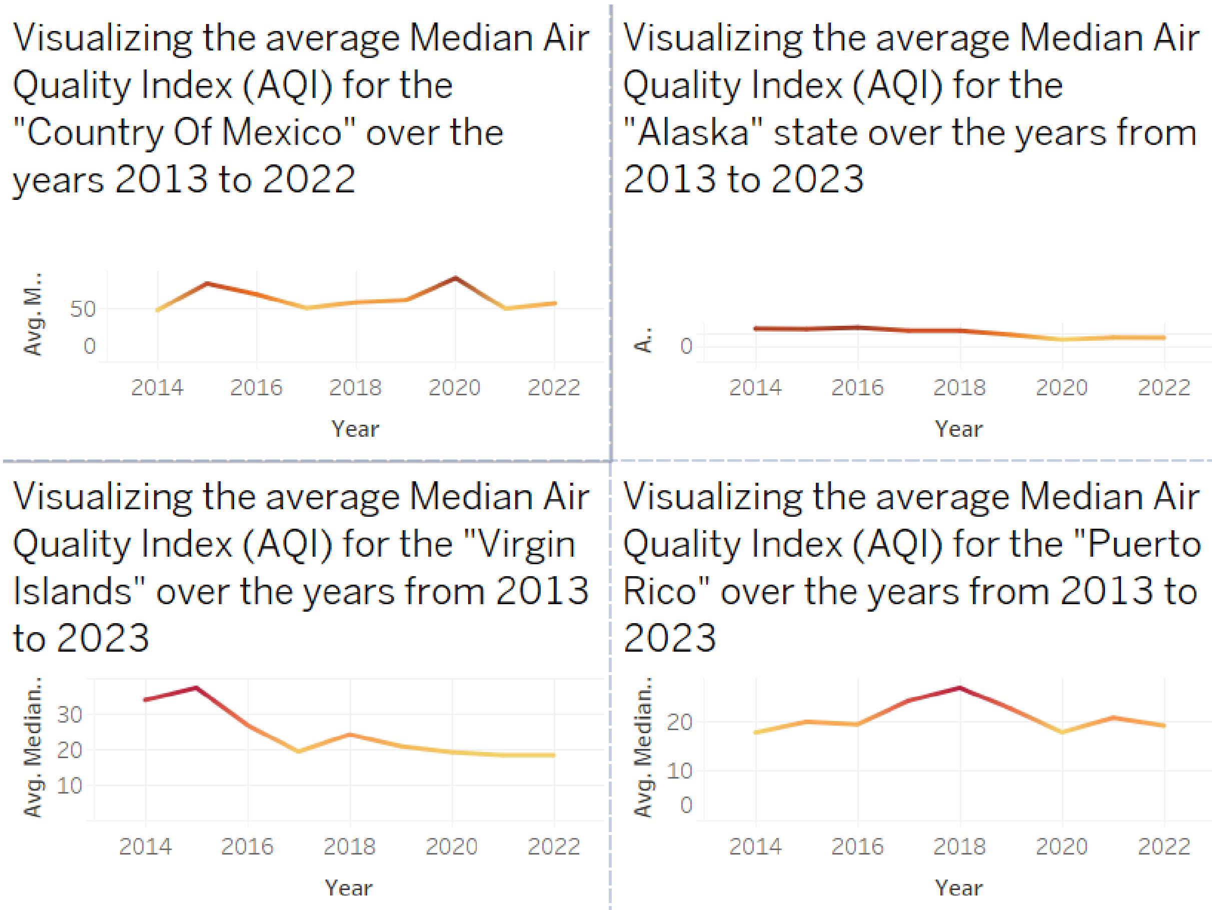
Median AQI Trends Over Time for States: How has the air quality changed over the years for specific states?

Overview:

Monitoring Air Quality Index (AQI) trends over time helps understand how pollution levels vary across different regions, providing insights into localized and systemic challenges. This report specifically examines median AQI trends across several states in the United States and compares them with other regions, particularly the Country of Mexico, to understand these variations better.

Key Observations:

Below are few observations based on some states in the United States:



Range of AQI Values: The average Median AQI ranges from approximately 48 to 78 over the observed period for the Country of Mexico. This range is significantly higher than those seen in the previous locations. This shows that the Country of Mexico has a poorer air quality overall.

Trend in AQI: There is a noticeable peak in AQI values around 2019 for the country of Mexico, which sharply rises and then falls. This might indicate a specific environmental event or change in air quality regulations and their enforcement around that time.

Comparisons with the other States and US territories:

Compared to Alaska: The "Country of Mexico" produces much higher AQI values than Alaska. It results in a significantly poorer air quality. Alaska's AQI values were the lowest and most stable among the regions visualized in the dashboard.

Compared to the Virgin Islands and Puerto Rico: While the Virgin Islands and Puerto Rico showed moderate AQI ranges, although Puerto Rico's AQI values are marginally lower on average, the "Country Of Mexico" shows higher values throughout the years. This highlights greater air quality challenges in this region.

Variability: The "Country of Mexico" shows a more variability and a trend towards higher pollution levels than the other regions discussed.

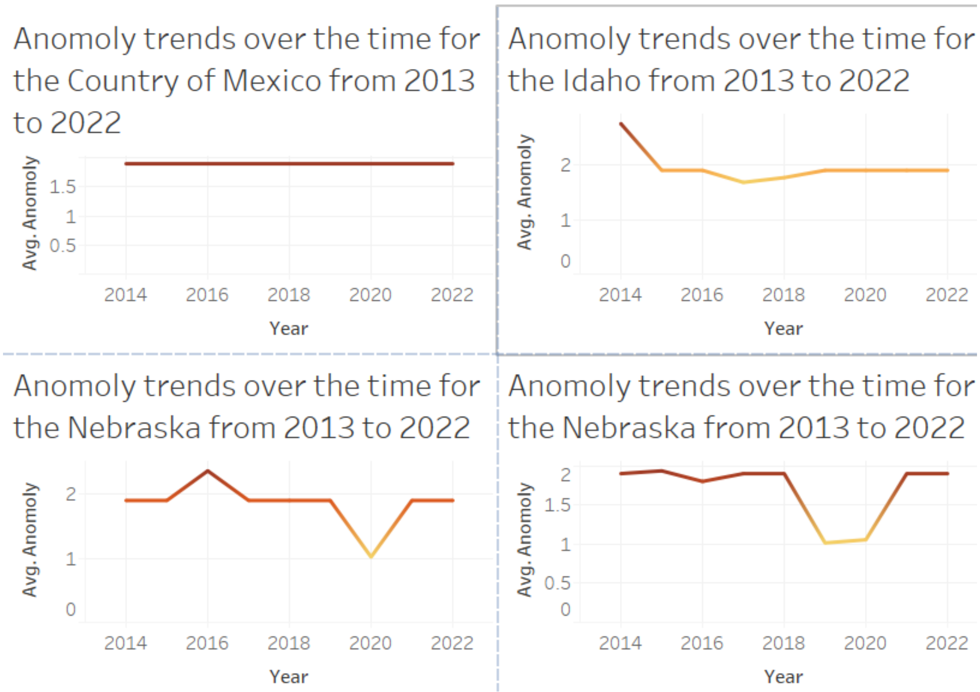
Conclusion

- Country of Mexico experiences significant AQI fluctuations, with the 2019 peak suggesting a specific environmental event.
- Compared to other regions, Country of Mexico has higher AQI values, indicating poorer air quality.
- Certain counties in Country of Mexico are particularly affected by unhealthy air quality days.

Average Anomaly trend for States: Are there any noticeable trends in air quality anomalies over the years?

Observations:

Below are few observations for Anomality based on some states in the United States:



Stable Anomaly Values: The anomaly values are stable at 1.9 throughout the period, suggesting that there hasn't been significant deviation from the expected norms based on the historical data used to calculate these anomalies. This constancy implies that while conditions may not be improving, they are not worsening either.

High and Fluctuating AQI Values: Contrary to the stability in the anomaly values, the AQI values as previously seen fluctuate significantly, with a range from about 48 to 78. This indicates that air quality in terms of AQI is not stable and experiences significant highs and lows.

Analytical Insights:

Disparity in Stability: The stability in anomaly values despite fluctuating AQI suggests that the baseline or the norm from which anomalies are calculated might be set at a level that expects higher pollution. This means the expected average conditions may already account for relatively poor air quality.

Implications for Environmental Policy: The stable anomaly against a backdrop of fluctuating AQI values could indicate that the measures in place to manage air quality are not effectively mitigating the most significant pollution events. This could suggest a need for revising air quality standards or implementing more stringent environmental management strategies.

Health and Environmental Concerns: High AQI values often correlate with increased risks to health and the environment. The acceptance of these levels as 'normal' (reflected by the stable anomaly) could pose ongoing risks to public health and necessitate targeted interventions.

Key Observations from Each Region:

Country of Mexico:

- The average anomaly appears consistent around 1.9 to 2 throughout the period, showing minimal fluctuations.
- This stability could suggest a persistent condition or unchanged regulatory baseline in air quality.

Idaho:

- There is a sharp decrease in anomaly values around 2014, after which the values stabilize.
- This decline might suggest an event or policy change that altered the air quality baseline, resulting in lower anomalies.

Nebraska:

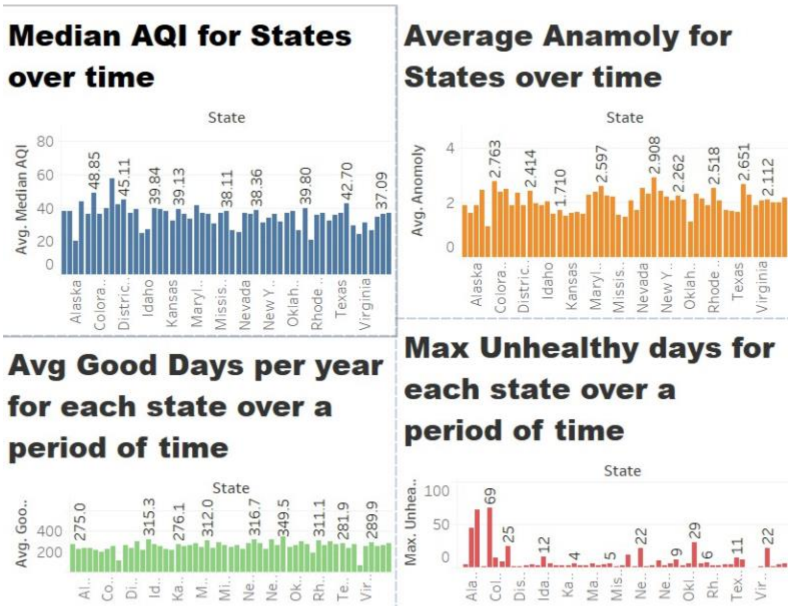
- The anomaly values are relatively higher around 1.5 to 2 and show fluctuations.
- There's a notable dip around 2018, which indicates a temporary improvement or environmental change, followed by a return to higher values.

Geographical Analysis: Which states have consistently experienced the best or worst air quality?

Overview:

By analysing median AQI trends over time, average anomalies, the number of "Good" days, and the maximum unhealthy days, we can identify states that consistently experience the best or worst air quality.

Key Observations:



Median AQI for States Over Time

- **High AQI States:** States with consistently high AQI values have noticeably poorer air quality. Such states face challenges related to pollution from industrial activities, dense traffic, or natural events like wildfires.

- **Low AQI States**: States like Alaska and West Virginia have lower AQI values, suggesting better air quality. These states likely benefit from less industrial pollution and lower population densities.

Average Anomaly for States Over Time

- The bar chart of average anomaly reveals the variation from the historical average for each state.
- States with the lowest anomaly values suggest either consistent environmental conditions or significant policy changes that have improved air quality.

Average "Good" Days Per Year

- The average number of days classified as "Good" for air quality varies by state.
- States like California and Pennsylvania have fewer "Good" days, indicating potential air quality challenges due to factors like traffic, industrial pollution, or natural events like wildfires.

Maximum Unhealthy Days

- This metric identifies states with the highest recorded number of unhealthy air quality days.
- States with significantly more unhealthy days might require stronger environmental regulations or interventions to manage pollution levels.

Conclusion:

From this geographical analysis, several conclusions can be drawn:

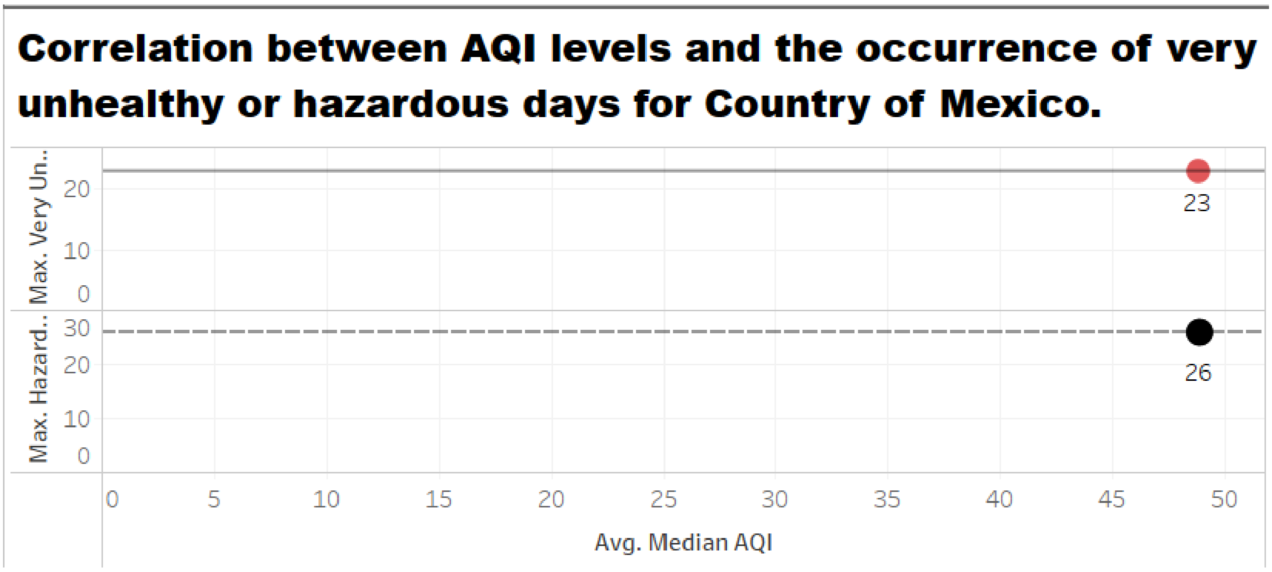
1. **States with Poor Air Quality:** States that consistently have high AQI values, a low number of "Good" days, and more unhealthy days face significant air quality challenges. They need targeted policy measures and intervention to reduce pollution sources.
2. **States with Good Air Quality:** States like Alaska and West Virginia maintain relatively low AQI values and fewer unhealthy days, reflecting the benefits of less industrialization and lower population densities.
3. **Policy Implications:** The findings highlight the need for differentiated environmental policies across states. States with poor air quality should focus on mitigating the key sources of pollution, while those with better air quality should continue current practices to maintain their air quality.
4. **Health and Environment Impact: States** with consistently high AQI values and unhealthy days may see increased health risks and environmental degradation. Addressing these concerns requires comprehensive strategies involving regulation, monitoring, and public awareness.

Health and Safety Insights

OvervieW

Correlation between AQI Levels and Unhealthy or Hazardous Days:

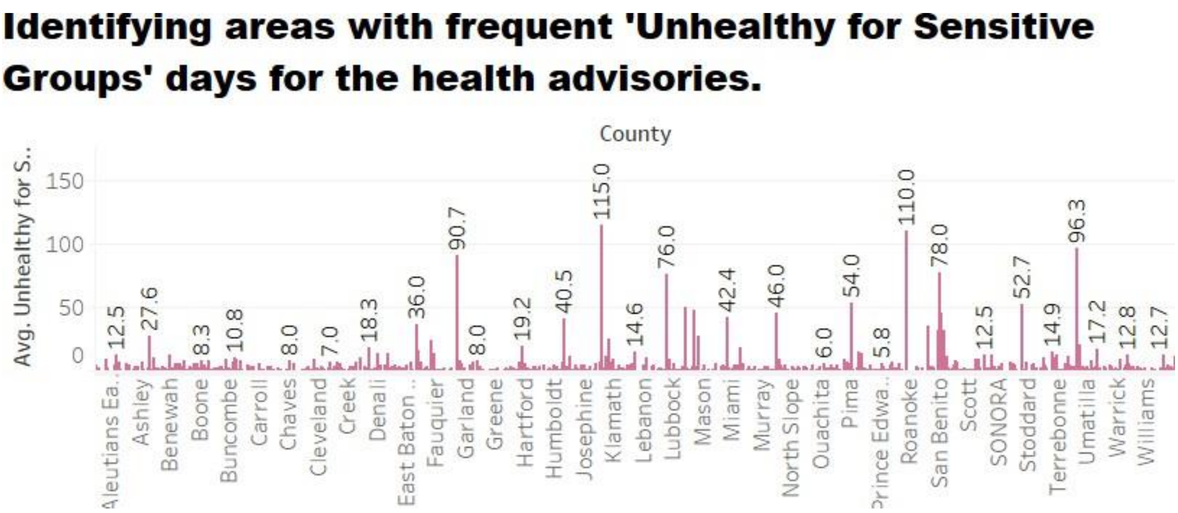
- The scatter plot indicates a clear relationship between average AQI levels and the maximum number of very unhealthy or hazardous days.



- In the Country of Mexico, higher average AQI levels are associated with significantly more days classified as very unhealthy or hazardous.
- The state of California demonstrates this correlation well, with data points showing a maximum of 23 very unhealthy days and 26 hazardous days, reinforcing the trend.

Identifying Areas with Frequent 'Unhealthy for Sensitive Groups' Days

- The bar chart below highlights the variation in the average number of days classified as "Unhealthy for Sensitive Groups" across different counties.



- Notably, counties such as Denali (183 days), Houston (192 days), and La Plata (146 days) exhibit a high frequency of such days, concluding potential health risks for sensitive groups in these regions.
- The bar chart thus serves as a tool for identifying regions where residents are more frequently exposed to unhealthy air quality, guiding health advisories for possible interventions.

Conclusion:

Correlation of AQI with Hazardous Days: and Unhealthy days

Higher AQI levels are associated with more unhealthy or hazardous days. In the analysed data, one data point (Unhealthy days) for California stands out with a much higher count of such days, possibly due to higher pollution, dense population or other localized or socio-economic factors.

Counties with High Sensitivity when exposed to Unhealthy days for Medical Advisors:

Certain counties are more prone to having days that are unhealthy for sensitive groups, as disclosed before, indicating possible localized pollution issues or specific environmental conditions, and might possibly need medical intervention.

Policy Implications

1. **Targeted Regulation:** Regions like California, with high AQI levels and frequent unhealthy days, require stricter regulation on pollution sources such as traffic, industrial emissions, and wildfires. Policies should emphasize emission reductions, cleaner technologies, and improved urban planning.
2. **Localized Monitoring:** Counties identified as having frequent "Unhealthy for Sensitive Groups" days should establish localized air quality monitoring to better understand pollution sources and trends. This data can inform targeted interventions to address specific issues.

3. **Climate Adaptation:** Given the impact of climate variability on air quality, policies should integrate climate adaptation strategies, including enhancing air quality standards in anticipation of changing climate patterns.
4. **Public Health Initiatives:** Implement community-wide health initiatives in regions with poor air quality to increase awareness, guide protective measures for sensitive populations, and ensure timely public advisories on hazardous days.
5. **Incentivize Green Practices:** Encourage businesses and industries to adopt greener practices through incentives such as tax breaks or grants for reducing emissions. This will help lower overall pollution levels and improve air quality.

Future Research

While I believe this just touches the bare minimum and overview that will indeed help us perform deeper analysis into this matter, here are some more research questions and concerns that can be addressed with this information.

1. **Longitudinal Studies:** Conduct longitudinal studies to understand the long-term health effects of fluctuating AQI levels, focusing on sensitive groups like children, the elderly, and those with pre-existing health conditions.
2. **Climate and Air Quality Models (Considering the rural and urban regions):** Develop advanced models to predict the impact of climate variability on air quality. This will help in designing better predictive tools for managing air quality in the face of changing climate patterns.
3. **Source Attribution Studies:** Investigate the major sources contributing to high AQI in specific regions, enabling the development of targeted reduction strategies and policies.
4. **Impact of Wildfires:** Research the influence of wildfires on AQI levels in affected regions, particularly on very unhealthy and hazardous days, to inform firefighting agencies to undertake preventive strategies.
5. **Effectiveness of Regulations:** Assess the effectiveness of existing air quality regulations in improving AQI levels and reducing unhealthy days, identifying gaps where policies can be strengthened.
6. **Sociodemographic Analysis:** Explore the relationship between air quality and sociodemographic factors to identify communities disproportionately affected by poor air quality and guide equitable policy interventions.