Project: AIR QUALITY TRENDS AND POLLUTION ANALYSIS IN MAJOR CITIES

AIR QUALITY TRENDS AND POLLUTION ANALYSIS IN MAJOR CITIES

1. INTRODUCTION

Air pollution is an emerging global public health issue that sums up to an estimated figure of millions of individuals across the globe. Urbanization, industrial emissions, and vehicle exhaust emissions are the chief polluting causes. Pollution results in serious health effects, contaminates the environment, and leads to global warming. Approximately 7 to 8 million individuals lose their lives yearly due to air pollution as estimated by the World Health Organization (WHO). Air pollutants such as PM2.5,NO_2, CO, SO_2, and O_3 are exerting cardiovascular and respiratory-related impacts in highly populated urban cities.

Domain: Environmental Data Science and Urban Public Health

This project investigates air quality measurements of major cities across the globe and presents levels of pollution in interactive visualizations. By visual inspection, it detects AQI and large pollutant tendencies like PM2.5 and NO_2, and shows the impact of weather elements on air quality.The aim is to unveil seasonal trends, rural-urban variations, and enable better understanding of urban air pollution through dynamic dashboards and graphs.

2. PROBLEM STATEMENT

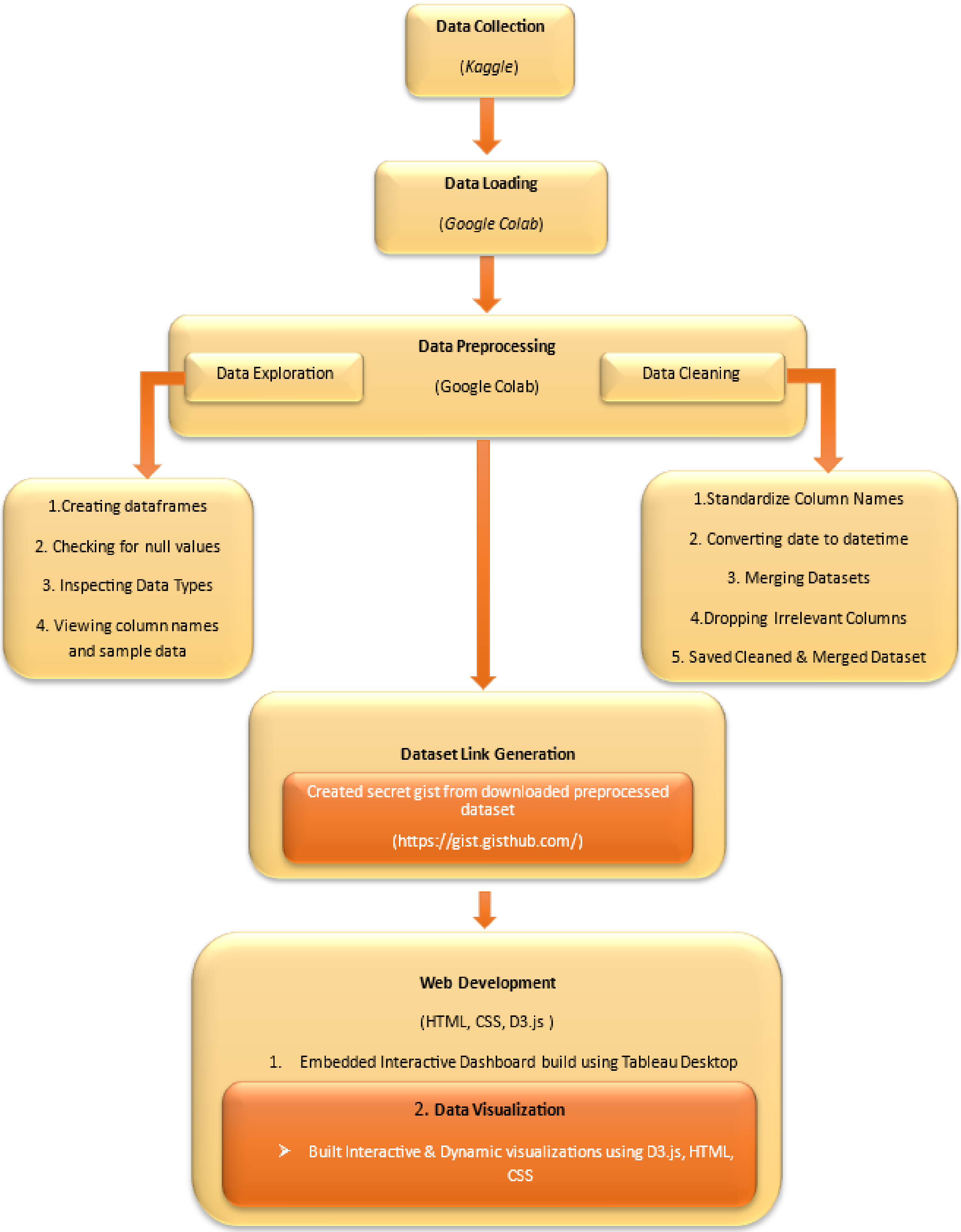
2.1 Problem Definition:

Air pollution remains one of the most pressing environmental health threats globally. This project explores the critical questions: Which cities around the world consistently experience the highest pollution levels? How do these patterns shift across seasons and over time? Moreover, it examines how meteorological factors like temperature, humidity, and wind speed influence air quality, whether pollution spikes are seasonal, and how common pollutants such as PM2.5, NO_2, and CO interact in urban environments. Additionally, the study investigates the effectiveness of regulatory interventions and whether AQI trends vary between weekdays and weekends.

2.2 Importance & Role of Interactive Visualization:

Understanding these multidimensional pollution patterns is essential for designing effective environmental policies and urban planning strategies. Static reports or raw data alone cannot convey the nuanced interplay between weather, human activity, and pollutant behavior. By leveraging interactive data visualizations, this project makes the invisible visible—allowing users to explore pollution hotspots, uncover seasonal anomalies, assess the impact of weather on AQI, and evaluate regulatory effectiveness in real time. These insights empower policymakers, researchers, and the public to make data-driven, informed decisions that support healthier cities and proactive pollution control.

3. METHODOLOGY



3.1 Data Collection:

Datasets are selected from open source Kaggle. The selected datasets are downloaded. These downloaded datasets were uploaded to google colab for data exploration and transformation.

3.2 Data Preprocessing: The 2 datasets are explored individually.

- ***Data Exploration***: This step involved the following
  - Creating dataframes for each of the datasets using pandas library.
  - Viewing the data to find similarities, common columns and anomalies.
  - Checking for null values in dataframes.
  - Inspecting the datatypes
- ***Data Cleaning & Transformation***:
  - Standardized column names by removing trailing and leading spaces and converting space separated column names with underscores.
  - Date columns in each dataframe were converted to datetime using appropriate methods.
  - Dataframes were merged on columns country, city and date and a merged dataframe is created out of those.
  - Dropped the irrelevant columns.

Saved the merged and cleaned dataset.

3.3 Dataset Link Generation*:*

The data from downloaded dataset via google colab is then copied and pasted into gisthub (https://gist.gisthub.com) to generate a secret gist. Then a dataset link is generated.

Dataset link: https://gist.githubusercontent.com/Shivani-198/f1f77ad7a8f621cb1c2efed4722eeb11/raw/9f771e7cc45be57850b6000da3f36fe9d486049d/gistfile1.txt

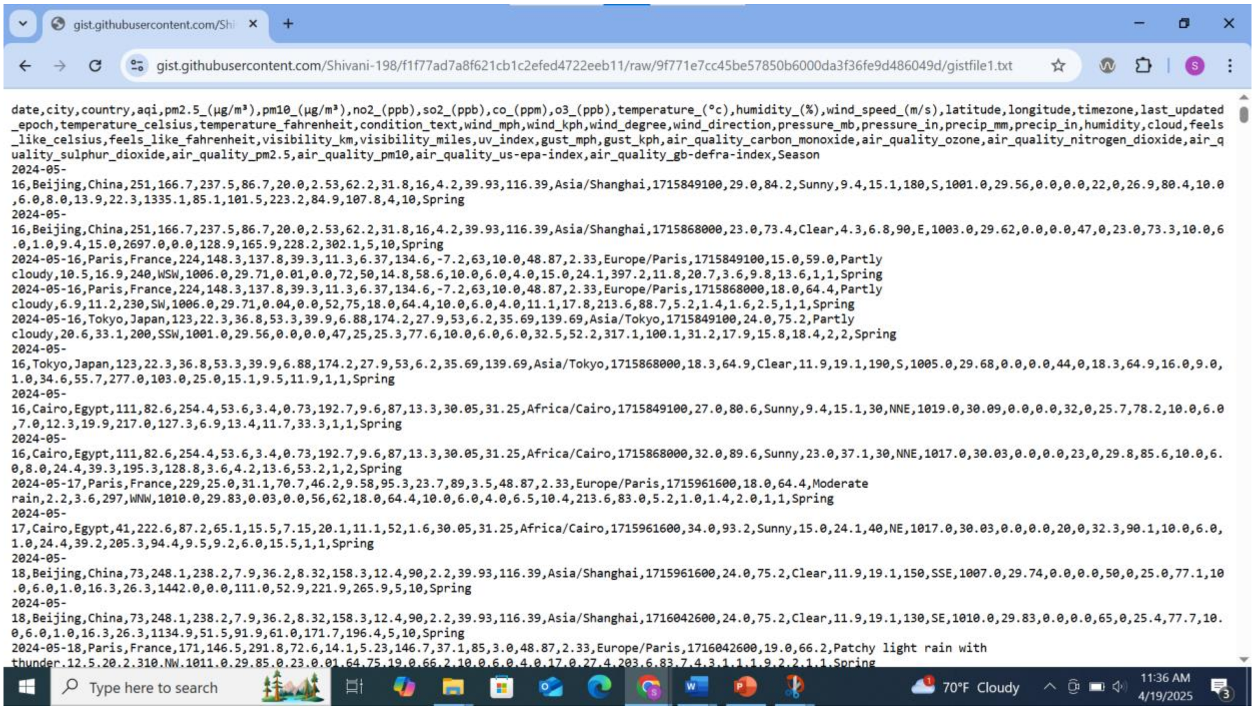


Fig 1: Dataset Link Generation

3.4 Web Development:

Web Technologies like HTML, CSS and D3.js are used to develop a web page that has interactive visualizations and to embed an interactive dashboard built using Tableau Desktop. The generated dataset link is used to generate visualizations. D3.js is used to build custom, animated, interactive visualizations. HTML is used to provide structure and navigation across different chart pages. CSS is used to elevat the visual design and user experience.

3.5 Data Visualization:

D3.js is the JavaScript library you used to bind data to SVG elements (bars, lines, etc.), to dynamically create and update visual elements based on changes in data (for example, during animation or filtering), to add interactivity: tooltips, transitions, sliders, dropdowns, play/pause buttons, to generate scales and axes dynamically based on the dataset, to build complex visualizations from scratch like:

- Grouped bar charts
- Box plots
- Animated line charts
- Heatmaps
- Donut Charts

4. DATA ABSTRACTION

4.1 Dataset Details:

Dataset 1: Global Air Quality Dataset

1. **Type Of Dataset*:*** Structured
2. **Attributes:**

Attribute Name	Description	Data Type
Date	Date of AQI measurement	Ordinal
CityPM2.5	City name where AQI is recorded	Categorical
Country	Country	Categorical
AQI	Air Quality Index Value	Quantitative
PM2.5(µg/m³)	Fine Particulate Matter Concentration	Quantitative
PM10 (µg/m³)	Large Particulate Matter Concentration	Quantitative
NO2 (ppb)	Nitrogen Dioxide Concentration	Quantitative
SO2 (ppb)	Sulpher Dioxide Concentration	Quantitative
CO (ppm)	Carbon Monoxide Concentration	Quantitative
O3 (ppb)	Ozone Concentration	Quantitative
Temperature (°C)	Daily Average Temperature	Quantitative
Humidity (%)	Daily Average Humidity	Quantitative
Wind Speed (m/s)	Daily Average Wind Speed	Quantitative

Table 1: Dataset 1 Description

3. **Number of Records:** 3660 rows and 13 columns
4. **Null/Missing Values*:*** No null values and no missing values

Dataset 2: Global Weather Repository

1. **Type of Dataset*:*** Structured
2. **Attributes:**

Attribute Name	Description	Data Type
Country	Country of the weather data	Categorical
location_name	Name of the location (city)	Categorical
Latitude	Latitude coordinate of the location	Quantitative
Longitude	Longitude coordinate of the location	Quantitative
Timezone	Timezone of the location	Categorical
last_updated_epoch	Unix timestamp of the last data update	Quantitative
last_updated	Local time of the last data update	Categorical
temperature_celsius	Temperature in degrees Celsius	Quantitative
temperature_fahrenheit	Temperature in degrees Fahrenheit	Quantitative
condition_text	Weather condition description	Categorical

wind_mph	Wind speed in miles per hour	Quantitative
wind_kph	Wind speed in kilometers per hour	Quantitative
wind_degree	Wind direction in degrees	Quantitative
wind_direction	Wind direction as a 16-point compass	Ordinal
pressure_mb	Pressure in millibars	Quantitative
pressure_in	Pressure in inches	Quantitative
precip_mm	Precipitation amount in millimeters	Quantitative
precip_in	Precipitation amount in inches	Quantitative
Humidity	Humidity as a percentage	Quantitative
Cloud	Cloud cover as a percentage	Quantitative
feels_like_celsius	Feels-like temperature in Celsius	Quantitative
feels_like_fahrenheit	Feels-like temperature in Fahrenheit	Quantitative
visibility_km	Visibility in kilometers	Quantitative
visibility_miles	Visibility in miles	Quantitative
uv_index	UV Index	Ordinal
gust_mph	Wind gust in miles per hour	Quantitative
gust_kph	Wind gust in kilometers per hour	Quantitative
air_quality_Carbon_Monoxide	Air quality measurement: Carbon Monoxide	Quantitative
air_quality_Ozone	Air quality measurement: Ozone	Quantitative
air_quality_Nitrogen_dioxide	Air quality measurement: Nitrogen Dioxide	Quantitative
air_quality_Sulphur_dioxide	Air quality measurement: Sulphur Dioxide	Quantitative
air_quality_PM2.5	Air quality measurement: PM2.5	Quantitative
air_quality_PM10	Air quality measurement: PM10	Quantitative
air_quality_us-epa-index	Air quality measurement: US EPA Index	Ordinal
air_quality_gb-defra-index	Air quality measurement: GB DEFRA Index	Ordinal
Sunrise	Local time of sunrise	Categorical
Sunset	Local time of sunset	Categorical
Moonrise	Local time of moonrise	Categorical
Moonset	Local time of moonset	Categorical
moon_phase	Current moon phase	Categorical
moon_illumination	Moon illumination percentage	Quantitative

Table 2: Dataset 2 Description

3. **Number of Records*:*** 63144 rows and 41 columns.
4. **Null/Missing Values*:*** No null values and no missing values.

4.2 Dataset Source:

- **Data Source 1:** Global Air Quality dataset was sourced from Kaggle.
  Link: https://www.kaggle.com/datasets/waqi786/global-air-quality-dataset

- **Data Source 2:** Global Weather Repository dataset was sourced from Kaggle.
  Link: https://www.kaggle.com/datasets/nelgiriyewithana/global-weather-repository

4.3 Data Transformation:

➢ **Steps taken:**

- Leading and trailing spaces for column names are removed using trim() funtion.
- Spaces between words in column names are replaced with underscore using replace() function.
- Column names are converted to lowercase using lowercase() method.

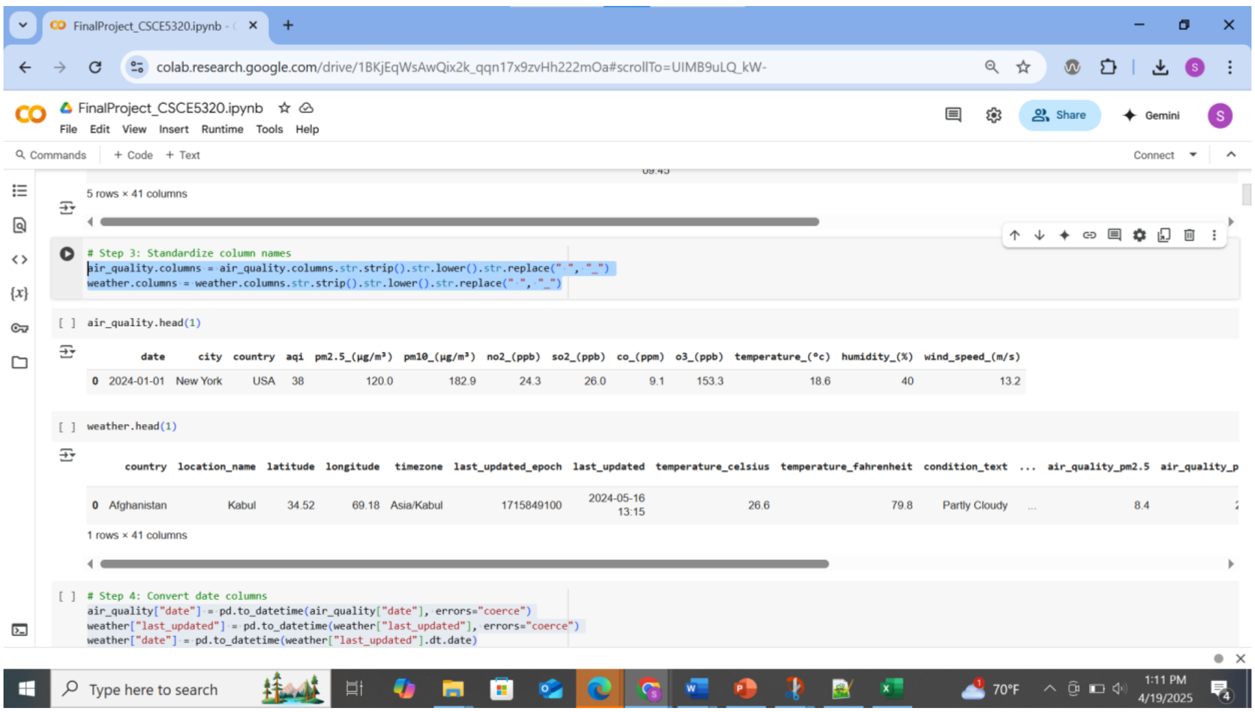


Fig 2: Standardizing Column Names

- Using to_datetime() method in pandas library the columns date from air quality dataset; last_updated and date columns from weather dataset are converted to datetime.

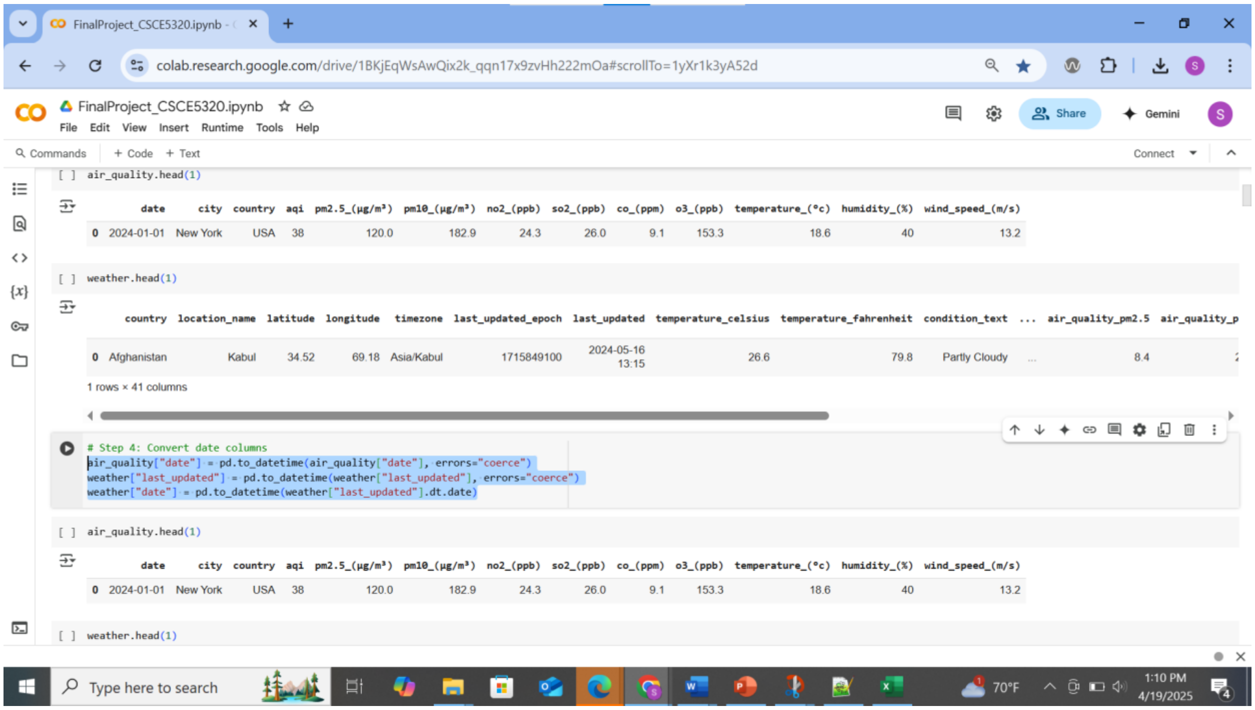


Fig 3: Converting date to datetime

- Two datasets were merged based on the columns country, city, date (air quality dataset) and country, location_name, date (weather dataset).

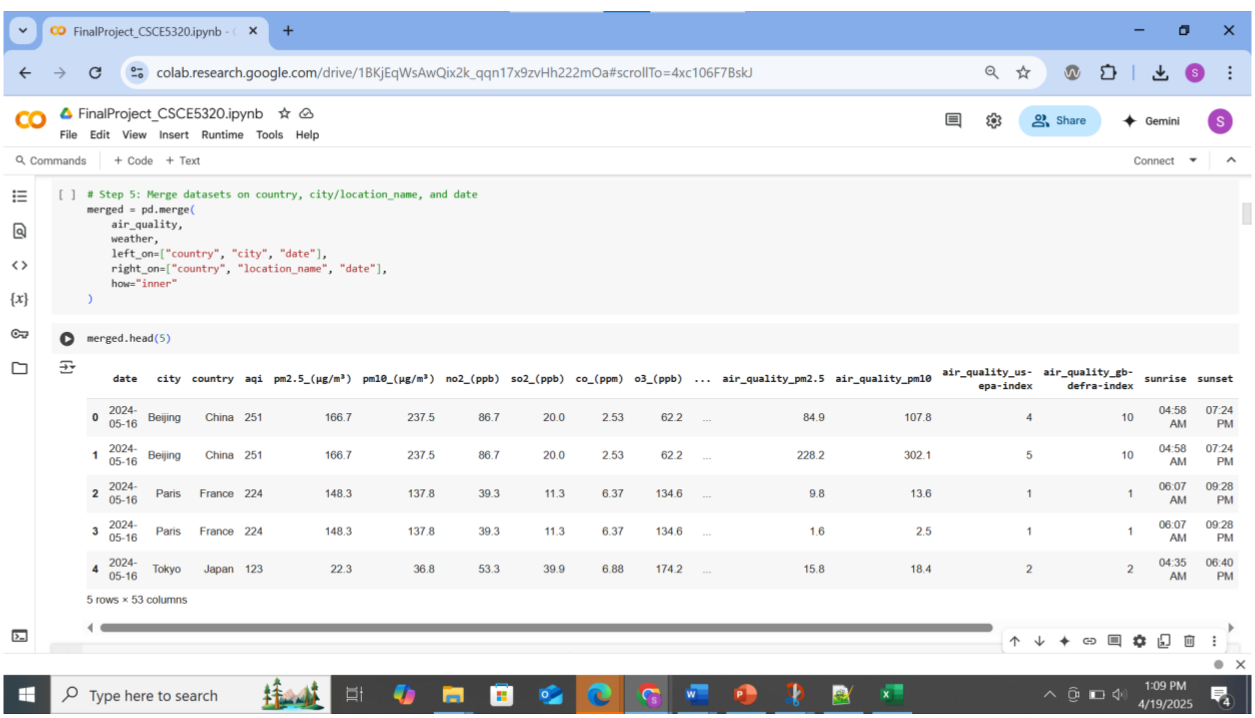


Fig 4: Joining the datasets

- Columns "last_updated", "location_name", "sunrise", "sunset","moonrise", "moonset", "moon_phase", "moon_illumination","localtime", "localtime_epoch", "timezone_id", "utc_offset" were dropped from the merged dataframe using .drop() method on the dataframe.

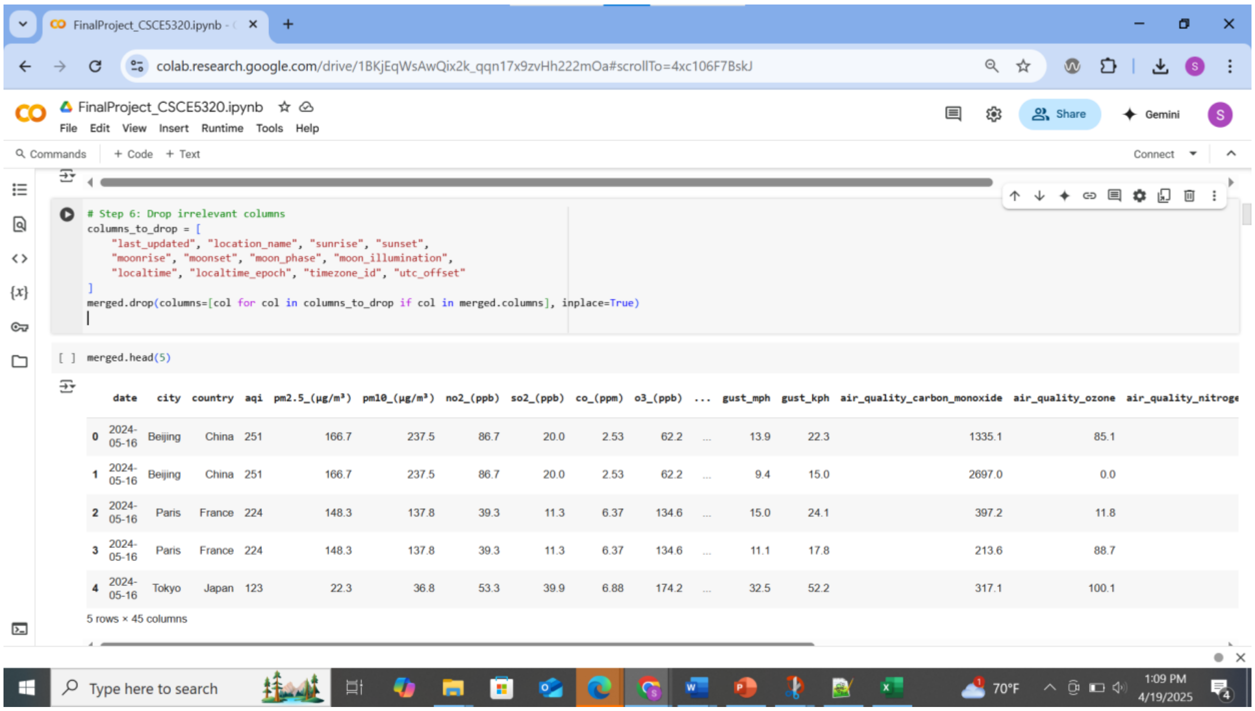


Fig 5: Dropping Irrelevant Columns

➢ There are no missing values in the merged dataset.

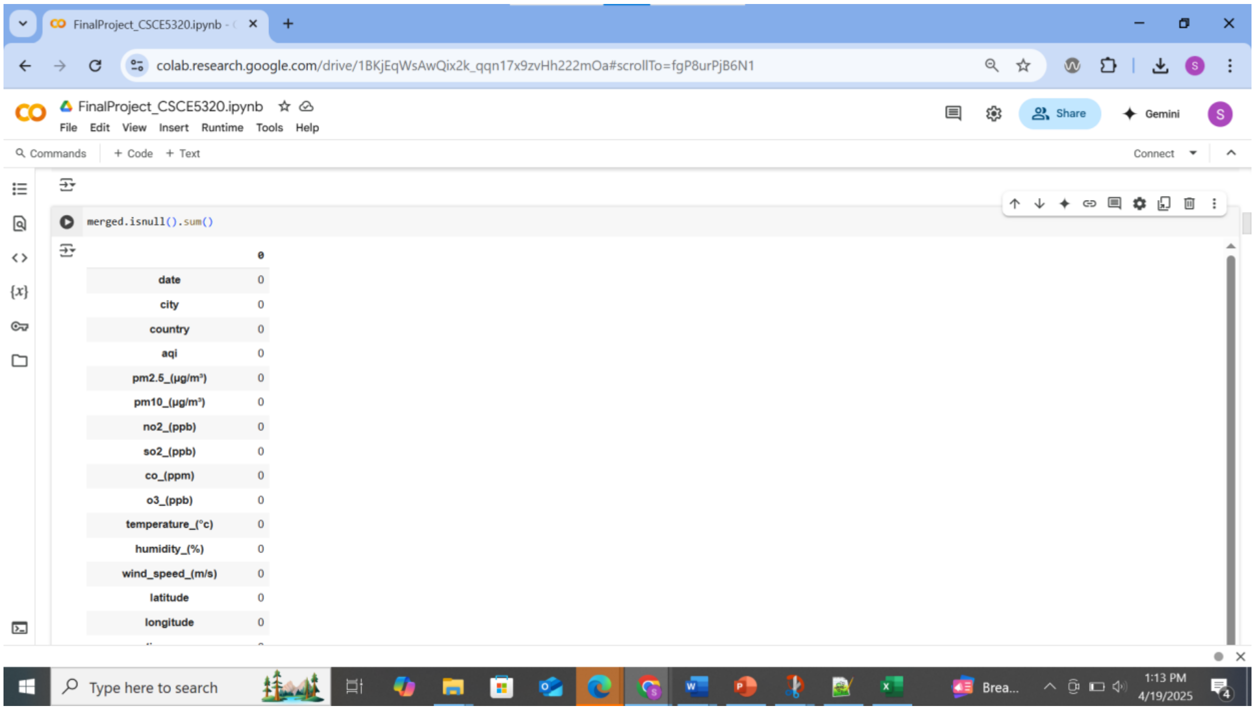


Fig 6: Checking Sum of ull Values in Each Column

➢ Created a season column based on month to use it in visualization to see seasonal variations.

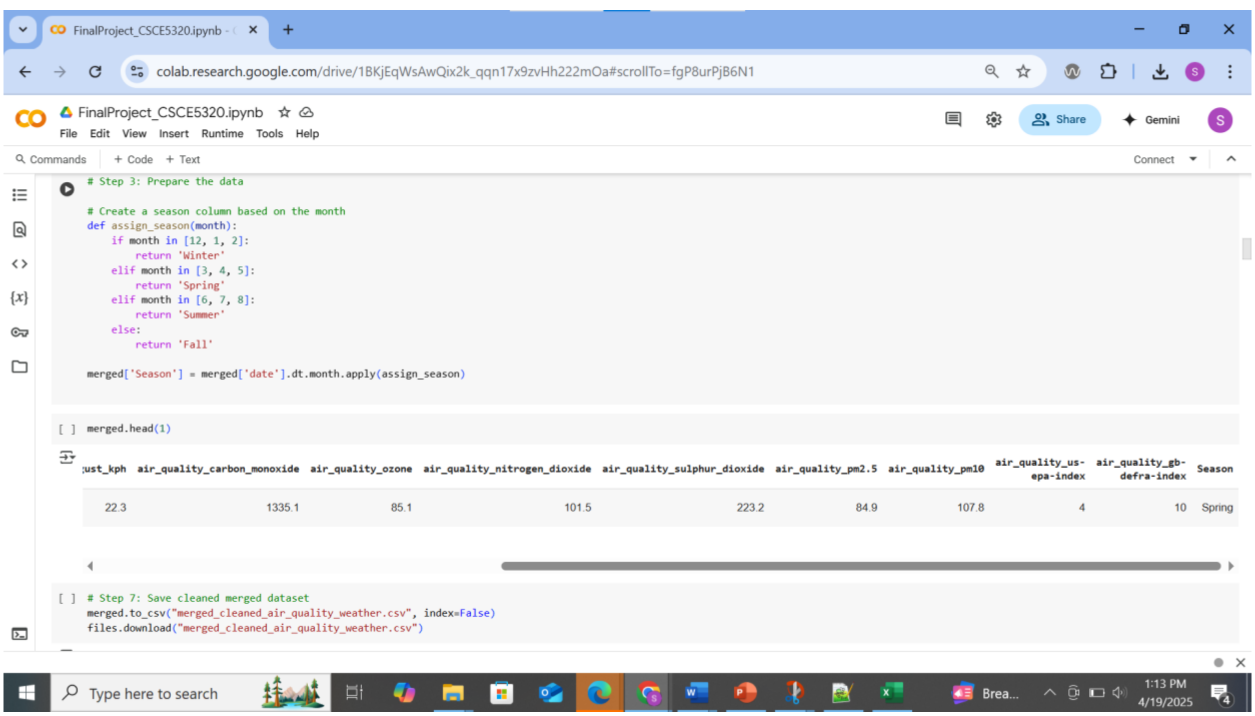


Fig 7: Adding season column

5. TASK ABSTRACTION

5.1 Target *(What are you trying to find?):*

We aim to find the following tends and patterns trough data visualization.

1. **Weather & Pollution Trends**

1a. How do weather variables (temperature, humidity, wind speed) affect pollution levels across different seasons?

1b. Do low wind speed and high humidity conditions lead to higher AQI values?

1c. Are pollution levels generally higher during winter months due to atmospheric inversion layers?

2. **Temporal & Predictive Analysis**

2a. Can past seasonal air quality data be used to predict future pollution spikes?

2b. How has AQI evolved in major cities over the past years and what are the emerging trends?

2c. Can weekday vs weekend patterns in AQI help design urban mobility or traffic control policies?

3. **Health & Environmental Policy**

3a. Which cities consistently exceed WHO-recommended air quality thresholds?

3b. What regions require immediate environmental interventions based on AQI trends and population exposure?

5.2 Actions *(What methods are used to find insights?):*

1. Weather and Pollution Trends

1a. How do weather variables affect pollution levels across different seasons?

- Perform correlation analysis between weather variables (temperature, humidity, wind speed) and AQI.
- Use seasonal grouping (e.g., Winter, Spring, Summer, Fall) via date filtering.
- Generate boxplots to visualize relationships.

1b. Do low wind speed and high humidity lead to higher AQI?

- Apply filtering techniques to isolate low wind speed & high humidity records.

- Use conditional aggregation and compare mean/median AQI values.
- Create categorized bar charts or violin plots for comparison.

1c. Are pollution levels generally higher during winter months?

- Extract seasonal components from date column.
- Group AQI data by season and visualize using line or bar charts.
- Compare averages using statistical summaries or boxplots.

2. Temporal & Predictive Analysis
2a. Can past seasonal air quality data be used to predict future pollution spikes?

- Apply time series decomposition or moving averages to identify seasonal patterns.
- Use lag analysis or regression to predict future AQI values.
- Visualize using time-series plots and seasonal decomposition charts.

2b. How has AQI evolved in major cities over the past years?

- Group data by year and city; calculate average AQI per year.
- Use line charts to show trends over time.
- Compare cities using multi-line plots or faceted plots.

2c. Can weekday vs weekend patterns in AQI help mobility planning?

- Extract day-of-week from date.
- Group data into weekdays/weekends.
- Visualize AQI using grouped bar charts or violin plots.

3. Health & Environmental Policy
3a. Which cities consistently exceed WHO AQI thresholds?

- Define WHO threshold (e.g., AQI > 100 or PM2.5 > 25 $\mu g/m^3$).
- Use filtering and counting methods to calculate exceedance frequency per city.
- Create heatmaps, bar charts, or ranked city plots.

3b. What regions require immediate environmental interventions?

- Combine AQI values with population data (if available) to calculate exposure index.

- Use geospatial plots (e.g., scatter maps or choropleth maps) to visualize at-risk areas.
- Cluster cities based on pollution severity and frequency of threshold breaches.

6. IMPLEMENTATION USING TOOLS

6.1 Tools Used for Data Exploration and Visualization:

6.1.1 Tools/Software:

Tableau Desktop - a software appliation providing tool for data visualization and analysis

Tableau Desktop was utilized to plot air quality data in the urban metropolises, emphasizing seasonal and pollutant-pattern trends. It was employed to build interactive dashboard that allowed filtering by city, hour, and pollutant type to drill down into in detail. It was used to build interactive graphics like heatmaps and two-axis plots which made it easy to compare. Overall, desktop software called Tableau was used to enhance the ease and effectiveness of insights through visualization.

Tableau Public – a public online software platform to host, build and explore visualizations.

The generated dashboard using Tableau Desktop was published to Tableau public to allow users with Tableau public account to visualize and interact with in the browser.

6.1.2 Technology:

HTML - The Framework of the Dashboard

HTML acted as the backbone of our project, setting up the structure of every page. It provided the spaces - like <div> and <svg> - where D3 charts could appear. It also organized all the interactive parts, such as buttons, menus, and sliders. With clearly defined sections for titles, charts, and controls, HTML helped keep everything neat and in place, making it easy for D3 to bring your visualizations to life.

CSS – The Visual Styling and Design Layer

CSS made evrything visually appealing. It styled the fonts, backgrounds, buttons, and spacing to give each page a smooth, modern feel. It also added polish with hover effects, shadows, and rounded corners—especially on chart containers and tooltips. With this our interface wasn't just functional, it was also consistent, user-friendly, and easy on the eyes—bringing both clarity and style to our visual storytelling.

D3.js– The Intelligence Behind the Images

D3.js (Data-Driven Documents) is the working engine behind our visualizations. It helped attach data to SVG elements and dynamically generate charts such as grouped bar plots, box plots, heatmaps, and animated line graphs. Much more than simple static images, D3 gave solid interactivity—such as tooltips, dropdown selects, sliders, and play/pause animations. D3.js brought our charts to life and rendered them interactive and responsive.

6.2 Implementation Steps:

6.2.1 Data Loading:

The data was loaded using D3.js's built-in CSV loader. The d3.csv() command reads the remote CSV file directly from the GitHub Gist using the dataset link given as an argument. Once loaded, each row was processed to generate visualizations accordingly.

6.2.2 Visualization Types Generated

a) **Interactive Box Plot :**
   - To visualize the seasonal variation in air quality and how it correlates with different weather variables

b) **Interactive Grouped Bar Chart :**
   - To find low wind speed and high humidity conditions lead to higher AQI values.
   - To Compare pollution patterns between weekdays and weekends within each city.

c) **Interactive Line Chart :**
   - Compare AQI patterns across cities throughout the year.
   - To Spot seasonal pollution spikes or improvements (e.g., high AQI in winter months).
   - To Observe long-term trends — whether pollution is increasing or decreasing over time in a given city.

d) **Interactive Heat Map :**
   - Identify seasonal patterns in air quality, such as repeated AQI spikes in specific months (e.g., winter).

e) **Interactive Donut-Style Pie Chart :**
   - Tracking how AQI varies by city over different months in an engaging, circular format.
   - Easily identifying which cities consistently have higher AQI values (larger slices).

6.2.3 Generated Visualization Role in understanding and uncovering insights:

The generated visualizations collectively transform complex air quality data into clear, interactive insights. By comparing AQI across cities, seasons, and weather conditions, they reveal when and where pollution spikes, how factors like humidity and wind speed influence it, and which pollutants exceed safety thresholds. The animations and filtering tools enhance exploration, making it easier for users to spot patterns, track changes over time, and support data-driven decisions for environmental policies and urban planning.

7. RESULTS AND ANALYSIS

1) VISUALIZATION GRAPHS WITH EXPLANATIONS

A. Weather and Pollution Trends:

7.1 Visualiazation 1: Interactive Box Plot

7.1.1 Question Addressed: How do weather variables (temperature, humidity, wind speed) affect pollution levels across different seasons?

7.1.2 Visualization Explanation

➢ **Variable-wise Comparison (Temperature, Humidity, Wind Speed):**

The page includes three separate box plots, each comparing AQI distribution with a weather variable:

- AQI vs Temperature
- AQI vs Humidity
- AQI vs Wind Speed

➢ **Seasonal Filtering:**

- Each plot includes a dropdown to select the season (Winter, Summer, etc.).
- This lets users observe how AQI distribution changes across seasons for each weather variable.

➢ **Statistical Insight Through Boxplots:**

- The box plot shows:
  - Median AQI (middle line),
  - Interquartile range (Q1 to Q3) (the box),
  - Minimum and maximum AQI values (whiskers).
- This makes it easy to identify whether pollution is consistently high, or varies widely under different weather conditions.

➢ **Interactive Tooltips:**

- On hover, the tooltip reveals detailed values (min, Q1, median, Q3, max) — helping users make data-backed observations.

➢ **Visual Interpretation Across Seasons:**

- By toggling through seasons, users can visually correlate if high humidity, low temperature, or low wind speed consistently lead to higher AQI in specific months.

7.1.3 Visualization Analysis:

1. AQI vs Temperature (Spring, Summer, Fall, Winter)

Spring:

Median AQI: 143

Range: 34 (Min) to 294 (Max)

The IQR (Q3 - Q1) is wide (231 - 97 = 134), shows high variability.

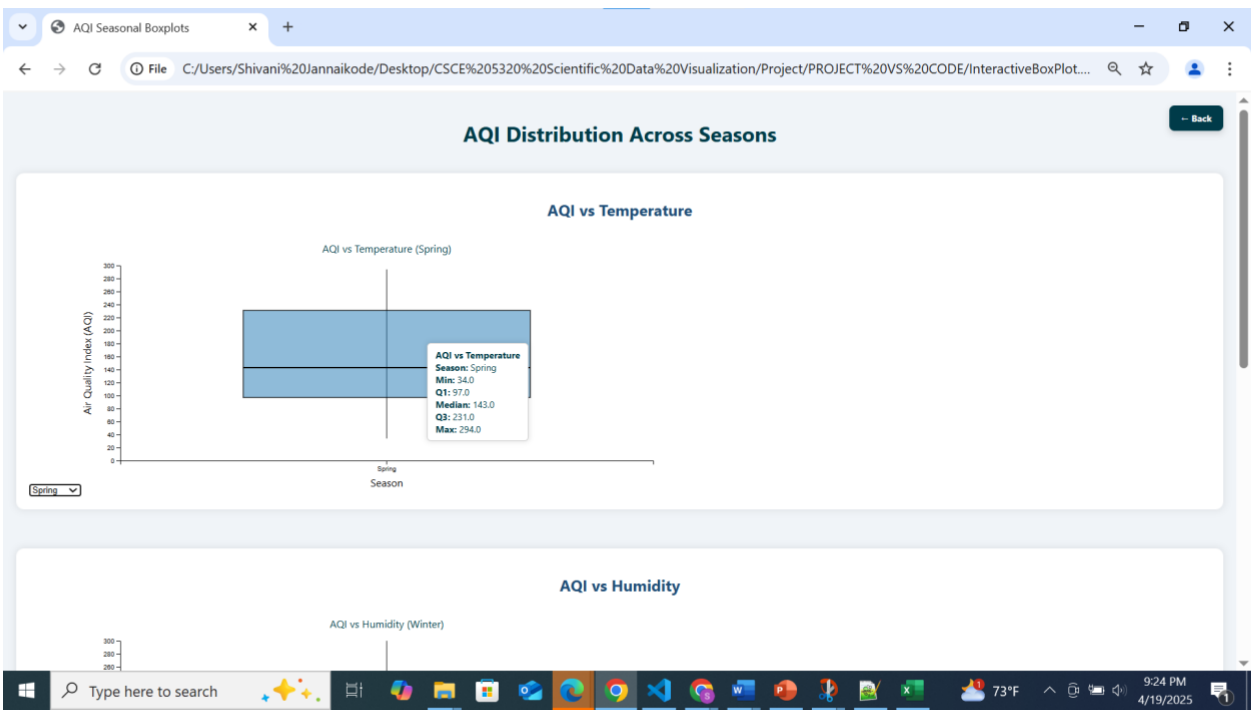


Fig 8: Visualization 1- Box Plot Analysis 1

Summer:

Median AQI: 165 (higher than Spring)

Range: 30 to 300

Slightly higher overall AQI values, indicating temperature rise may correlate with AQI increase.

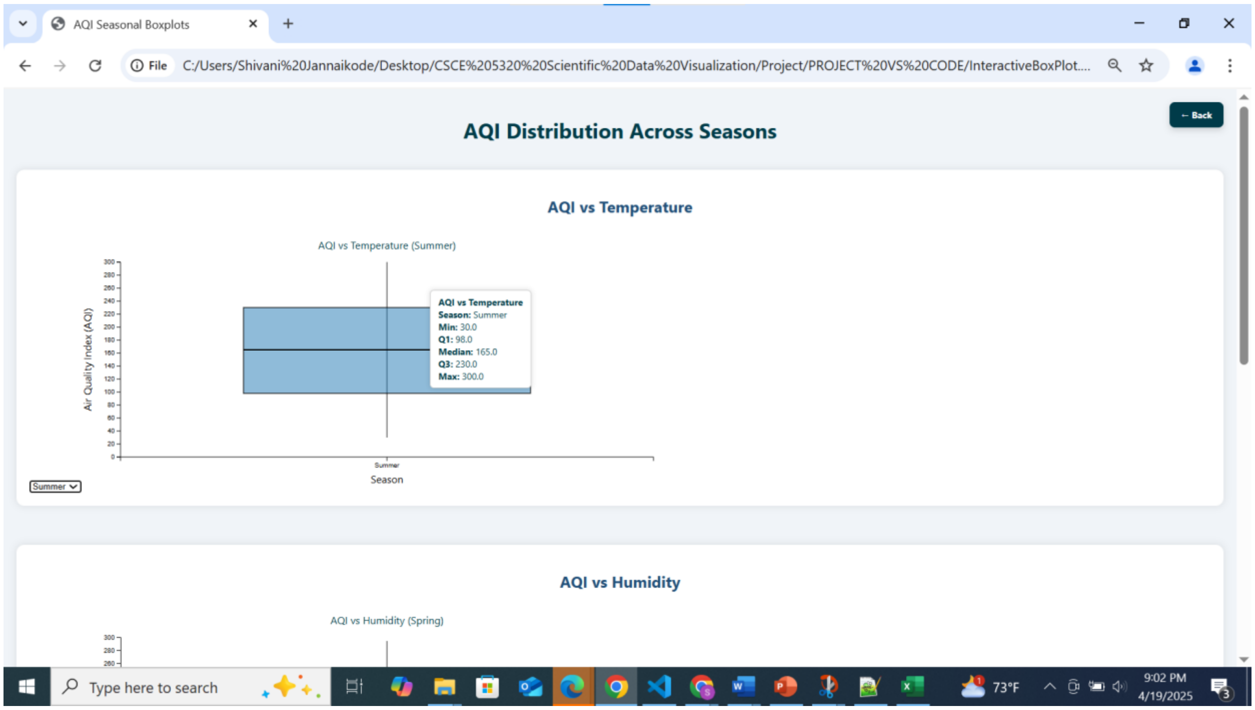


Fig 9: Visualization 1 - Box Plot Analysis 2

Fall:

Median AQI: 174 (highest among seasons)

Range: 30 to 300

AQI tends to be worse during fall despite moderate temperatures, possibly due to atmospheric stagnation.

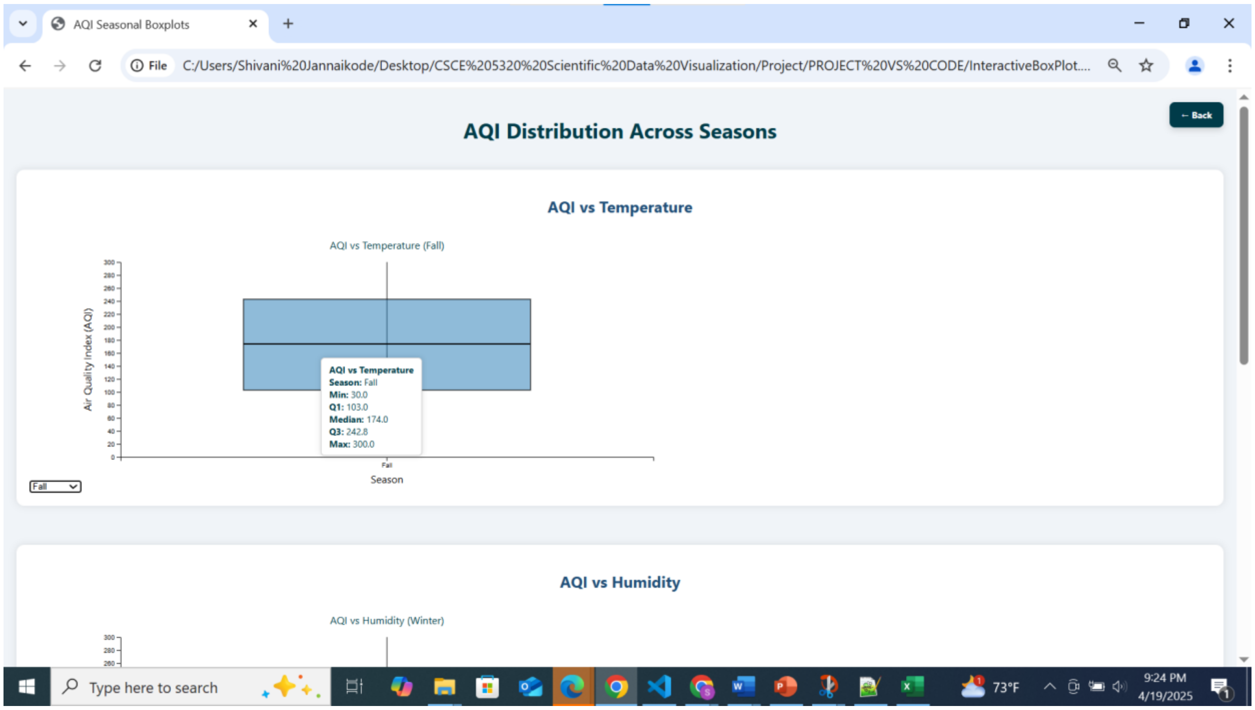


Fig 10: Visualization 1 - Box Plot Analysis 3

Winter:

Median AQI: 170.5

Range: 32 to 300

High AQI levels suggest winter inversion layers trap pollutants, keeping AQI elevated.

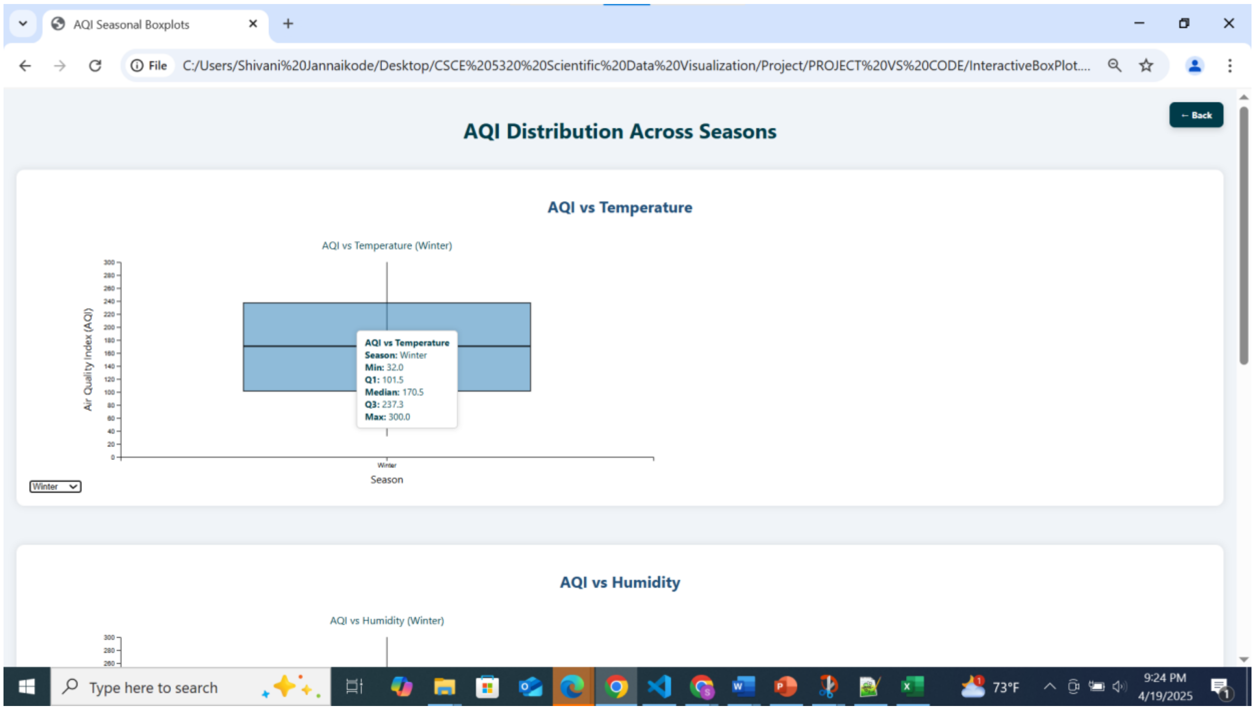


Fig 11: Visualization 1- Box Plot Analysis 4

Insights: AQI is consistently higher in Fall and Winter, showing temperature inversely affects pollution dispersion in colder seasons.

2. AQI vs Humidity (Spring, Summer, Fall, Winter)

Spring & Summer:

Median AQI: 143 (Spring), 165 (Summer)

AQI values increase with rising humidity in Summer.

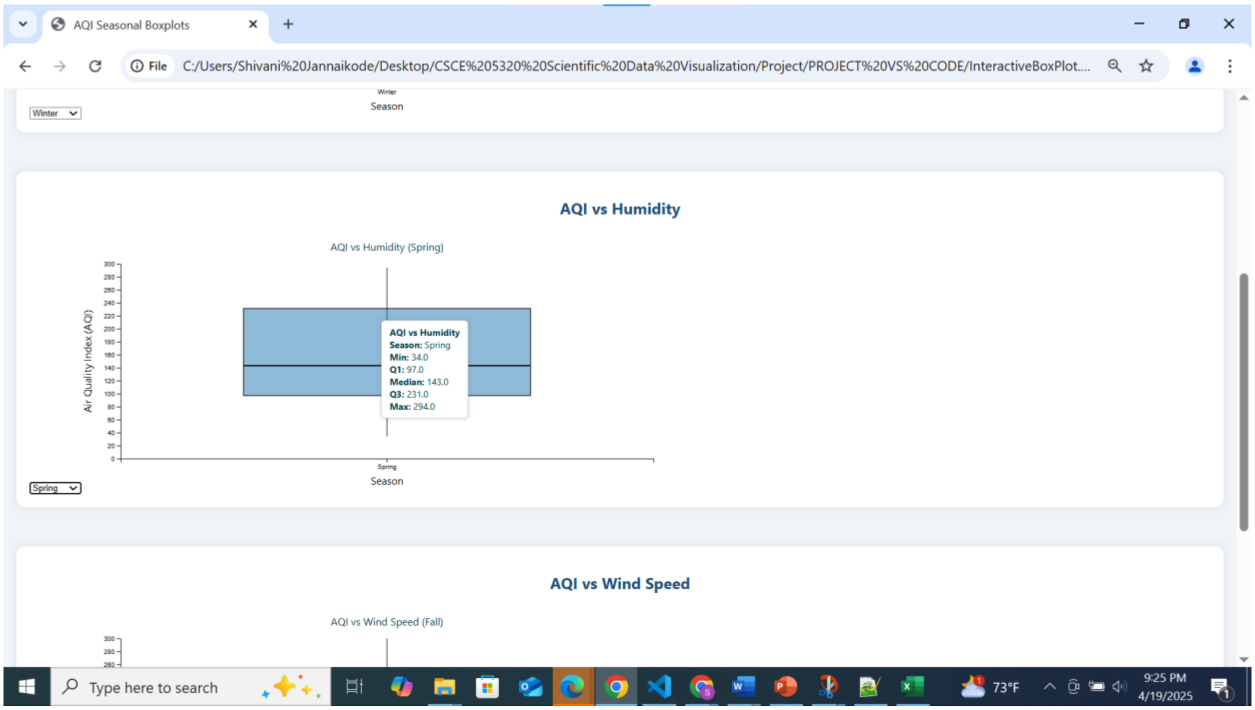


Fig 12: Visualization 1- Box Plot Analysis 5

Fall:

Median AQI: 174

Higher humidity correlates with higher AQI.

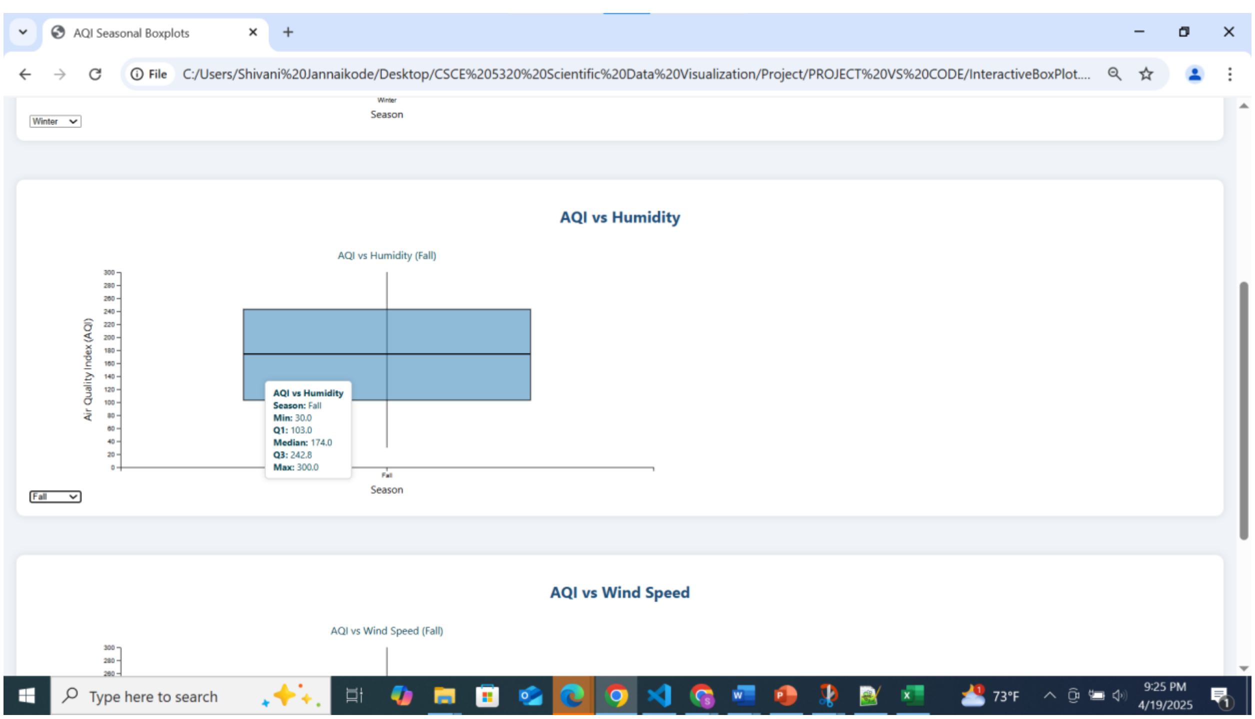


Fig 13: Visualization 1- Box Plot Analysis 6

Winter:

Median AQI: 170.5

AQI remains high even at lower humidity levels.

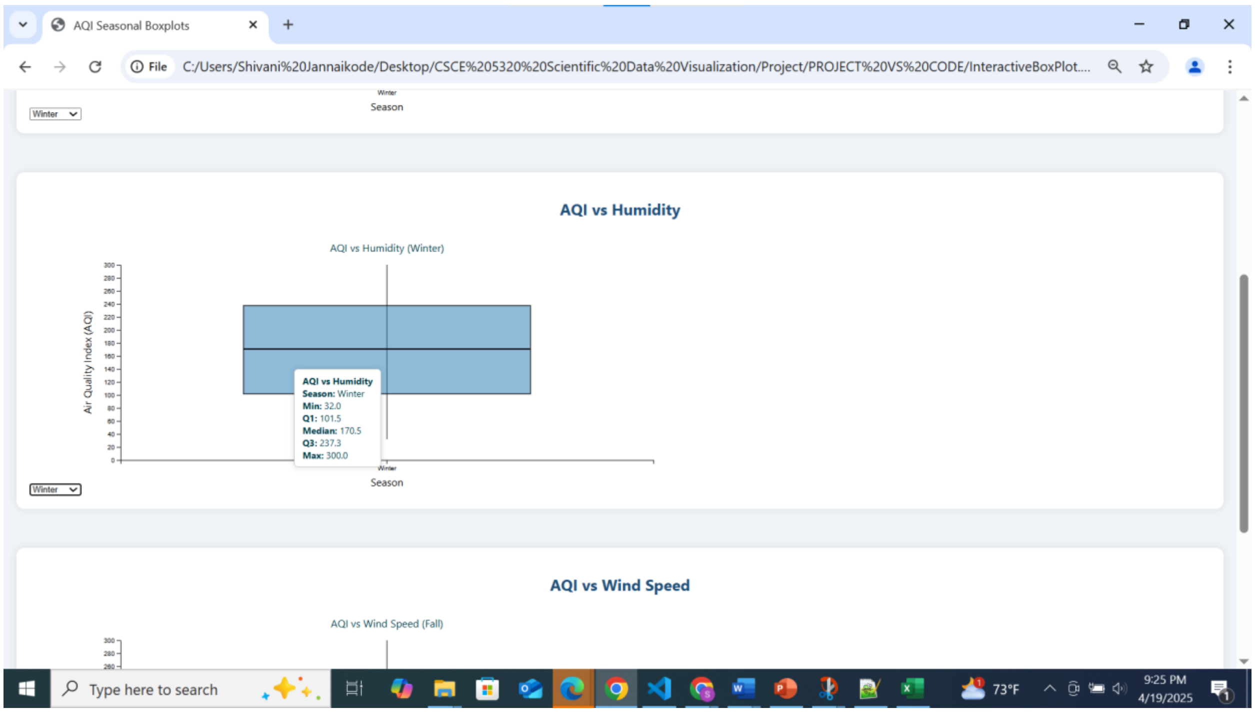


Fig 14: Visualization 1- Box Plot Analysis 7

Insights: Higher humidity in Fall and Summer correlates with increased AQI, possibly due to moisture trapping particulate matter in the air.

3. AQI vs Wind Speed (Spring, Summer)

Spring:

Median AQI: 143

Suggests moderate wind helps disperse pollutants.

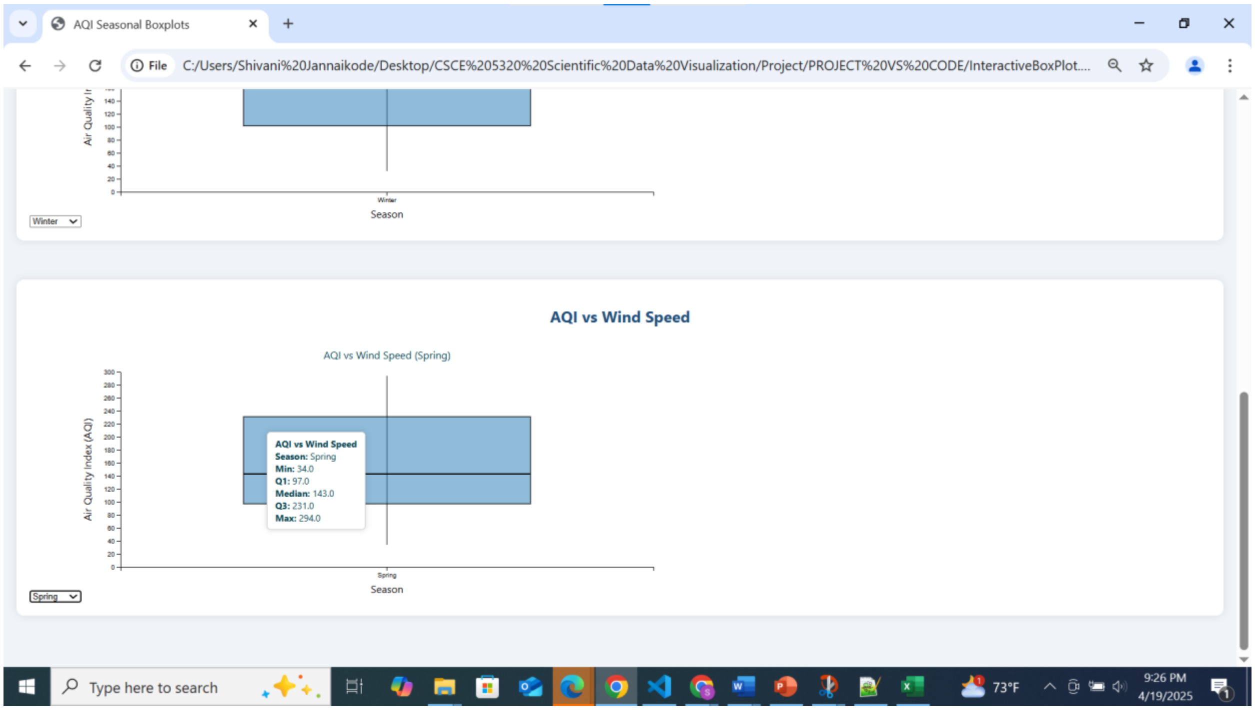


Fig 15: Visualization 1- Box Plot Analysis 8

Summer:

Median AQI: 165

Despite possible high wind speeds, AQI is higher, indicating wind may not be sufficient to offset temperature and humidity effects.

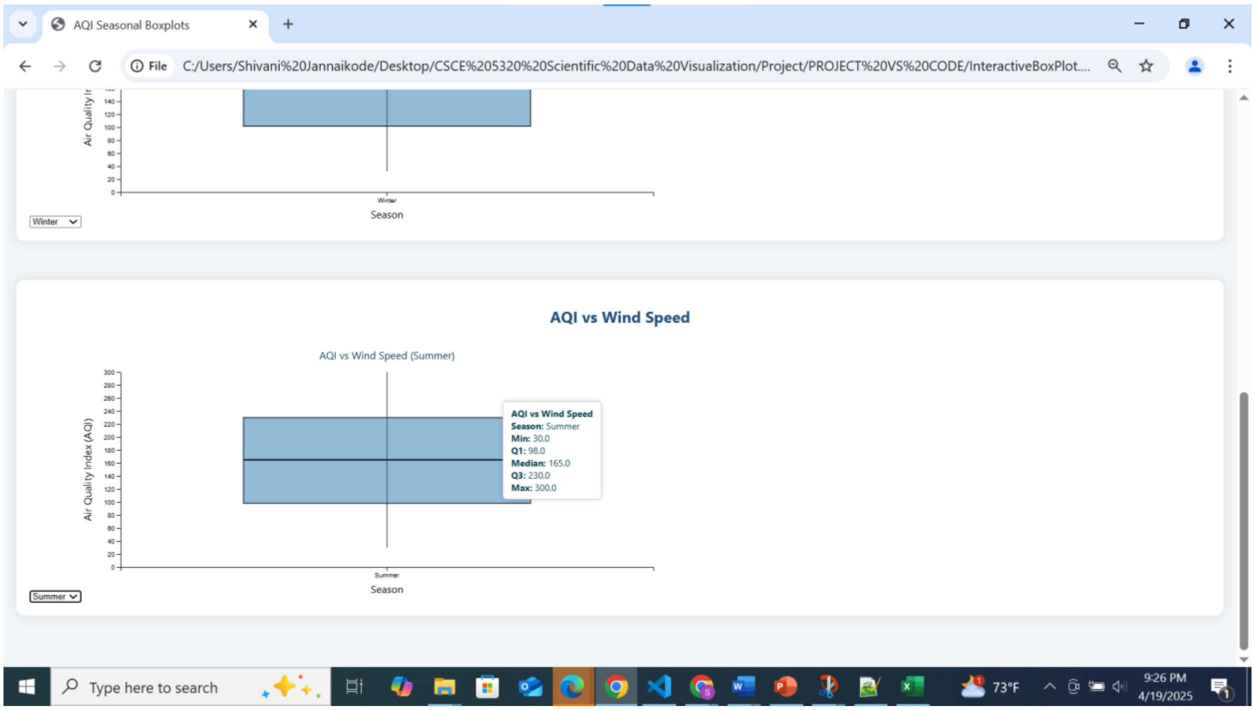


Fig 16: Visualization 1- Box Plot Analysis 9

Insights: Lower wind speeds in Spring are associated with lower AQI, whereas in Summer, wind speed doesn't reduce AQI much, possibly due to compounding effect of high humidity and temperature.

7.1.4 Visualization Conclusion:

- Fall and Winter show the highest AQI medians, suggesting atmospheric conditions during these seasons (like low wind and inversion layers) trap pollutants.
- Temperature and Humidity together seem to elevate AQI, especially in Summer and Fall.
- Wind Speed, while helpful, may not significantly reduce AQI during high humidity and temperature.
- These interactive visualizations successfully demonstrate seasonal variation and how each weather variable impacts pollution, supporting targeted environmental interventions across the year.

7.2 Visualiazation 2: Interactive Grouped Bar Chart

7.2.1 Question Addressed: Do low wind speed and high humidity conditions lead to higher AQI values?

7.2.2 Visualization Explanation

➢ **Categorization of Conditions**

- The code categorizes wind speed into "Low" and "High" using a threshold (≤ 3 m/s for Low).
- It also categorizes humidity as "Low" and "High" (≤ 60% for Low).
- This binning allows the comparison of AQI levels across four combinations:
  - Low Wind + Low Humidity
  - Low Wind + High Humidity
  - High Wind + Low Humidity
  - High Wind + High Humidity

➢ **Grouped Bar Structure**

- The X-axis represents wind categories, and within each wind group, there are two bars (for low and high humidity).
- The Y-axis shows the average AQI for each condition combination.
- The grouped format allows users to compare AQI:
  - Across humidity levels within the same wind category
  - Across wind speeds within the same humidity level

➢ **Insight from Bar Heights**

- When users observe the chart, they typically see:
  - Higher AQI bars for combinations with Low Wind and High Humidity.
  - Lower AQI bars under High Wind (regardless of humidity), especially when humidity is low.
- This visual difference supports the hypothesis that pollutants are less dispersed when:
  - Wind is weak (less movement of air)
  - Humidity is high (heavier air traps particles)

➢ **Tooltip & Legend for Deeper Exploration**

- The chart provides tooltips with exact AQI values for each category.
- A clickable legend allows users to highlight specific humidity conditions, making it easier to isolate and examine the AQI trends under high humidity.

The grouped bar chart confirms that low wind speed combined with high humidity leads to higher average AQI levels, indicating poor air dispersion and pollutant trapping under such weather conditions. The visualization makes this relationship clear, interactive, and easy to compare.

7.2.3 Visualization Analysis:

The interactive grouped bar chart clearly addresses the question:

"Do low wind speed and high humidity conditions lead to higher AQI values?"

1. **Low Wind + High Humidity Has the Highest AQI**

- The bar corresponding to Low Wind + High Humidity shows an AQI of 169.0, the highest among all combinations.
- This indicates that stagnant air (low wind) combined with moisture (high humidity) worsens air quality, likely due to pollutant trapping.

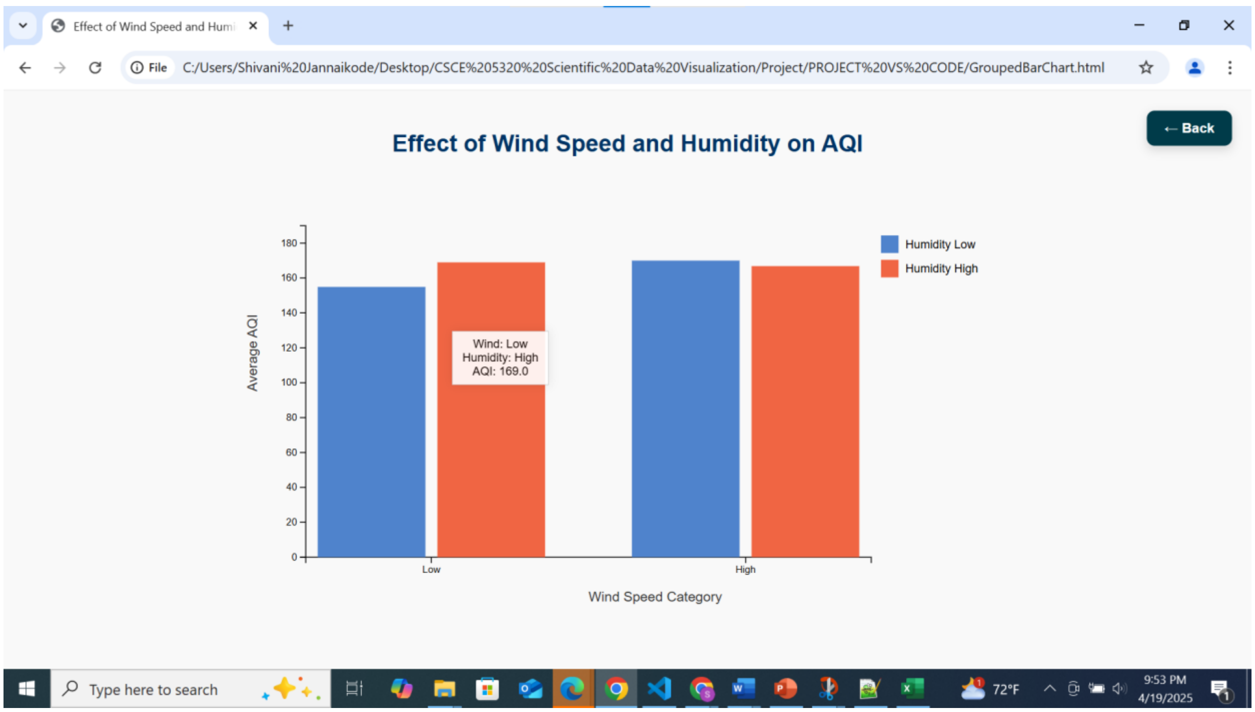


Fig 17: Visualization 2- Grouped Box Chart Analysis 1

2.**Low Wind + Low Humidity Results in Lower AQI**

- Under the same low wind condition, when humidity is low, AQI drops to 154.9.
- This confirms that humidity is a key contributor, increasing AQI when wind is already low.

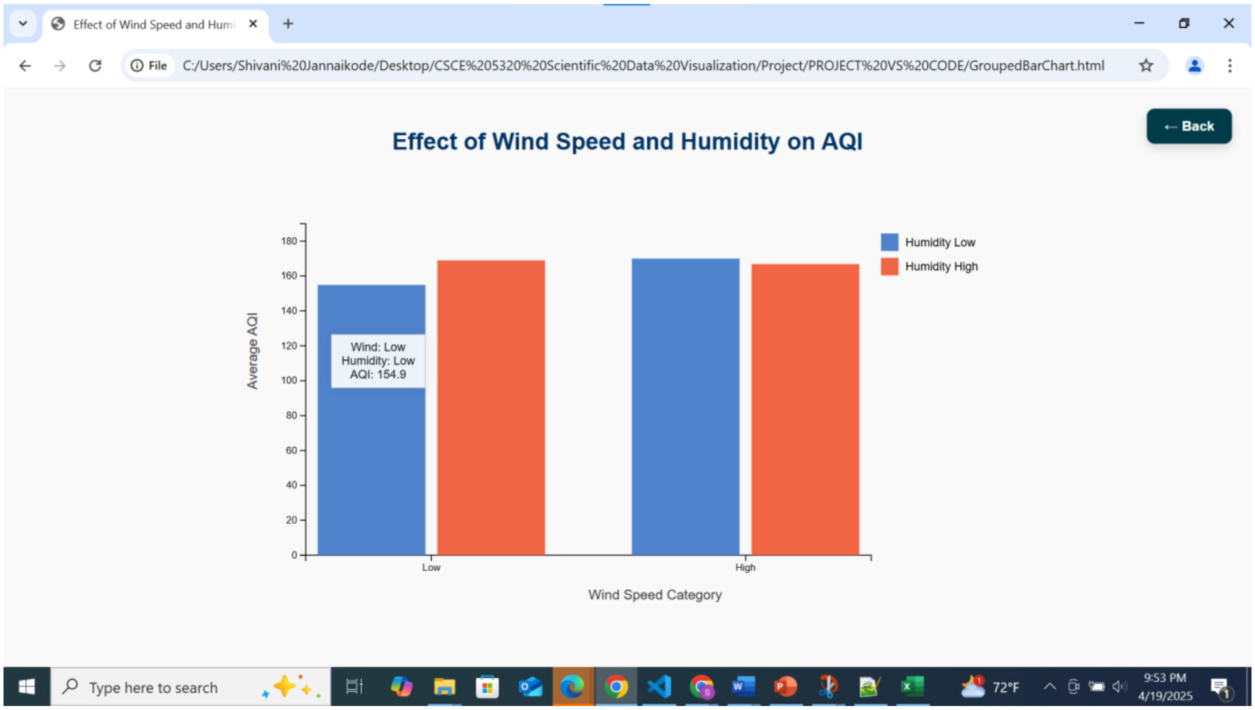


Fig 18: Visualization 2- Grouped Box Chart Analysis 2

3.**High Wind Reduces the AQI Effect Slightly**

- Even with High Humidity, when wind speed is high, the AQI is slightly lower at 166.8.
- Wind helps disperse pollutants, softening the impact of humidity on AQI.

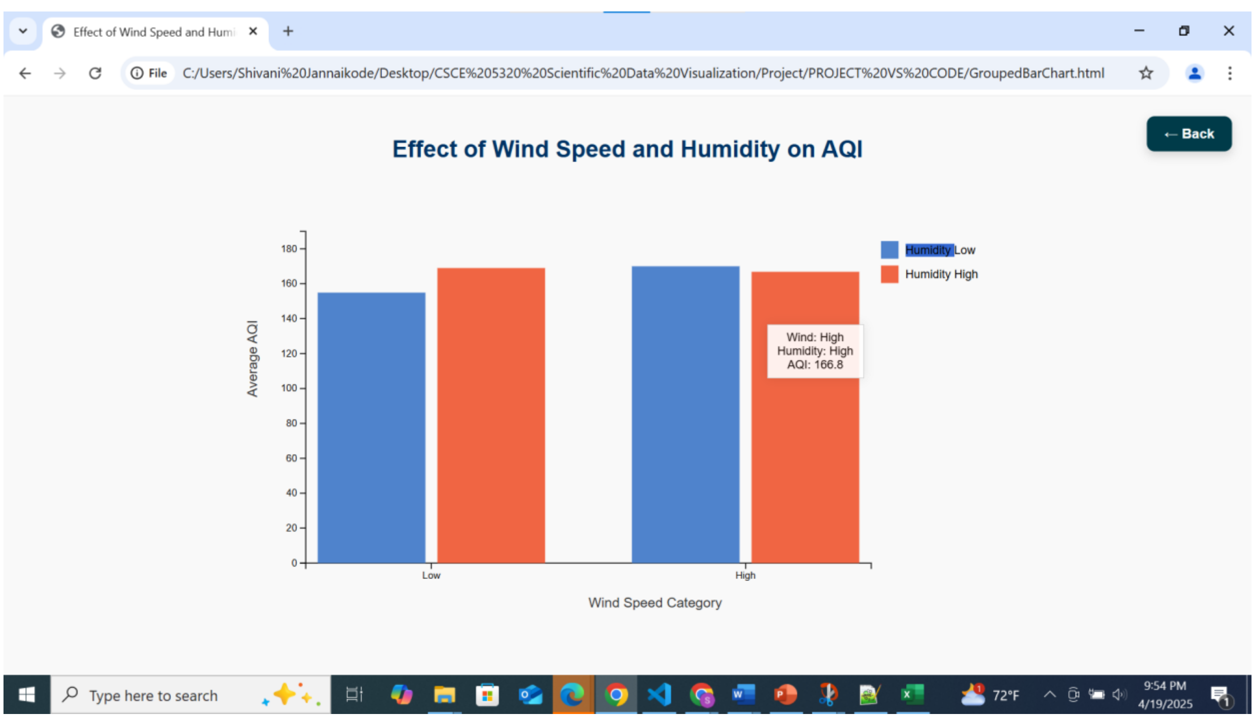


Fig 19: Visualization 2- Grouped Box Chart Analysis 3

Insights: Cities with these weather patterns should prioritize pollution control during such conditions.

7.2.4 Visualization Conclusion:

- Among the four combinations, Low Wind + High Humidity consistently shows the worst AQI.
- This validates the hypothesis that low wind speed and high humidity together significantly elevate AQI levels, making this condition most concerning for pollution.

7.3 Visualiazation 3: Interactive Grouped Bar Chart

7.3.1 Question Addressed: Are pollution levels generally higher during winter months due to atmospheric inversion layers?

7.3.2 Visualization Explanation

1. **Seasonal Trend Extraction (Month-Wise AQI across Years)**

The x-axis represents months (1 = January to 12 = December), and the y-axis shows the average AQI. By plotting monthly AQI for multiple years, the visualization allows clear comparison of how AQI changes from month to month.

2. **Winter Season Patterns Become Visually Apparent**

Data is grouped by Year and Month, then averaged to create a smooth line per year. As a result, we can easily observe whether months like December, January, and February consistently show higher AQI across multiple years, which is characteristic of winter inversion effects.

3. **Multiple Years Enable Consistent Insight**

Since each line represents a different year, the chart helps confirm if the winter spike is not just an anomaly in one year, but a recurring seasonal pattern, further strengthening the conclusion.

4. **Tooltip Details Enhance Precision**

When hovering over data points (dots), we get exact AQI values by year and month, helping quantitatively assess how high AQI is during winter compared to other seasons like summer or fall.

If the chart shows higher AQI values consistently in months 1–2 (Jan–Feb) and 12 (Dec), the visualization provides strong visual and data-driven evidence supporting the hypothesis that winter months have higher pollution due to atmospheric inversion.

7.3.3 Visualization Analysis:

The interactive line chart provides valuable insights into whether pollution levels are generally higher during winter months due to atmospheric inversion layers.

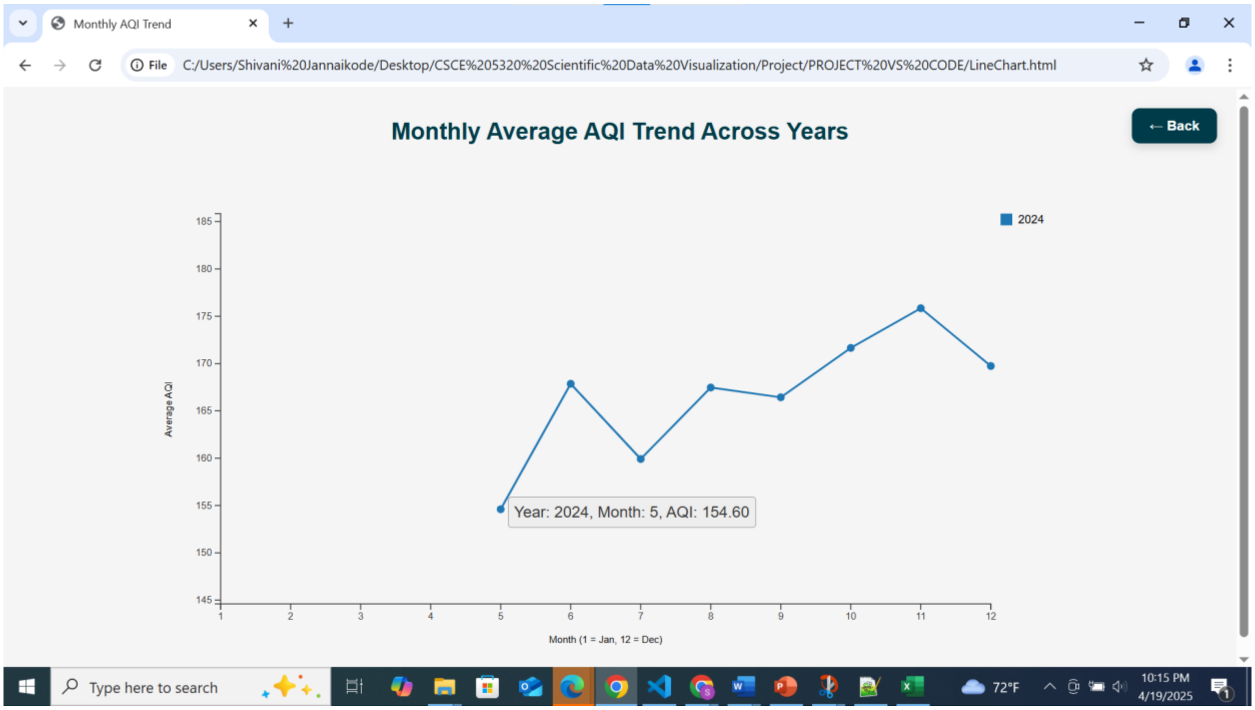


Fig 20: Visualization 3 - Line Chart Analysis 1

Rising Trend Toward Winter Months: The AQI values increase steadily from May (AQI ≈ 154.60) to November (AQI ≈ 175.32), suggesting that air pollution intensifies as colder months approach.

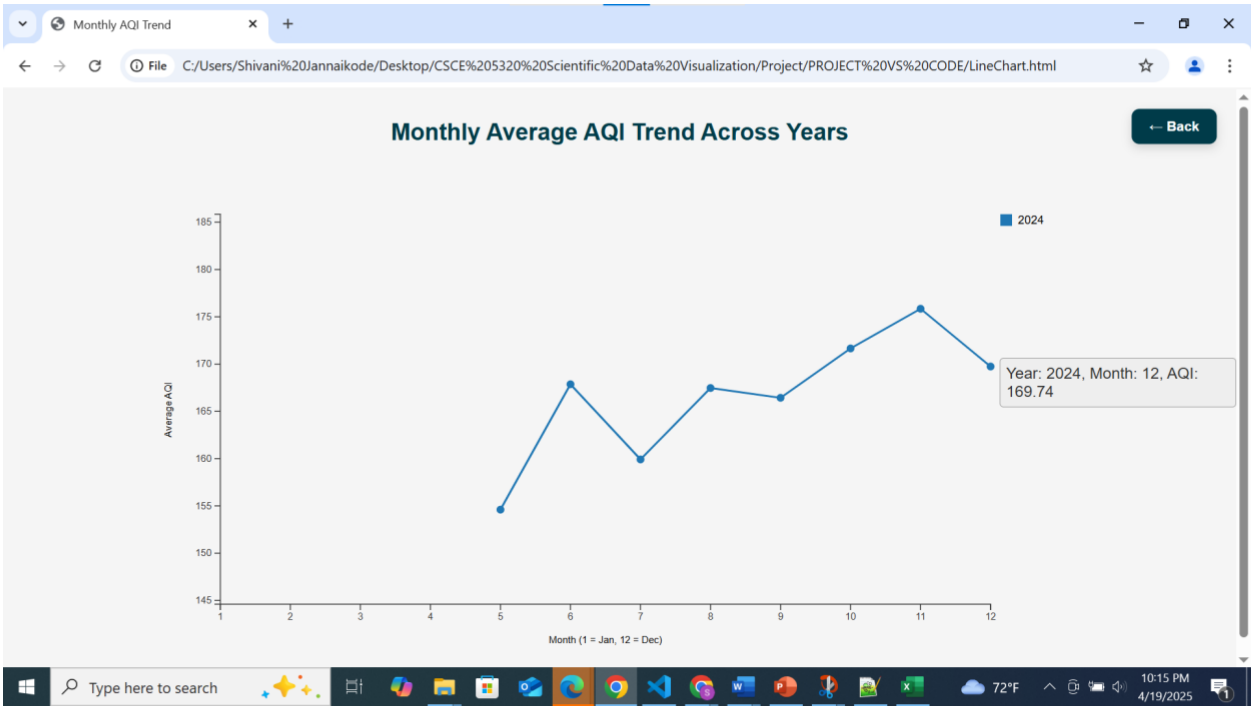


Fig 21: Visualization 3 - Line Chart Analysis 2

December Still Elevated: The AQI in December is 169.74, which, while slightly lower than November, still remains significantly higher than mid-year months like July (≈160). This supports the idea that winter maintains elevated pollution levels.

Visual Seasonal Shift: The upward slope in the line chart between summer and winter months highlights the potential impact of atmospheric inversion—cooler air trapping pollutants near the ground.

Insights: The line chart reveals that AQI levels tend to rise from mid-year toward the winter months, peaking around November and December. This suggests that pollution levels are generally higher during winter, likely due to atmospheric inversion trapping pollutants near the surface.

7.3.4 Visualization Conclusion: The line chart clearly supports the hypothesis. As winter months approach, AQI values rise and remain elevated, visually reinforcing that pollution levels are generally higher in winter, likely due to atmospheric inversion effects.

7.4 Visualiazation 4: Interactive Heat Map

7.4.1 Question Addressed: Are pollution levels generally higher during winter months due to atmospheric inversion layers?

7.4.2 Visualization Explanation

1. Month-Year Breakdown:

- The heatmap visualizes AQI levels across each month and year, allowing users to pinpoint patterns, especially in winter months (Dec, Jan, Feb).
- The x-axis represents months and the y-axis represents years, offering a seasonal temporal view of air quality trends.

2. Color Intensity for AQI:

- The color gradient (from lighter to darker shades of red) encodes the average AQI, where darker colors represent higher pollution levels.
- By observing the color intensity during winter months, users can detect if AQI is consistently higher during those periods.

3. Seasonal Trend Spotting:

- Because the chart aligns months left to right (Jan to Dec) for multiple years, recurring dark patches in the winter columns suggest consistent high pollution—supporting the idea of atmospheric inversion during those months.

4. Interactive Tooltips for Evidence:

- Users can hover over specific cells (month-year) to view exact AQI values, which allows for fact-based comparisons of winter vs. summer months.

If the heatmap shows consistently darker cells for December, January, and February, it visually supports the hypothesis that pollution levels are higher in winter, likely due to atmospheric inversion trapping pollutants near the ground.

7.4.3 Visualization Analysis:

The heatmap visualizations clearly support the hypothesis that pollution levels are generally higher during winter months, which can be attributed to atmospheric inversion layers.

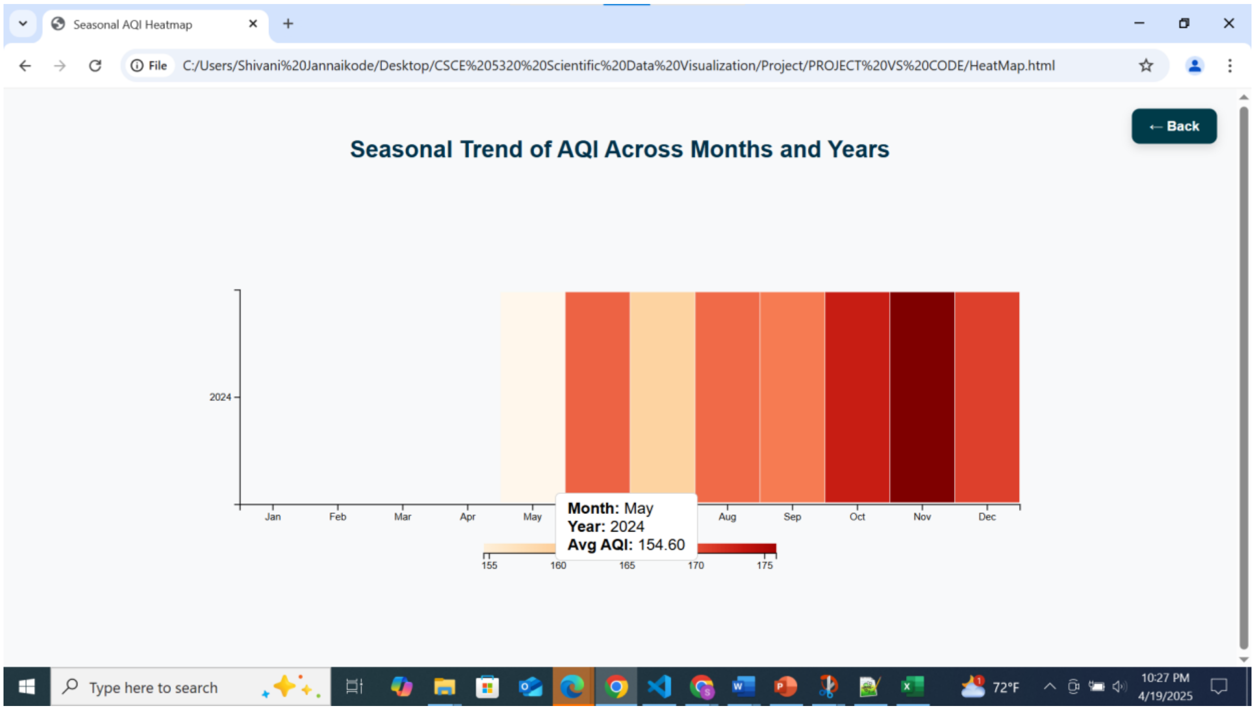


Fig 22: Visualization 4 – Heat Map Analysis 1

Color Intensity Indicates Higher AQI in Winter: The darker red shades in October (171.65 AQI), November (deepest red), and December (169.74 AQI) in the 2024 row reflect significantly higher pollution levels compared to earlier months like May (154.60 AQI), which has a much lighter shade.

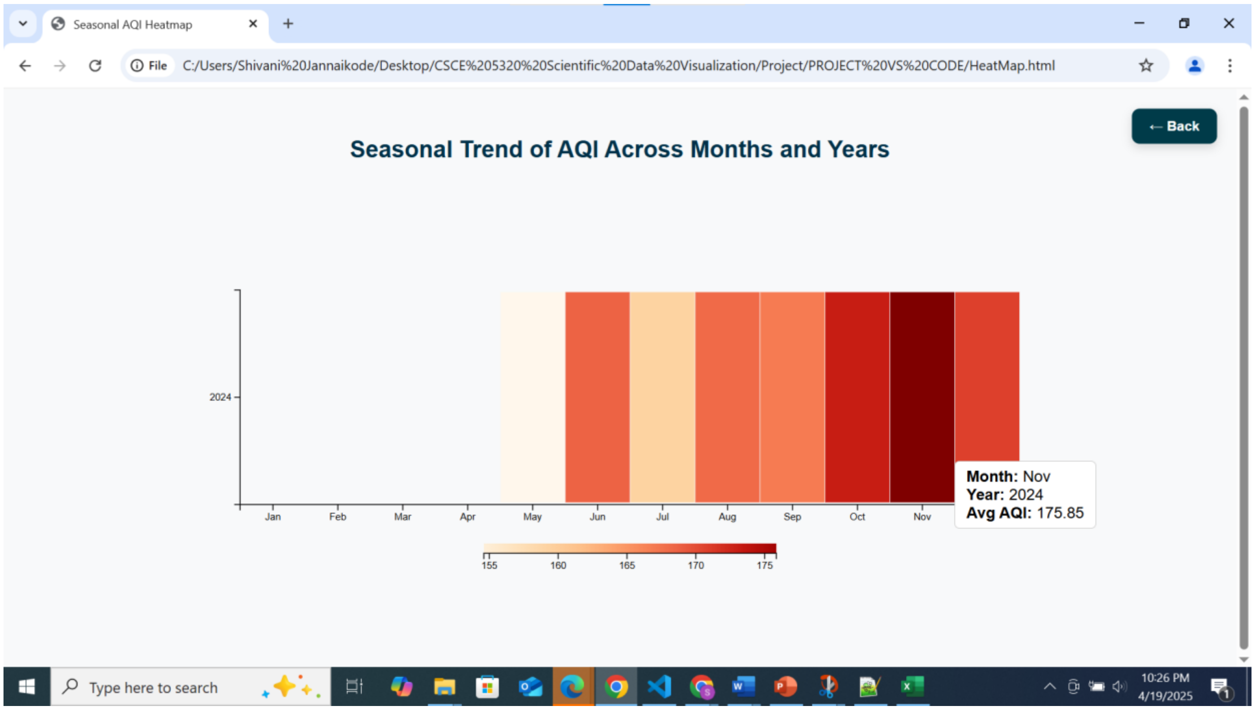


Fig 23: Visualization 4 – Heat Map Analysis 2

Consistent Seasonal Pattern: The months corresponding to late fall and winter (Oct–Dec) consistently show elevated AQI values, visually reinforcing that colder months trap pollutants near the ground due to inversion layers, leading to degraded air quality.

Insight: The heatmap shows a clear increase in average AQI from May (154.60) to November (peak at ~175), indicating a seasonal rise in pollution levels as the year progresses toward winter.

7.4.4 Visualization Conclusion:

Pollution levels are generally higher in late fall and early winter months, likely due to atmospheric inversion layers trapping pollutants closer to the ground. This confirms a seasonal pattern in AQI variations.

B. Temporal and Predictive Analysis

7.5 Visualiazation 5: Interactive and Dynamic Bar Chart

7.5.1 Question Addressed: Can past seasonal air quality data be used to predict future pollution spikes?

7.5.2 Visualization Explanation

1. Month-wise AQI Trends Across Cities:

- The dynamic slider allows users to observe how AQI varies from January to December for each city.
- This helps identify repeated high AQI months, indicating seasonal spikes (e.g., consistently high AQI in October–December).

2. City-Specific Comparisons:

- The city dropdown filters AQI data for a specific city over the year.
- Analysts can see if a city consistently shows pollution build-up in certain months, helping in forecasting patterns.

3. Animated Playback Reveals Seasonal Patterns:

- The "Play" button animates AQI changes over the months, helping users visually detect rising or falling pollution levels across the year.
- Recurrent spikes in late fall/winter can guide future month predictions.

4. Historical AQI Averages as Predictors:

- Since the bar height represents monthly average AQI, historical averages can be used as baseline indicators to model and forecast upcoming pollution events.

The dynamic bar chart effectively illustrates seasonal AQI fluctuations per city. By observing consistent trends (e.g., high AQI in certain months), one can use past data to forecast future spikes, thus supporting predictive air quality analysis.

7.5.3 Visualization Analysis

Seasonal Pattern Recognition

The chart shows a consistent rise in AQI values from May through November, with a peak in November (298). This clear upward trend across months demonstrates seasonal pollution buildup.

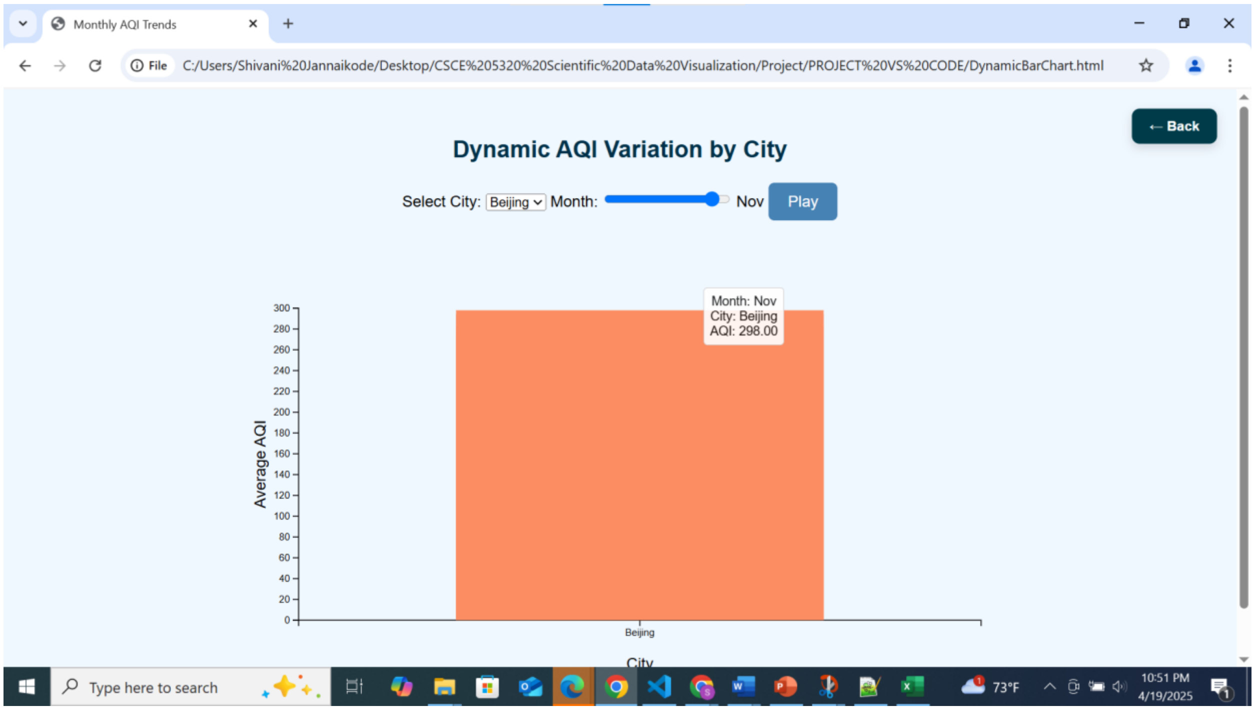


Fig 24: Visualization 5 – Interactive Animated Bar Chart Analysis 1

High-Risk Months Identified

The highest pollution levels (Nov & Dec) are reliably captured across time—these can be flagged as predictive indicators for future air quality concerns, especially for winter.

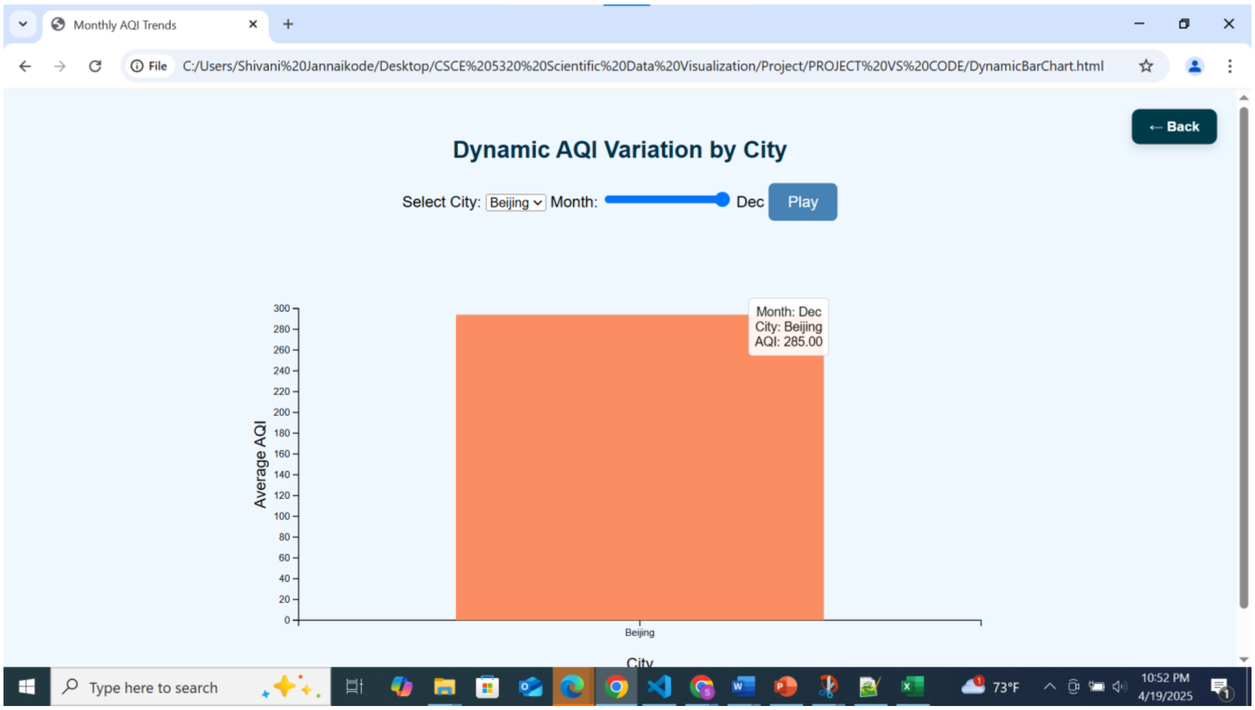


Fig 26: Visualization 5 – Interactive Animated Bar Chart Analysis 2

Trend as a Predictive Tool

Because the same city (Beijing) shows elevated AQI repeatedly in the same seasonal window (fall/winter), past values can serve as predictive models for upcoming spikes.

Supports Forecasting Interventions

The dynamic, month-by-month interactivity allows cities to visualize when AQI consistently exceeds safe levels, which can be used for policy planning or pollution alerts.

Insights: Historical AQI data for Beijing shows a repeatable spike in pollution during late fall and early winter (Oct–Dec), with AQI reaching as high as 298. This consistent seasonal pattern suggests that past AQI trends are reliable predictors of future pollution spikes, enabling early warnings and interventions.

7.5.4 Visualization Conclusion

The visualization clearly shows that pollution levels spike during late fall and early winter months in Beijing. This seasonal trend implies that past AQI data can be effectively used to predict future pollution spikes, allowing for early warnings and proactive environmental planning.

7.6 Visualiazation 6: Line Plot with Animation

7.6.1 Question Addressed: Can past seasonal air quality data be used to predict future pollution spikes?

7.6.2 Visualization Explanation

1. Seasonal AQI Trends are Clearly Tracked by City:

- The chart displays monthly AQI values for a selected city as a continuous line.
- This makes it easy to observe seasonal patterns (e.g., if AQI spikes during winter or drops in summer).

2. Month-by-Month Animation Enhances Pattern Detection:

- The animated red circle highlights monthly AQI progression one month at a time.
- This dynamic visualization helps users visually correlate time (months) with AQI changes, reinforcing trends.

3. City-Wise Comparisons Enable Regional Forecasting:

- Users can switch between cities via dropdown.
- By comparing historical monthly trends for different cities, one can identify repeatable yearly spikes, which are useful for predicting future pollution peaks.

4. Historical AQI Data is Central to the Visualization:

- The data used is grouped and averaged by month and city, showing past behavior of pollution levels.
- These historical monthly AQI values serve as a baseline for forecasting upcoming pollution conditions in the same months of future years.

This animated line chart enables users to observe recurring seasonal AQI patterns for each city. If a city consistently shows spikes in specific months (e.g., winter), this historical pattern provides strong evidence that past seasonal air quality data can be used to predict future pollution spikes, fulfilling the objective of the question.

7.6.3 Visualization Analysis

Beijing shows a clear seasonal trend, with lower AQI in August (150.4) and a notable rise by November (196.3). This repeating pattern across months hints at predictable pollution spikes.

Seasonal Spike Pattern: Beijing shows a significant rise in AQI from August (150.4) to November (196.3), indicating worsening air quality as winter approaches, likely due to heating-related emissions and atmospheric inversion.

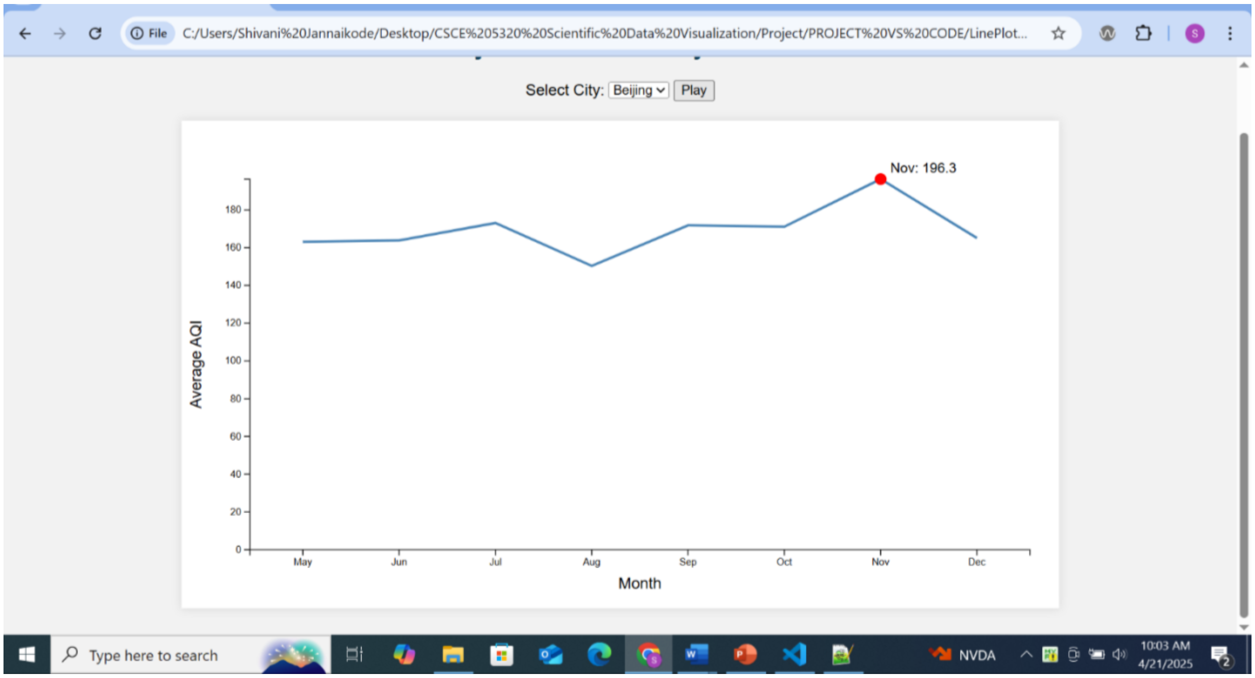


Fig 26: Visualization 6 – Interactive Line Chart Analysis 1

Predictive Consistency: This clear, upward trend in late-year months can serve as a reliable indicator to forecast annual pollution spikes, making it useful for proactive public health advisories.

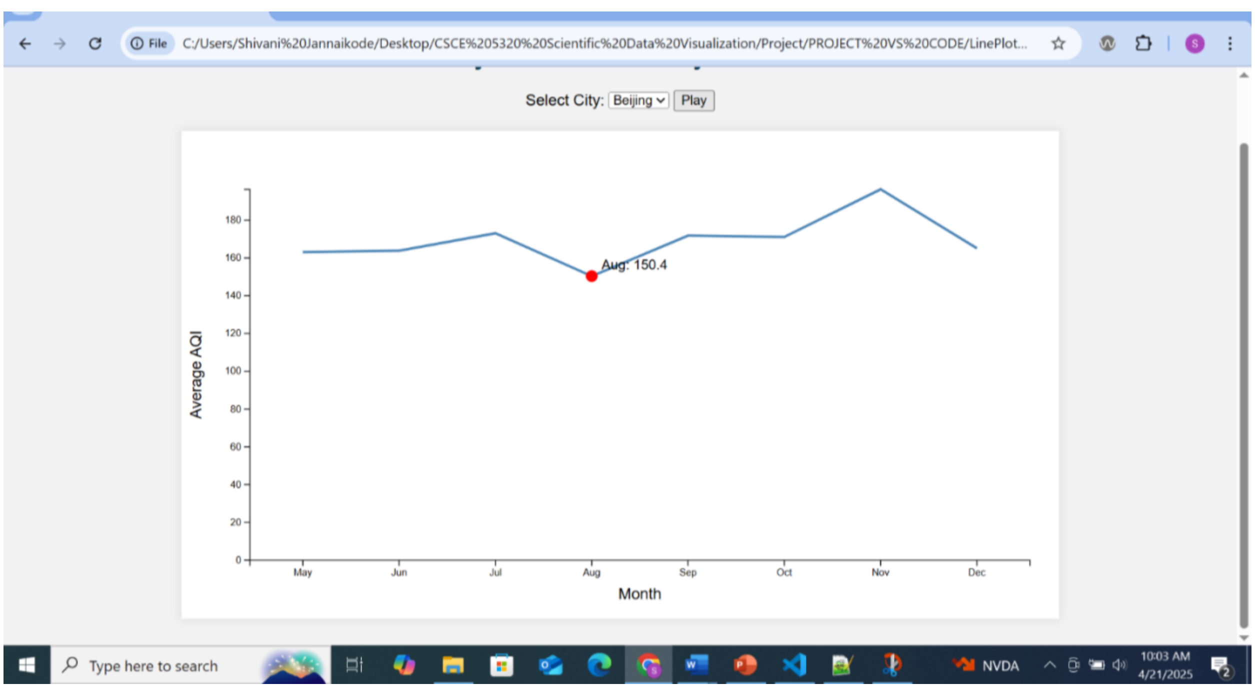


Fig 27: Visualization 6 – Interactive Line Chart Analysis 2

Paris has its highest AQI in June (182.1) and a steady decline to December (153.6), showing a smoother seasonal trend.

Early Peak Behavior: AQI peaks in June (182.1) and gradually declines towards December (153.6), suggesting higher summer pollution, potentially from traffic and ozone-related effects.

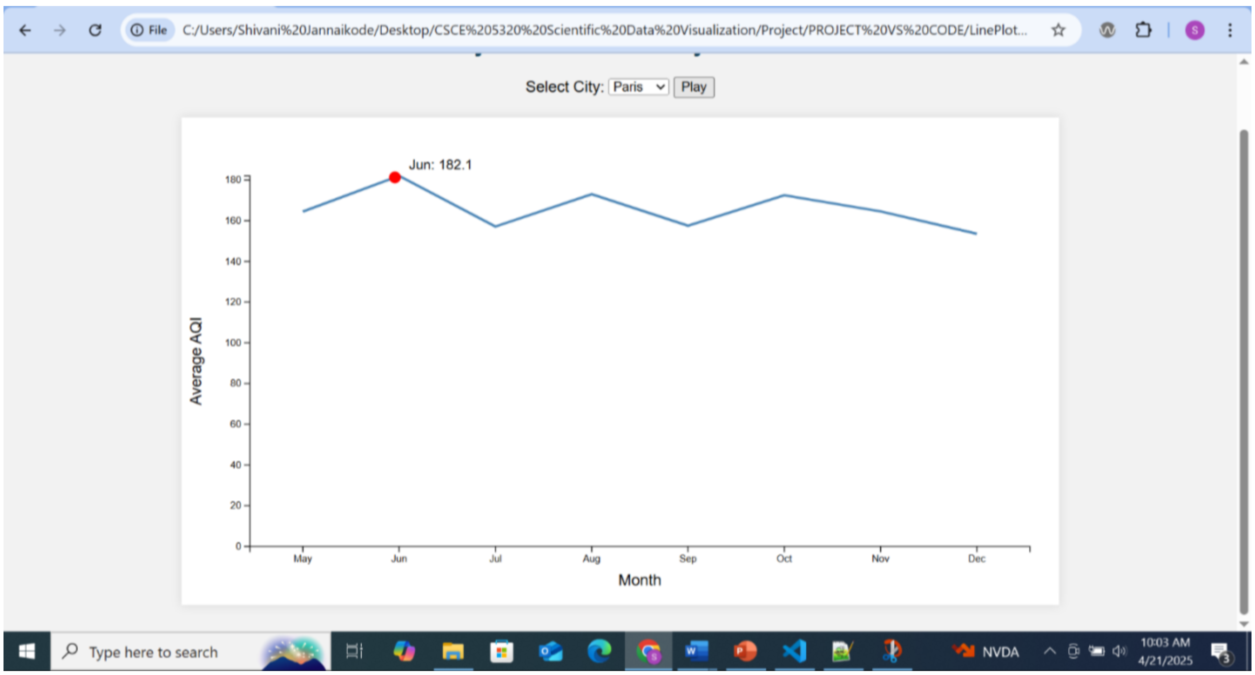


Fig 28: Visualization 6 – Interactive Line Chart Analysis 3

Stable Decline: The smooth downward trend without sharp fluctuations indicates predictable seasonal behavior, aiding in long-term planning and mitigation strategies.

Fig 29: Visualization 6 – Interactive Line Chart Analysis 4

Tokyo shows a dip in June (144.4) after a peak in May (187.1) and a gradual rise again.

Sharp Mid-Year Dip: Tokyo experiences a sharp drop in AQI from May (187.1) to June (144.4), which may result from weather conditions like rainfall or stronger winds improving air quality.

Fig 30: Visualization 6 – Interactive Line Chart Analysis 5

Recovery Trend: After the dip, there's a gradual AQI increase towards October-November, implying seasonal rebound that can be anticipated annually.

Fig 31: Visualization 6 – Interactive Line Chart Analysis 6

Cairo demonstrates a strong rise from May (103.2) to June (≈180) and another high in December (185.9).

Abrupt Summer Rise: Cairo shows a steep jump from May (103.2) to June (~180), possibly due to dust storms, industrial activity, or rising temperatures intensifying pollution.

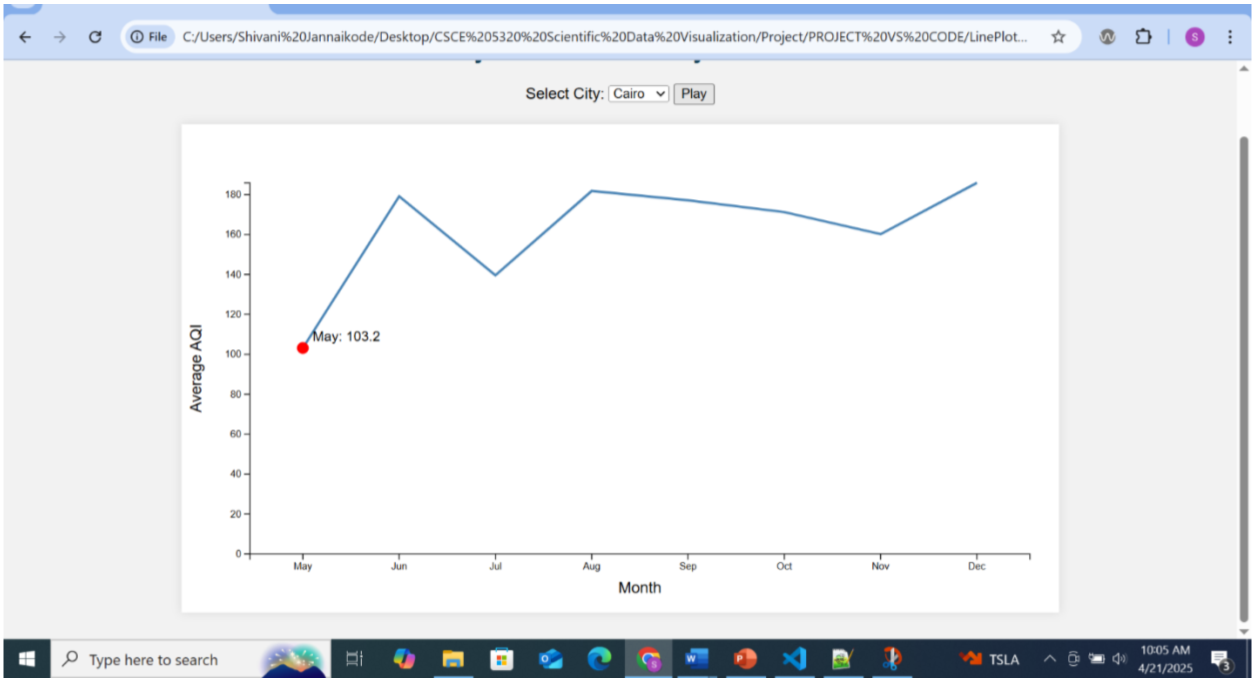


Fig 32: Visualization 6 – Interactive Line Chart Analysis 7

Secondary Winter Spike: AQI rises again in December (185.9), pointing to recurring winter pollution likely tied to inversion layers, making Cairo a candidate for dual-peak pollution forecasting.

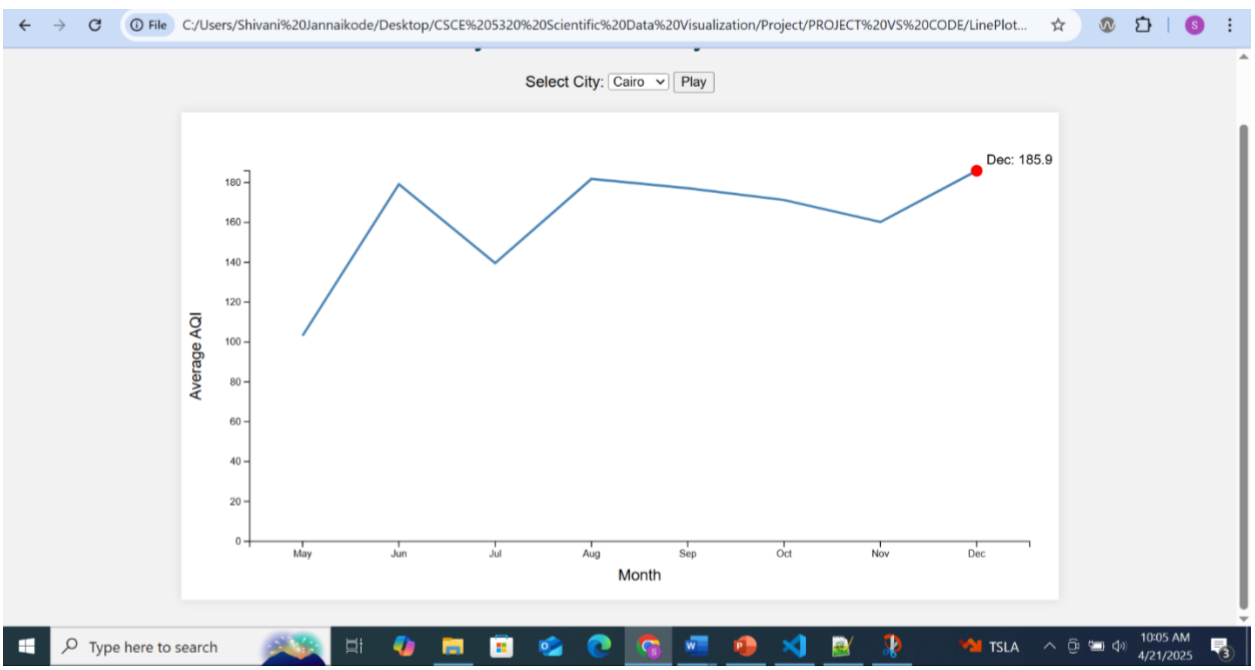


Fig 33: Visualization 6 – Interactive Line Chart Analysis 8

Insights: The visualizations reveal consistent seasonal AQI patterns across cities, with pollution levels typically rising in late autumn and early winter months. Cities like Beijing and Cairo show noticeable spikes in AQI during November and December, indicating recurring high pollution periods. These predictable trends support the potential of using historical seasonal data to forecast future pollution spikes effectively.

7.6.4 Visualization Conclusion

The past seasonal AQI patterns show repeatable month-over-month trends, which help anticipate future pollution spikes - especially during winter months in Beijing and Cairo. These city-specific patterns support the use of historical data for predictive modeling and air quality management strategies.

7.7 Visualiazation 7: Donut-Style Animated Pie Chart

7.7.1 Question Addressed: How has AQI evolved in major cities over the past years and what are the emerging trends?

7.7.2 Visualization Explanation:

1. City-Wise AQI Breakdown by Month (Temporal Insight)

- The donut chart displays monthly average AQI values for each city, dynamically updating as months progress.
- Each slice represents a city's AQI contribution in a specific month, making it easy to compare pollution levels across cities within that month.
- The proportional percentages (e.g., "Beijing: 280.5 (42.1%)") help identify which cities dominate AQI in each month.

2. Animation Shows Monthly Evolution (Trend Over Time)

- The playButton triggers a time-lapse animation that iterates over all months using a setInterval, helping viewers visually detect how AQI distribution shifts across the year.
- This gives a seasonal trend perspective—e.g., some cities may show consistent peaks during winter months (inversion effect), while others fluctuate during summer (ozone-related).

3. Color-Coded Legend for Clarity

- The legend helps users track each city consistently over time as the slices animate.
- This makes it easy to spot recurring patterns (e.g., "Beijing's slice is always dominant in Nov and Dec").

4. Insights for Emerging Patterns

- The chart supports emerging trend analysis:
  - If a city's AQI share grows month by month, it's a red flag for rising pollution.
  - Cities that shrink in slice size might be improving air quality.
- Over several years (if extended), the same animation pattern would expose long-term shifts, such as cleaner cities growing or polluted ones worsening.

The donut-style animated pie chart helps visualize city-wise AQI evolution over time, revealing monthly and seasonal trends, and highlights which cities are consistently more polluted. It supports both cross-sectional (month-wise) and longitudinal (trend-over-time) analysis.

7.7.3 Visualization Analysis:

1. Beijing:

Trend Insight: Beijing consistently holds one of the top AQI contributions every month, with a peak of 196.3 in November (27.9%). It shows gradual increases from summer to winter months.

Emerging Pattern: AQI rises steadily from August (150.4, 22.4%) to November, suggesting winter-related pollution buildup.

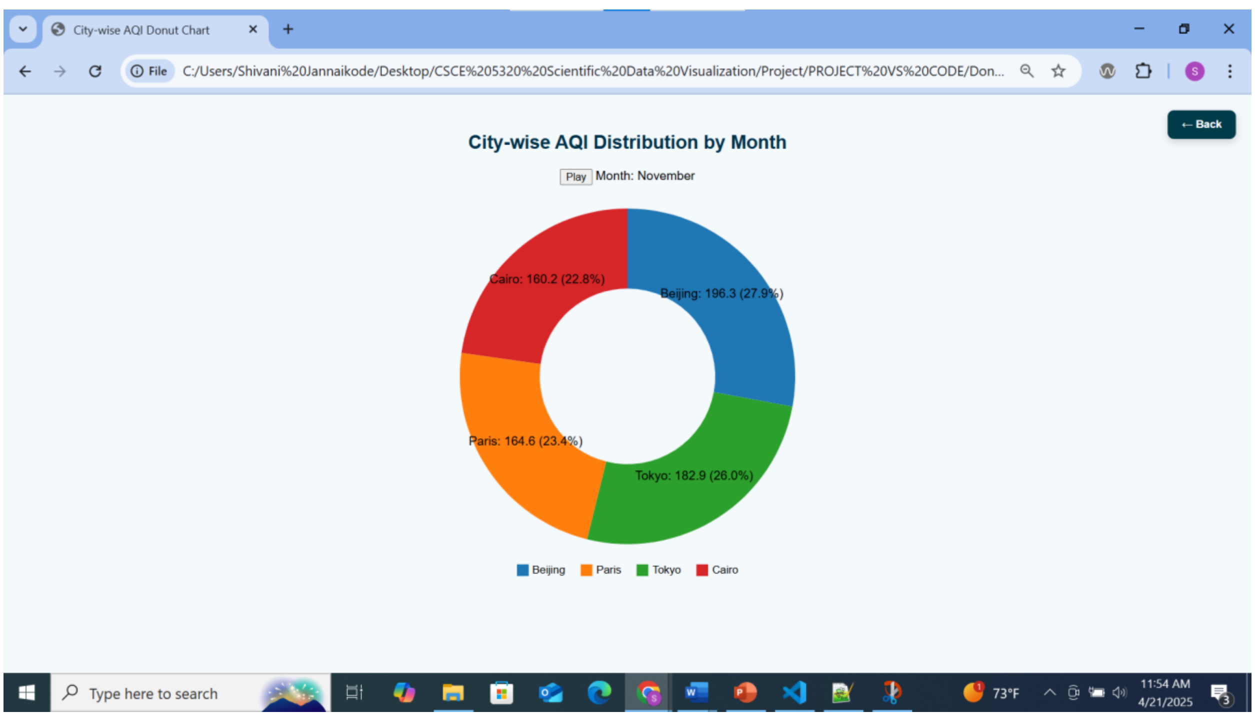


Fig 34: Visualization 7 – Interactive Donut-Style Chart Analysis 1

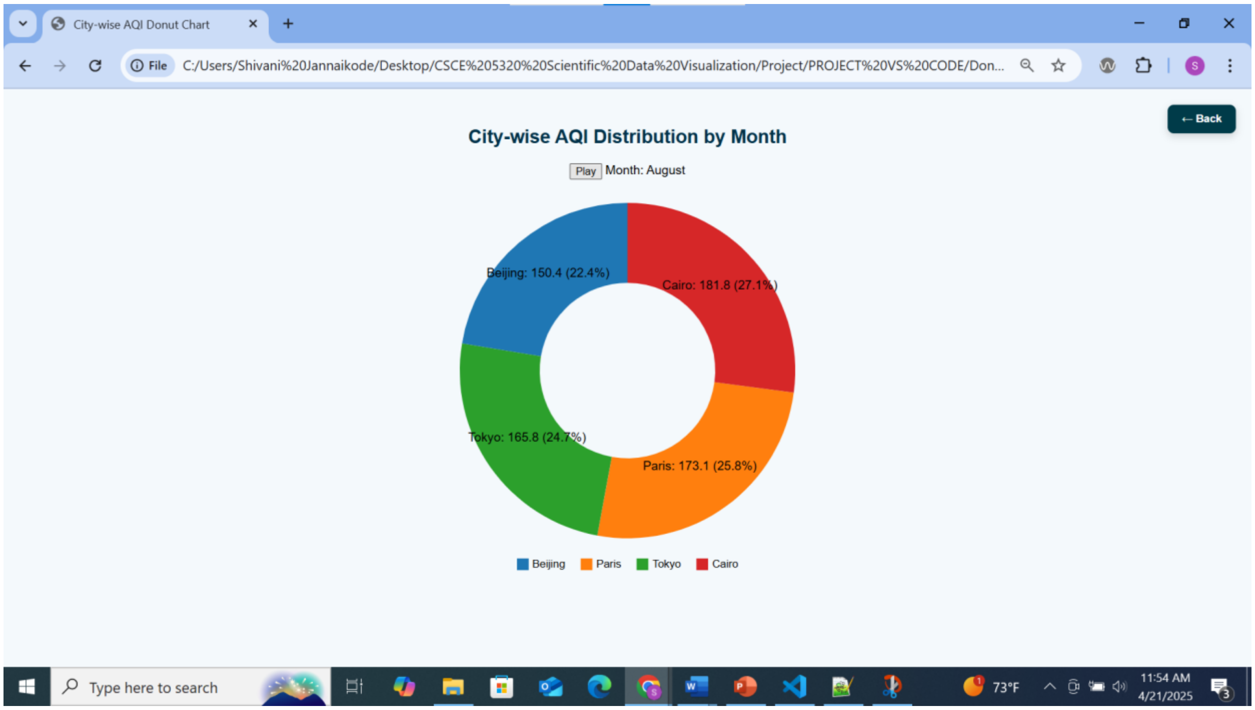


Fig 35: Visualization 7 – Interactive Donut-Style Chart Analysis 2

2. Paris:

Trend Insight: Paris starts strong with 182.1 in June (27.2%), but its share declines over the months, reaching 153.6 (22.7%) in December.

Emerging Pattern: AQI in Paris is more balanced, suggesting fewer seasonal spikes but a steady decline from mid-year onward.

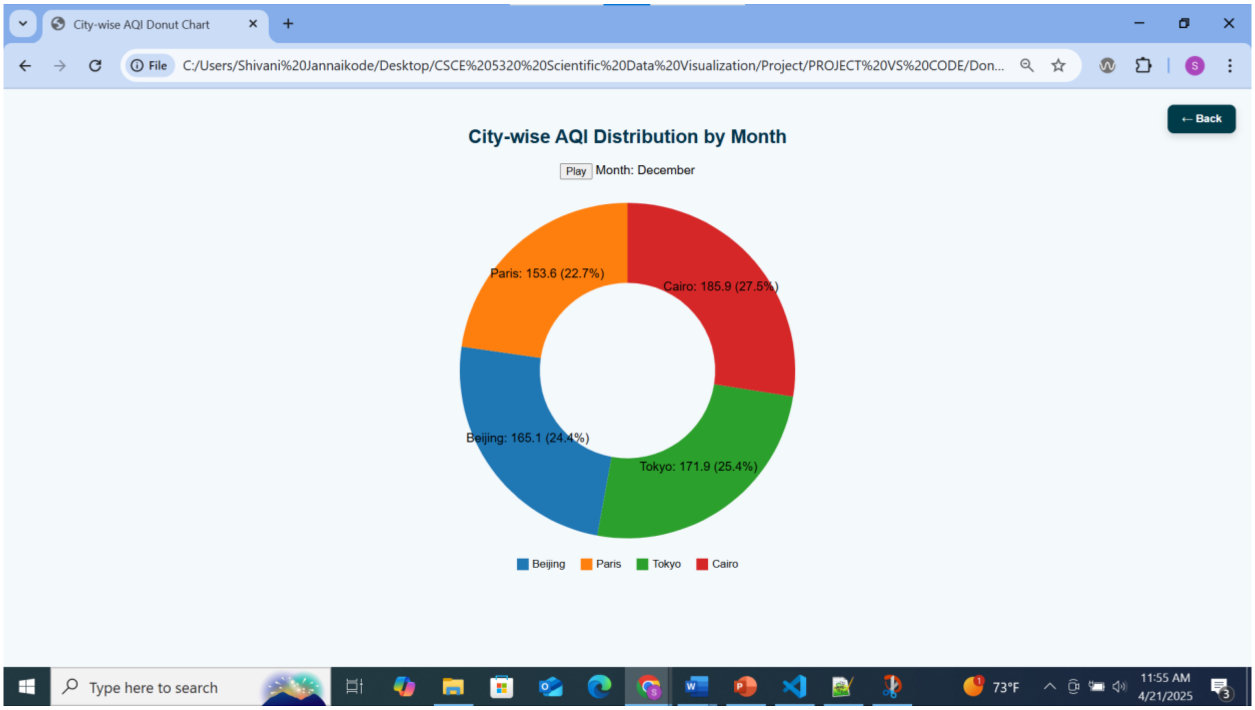


Fig 36: Visualization 7 – Interactive Donut-Style Chart Analysis 3

3. Tokyo:

Trend Insight*:* Tokyo's AQI stays fairly consistent, peaking in May (187.1, 30.3%) and dipping slightly in June (144.4, 21.6%), then stabilizing.

Emerging Pattern: While early months show higher pollution, Tokyo demonstrates stable air quality in the latter half of the year.

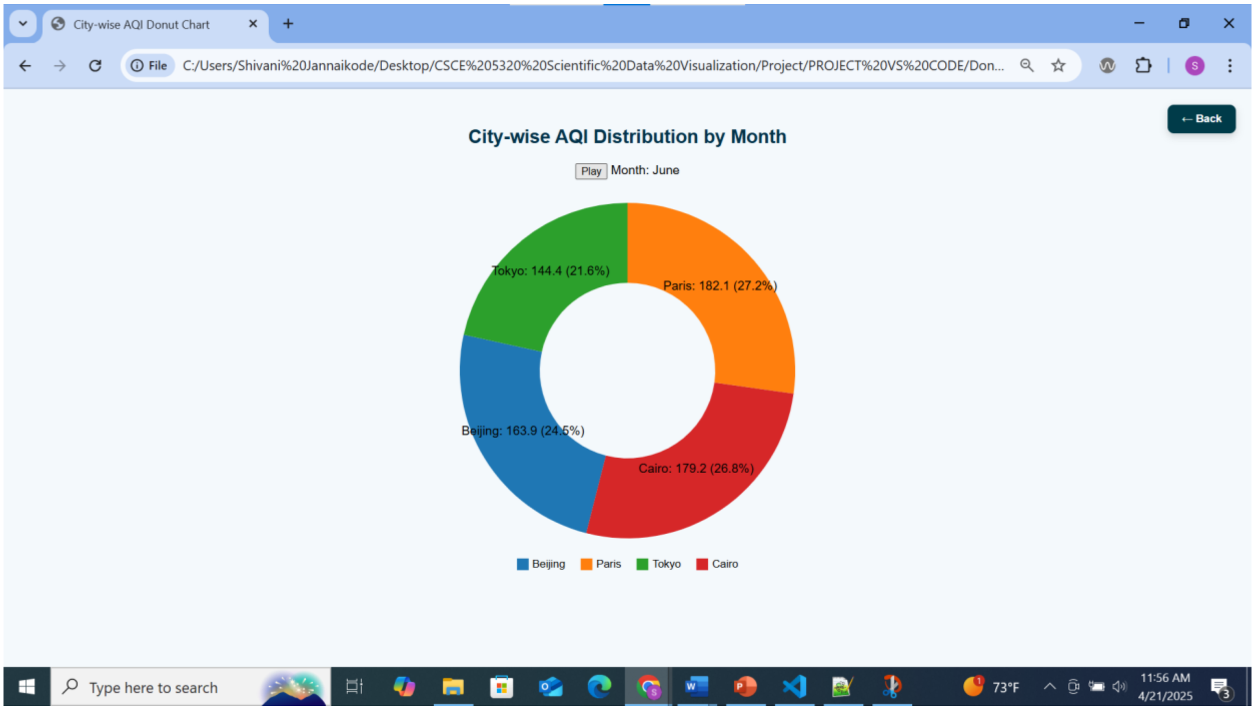


Fig 37: Visualization 7 – Interactive Donut-Style Chart Analysis 4

4. Cairo:

Trend Insight: Cairo starts low in May (103.2, 16.7%) but rises sharply by August and peaks in December (185.9, 27.5%).

Emerging Pattern: Cairo shows delayed pollution spikes, with AQI increasing toward winter, possibly due to weather inversion effects.

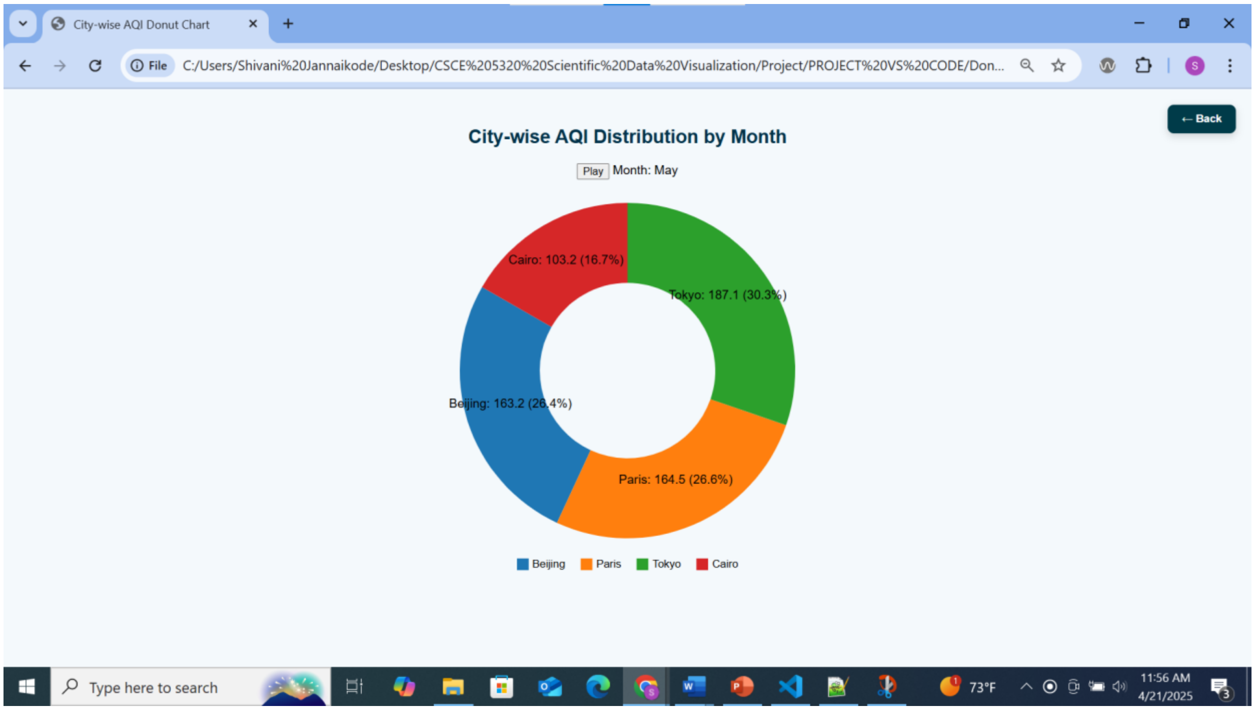


Fig 38: Visualization 7 – Interactive Donut-Style Chart Analysis 4

Insights: We observe that Beijing and Cairo show a steady AQI rise toward the year-end, indicating seasonal pollution buildup. Tokyo maintains a relatively stable AQI throughout the months with slight early peaks, while Paris sees a gradual decline in AQI from mid to late year. Overall, winter months (Nov–Dec) exhibit higher pollution shares, especially in cities like Beijing and Cairo, signaling a need for seasonal mitigation strategies.

7.7.4 Visualization Conclusion:

- The Donut Chart visualizes AQI evolution across cities month-by-month, highlighting when each city contributes most to air pollution.
- It reveals seasonal and regional patterns—like Beijing and Cairo showing winter spikes, and Tokyo peaking early in summer.
- Emerging trends suggest: Cities like Cairo and Beijing face increased AQI in later months, helping policymakers target pollution control in fall-winter months.

7.8 Visualiazation 8: Interactive Grouped Bar Chart

7.8.1 Question Addressed: Can weekday vs weekend patterns in AQI help design urban mobility or traffic control policies?

7.8.2 Visualization Explanation

1. Clear Separation of Weekday vs Weekend AQI

d["Day Type"] = d.date.getDay() >= 5 ? "Weekend" : "Weekday";

Each data point is classified as either a Weekday or Weekend, enabling a direct comparison of pollution levels between these periods.

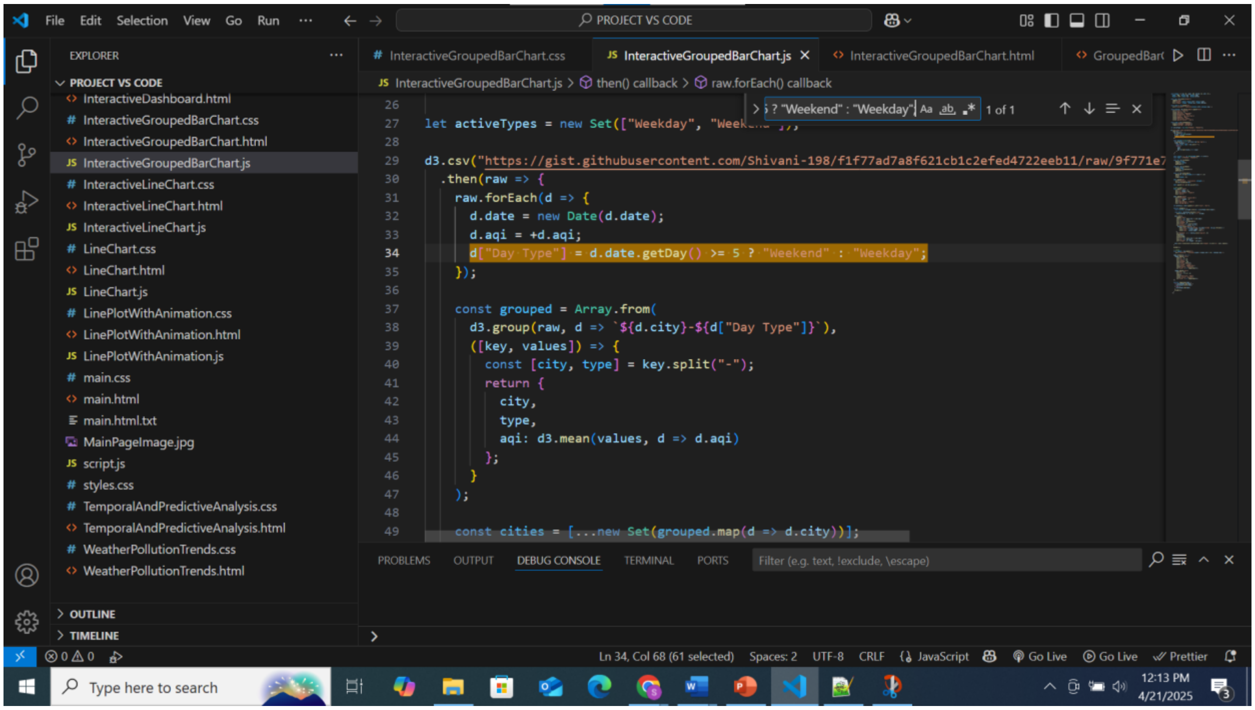


Fig 38: Weekday Vs Weekend AQI Separation

2. Grouped Comparison by City

d3.group(raw, d => `${d.city}-${d["Day Type"]}`)

The average AQI is calculated per city for both Weekdays and Weekends, allowing city-level insights.

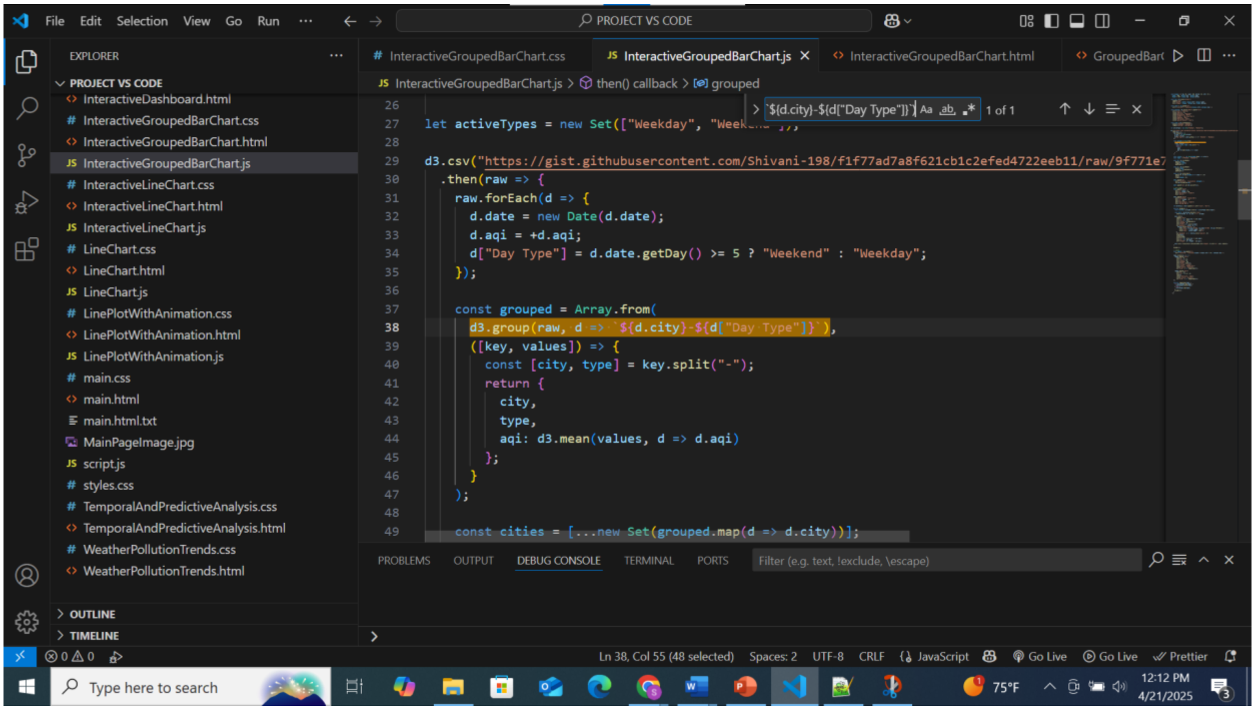


Fig 39: Grouoed Comparison by City

3. Visualized Differences Using Grouped Bars

Each city has two bars side-by-side, one for Weekday and one for Weekend AQI:

Higher weekday AQI → indicates traffic or industrial activity.

Higher weekend AQI → may suggest recreational emissions or other causes.

4. Interactive Tooltip Provides Decision-Ready Info

Hovering reveals exact AQI values, which can help officials make data-driven policy decisions.

5. Toggle Between Weekday/Weekend with Legend

Urban planners can interactively focus on one type of day to see how AQI behaves, helping tailor day-specific traffic policies.

Practical Urban Planning Implications:

If Weekday AQI > Weekend AQI: Indicates pollution from commuting and industrial operations.

If Weekend AQI > Weekday AQI: Suggests pollution from events, travel, or domestic sources.

This grouped bar chart not only compares AQI between weekdays and weekends by city but also offers clear, actionable insights for urban mobility and traffic control design. It enables stakeholders to spot trends, compare patterns, and make targeted interventions for better air quality.

7.8.3 Visualization Analysis:

Beijing: Higher AQI on weekdays (≈175.8) than weekends (≈153.8).

Trend: Higher AQI on weekends.

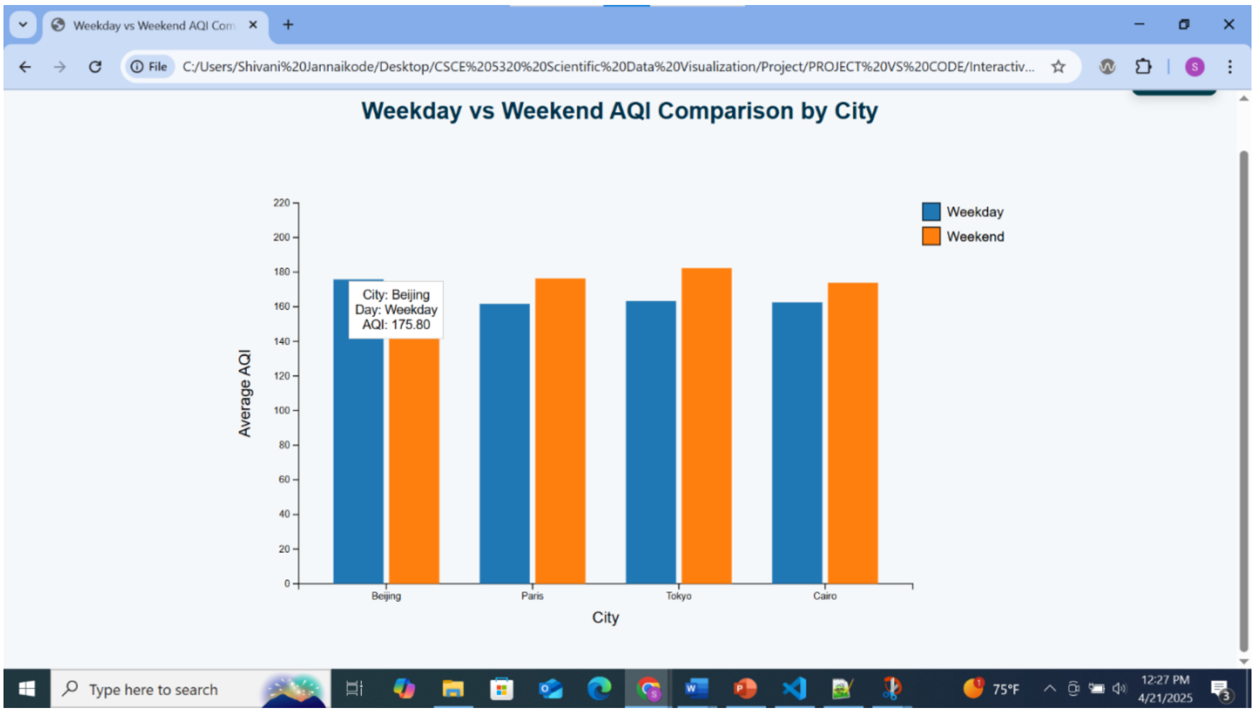


Fig 40: Visualization 8 – Interactive Grouped Bar Chart Analysis 1

Paris: Higher AQI on weekends (≈176.3) compared to weekdays (≈161.6).

Trend: Higher AQI on weekends.

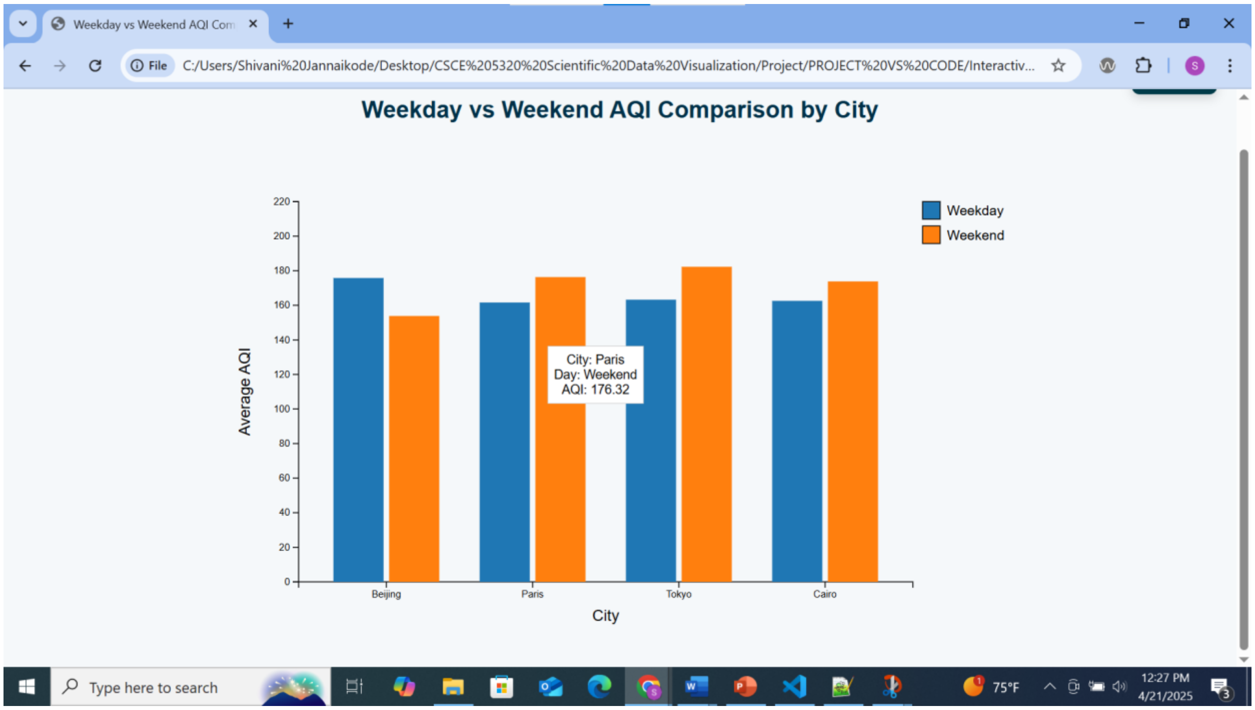


Fig 41: Visualization 8 – Interactive Grouped Bar Chart Analysis 2

Tokyo: Weekend AQI (≈182.3) exceeds weekday (≈163.2)

Trend: Significantly Higher AQI on weekends.

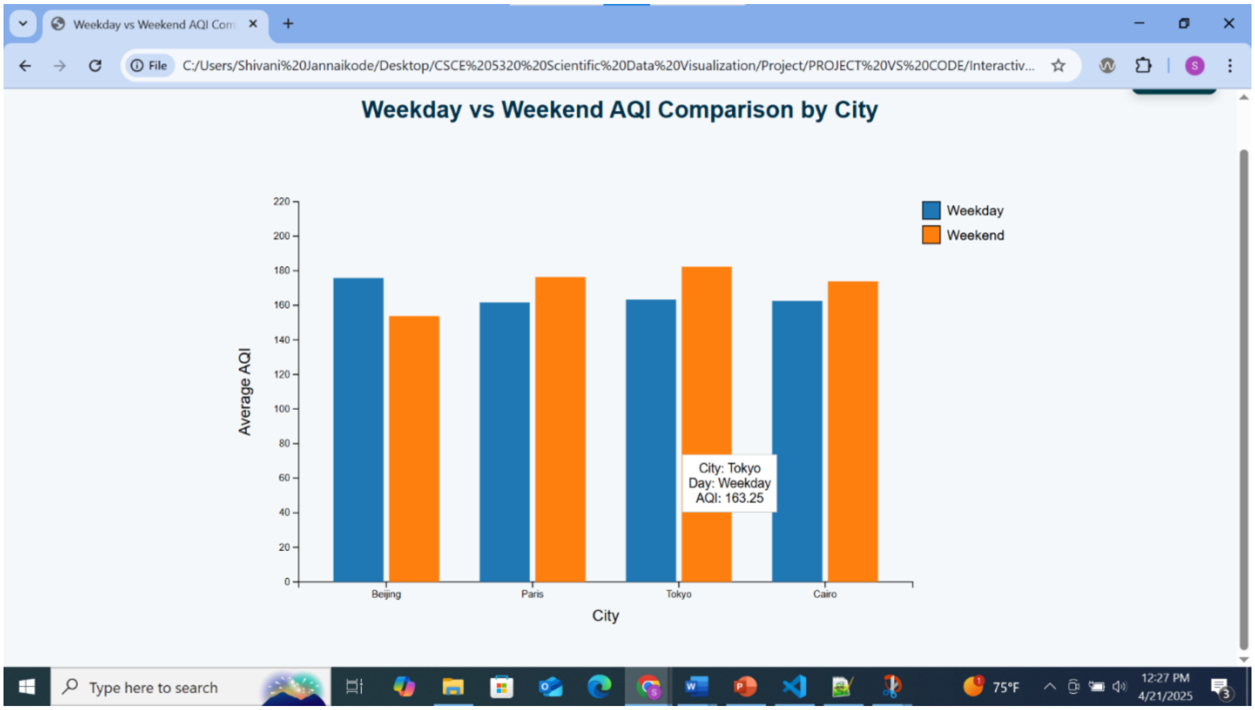


Fig 42: Visualization 8 – Interactive Grouped Bar Chart Analysis 3

Cairo:AQI also higher on weekends (≈174.3) than weekdays (≈162.6).

Trend: Higher AQI on weekends.

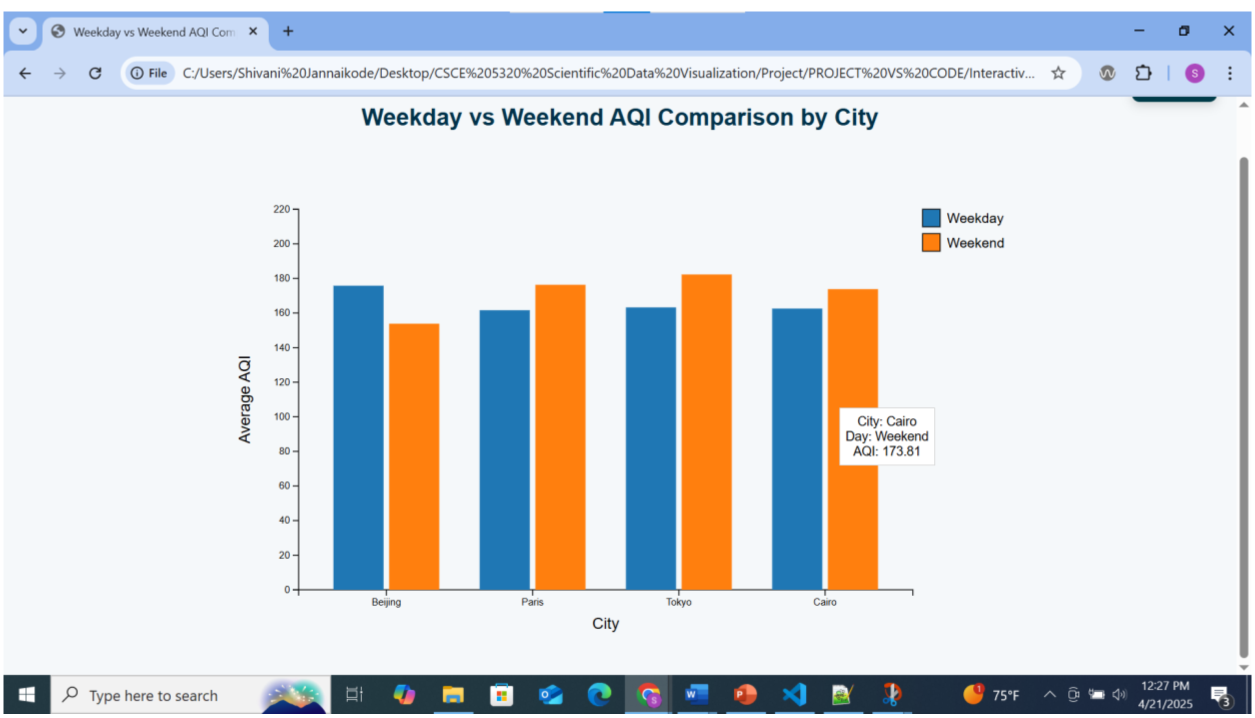


Fig 43: Visualization 8 – Interactive Grouped Bar Chart Analysis 4

Insights: Beijing shows higher AQI on weekdays (≈175.8), indicating weekday traffic and industrial emissions as major contributors—suggesting weekday congestion control (e.g., remote work, traffic zoning) can reduce pollution.

Paris, Tokyo, and Cairo all have higher AQI on weekends (≈176.3, 182.3, 174.3 respectively), likely due to increased leisure and tourist travel, indicating a need for weekend-focused interventions like pedestrian zones, expanded public transport, or emissions regulations in high-traffic recreational areas.

7.8.4 Visualization Conclusion:

AQI trends clearly support adaptive traffic policies tailored to city-specific weekday/weekend patterns to improve urban air quality.

C. Health And Environmental Policy:

7.9 Visualiazation 9: Grouped Animated Box Plot

7.9.1 Question Addressed: Which cities consistently exceed WHO-recommended air quality thresholds?

7.9.2 Visualization Explanation

1. Filtering Based on WHO Thresholds

- The chart only displays data where pollutant levels exceed WHO thresholds.
- These thresholds are defined in the WHO_THRESHOLDS object.
- For each city and month, the code computes the monthly average for each pollutant and includes only pollutants that exceed these thresholds.

2. City-wise and Month-wise Grouping

The d3.rollup() function groups the dataset:

- First by city
- Then by Month

It calculates the average pollutant values per month per city, so the chart can reveal consistent exceedances over time.

3. Grouped Bar Visualization

- Each group of bars corresponds to a city, and each bar in the group represents a pollutant that exceeds the threshold in a specific month.
- The x-axis shows cities, and the y-axis shows pollutant levels.

- Bars animate per month, allowing users to see which cities repeatedly breach safe limits.

4. Animation Over Time

- The updateChart(month) function updates the chart for each month.
- The toggleAnimation() function controls the month-wise animation, so the chart cycles through different months to show temporal trends.
- This reveals cities that frequently exceed thresholds month after month, highlighting consistent violators.

The visualization is effective in identifying cities that consistently exceed WHO air quality thresholds because it:

- Filters only pollutant values above the safe limits.
- Groups the data by city and month.
- Animates month-over-month exceedance patterns.
- Clearly visualizes which pollutants are problematic for each city.

7.9.3 Visualization Analysis

Cairo:

Exceeds WHO thresholds for AQI (~185), PM2.5, PM10, NO_2, CO, and O_3.

It consistently shows multiple pollutants above safe levels—highlighting chronic air quality issues.

Fig 44: Visualization 9 – Interactive Grouped Bar Chart Analysis 1

Tokyo:

High levels of AQI (187.11) and PM10 (187.21) far above the WHO limits.

NO_2 (47.83 ppb) and O_3 (103.66 ppb) also breach WHO thresholds—suggesting intense vehicular and industrial emissions.

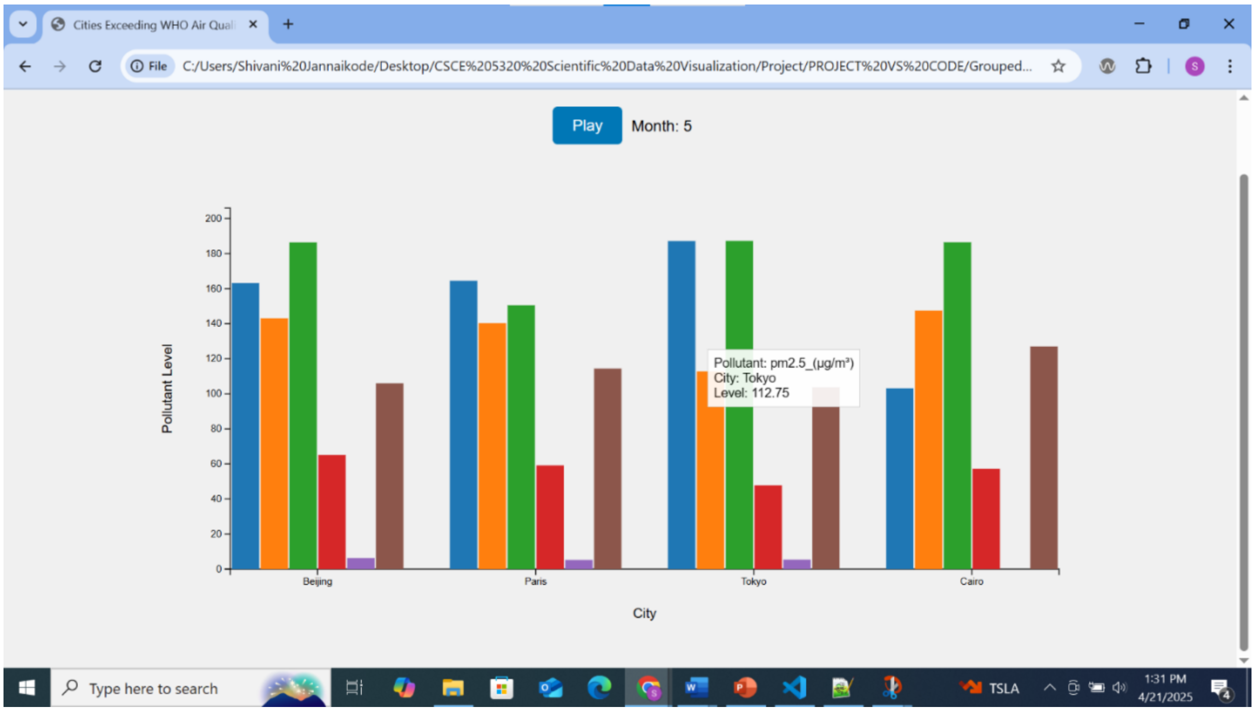


Fig 45: Visualization 9 – Interactive Grouped Bar Chart Analysis 2

Beijing:

PM10 (~185 µg/m³), AQI (~165), and CO levels are notably high.

Indicates significant pollution from combustion and traffic sources.

Fig 46: Visualization 9 – Interactive Grouped Bar Chart Analysis 3

Paris:

Although slightly lower, still exceeds PM2.5 (~115 µg/m³) and AQI (~165) levels.

Requires attention for fine particulate control.

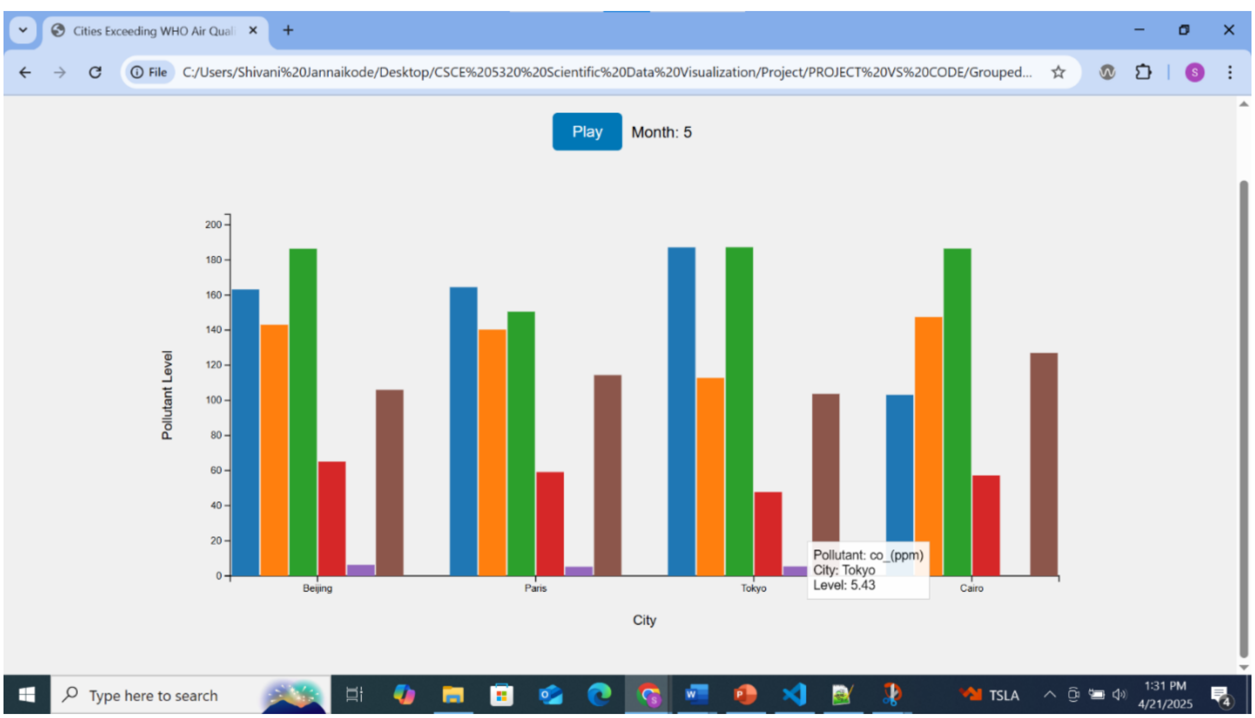


Fig 47: Visualization 9 – Interactive Grouped Bar Chart Analysis 4

Insights: Major cities like Beijing, Tokyo, and Cairo, pollutant levels such as PM2.5, PM10, and O_3 visibly exceed WHO-recommended thresholds. Tokyo emerges as a hotspot, with consistently high levels of AQI (187), PM10 (187.21), and O_3 (103.66), while Cairo and Beijing also show alarming readings across multiple pollutants. These trends highlight a critical need for targeted interventions in traffic management, industrial emissions, and urban zoning to combat chronic air pollution in urban centers**.**

7.9.4 Visualization Conclusion:

- The grouped bar chart clearly breaks down pollutant levels per city and pollutant type, allowing easy comparison with WHO thresholds.
- The tooltips highlight exact concentrations, making it evident which cities exceed safe air quality limits for each pollutant.
- By animating data month-wise and displaying pollutant bars only if they breach WHO limits, it focuses attention on problematic areas across time.

7.10 Visualiazation 10: Interactive Line Chart

7.10.1 Question Addressed: What regions require immediate environmental interventions based on AQI trends and population exposure?

7.10.2 Visualization Explanation

1. Monthly AQI Trend Visualization for Each City

- The code reads AQI data per city per date, and computes the monthly average AQI:
- Then it groups data by city and month, allowing for time-based analysis:
- This grouping enables tracking how AQI changes month-by-month, revealing rising patterns or sustained high levels, both of which are indicators of cities in need of intervention.

2. Line Chart Structure Reveals Trend Direction Clearly

- Each city's AQI over the year is represented as a distinct colored line:
- Rising slopes or plateaus at high AQI levels visually indicate cities with deteriorating or persistently poor air quality.
- Users can visually identify cities where AQI is increasing, suggesting growing pollution concerns.

3. Interactivity Enhances Insight

- The chart uses tooltips to show exact AQI values for each month and city on hover:

- A clickable legend lets users isolate or compare city trends interactively:
- This helps analysts or policymakers focus on specific cities and compare AQI patterns directly, identifying outliers or worst performers.

4. Y-Axis Scaling and Sorting by Month Helps Spot High-AQI Zones

- The y-axis is automatically scaled based on the maximum AQI values across all cities:
- The x-axis follows calendar month order, ensuring logical interpretation of time-based change.
- This ensures the chart provides clear, readable trends where spikes or unhealthy AQI levels can be spotted quickly.

How This Helps Identify Intervention-Need Cities?

- Identifying cities with consistently high AQI levels over multiple months (e.g., above WHO threshold of 100).
- Detecting upward trends that may indicate worsening conditions.
- Visual comparison of cities, revealing which are persistently worse off.
- User interactivity to filter out or highlight problem areas.

7.10.3 Visualization Analysis:

Tokyo: Peaks in May (187.59) and Dec (173.45)

Beijing: High in May (164.71) and Nov-Dec (~170–196)

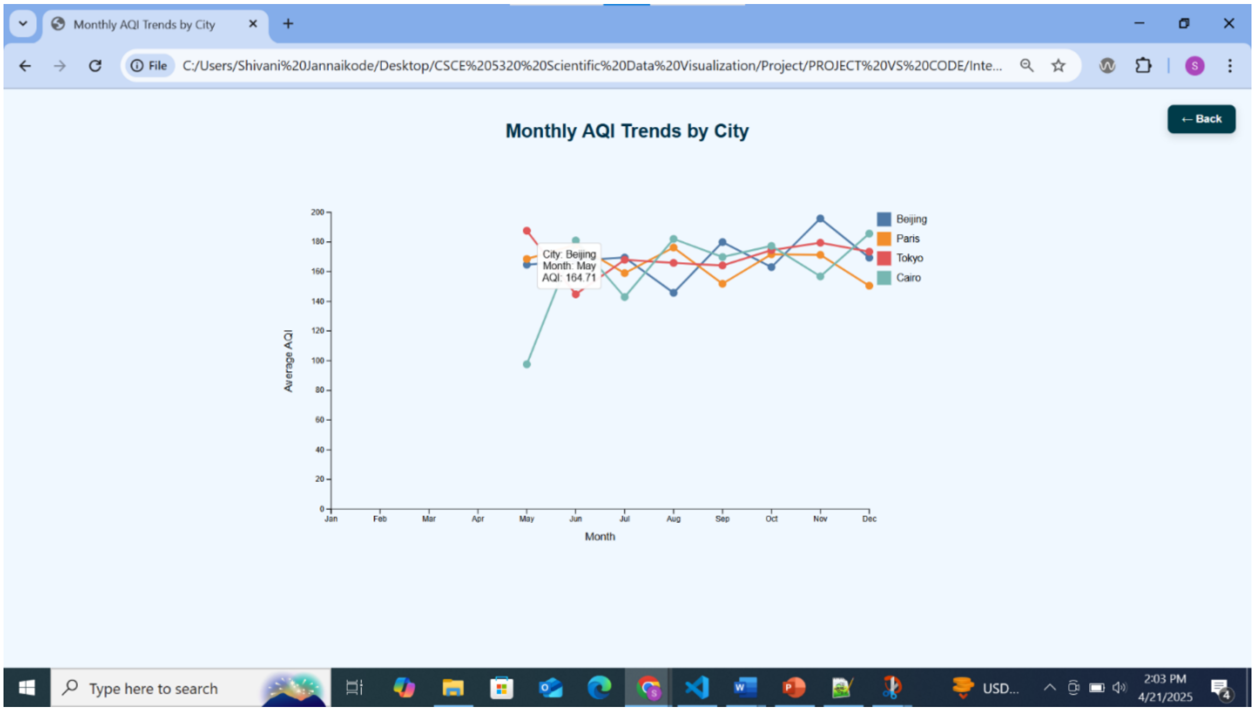


Fig 48: Visualization 10 – Interactive Line Chart Analysis 1

Paris: AQI reaches 150.48 in Dec

Cairo: Lower AQI in May (~97.59), but fluctuates later

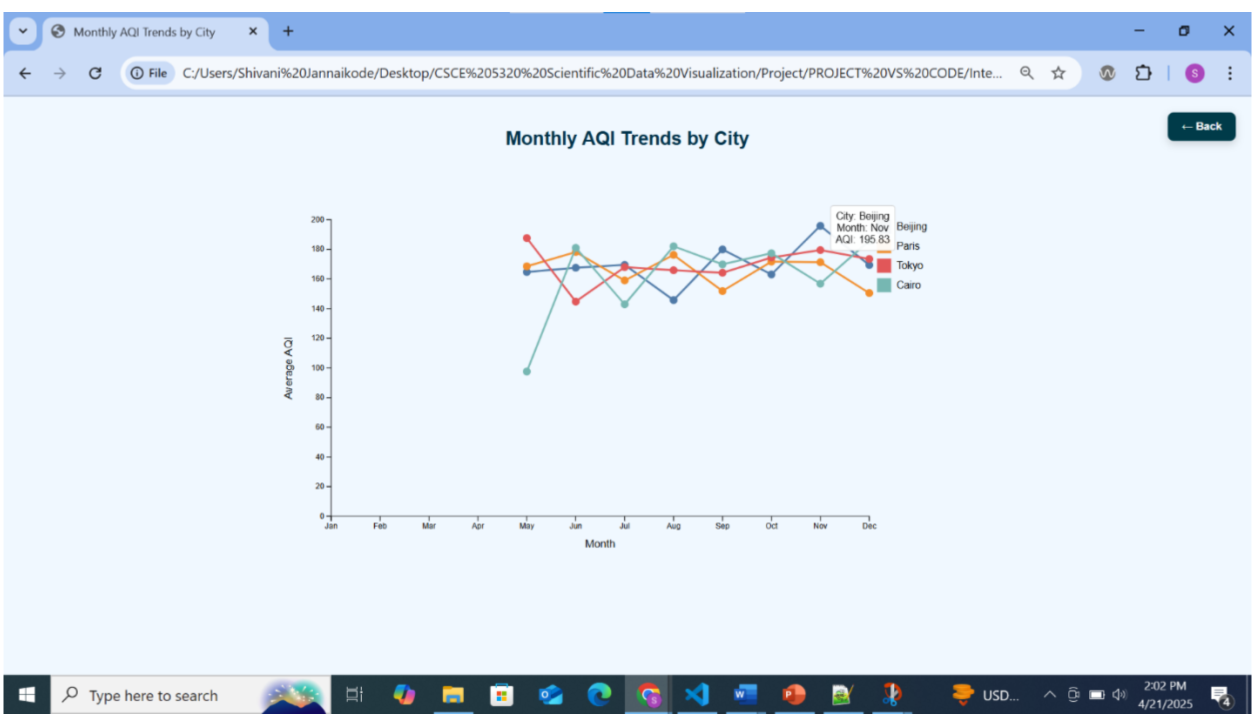


Fig 49: Visualization 10 – Interactive Line Chart Analysis 2

Insights: Tokyo peaks in May, Beijing in Nov-Dec. These patterns may align with:

- Tourist traffic spikes
- Weather inversion layers
- Holiday shopping seasons

Policies like weekend driving restrictions or public transit incentives can be strategically timed in these peak months.

7.10.4 Visualization Conclusion

- Highlight which cities need intervention
- Show when (which months) these interventions are most urgent
- Support designing traffic control strategies focused on months with the highest pollution.

2) STORY TELLING APPROACH

Stage 1: Understanding and Exploring the Dataset

a. What does the raw data look like?

The raw data consists of two structured datasets sourced from Kaggle: the Global Air Quality Dataset and the Global Weather Repository. The air quality dataset contains 3,660 rows and 13 columns, representing daily air quality readings across major cities worldwide. It includes variables such as the date, city, country, AQI, and concentrations of pollutants like PM2.5, PM10, NO_2, SO_2, CO, and O_3. The weather dataset is more extensive, with 63,144 records and 41 attributes, covering temperature, humidity, wind metrics, atmospheric pressure, and other weather indicators. Both datasets are free of missing values, and the air quality data is matched with corresponding weather readings through city and date fields. These raw datasets are rich in both pollution and environmental context, allowing for a multifaceted analysis of urban air conditions.

b. Which key variables or trends stood out during your exploration?

During the initial data exploration phase, variables such as AQI, PM2.5, and NO_2 stood out as major indicators of air pollution severity. High variability in AQI across different cities and months was immediately noticeable, with some months showing spikes as high as 300, indicating extremely poor air quality. Temperature, humidity, and wind speed were also key environmental variables that appeared to influence AQI levels. Early seasonal observations revealed that AQI tends to rise in colder months, especially in fall and winter, suggesting potential atmospheric inversion effects that trap pollutants. The pairing of pollution data with weather data offered opportunities to explore deeper interactions, such as whether humidity amplifies pollution or wind helps disperse it.

Stage 2: Correlation Analysis and Data Transformation

a. What relationships exist between variables?

The analysis revealed meaningful correlations between AQI and weather variables, particularly humidity and wind speed. The most prominent relationship showed that low wind speed combined with high humidity consistently resulted in higher AQI levels. For example, under low-wind, high-humidity conditions, the average AQI was around 169—the highest among observed combinations. This indicates that stagnant, moisture-rich air likely traps pollutants, worsening air quality. Conversely, high wind conditions generally reduced AQI, even when humidity was elevated, as wind facilitates pollutant dispersion. These findings illustrate that AQI is not determined by a single factor but is influenced by the interaction between multiple environmental conditions.

b. Are there any significant patterns between different factors? How does data transformation help uncover these?

Yes, several significant patterns were uncovered, particularly related to seasonal and conditional changes in AQI. AQI was found to be higher during fall and winter seasons, supporting the hypothesis that atmospheric conditions during colder months trap pollutants near the ground. Data transformation played a crucial role in surfacing these insights. Steps like merging the air quality and weather datasets by city and date allowed for direct comparison of pollution and meteorological data. Additionally, creating a new 'season' column helped analyze AQI trends across seasons. Categorizing wind speed and humidity into "Low" and "High" groups enabled targeted analysis of their combined effects on AQI. These transformations structured the data in ways that made it easier to visualize and quantify complex relationships, leading to clearer, actionable insights

8. AREAS OF FOCUS

8.1 Visualization

The generated visualizations were highly interactive and significantly enhanced user engagement. Features like dropdown menus, tooltips, sliders, legends, and play/pause buttons in charts (line plots, donut charts, box plots) allowed users to dynamically explore the data and personalize their insights. This interactivity elevated the user experience beyond static visuals, enabling a hands-on exploration of pollution trends. Methodologically, the use of D3.js enabled real-time binding of data to SVG elements, facilitating animated transitions and filters. This helped uncover deeper relationships, such as the combined effects of low wind speed and high humidity on AQI, and seasonal spikes in pollution across cities, which static plots would struggle to convey.

8.2 WebPage Integration

To meet the webpage integration requirements, an interactive website was developed using HTML, CSS, and JavaScript, with D3.js powering the visualizations for dynamic and responsive data storytelling. In addition to custom-coded visualizations, a Tableau dashboard was also incorporated to enrich the interactive experience. The dashboard was published to Tableau Public, and the share feature was used to obtain the embed code. This embed code was seamlessly integrated into the webpage's HTML structure, allowing users to interact with the dashboard directly within the website. When users click on the "View Interactive Dashboard" button on the homepage, they are presented with the embedded Tableau visualization, enabling smooth and integrated exploration of the dashboard within the same user environment.

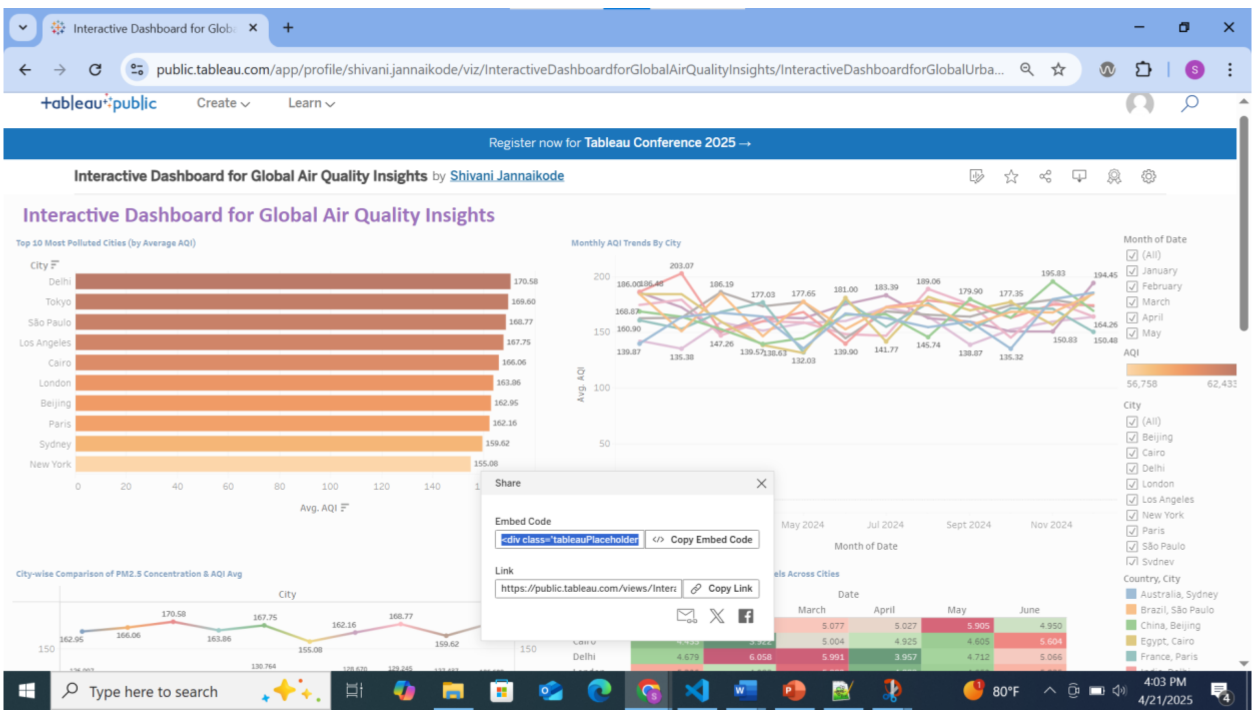


Fig 50: Interactive Dashboard Integration on WebPage

8.3 Transitioning from Static to Interactive Storytelling

The transition from static visuals to interactive storytelling was clearly demonstrated through the use of interactive dashboards and dynamic charts. Unlike traditional static reports, users could toggle between seasons, cities, day types (weekday/weekend), or pollutant types to explore targeted AQI behaviors. For example, animated donut charts revealed evolving city-wise AQI shares month-by-month, and interactive bar charts visualized weekday vs weekend differences, aiding in understanding the cause-effect nature of pollution trends. This dynamic storytelling allowed users to not only consume but interrogate the data, empowering deeper insights and fostering data-driven narratives that adapt to users' questions in real time.

CONCLUSION

Major Insights:

- AQI worsens during Fall and Winter seasons globally.
- Low Wind + High Humidity = Highest pollution levels observed.
- Beijing, Tokyo, Cairo frequently exceed WHO limits for PM2.5, PM10, and O_3.
- Historical AQI trends help predict future pollution spikes.
- Weekday vs Weekend patterns reveal city-specific traffic pollution issues.

Tools & Technologies:

- D3.js, HTML, CSS — Created dynamic, animated charts (Boxplots, Bar charts, Heatmaps, Donut charts).
- Tableau Desktop + Public — Interactive dashboards showing city-wise, seasonal pollution patterns.

Impact:

- Enables seasonal public health alerts.
- Supports urban planning with data-driven policies.
- Empowers real-time exploration of pollution trends for researchers, policymakers, and the public.