CSE 564

Mini Project #1

By

Ravi Praveen, Gali

SBUID: 115787605

Information about Dataset

The dataset encompasses daily weather records for major capital cities worldwide, commencing from August 29, 2023. It furnishes an extensive array of attributes, exceeding 40 in number, comprising temperature, wind speed, atmospheric pressure, precipitation levels, humidity levels, visibility range, air quality metrics, and numerous others. Distinguished from mere prognostic data, this dataset furnishes a comprehensive depiction of prevailing weather conditions across the globe. Its breadth and depth render it invaluable for scrutinizing global weather dynamics, unraveling climate variations, and elucidating interconnections among diverse meteorological parameters.

url: https://www.kaggle.com/datasets/nelgiriyewithana/global-weather-repository

Potential Use Cases:

- **Climate Analysis:** Study long-term climate patterns and variations in different regions.
- **Weather Prediction:** Build models for weather forecasting based on historical data.
- **Environmental Impact:** Analyze air quality and its correlation with various weather parameters.
- **Tourism Planning:** Use weather data to help travelers plan their trips more effectively.
- **Geographical Patterns:** Explore how weather conditions differ across countries and continents.

Descriptions for each of the selected attributes:

1. **temperature_celsius:** The temperature at the location in degrees Celsius.

2. **condition_text:** A textual description of the current weather conditions (e.g., "Sunny", "Cloudy", "Rainy", etc.).

3. **wind_kph:** The speed of the wind at the location in kilometers per hour.

4. **wind_direction:** The direction from which the wind is blowing, expressed as a 16-point compass direction (e.g., "N" for north, "NE" for northeast, etc.).

5. **pressure_in:** The atmospheric pressure at the location, measured in inches of mercury (inHg).

6. **humidity:** The relative humidity at the location, expressed as a percentage.

7. **feels_like_celsius:** The "feels-like" temperature, which takes into account factors such as wind chill and humidity, in degrees Celsius.

8. **visibility_km:** The visibility range at the location, measured in kilometers.

9. **uv_index:** The UV (Ultraviolet) Index, which indicates the strength of ultraviolet radiation from the sun.

10**. gust_kph:** The maximum wind gust speed at the location in kilometers per hour.

11. **air_quality_Carbon_Monoxide:** Measurement of Carbon Monoxide (CO) concentration in the air.

12. **air_quality_Ozone:** Measurement of Ozone (O3) concentration in the air.

13. **air_quality_Nitrogen_dioxide:** Measurement of Nitrogen Dioxide (NO2) concentration in the air.

14. **air_quality_Sulphur_dioxide:** Measurement of Sulphur Dioxide (SO2) concentration in the air.

15. **moon_phase:** A description of the current phase of the moon (e.g., "New Moon", "First Quarter", "Full Moon", etc.).

These attributes provide a comprehensive overview of the current weather conditions, including temperature, wind, atmospheric pressure, humidity, visibility, UV index, air quality, wind gusts, and moon phase.

Why did I think this data is interesting?

The selected data attributes offer a captivating insight into a diverse array of factors shaping current weather conditions and environmental dynamics. From the fundamental measure of temperature, providing a baseline understanding of atmospheric warmth, to the descriptive account of weather conditions, detailing the atmospheric state in real-time, each attribute holds significance. Wind speed and direction illuminate the dynamic forces at play, influencing weather patterns and transportation systems alike, while atmospheric pressure serves as a key indicator of weather stability and imminent changes. Humidity levels provide crucial insights into human comfort, agricultural productivity, and disease propagation, while the "feels-like" temperature incorporates nuanced factors like wind chill and humidity, enhancing our understanding of weather's tangible impact. Visibility data sheds light on atmospheric clarity and potential hazards, and the UV index alerts sun exposure risks. Air quality measurements, encompassing carbon monoxide, ozone, nitrogen dioxide, and sulfur dioxide levels, are pivotal for assessing environmental health and human well-being. Lastly, the moon phase, while not directly meteorological, adds an intriguing celestial aspect to our comprehension of daily rhythms and cultural practices. Collectively, these attributes offer a comprehensive snapshot of current weather conditions, guiding our understanding of weather patterns, environmental health, and societal impacts.

About Implementation:

I began the implementation by creating an index.html file, which serves as the main interface for our weather data visualization application. To facilitate navigation, I designed a menu on the left-hand side of the page, organizing the available data into submenus based on their nature.

For categorical data visualization, I utilized JavaScript along with the D3 library to construct informative bar charts. Each bar chart includes labeled x-axis and y-axis, providing clear context, and a relevant title to enhance comprehension.

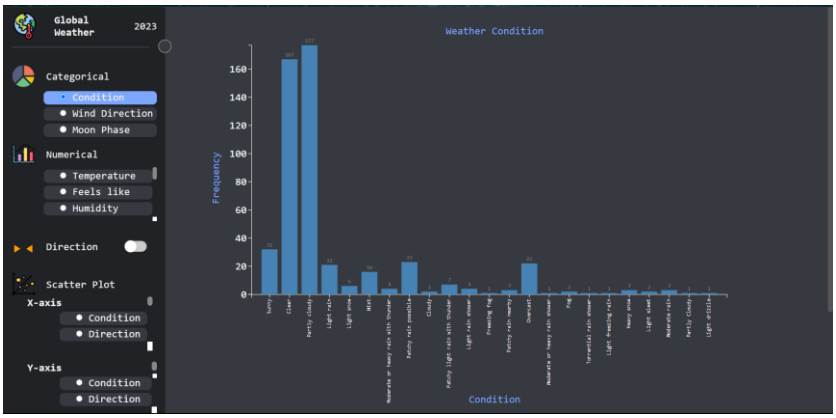

Figure 1. Categorical Data visualization

Transitioning to numerical data visualization, I employed the same tools to generate histograms. These histograms offer a comprehensive view of the distribution of numerical weather attributes, aiding in the identification of data patterns and variability.

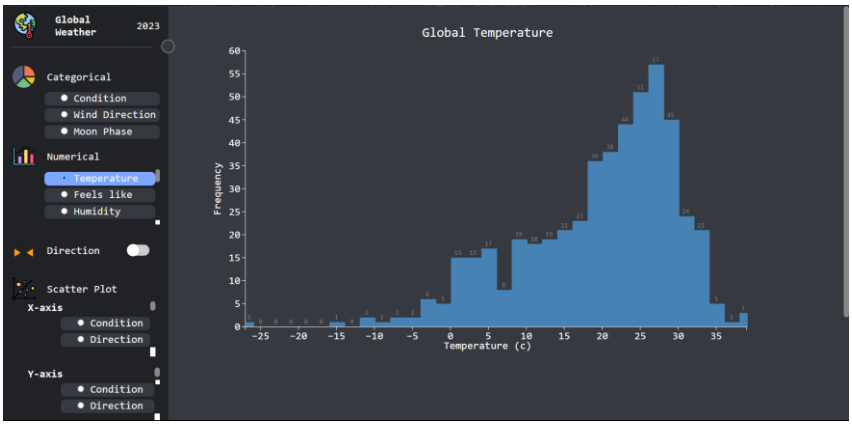

Figure 2. Numerical Data visualization

To explore relationships between different weather attributes, I implemented scatter plots. These scatter plots enable visual correlation analysis between numerical attributes and comparison of categorical and numerical data pairs.

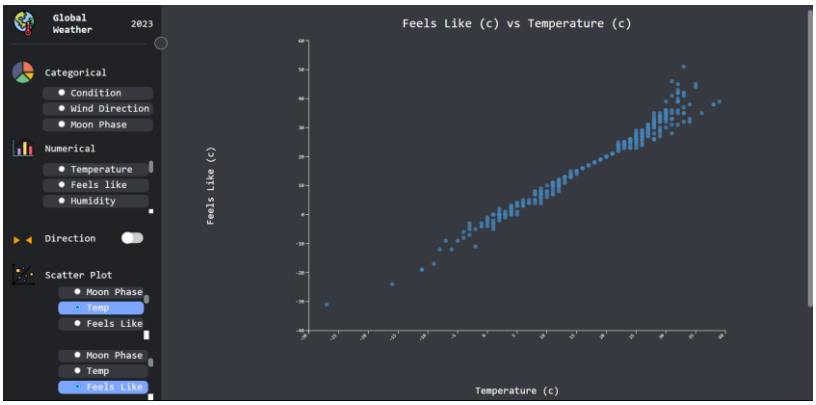

Figure 3. Scatter Plot Visualization

In the case of categorical-categorical comparisons, I creatively represented data points as mini-clusters, with each cluster comprising 25 points. This approach adds depth to the analysis, as denser clusters indicate higher frequencies within specific categories.

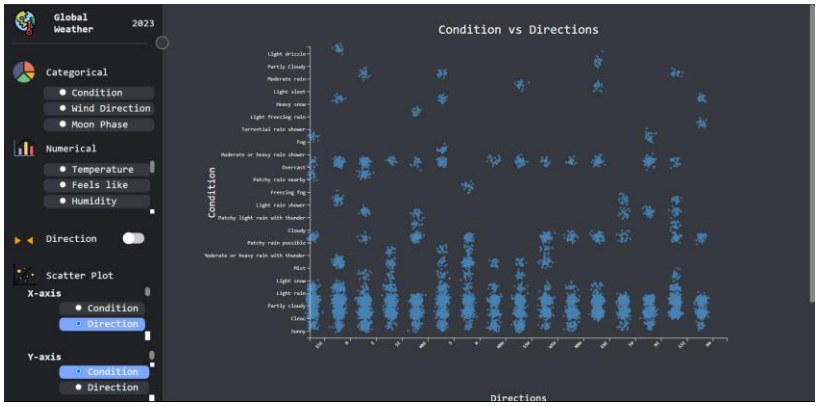

Figure 4. Scatter Plot (categorical vs categorical)

To ensure maintainability and modularity, I structured the codebase with separate functions for each visualization type: bar charts, histograms, and scatter plots. This organization enhances code clarity and facilitates future scalability and enhancements.

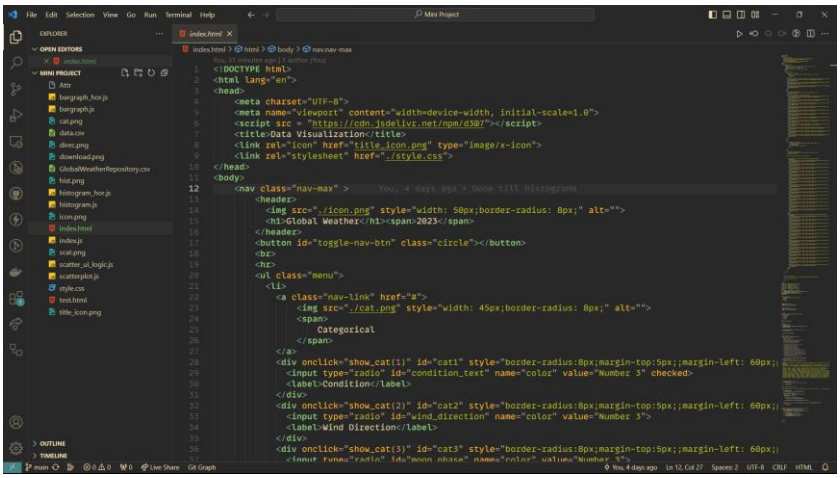

Figure 5. Code Base

Overall, the implemented visualization interface provides an intuitive and informative platform for exploring and understanding complex weather data. This approach allows for efficient data analysis and enables users to derive meaningful insights from the weather datasets.