Milestone 2: Core Analysis and Visualization Design

Introduction

This milestone focuses on analyzing the cleaned Global Weather Repository dataset to understand global weather behavior and to plan effective visualizations for an interactive dashboard. The aim is to study weather patterns, seasonal trends, regional variations, and extreme weather events in a simple and understandable way. The insights derived here will guide the dashboard development in the next milestone.

Statistical Analysis and Trends

Basic statistical analysis was performed on key numerical weather variables such as temperature, humidity, precipitation, and wind speed. Measures like mean, minimum, and maximum values were used to understand the general distribution of weather conditions across the dataset.

From this analysis, it was observed that temperature values vary significantly across regions, indicating strong geographic influence. Humidity levels are generally higher in coastal and tropical regions, while drier regions show lower humidity. Precipitation data is unevenly distributed, with certain regions receiving much higher rainfall compared to others. Wind speed also shows noticeable variation depending on location and terrain.

To analyze trends over time, the data was grouped by month using the date information. Monthly averages revealed clear seasonal patterns. In many regions, temperatures increase during mid-year months and decrease toward the beginning and end of the year. These seasonal trends help in understanding long-term climate behavior rather than short-term weather changes.

Correlation analysis was also performed to understand relationships between weather parameters. Temperature and humidity show a noticeable relationship in many regions, while precipitation tends to increase with higher humidity. Wind speed shows weaker correlation with other parameters, suggesting it is influenced more by geographic and atmospheric factors.

Identification of Extreme Weather Events

Extreme weather events were identified using percentile-based thresholds to keep the approach simple and consistent. Events falling within the top 5% of temperature values were considered extreme heat events. Similarly, very high precipitation values were treated as extreme rainfall events, and unusually high wind speeds were considered extreme wind events.

This method helped in isolating a small subset of data representing rare but significant weather conditions. These extreme events are important for understanding climate risks and anomalies. The extracted samples show that extreme heat events are more common in arid and desert regions, while extreme rainfall events are frequent in tropical and monsoon-affected areas. High wind events are often observed in coastal or open regions.

Regional Weather Comparison

To compare weather conditions across different regions, the data was grouped by country and average values of key weather parameters were calculated. This comparison highlights how weather behavior differs geographically.

Tropical regions generally show higher average temperatures and humidity levels. Temperate regions display stronger seasonal variation, with noticeable differences between summer and winter months. Some regions experience relatively stable weather throughout the year, while others show frequent extreme changes. These regional comparisons help in understanding global climate diversity and support meaningful geographic visualizations.

Selection of Visualization Types

Based on the analysis results, appropriate visualization types were selected to effectively represent the data and insights.

A **choropleth map** will be used to display average temperature or precipitation by country. This visualization allows users to easily compare weather conditions across different parts of the world.

A **line chart** will be used to represent monthly temperature trends. Line charts are well-suited for time-based data and clearly show seasonal patterns and long-term trends.

A **scatter plot** will be used to analyze the relationship between temperature and humidity. This helps in visually identifying correlations and understanding how changes in one variable affect another.

A **heatmap** will be used to display temperature variation across different months. Heatmaps are effective for highlighting seasonal intensity and recurring patterns in a compact visual form.

These visualization types were chosen because they are intuitive, widely used in data analysis, and suitable for the structure of the dataset.

Dashboard Layout Design (Wireframe Explanation)

The dashboard is planned to provide a clear and user-friendly flow from global overview to detailed insights. At the top of the dashboard, a metric selector will allow users to choose between temperature, humidity, precipitation, or wind speed. Filters for country and year will enable focused analysis.

The central area of the dashboard will feature a world map to show geographic patterns. Below the map, line charts and heatmaps will display trends and seasonal variations. Supporting charts such as scatter plots will help users explore relationships between weather parameters.

This layout ensures that users can quickly understand global patterns while also having the option to explore specific regions and time periods in more detail.

Conclusion

Milestone 2 successfully analyzed the cleaned weather dataset and identified key patterns, trends, and extreme events. The insights gained from statistical analysis and regional comparison provided a strong foundation for selecting suitable visualization types. A clear dashboard layout was also planned to effectively communicate these insights. This milestone prepares the project for the next phase, where the planned visualizations and interactivity will be implemented.