**Global Weather Data Visualization Project**

**1. Introduction**

This project explores global weather patterns using a dataset of capital cities worldwide. The dataset includes over 40 features, covering temperature, wind, pressure, precipitation, humidity, visibility, air quality, and more. The analysis focuses on the United States’ air quality, providing insights into trends and key factors affecting environmental conditions.

**2. Data Management**

2.1 Source Data

The dataset used in this project, GlobalWeatherRepository.csv, was sourced from kaggle repository. <https://www.kaggle.com/datasets/nelgiriyewithana/global-weather-repository/data>. The dataset is publicly available, and a cleaned version is included in the project repository.

2.2 Data Cleaning Process

The original dataset contained records from multiple countries; filtering was necessary to focus on US-based weather data. For Datetime conversion, the last\_updated field was transformed into a proper datetime format. For the selection of key features: Only relevant columns, including temperature\_celsius, wind\_kph, humidity, pressure\_mb, and air\_quality\_us-epa-index, were retained. In order to handle missing values, the dataset was checked for missing values, and no major gaps were found.

2.3 Sampling Methodology

As the dataset contains daily observations, no additional sampling was required. The analysis was limited to the US by filtering records based on the country column, ensuring relevant insights for air quality trends in the region.

**3. Methodology Explanation**

3.1 Visualization Selection Rationale

Line Chart (Temperature Trends) is selected to depict daily temperature variations in major cities, allowing for trend analysis over time.

A graph with colorful lines

AI-generated content may be incorrect.

Scatter Plot (Wind vs. Humidity) is chosen to explore correlations between wind speed, humidity, and air quality.

A grid with blue dots

AI-generated content may be incorrect.

Interactive Map (Air Quality Index) is used to visualize air pollution distribution across the US, leveraging color intensity and point size to indicate severity.

A map of the united states

AI-generated content may be incorrect.

3.2 Data Preparation Decisions

The dataset contained weather attributes from multiple time zones, requiring standardization of timestamps. Air quality data was selected as the primary focus, as it has direct environmental and health implications. Variables such as air\_quality\_PM2.5 and air\_quality\_us-epa-index were highlighted for their regulatory and scientific significance.

**4. Critical Analysis**

4.1 Self-Critique and Limitations

While comprehensive, the dataset lacks historical trends beyond the provided time frame (starting from August 29, 2023), limiting long-term climate analysis. The analysis does not incorporate external environmental factors such as wildfires or industrial emissions that may influence air quality. The dataset provides city-level observations, which may not capture microclimatic variations within states or specific urban areas.

4.2 Potential Improvements and Future Directions

Expanding the dataset with past weather data would enhance trend analysis. Future work could involve predictive modeling for air quality based on historical trends. Using satellite imagery and GIS integration to complement the weather data for a more granular air pollution assessment. Exploring relationships between air quality metrics and public health statistics, such as hospital admissions for respiratory conditions.

**5. Conclusion**

This project successfully visualized air quality trends, demonstrating the impact of wind and humidity on pollution levels. The combination of static and interactive visualizations allowed for a comprehensive understanding of weather patterns. Future enhancements could leverage predictive analytics and spatial data integration for deeper insights.