Weather Data Analysis Report

1. Introduction

This report provides a detailed analysis of global weather data, covering data cleaning, exploratory data analysis (EDA), forecasting models, and insights derived from the dataset. The goal is to understand weather patterns and make future predictions.

2. Data Cleaning & Preprocessing

- **Handling Missing Values:**

    - Checked for missing values and there weren't any.

- **Date Conversion:**

    - Converted the last_updated column to datetime format and sorted the data.

- **Normalization:**

    - Applied Min-Max Scaling to numerical columns for better model performance.

- **Outlier Detection & Removal:**

    - Used the IQR method to remove extreme values in temperature and precipitation. Also Used Z-score for analysis

3. Exploratory Data Analysis (EDA)

Correlation Matrix

- A heatmap was generated to visualize correlations among numerical variables.

- Key observation: Temperature and precipitation show minimal correlation.

Correlation Matrix
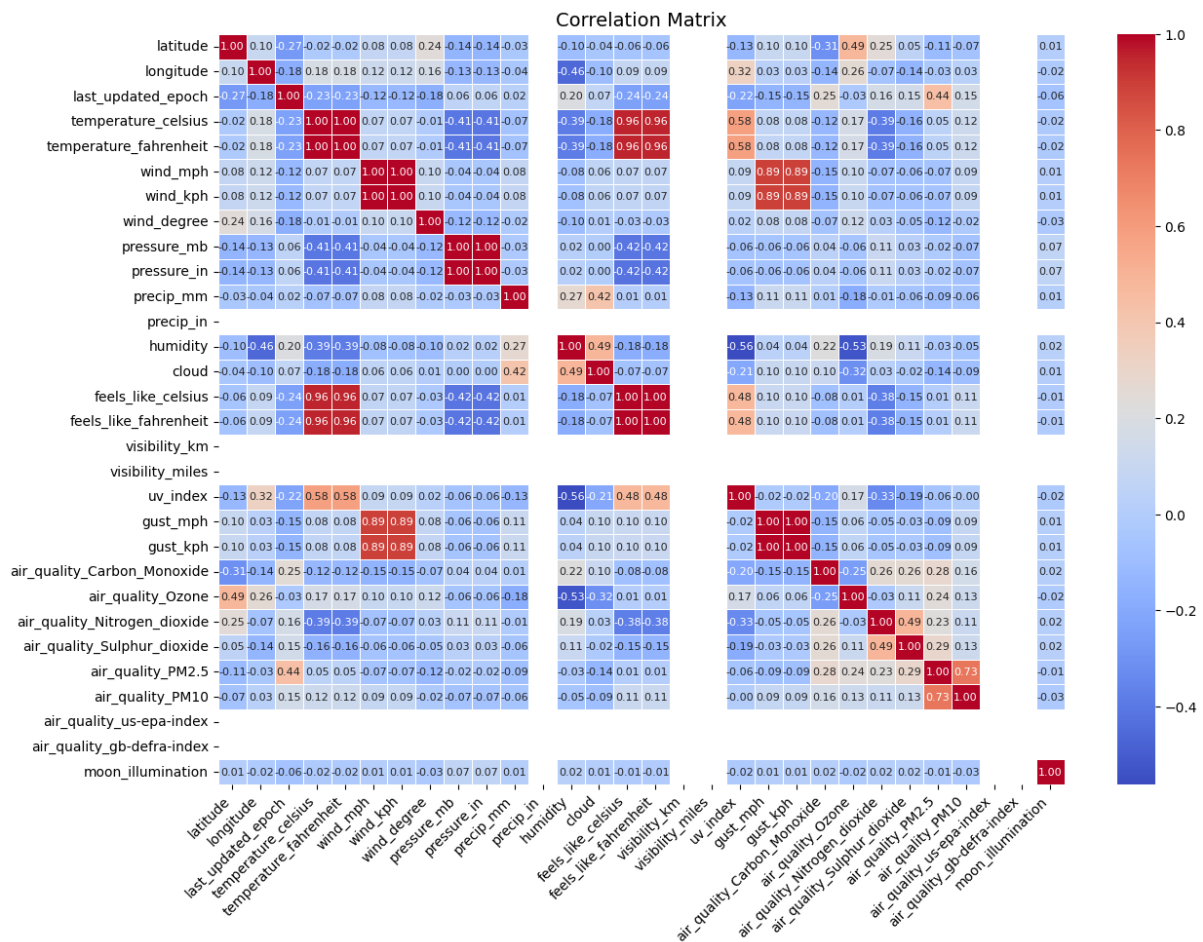
Data Distributions

- **Temperature Distribution:** A histogram and KDE plot show temperature trends.

- **Precipitation Distribution:** A histogram and KDE plot depict the spread of precipitation data.

- **Boxplots:** Used to detect and visualize outliers in the dataset.

Temperature Distribution
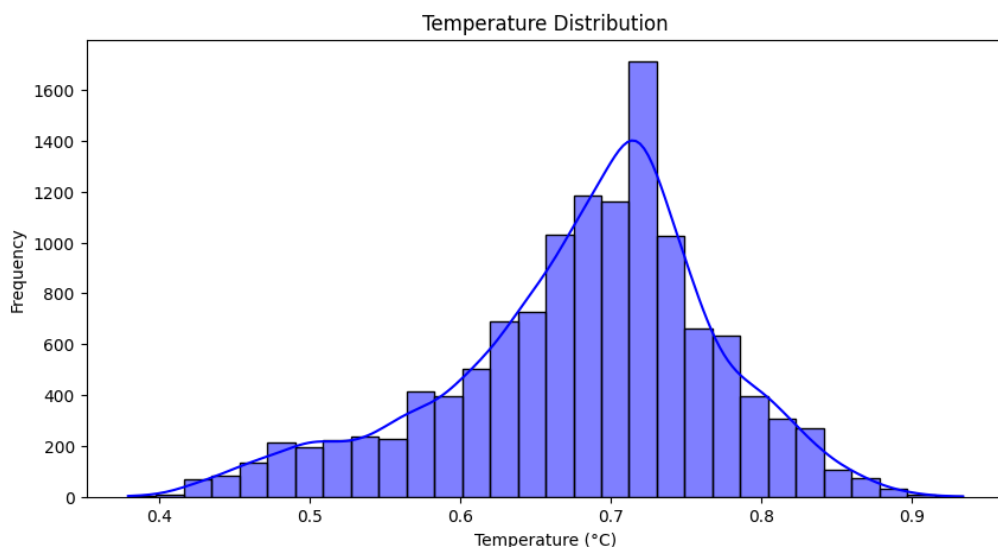
Time Series Trends

- **Temperature Over Time:** A line chart visualizing temperature variations over time.

- **Precipitation Over Time:** A line chart showing changes in precipitation patterns.
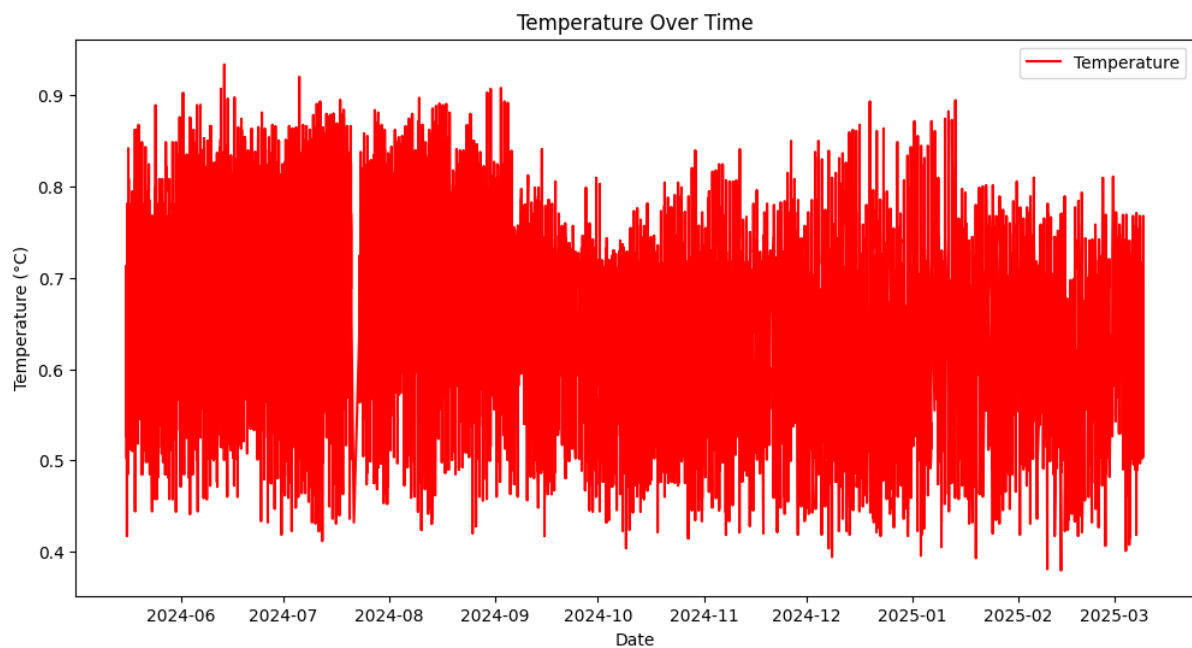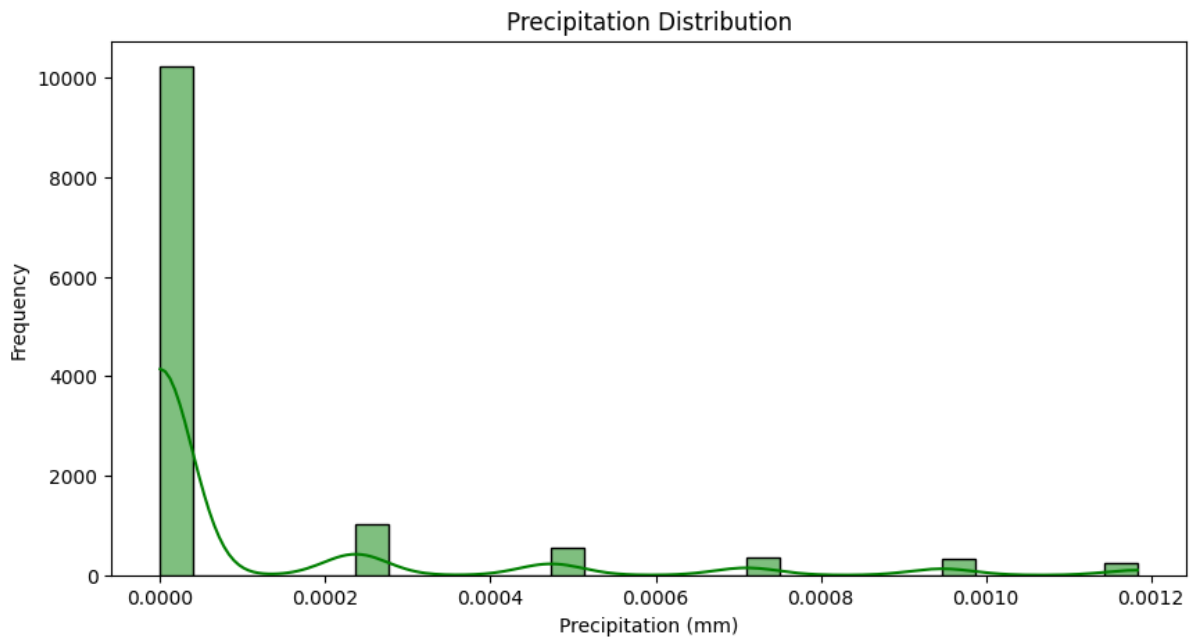
Anomaly Detection

- Outliers in key weather parameters such as temperature, humidity, wind speed, and pressure are detected using boxplots.

- Z-score-based filtering is used to highlight anomalies.

Visualization:

- Boxplot showing the distribution of weather parameters.

Boxplot of Weather Parameters
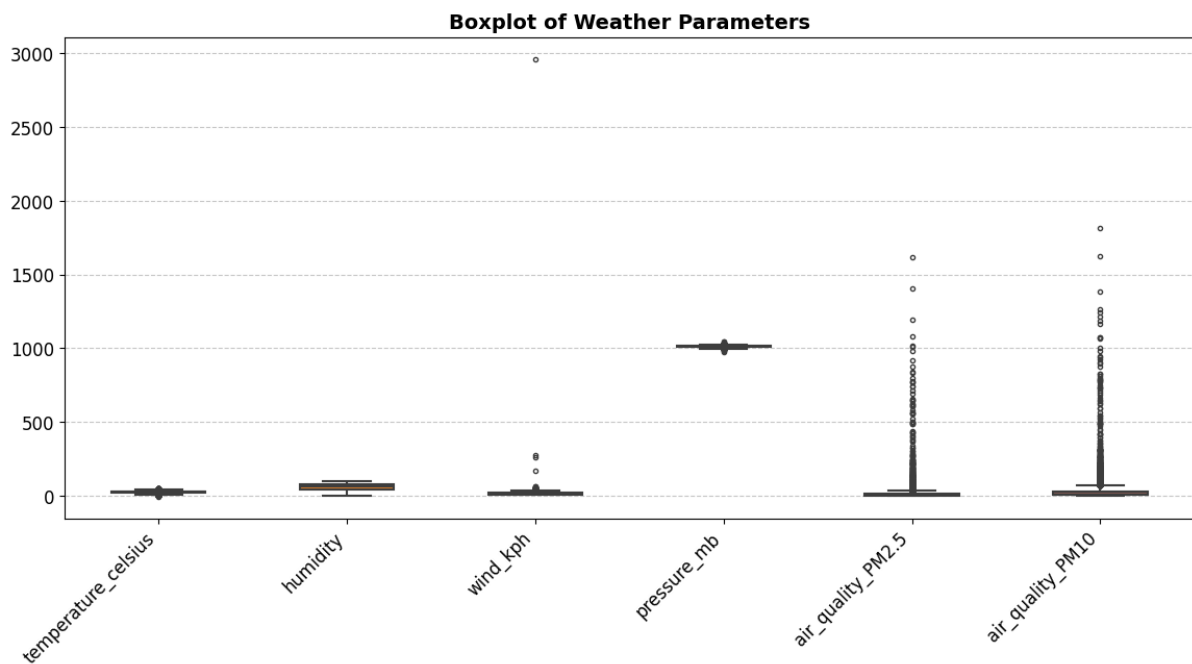
4. Forecasting Model

I)For Basic

- **Model Used:** ARIMA (AutoRegressive Integrated Moving Average)

- **Training & Testing:**

  o The dataset was split into 80% training and 20% testing.

  o ARIMA model was fitted on training data and tested on future temperature predictions.

- **Evaluation Metrics:**

  o Mean Absolute Error (MAE): 0.085

  o Root Mean Squared Error (RMSE): 0.101

II) For Advance

Model Selection

- Features: humidity, pressure_mb, wind_kph

- Target: temperature_celsius

- Model Used: **Random Forest Regressor**

Model Performance

- **Mean Absolute Error (MAE): 3.61444**

- **Mean Squared Error (MSE): 24.96483**

5. Advanced Insights

- **Temperature & Precipitation Trends:**
  - Observed seasonal variations in temperature and precipitation.

- **Potential Anomalies:**
  - Detected outliers that might indicate extreme weather events.

- **Recommendations:**
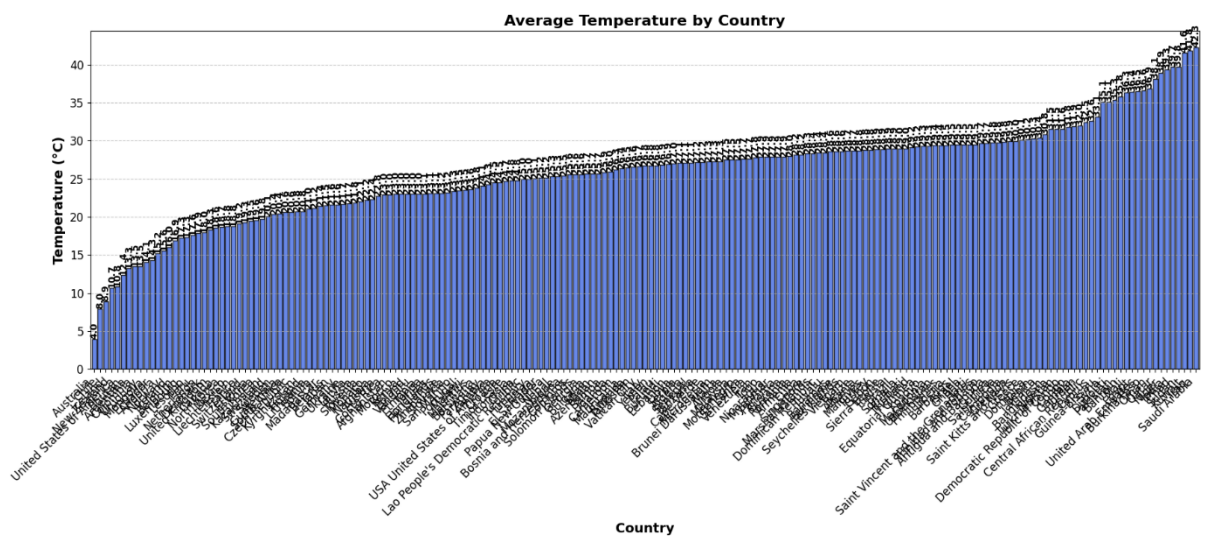 - Forecasting models can help in weather preparedness and resource allocation.

6. Advanced Analyses

Climate Trends

- The average temperature for each country is analyzed and visualized using a bar chart.
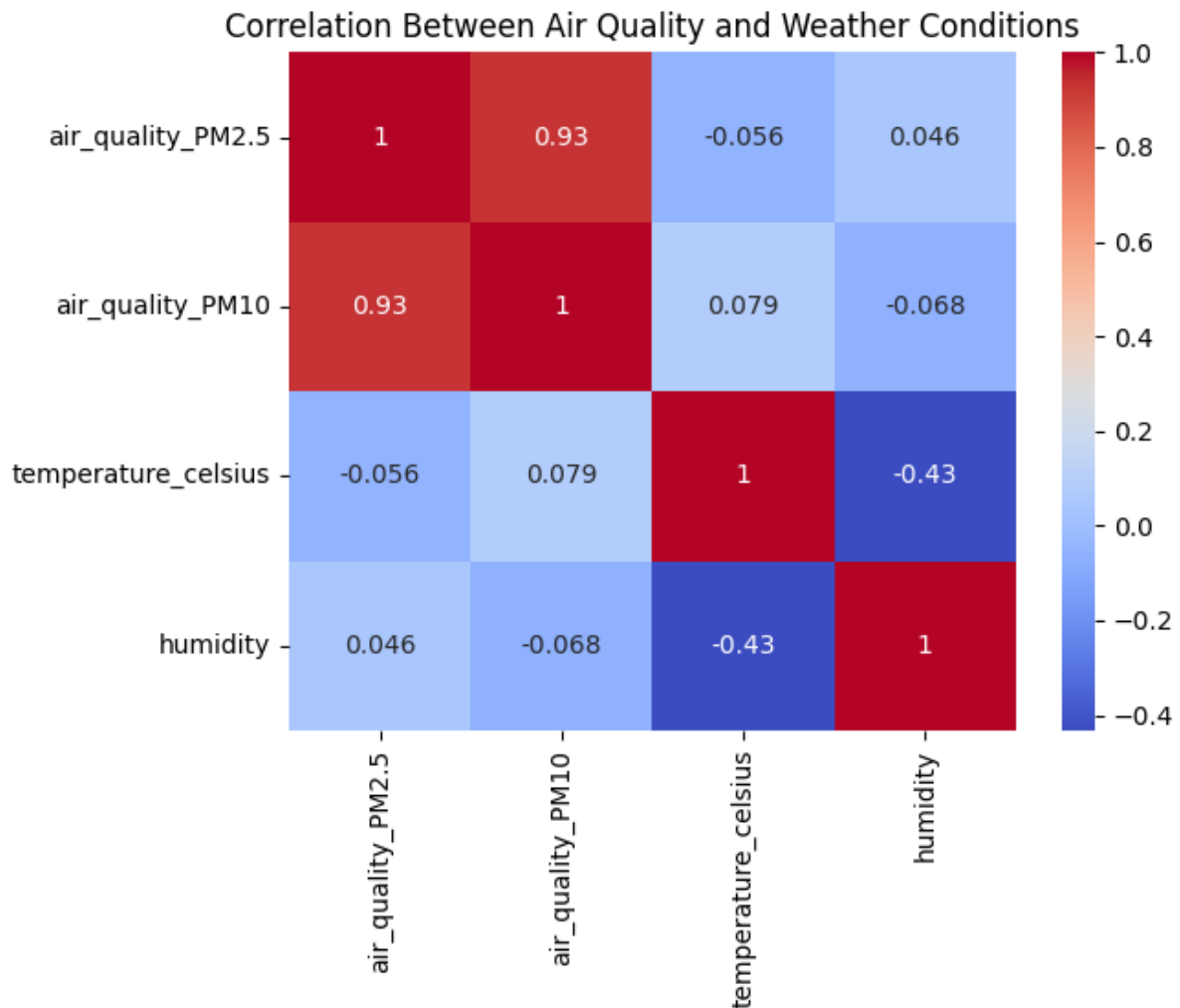
Visualization:

- Bar chart showing the average temperature by country.

Air Quality Impact

- Correlation between air quality (PM2.5, PM10) and weather conditions (temperature, humidity) is examined.

Correlation Between Air Quality and Weather Conditions

Visualization:

- Heatmap of correlations between air quality and weather parameters.

7. Conclusion

This report provides a comprehensive analysis of weather data, covering data preprocessing, visual exploration, forecasting, and insights. Future improvements can include integrating deep learning models for more accurate predictions.