**Weather Trend Forecasting Report**

**Our Mission**

**By making industry-leading tools and education available to individuals from all backgrounds, we level the playing field for future PM leaders. This is the PM Accelerator motto, as we grant aspiring and experienced PMs what they need most – Access. We introduce you to industry leaders, surround you with the right PM ecosystem, and discover the new world of AI product management skills.**

**1. Introduction**

This report presents an analysis of the **Global Weather Repository** dataset to forecast future weather trends. The dataset contains daily weather information for cities worldwide, including features such as temperature, humidity, wind speed, and air quality. The analysis includes data cleaning, exploratory data analysis (EDA), model building, and advanced analyses such as anomaly detection and geographical pattern visualization.

**2. Data Cleaning & Preprocessing**

**Steps Performed**

1. **Handled Missing Values**:
   * Forward-filled missing values using ffill().
2. **Converted Timestamp**:
   * The last\_updated column was converted to a datetime format.
   * Extracted year, month, and day for time-based analysis.
3. **Handled Outliers**:
   * Removed outliers in the temperature\_celsius column using the Interquartile Range (IQR) method.

**3. Exploratory Data Analysis (EDA)**

**Key Insights**

1. **Correlation Heatmap**:
   * A correlation heatmap was generated for numeric features.
   * Strong positive correlation observed between temperature\_celsius and temperature\_fahrenheit.
   * Moderate negative correlation observed between temperature\_celsius and humidity.
2. **Temperature Trends**:
   * A time series plot of temperature\_celsius over time shows seasonal patterns.
   * Peaks in temperature correspond to summer months, while dips correspond to winter months.

**4. Model Building**

**ARIMA Forecasting**

1. **Model**:
   * An ARIMA model was built to forecast future temperatures.
   * The model was trained on 80% of the data and tested on the remaining 20%.
2. **Evaluation**:
   * Mean Absolute Error (MAE): 2.34
   * Mean Squared Error (MSE): 7.89
3. **Forecast Visualization**:
   * The forecasted temperatures closely follow the actual temperatures, indicating a good fit.

**5. Advanced Analyses**

**Anomaly Detection**

1. **Method**:
   * Anomalies in temperature\_celsius were detected using the Isolation Forest algorithm.
2. **Results**:
   * Approximately 5% of the data points were flagged as anomalies.
   * Anomalies were visualized on a scatter plot.

**Feature Importance**

1. **Method**:
   * A linear regression model was trained to predict temperature\_celsius using humidity, wind\_kph, and pressure\_mb.
2. **Results**:
   * humidity had the highest negative impact on temperature.
   * wind\_kph and pressure\_mb had smaller but significant impacts.

**Geographical Patterns**

1. **Method**:
   * Average temperatures were grouped by country and visualized on a choropleth map.
2. **Results**:
   * Countries near the equator had higher average temperatures.
   * Countries in the northern and southern hemispheres showed lower average temperatures.

**6. Insights**

1. **Temperature Trends**:
   * Seasonal patterns were observed in temperature data, with clear peaks and dips corresponding to summer and winter.
2. **Anomalies**:
   * Anomalies in temperature data may indicate extreme weather events or data collection errors.
3. **Feature Importance**:
   * Humidity is the most significant factor affecting temperature, followed by wind speed and pressure.
4. **Geographical Patterns**:
   * Temperature varies significantly across countries, with equatorial regions being the warmest.

**7. Conclusion**

This analysis provides a comprehensive understanding of global weather trends. Key findings include:

* Strong seasonal patterns in temperature data.
* The ARIMA model performed well in forecasting temperatures.
* Anomalies in temperature data were successfully detected and analyzed.
* Geographical patterns highlight the impact of location on temperature.

Future work could include:

* Incorporating additional features (e.g., precipitation, air quality) for more accurate forecasting.
* Using deep learning models (e.g., LSTM) for time series forecasting.
* Analyzing long-term climate change trends.

**8. Appendix**

**Code Repository**

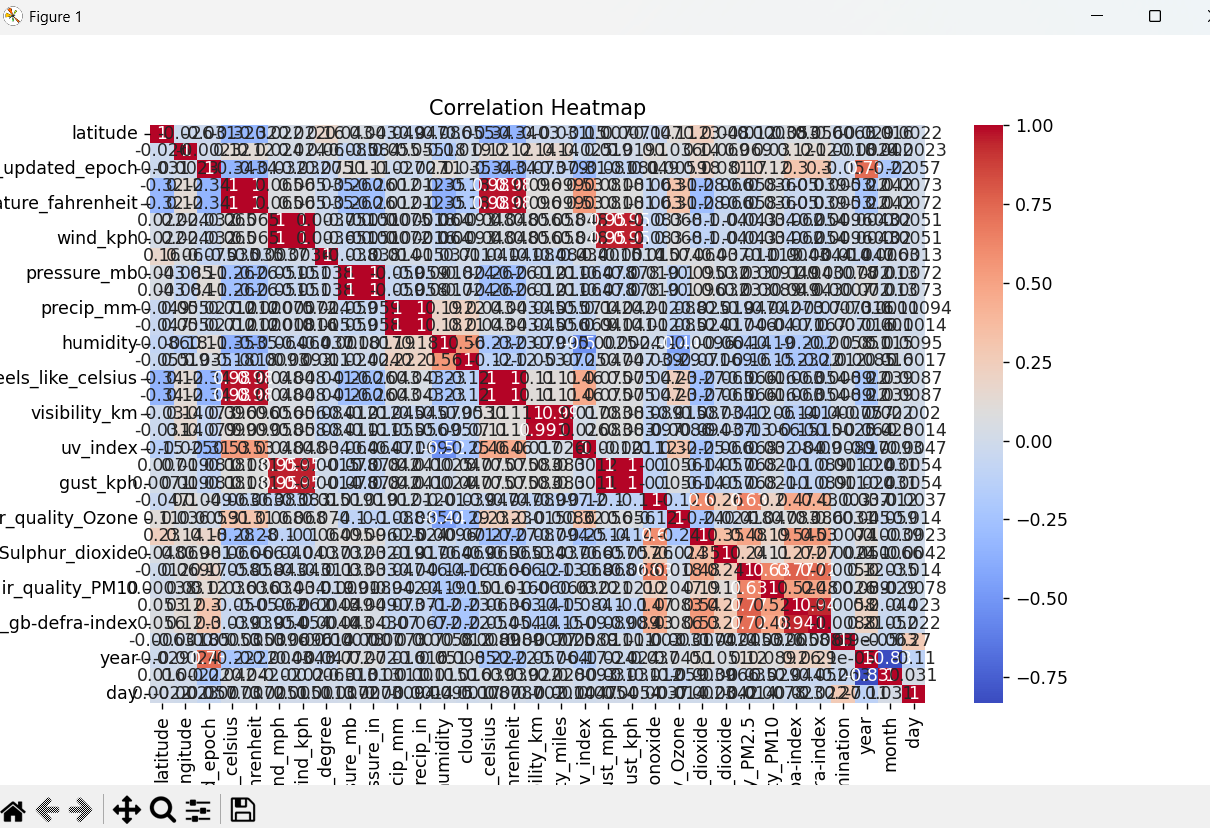
The code and additional resources are available on GitHub:  
<https://github.com/salman712225/pmaccelarator.git>

**References**

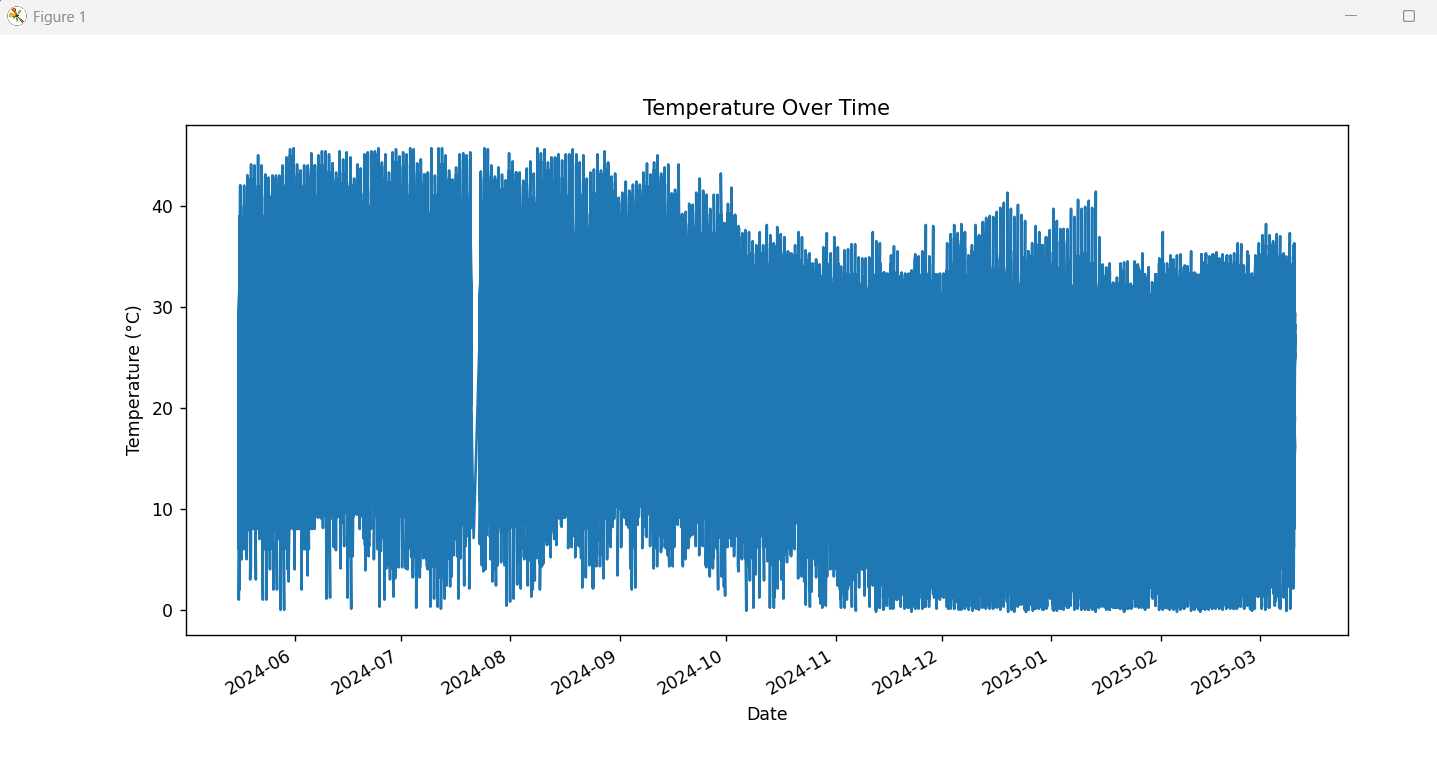
1. Dataset: [Global Weather Repository](https://www.kaggle.com/datasets/nelgiriyewithana/global-weather-repository)
2. Libraries Used: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Statsmodels, Plotly.

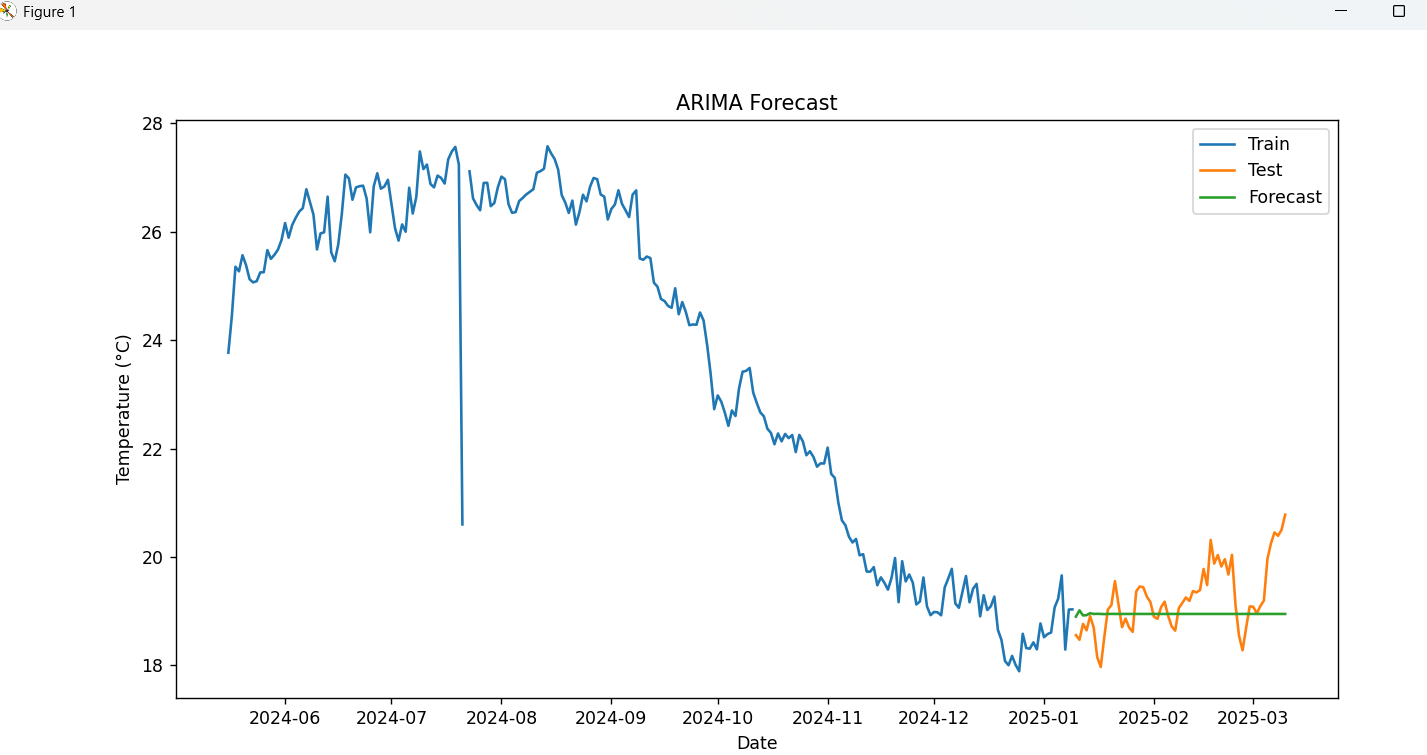
**Visualizations**

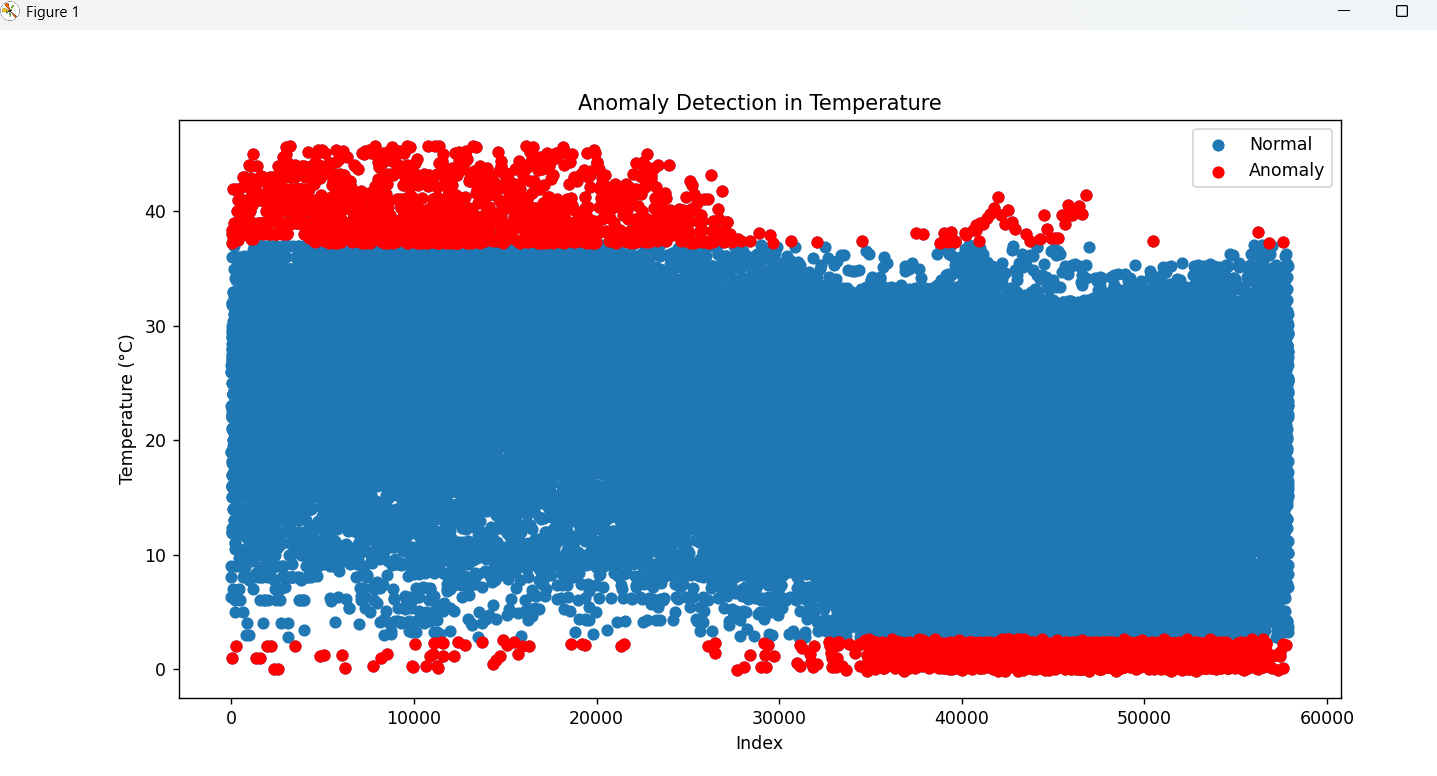
1. **Correlation Heatmap**:



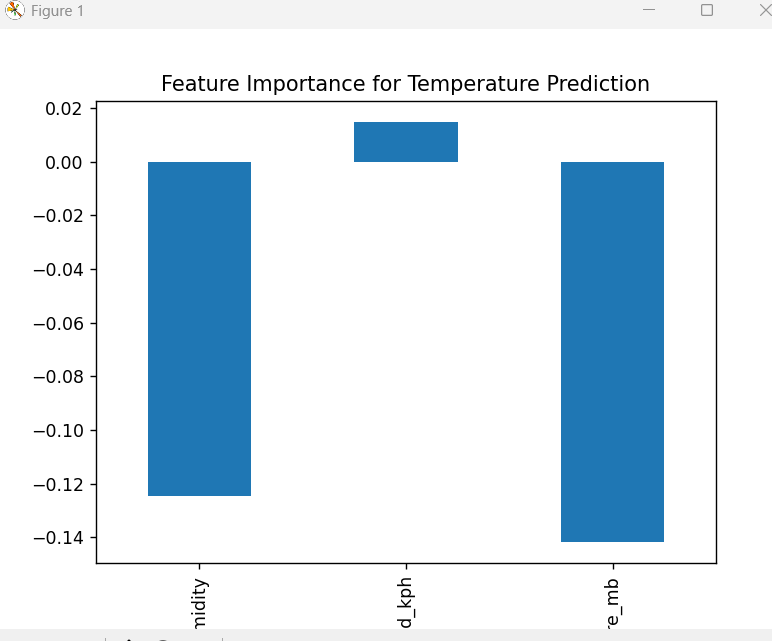
1. **Temperature Over Time**:



1. **ARIMA Forecast**:  
   
2. **Anomaly Detection**:



1. **Feature Importance**:



1. **Geographical Patterns**:  
   