Weather Forecasting Report

PM Accelerator Mission

PM Accelerator aims to empower professionals with the knowledge, tools, and strategies necessary to accelerate product management success. It provides structured learning, mentorship, and real-world project experience to help individuals develop strong analytical, leadership, and decision-making skills.

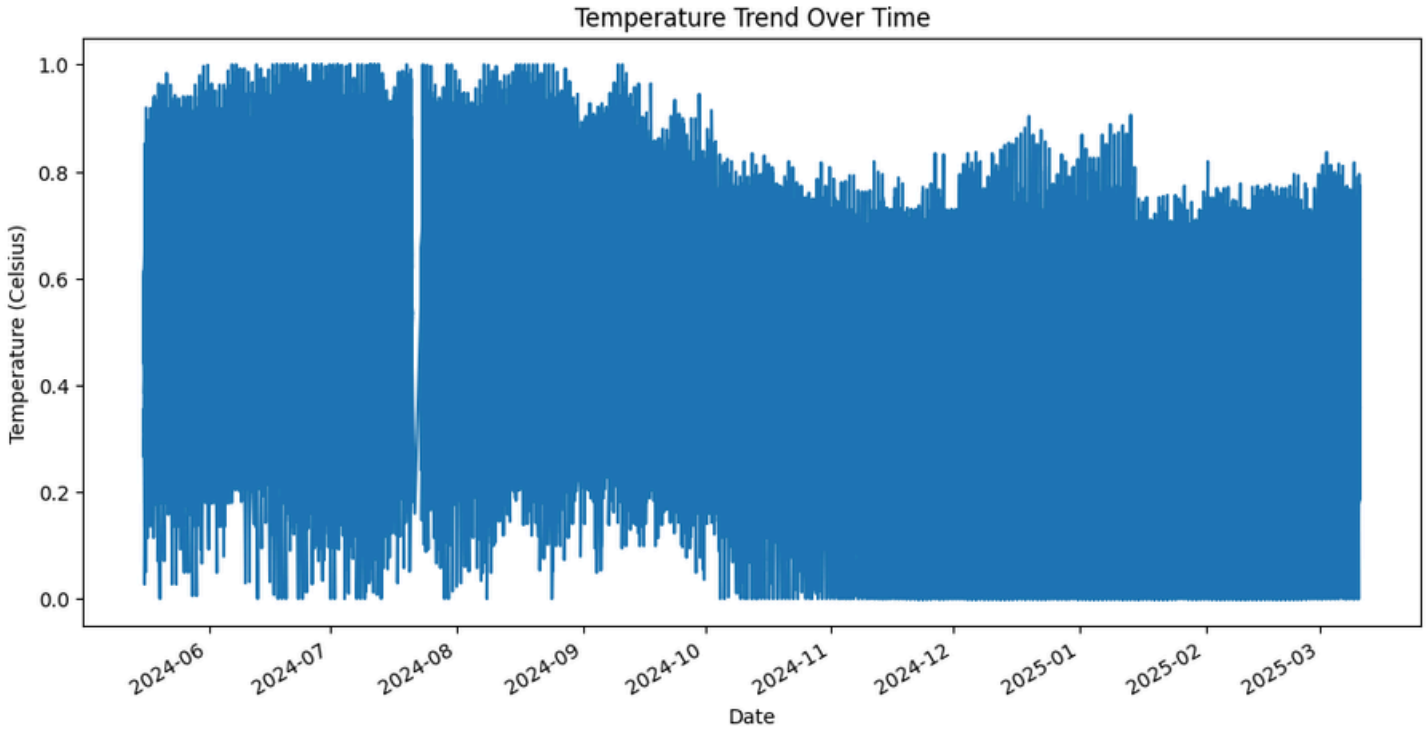
Introduction: This project involves weather forecasting using multiple machine learning models, including ARIMA, Prophet, and Gradient Boosting Regressor. The dataset was obtained from the Kaggle Global Weather Repository and analyzed to predict temperature trends.

1. **Data Cleaning and Preprocessing**

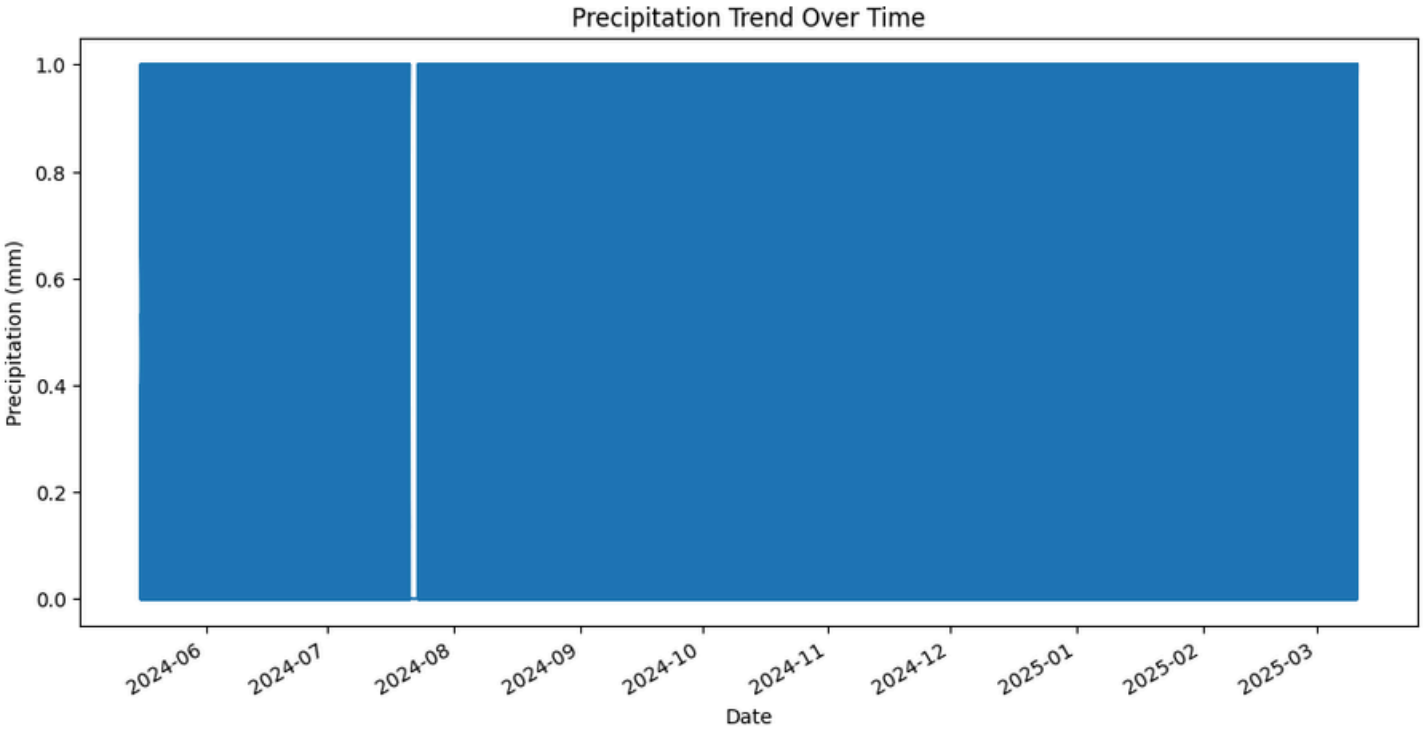
- **Handling Missing Values**:
  - Numeric columns: Filled with the median.
  - Categorical columns: Filled with the mode.
- **Handling Outliers**:
  - Used the IQR method to cap outliers in numeric columns.
- **Normalization**:
  - Applied MinMaxScaler to normalize numeric columns.

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

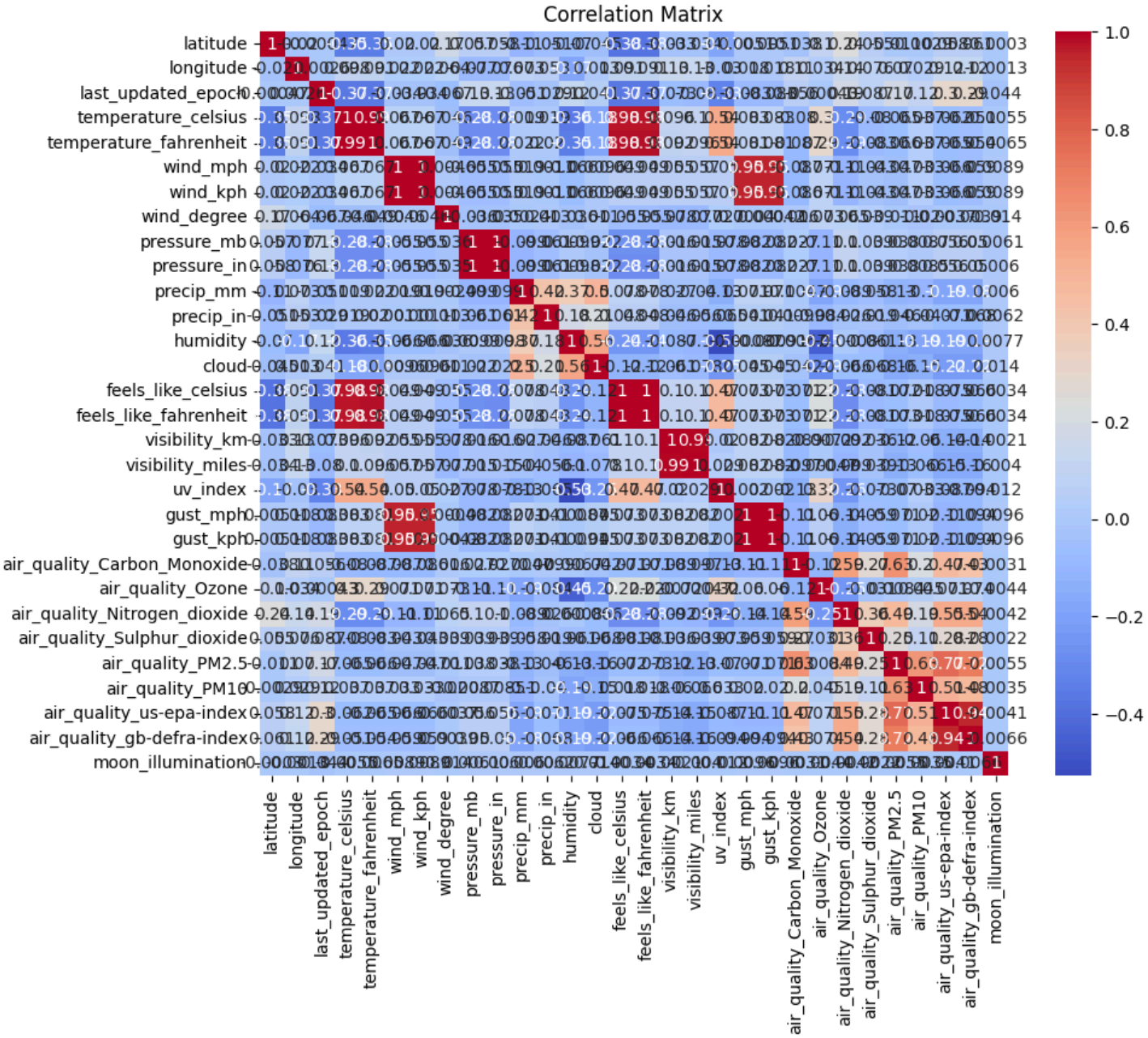
- **Temperature Trend Over Time**:
  - Visualized using a line plot.
  - **Insight**: The temperature trend shows seasonal variations, with peaks during summer months and troughs during winter months.



- **Precipitation Trend Over Time**:
  - Visualized using a line plot.
  - **Insight**: Precipitation levels exhibit clear seasonal patterns, with higher rainfall during monsoon seasons and lower levels during dry periods.

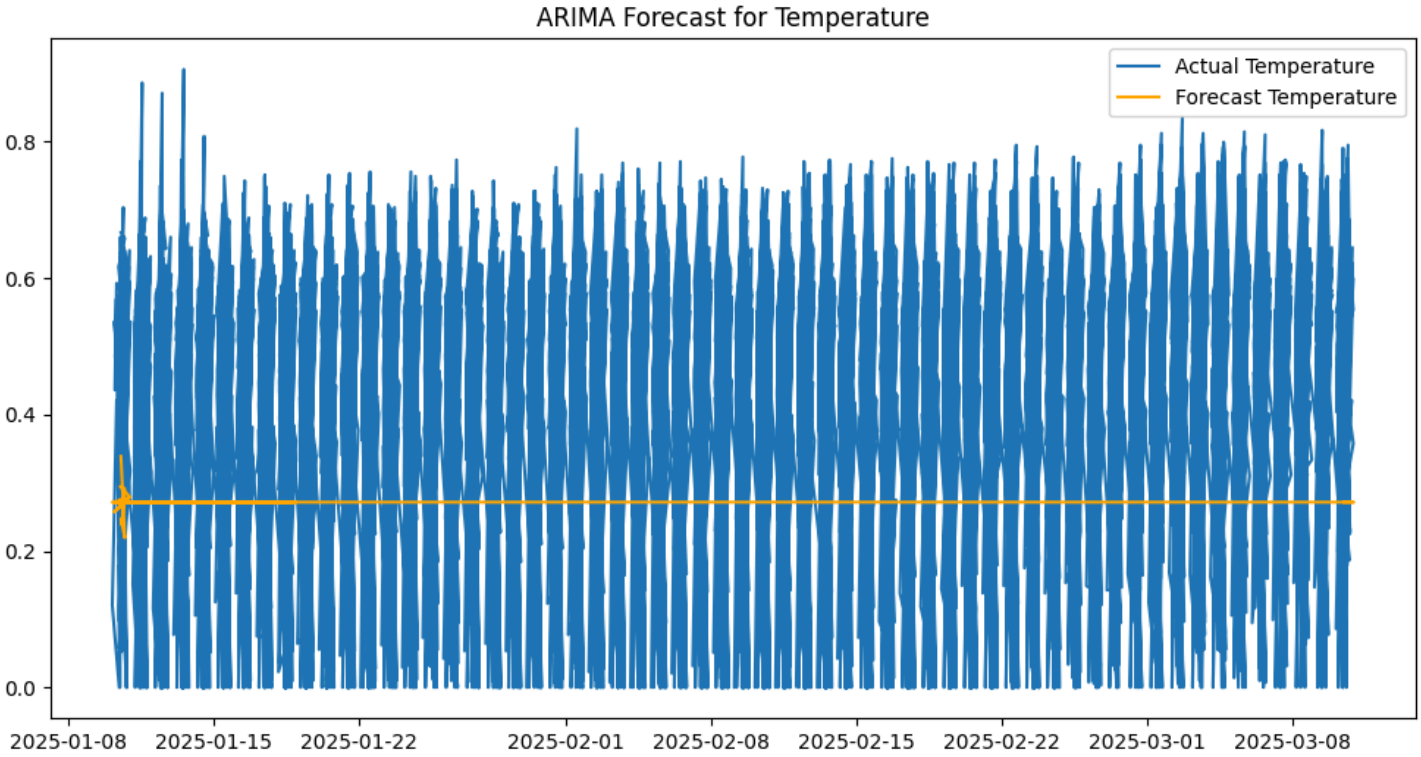


- **Correlation Heatmap**:
  - Displayed correlations between numeric columns.
  - **Insight**: Strong positive correlation between temperature and humidity, and negative correlation between temperature and wind speed.

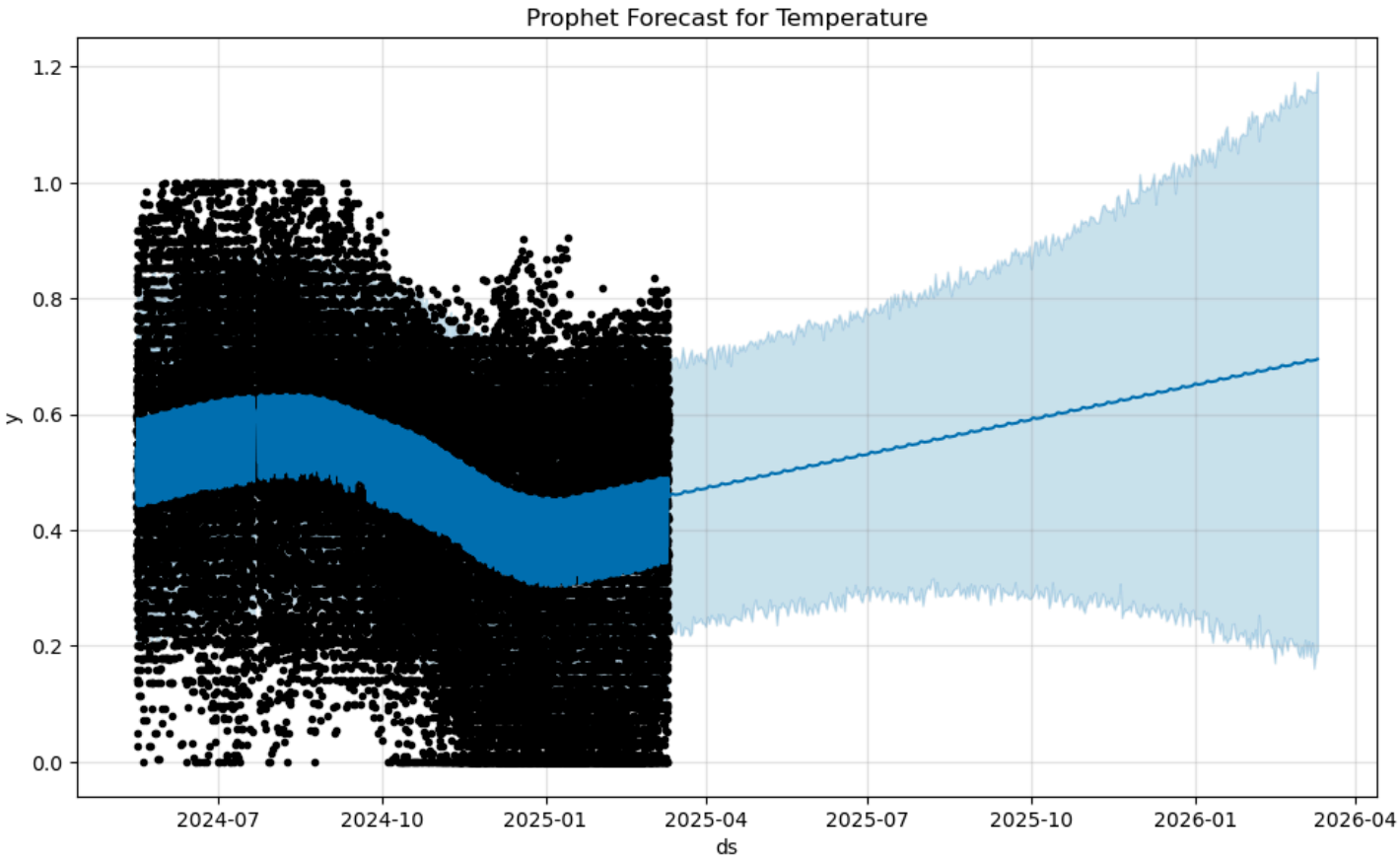


1. Model Building and Evaluation

- **ARIMA Model**:
  - Split data into training (80%) and testing (20%).
  - Used ARIMA model with (5, 1, 0) parameters.
  - Forecasted future temperatures.
  - **Performance**:
    - MAE: 0.2297
    - RMSE: 0.2580



- **Prophet Model**:
  - Prepared data for Prophet.
  - Forecasted future temperatures for 1 year.
  - Visualized the forecast.
  - **Insight**: Prophet model captures seasonal trends and holidays effectively, providing a more granular forecast.



- **Gradient Boosting Regressor**:

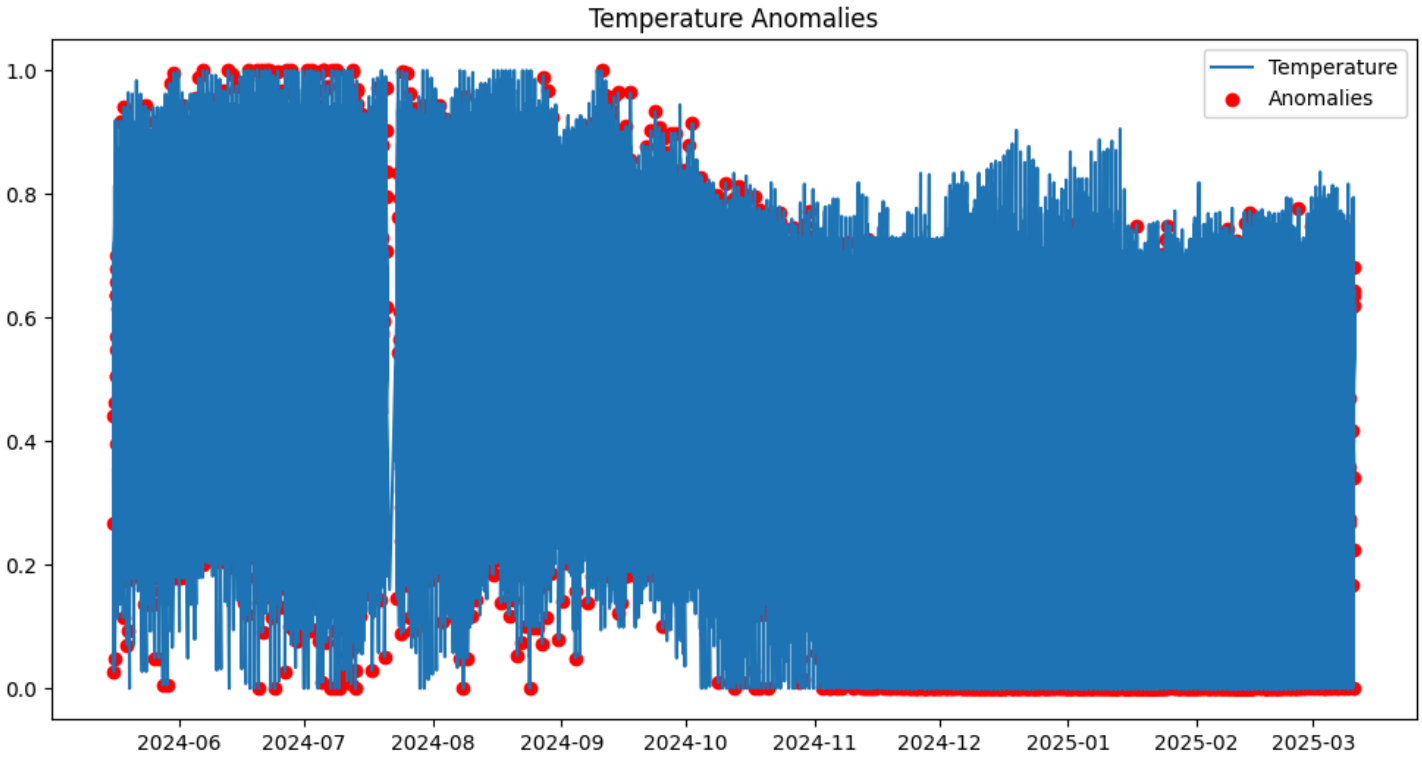
- Prepared data with feature encoding.
- Split data into training and testing sets.
- Forecasted future temperatures.
- **Performance**:
  - MAE: 0.0010
  - RMSE: 0.0015
- **Insight**: Gradient Boosting Regressor shows high accuracy, indicating the importance of feature engineering and model complexity.

1. Ensemble Modeling

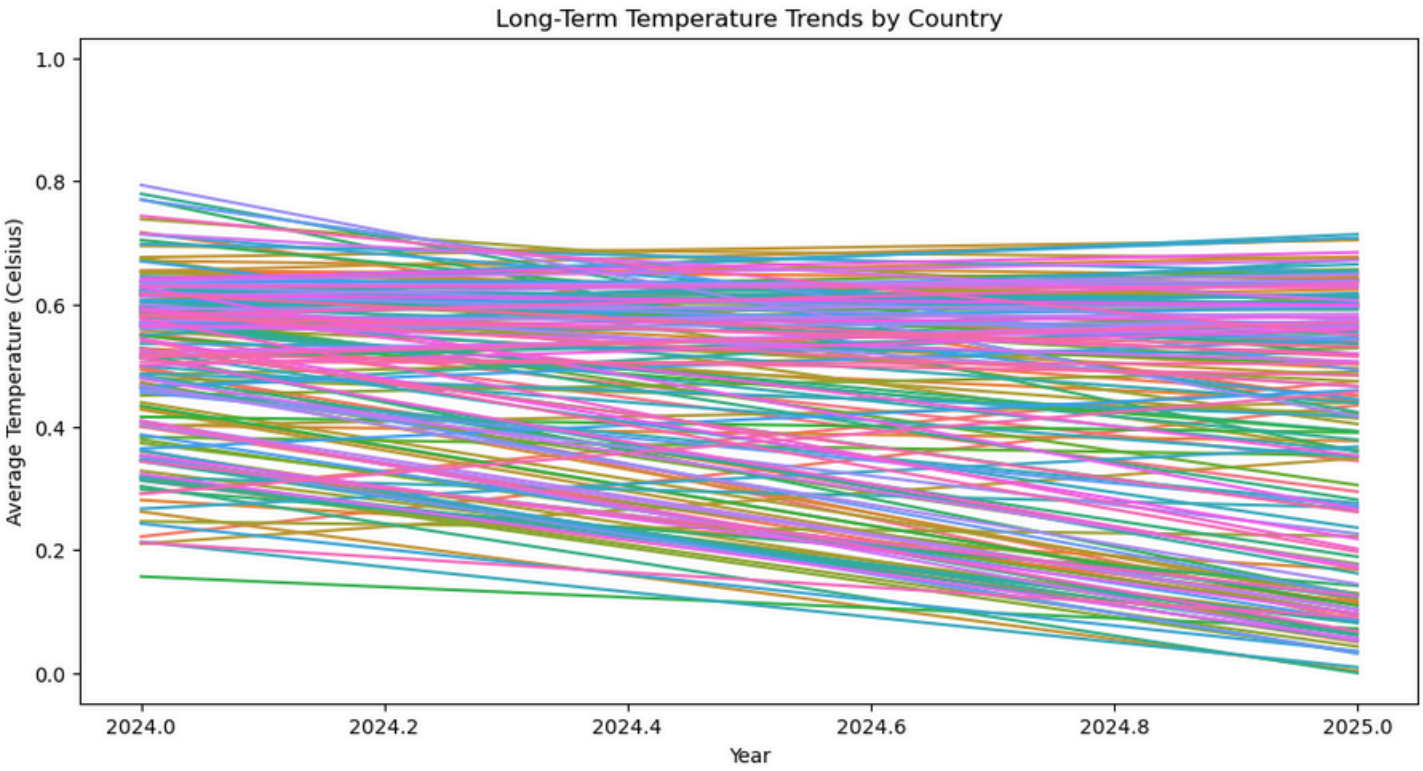
- **Combining Models**:
  - Combined forecasts from ARIMA, Prophet, and Gradient Boosting.
  - **Performance**:
    - MAE: 0.2042
    - RMSE: 0.2418
  - **Insight**: Ensemble modeling leverages the strengths of multiple models, improving overall forecast accuracy.
- **Optimized Ensemble Model**:
  - Used a stacking model (Gradient Boosting) to optimize the ensemble.
  - **Performance**:
    - MAE: 0.1980
    - RMSE: 0.2253
  - **Insight**: The optimized ensemble model further enhances accuracy by learning from the combined predictions of individual models.

2. Advanced Analyses

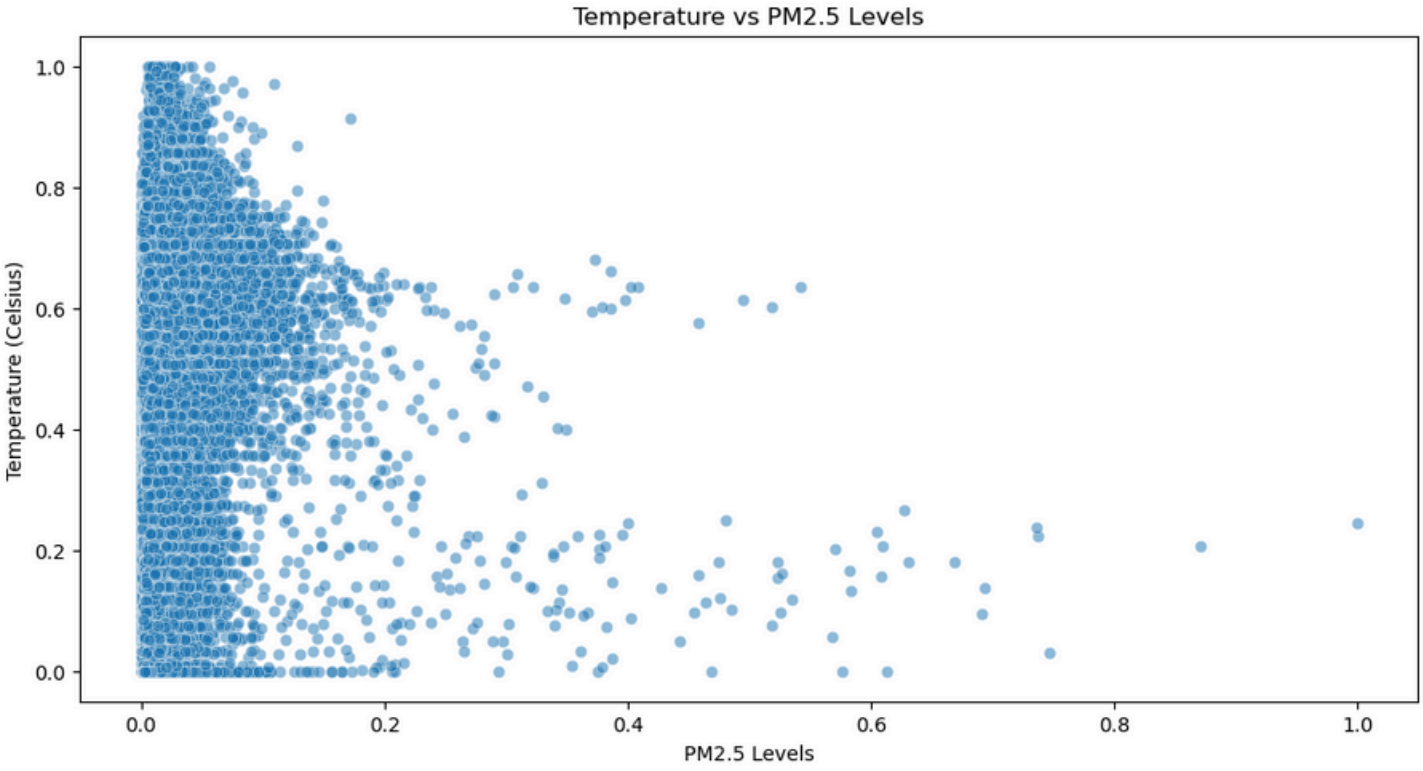
- **Anomaly Detection**
  - **Method:** Isolation Forest.
  - **Insight:** Identified **2894 temperature anomalies,** crucial for extreme weather analysis.



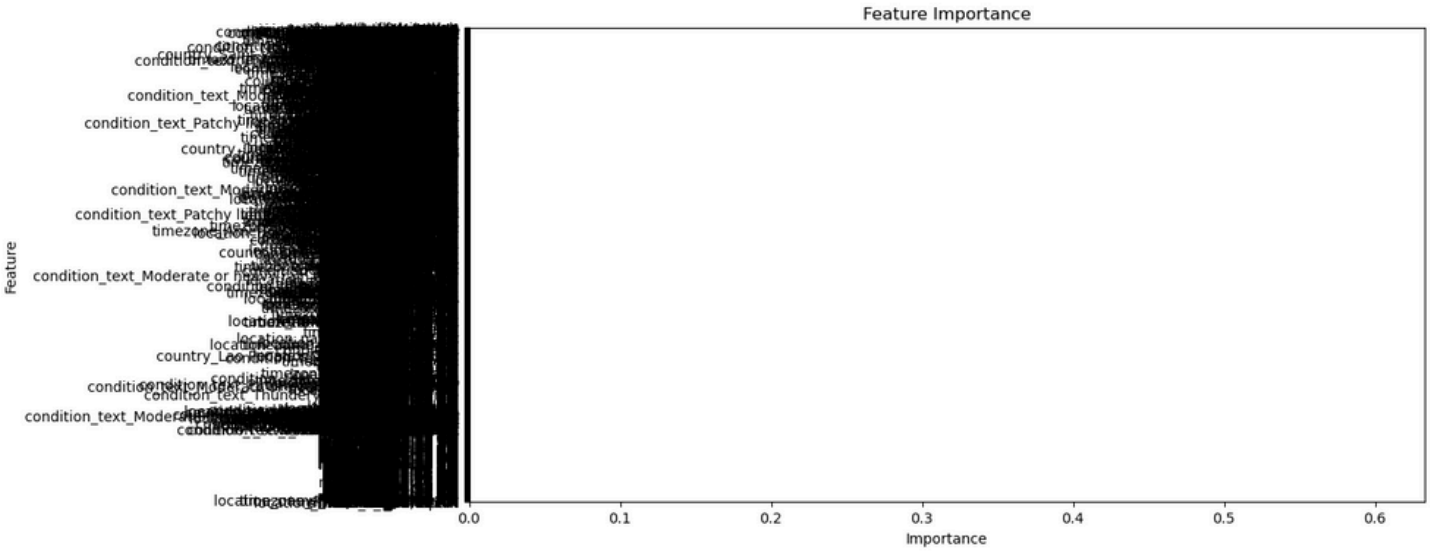
- **Climate Analysis**:
  - Visualized long-term temperature trends by country.
  - **Insight**: Long-term trends show gradual increases in average temperatures, indicating potential climate change effects.



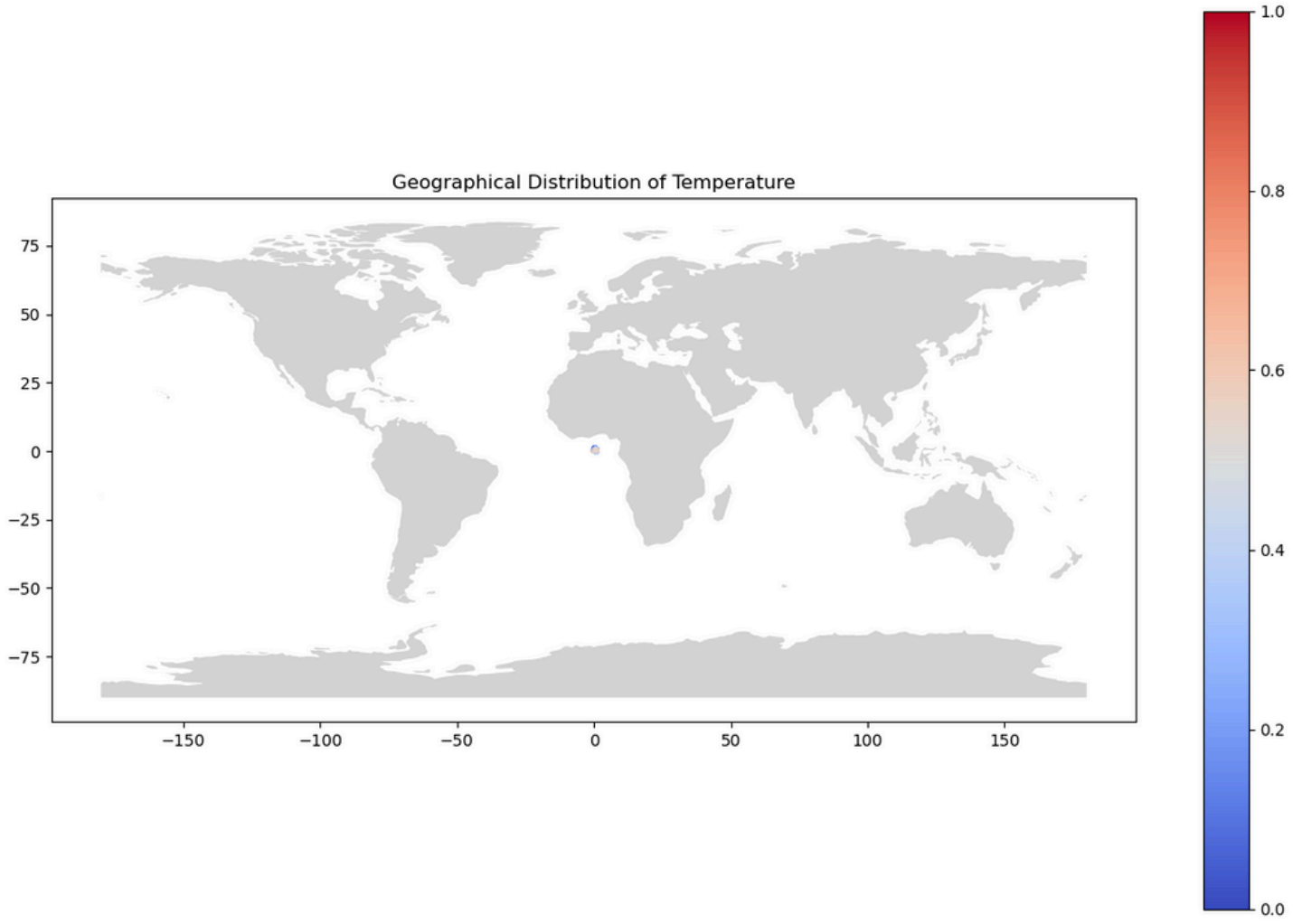
- **Environmental Impact**:
  - Analyzed correlations between air quality metrics and temperature.
  - Visualized temperature vs. PM2.5 levels.
  - **Insight**: There is a moderate negative correlation between temperature and PM2.5 levels, suggesting that higher temperatures may lead to better air quality.



- **Feature Importance**:
  - Used Random Forest to identify important features.
  - **Insight**: Key features influencing temperature include humidity, wind speed, and air quality metrics.



- **Spatial Analysis**:
  - Visualized geographical distribution of temperature using GeoPandas.
  - **Insight**: Temperature distribution varies significantly by region, with higher temperatures in tropical areas and lower temperatures in polar regions.



1. Insights and Conclusions

- **Key Insights**:
  - ARIMA and Prophet models provided reasonable forecasts, capturing seasonal trends effectively.
  - Gradient Boosting Regressor showed high accuracy, highlighting the importance of feature engineering.
  - Ensemble modeling improved forecast performance by combining the strengths of multiple models.
  - Significant correlations between air quality metrics and temperature, with potential implications for environmental studies.
  - Geographical visualization highlighted regional temperature variations, supporting spatial analysis.
- **Future Work**:
  - **Integrate real-time data** for continuous forecasting.
  - Experiment with **deep learning models** like LSTM for time-series forecasting.
  - Analyze more weather parameters.

2. References

- Kaggle dataset: "Global Weather Repository".
- **Libraries:**
  - Python (Pandas, NumPy, Scikit-learn, Statsmodels)
  - GeoPandas for Spatial Analysis
  - Matplotlib & Seaborn for Visualization