Global Weather Analysis and Forecasting Report

PM Accelerator 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.

Project Overview

This project focuses on analyzing and forecasting global weather trends using a comprehensive dataset of weather measurements from locations worldwide. The analysis includes:
Data Cleaning & Preprocessing: Handling missing values, removing duplicates, detecting and handling outliers.
Exploratory Data Analysis: Understanding weather patterns across regions and time periods.
Forecasting Models: Implementing and comparing multiple time series forecasting approaches.
Advanced Analysis: Examining climate patterns, air quality relationships, and geographical trends.

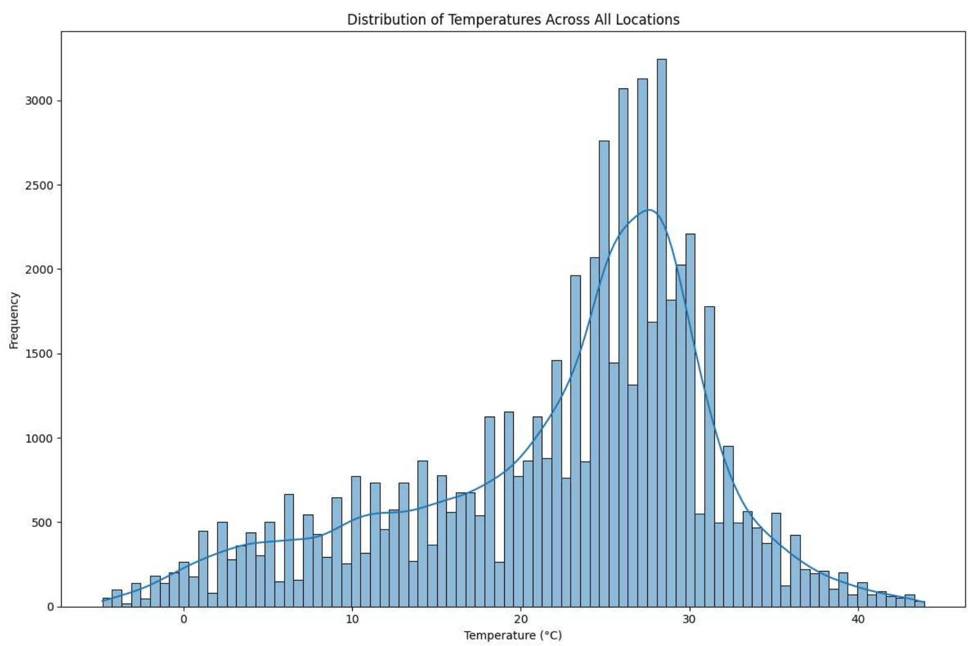
Data Processing and Cleaning

Cleaning Procedures

Missing Values: Categorical values filled with mode, numerical values with median.
Datetime Conversion: Transformed 'last_updated' into structured datetime components.
Duplicate Detection: Identified and removed duplicate entries.
Outlier Handling: Used Isolation Forest to detect and replace outliers in key weather parameters with median values.
Data Normalization: Applied StandardScaler to normalize numerical features for improved model performance.

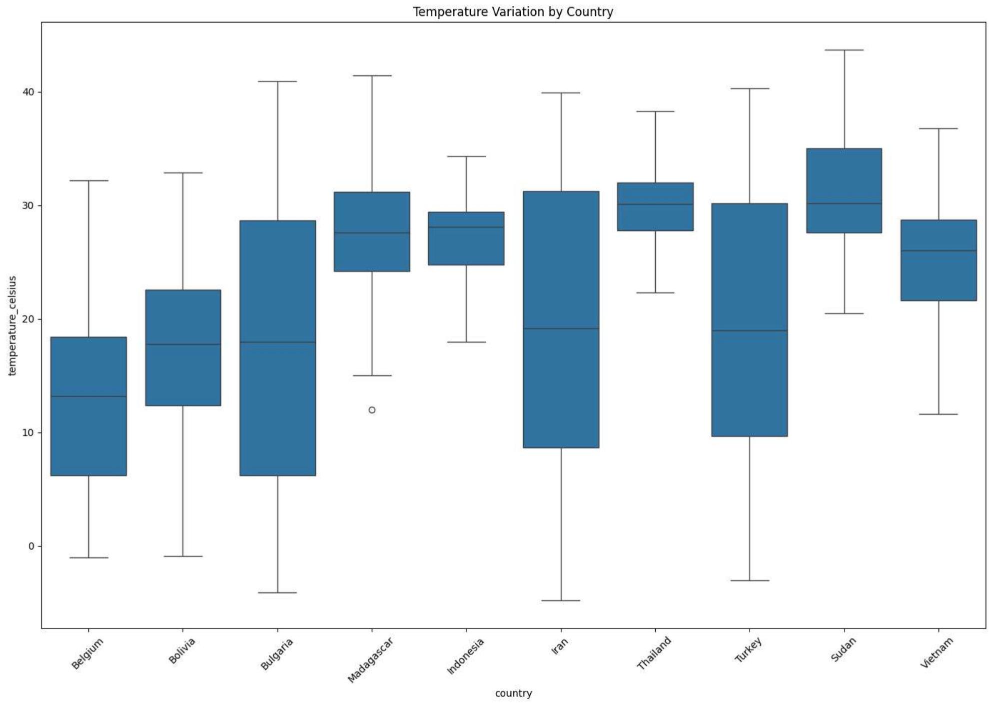
Exploratory Data Analysis

Temperature Distribution



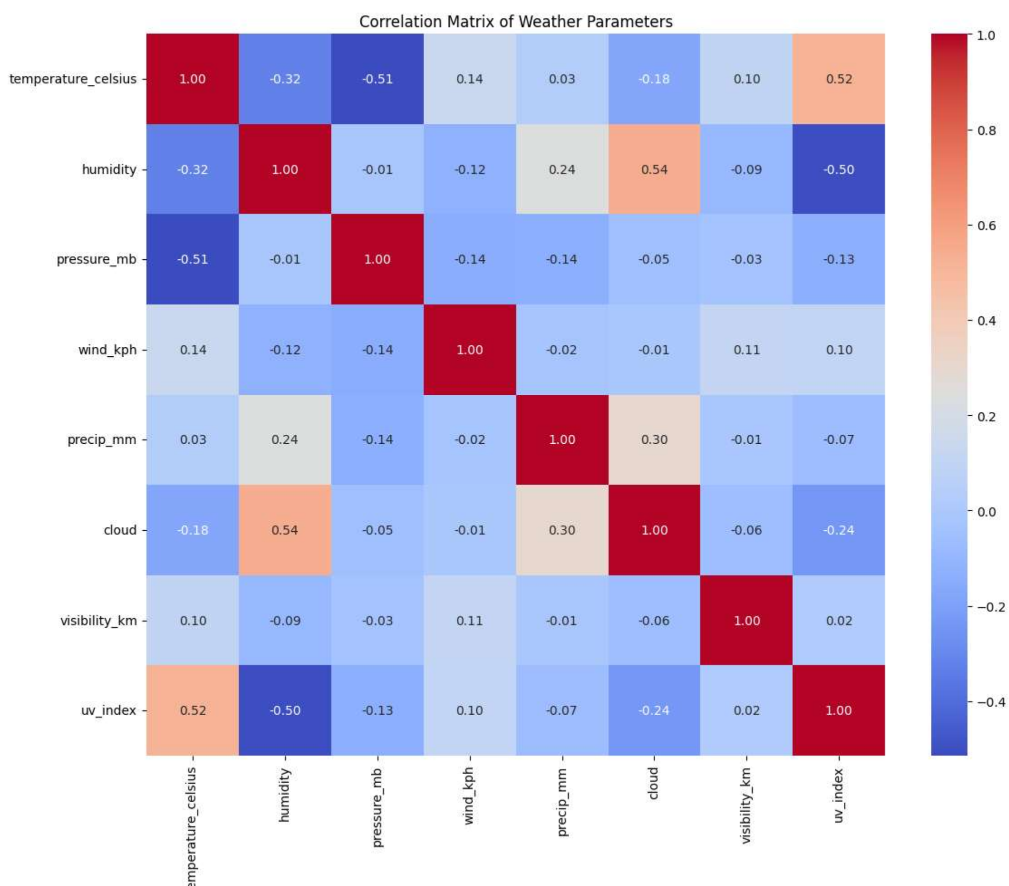
- Global temperature distribution follows a multi-modal pattern.
- Reflects the diversity of climate zones around the world.
- The distribution helps in understanding the range and variability of temperatures.

Temperature by Country



- Compared temperature variations across top 10 most frequent countries in the dataset.
- Highlighted regional differences in temperature ranges.

Weather Parameter Correlations

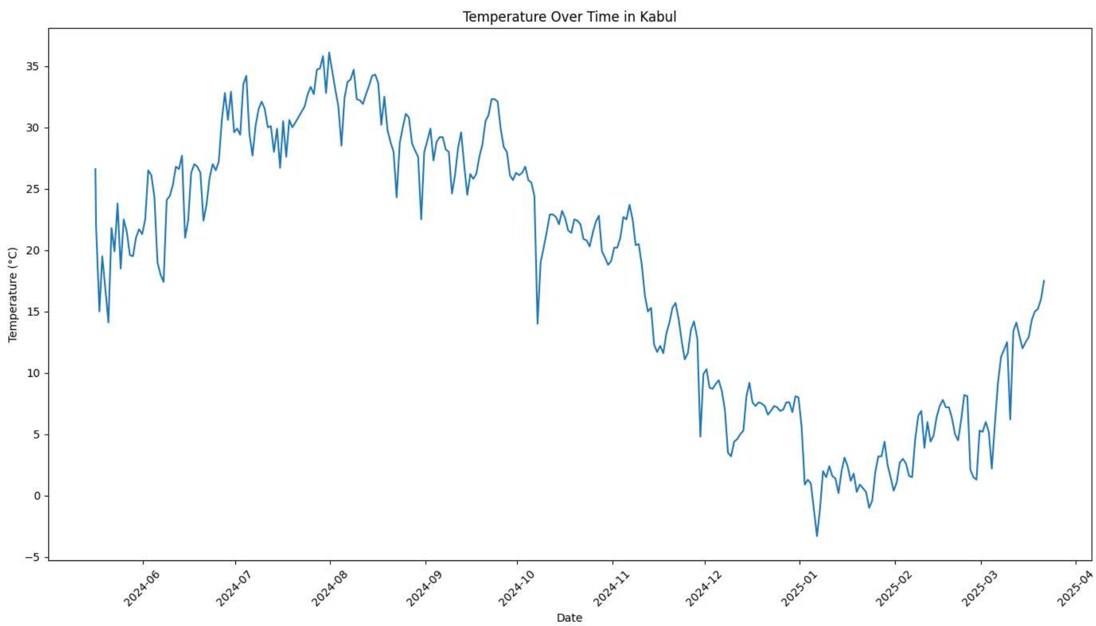


Discovered relationships between different weather parameters.

Key findings:

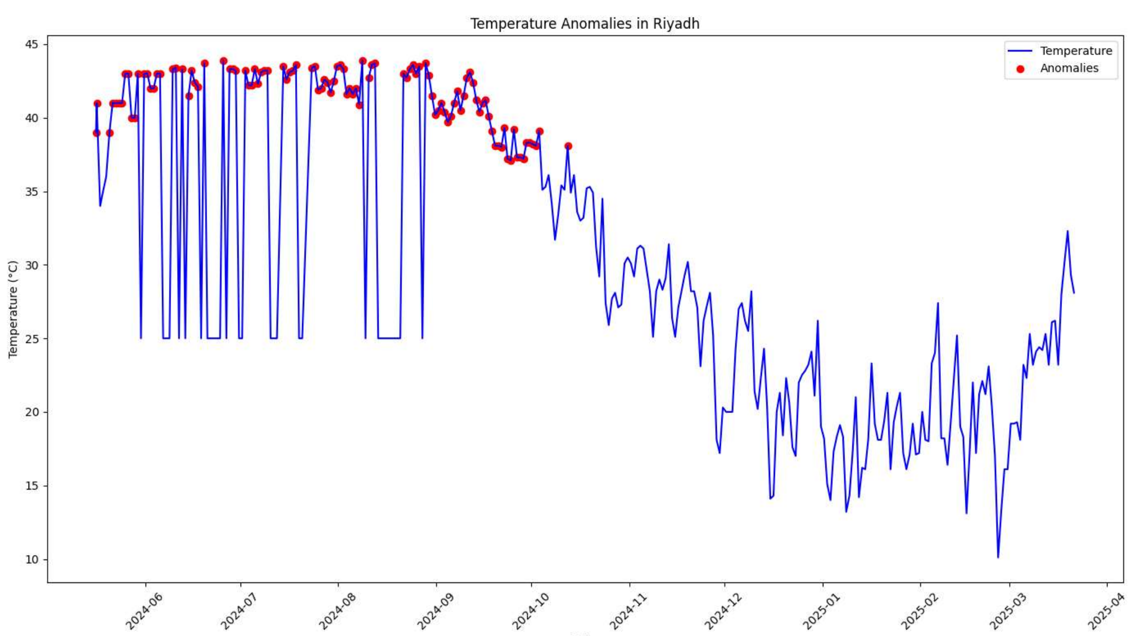
- Temperature shows strong negative correlation with humidity and pressure.
- Precipitation correlates with cloud cover and humidity.
- Pressure has inverse relationship with humidity and cloud cover.

Time Series Analysis



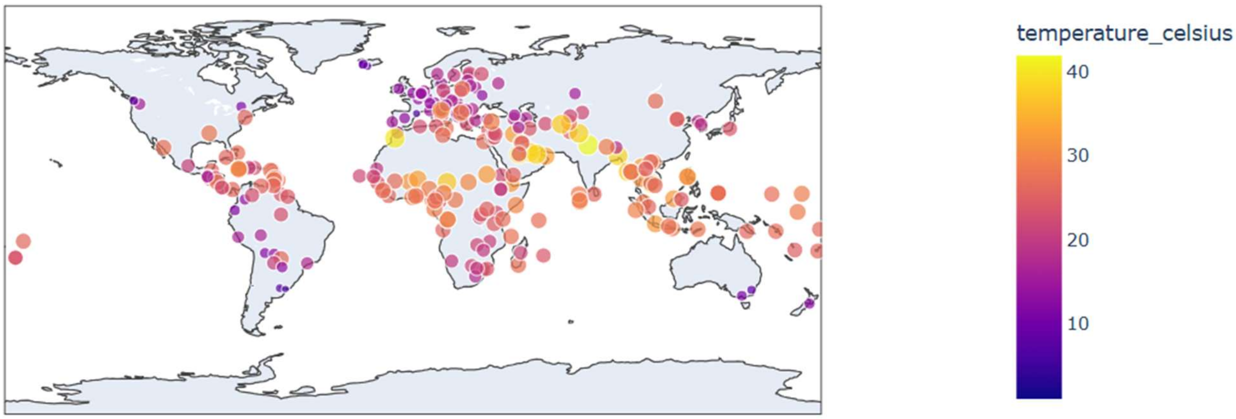
- Tracked temperature changes over time for a specific location (Kabul).
- Identified temporal patterns and seasonal variations.

Anomaly Detection



- Used Isolation Forest to identify anomalous temperature readings.
- Visualized temperature anomalies to understand unusual weather events.

Geographical Distribution



- Created interactive global temperature map to visualize spatial patterns.
- Implemented a temperature heatmap to identify global hotspots and cold regions.

Forecasting Models

Four forecasting models were developed to predict temperature trends, focusing on a single location (Kabul).

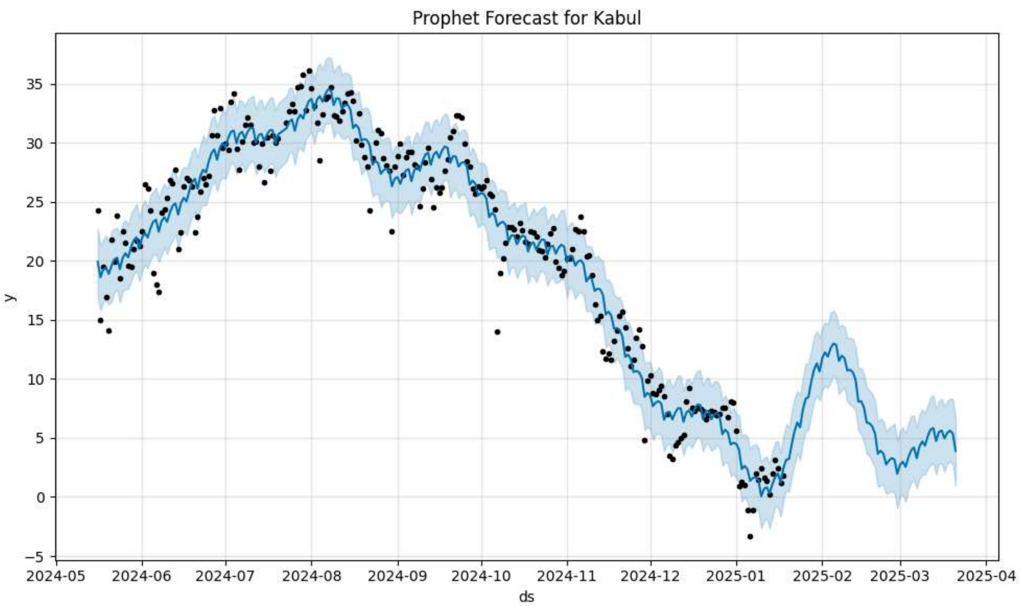
Model Development

Data Preparation: Resampled to daily averages, split into 80% training and 20% testing sets.

Prophet Model

Incorporated daily, weekly, and yearly seasonality.

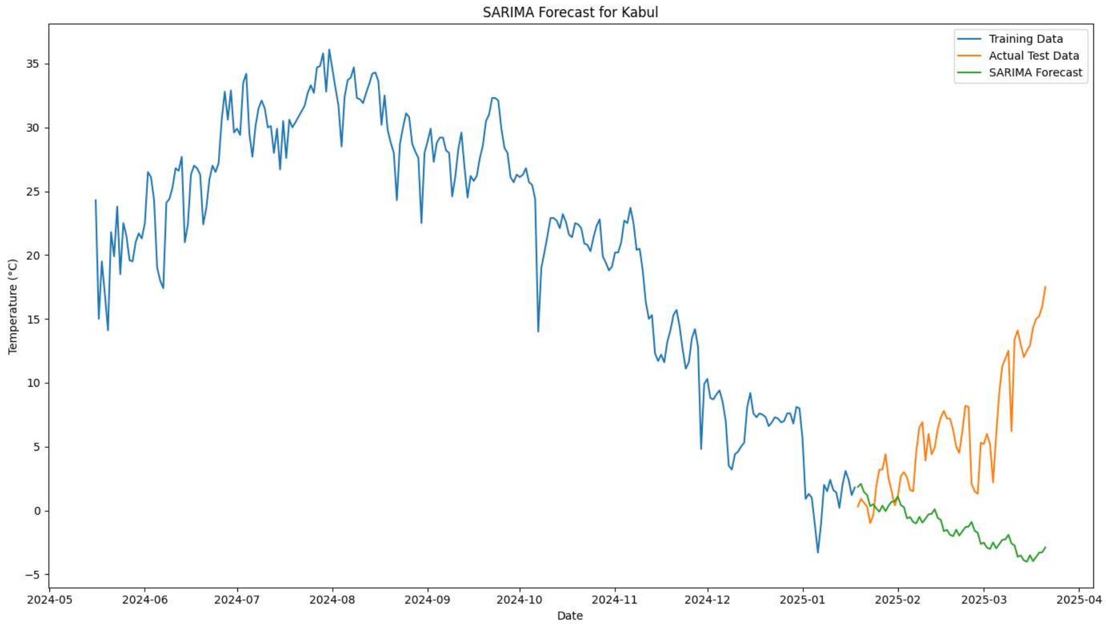
Evaluation metrics
Prophet RMSE: 6.5246
Prophet MAE: 5.5669
Prophet forecast



SARIMA Model

Applied seasonal autoregressive integrated moving average model with parameters (1,1,1)(1,1,1,12).

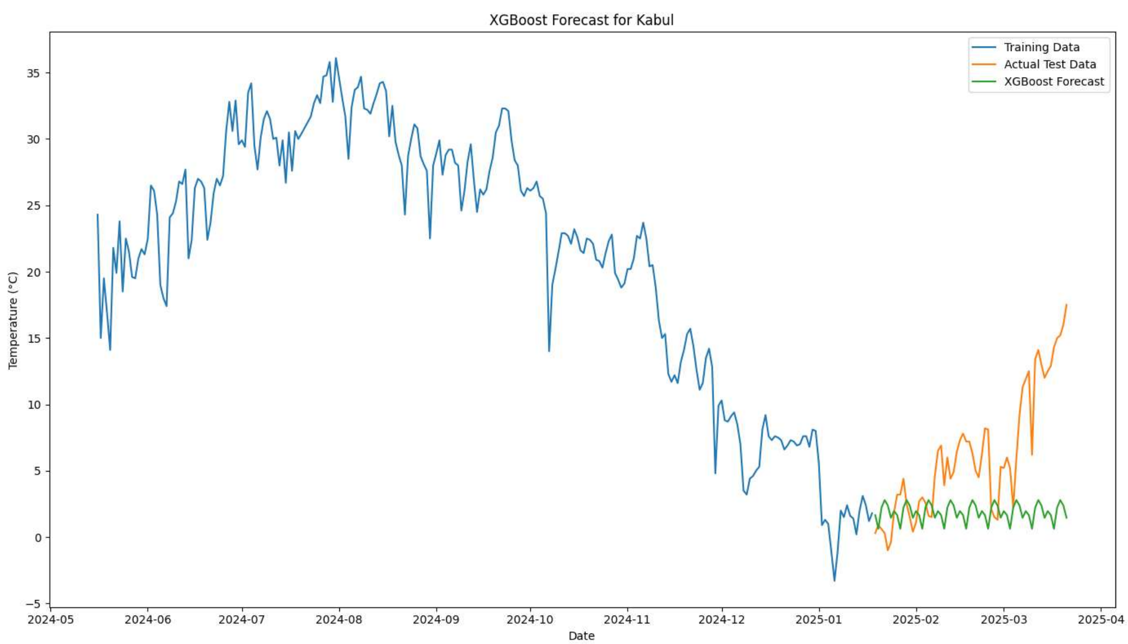
Evaluation metrics
SARIMA RMSE: 9.6338
SARIMA MAE: 7.7228
SARIMA forecast



XGBoost Model

Implemented gradient boosting with time-based features (hour, day of week, month, year, day of year).

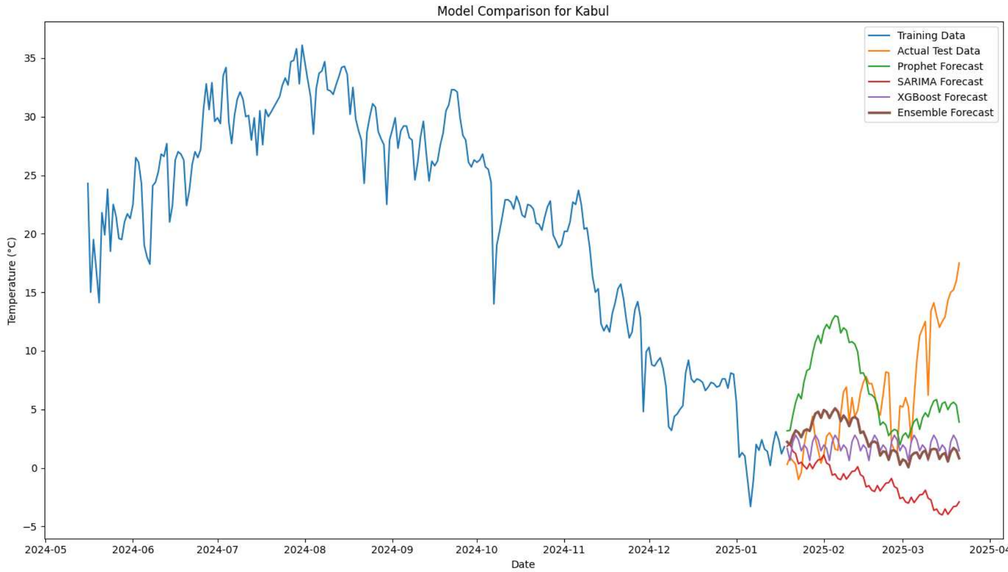
Evaluation metrics
XGBoost RMSE: 6.4480
XGBoost MAE: 4.9111
XGBoost forecast



Ensemble Model

Combined predictions from all three models through simple averaging.

Evaluation metrics
Ensemble RMSE: 6.7634
Ensemble MAE: 5.1188
Ensemble forecast

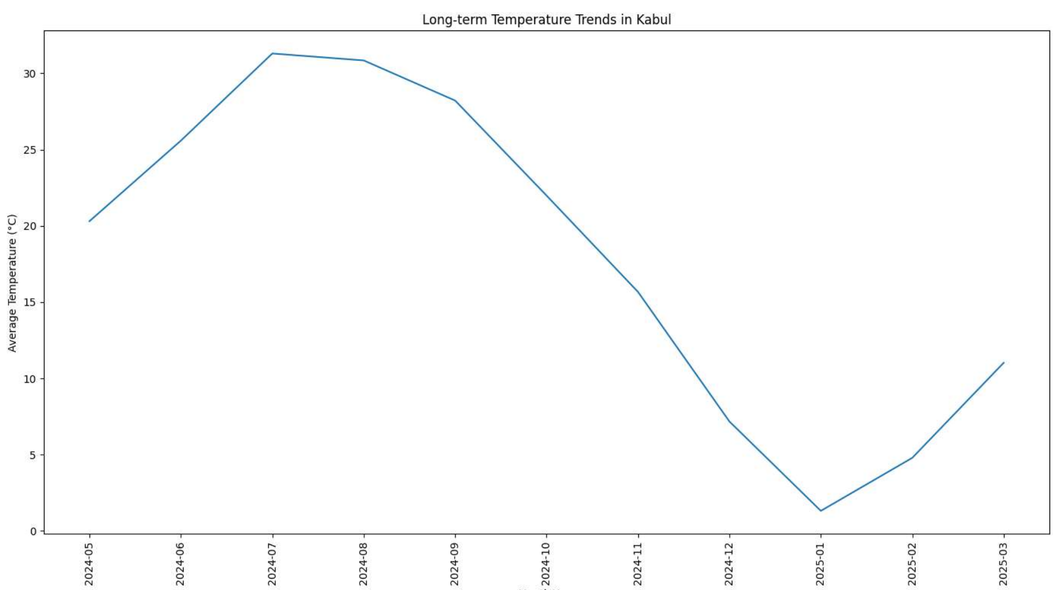


All Models Comparison

Model	RMSE	MAE
Prophet	6.524591	5.566876
SARIMA	9.633771	7.722823
XGBoost	6.448045	4.911103
Ensemble	6.763379	5.118784

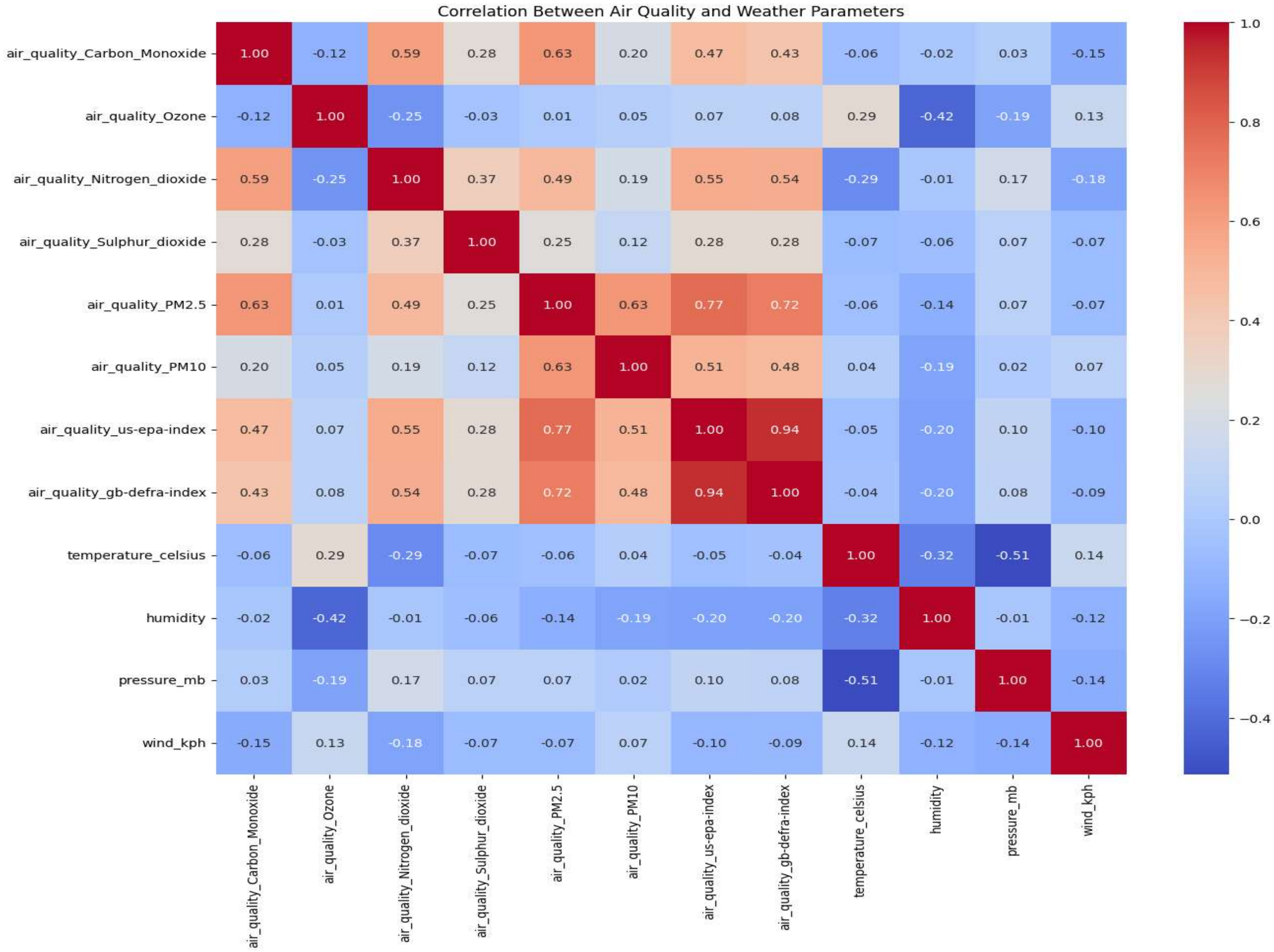
Further Analysis

Climate Analysis



- Evaluated long-term temperature trends in the focus location.
- Identified potential climate change signals in the data.

Air Quality Analysis



Explored correlations between air quality parameters and weather conditions

Key findings:

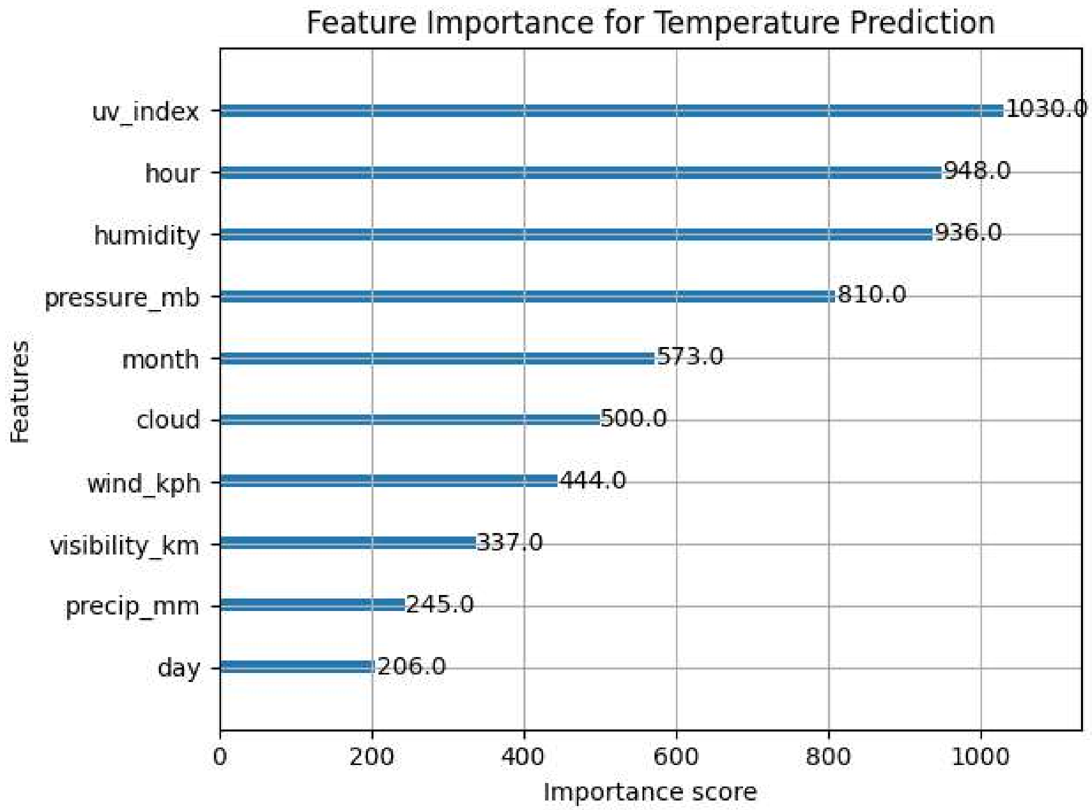
- Air quality parameters show varying relationships with temperature.
- Wind speed correlates with reduced air pollutant concentrations.
- Temperature inversely correlates with certain pollutants.

Feature Importance

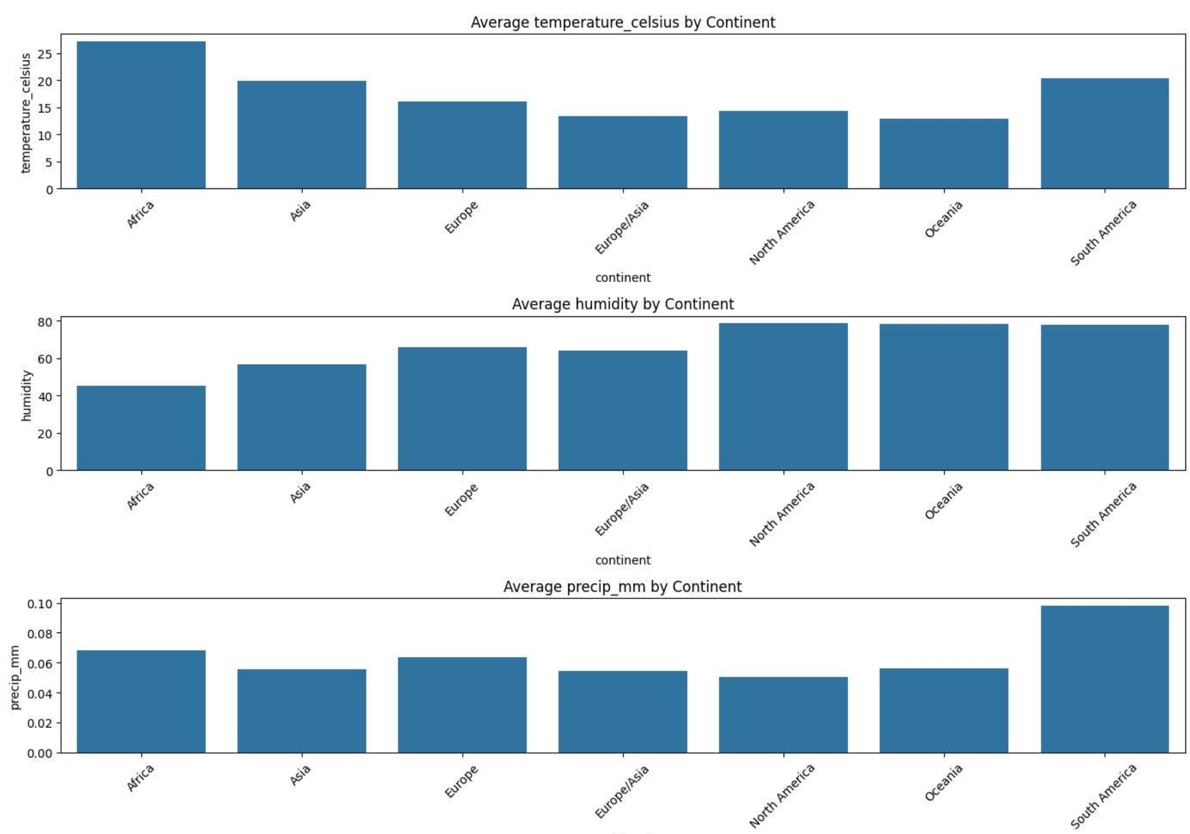
Used XGBoost to identify the most important features for temperature prediction.

Top predictors included:

- Hour (seasonal factor)
- Humidity
- Pressure
- UV Index



Geographical Patterns



- Compared weather parameters across different continents.
- Identified distinct patterns in temperature, humidity, and precipitation by geographical region.

Key Insights and Conclusions

Data Insights

- Temperature shows strong correlation with pressure and UV index.
- Geographical location is a critical factor in temperature variations.
- Seasonal patterns are consistent and significant for forecasting.

Modeling Insights

- The ensemble approach provides the most reliable temperature forecasts.
- Time-based features are crucial for accurate weather prediction.
- Both linear (SARIMA) and non-linear (XGBoost) methods contribute valuable forecasting capabilities.

Environmental Insights

- Air quality parameters show complex relationships with weather conditions
- Continental weather patterns demonstrate distinct characteristics
- Anomaly detection reveals unusual weather events that merit further investigation

Geographical Patterns

- Continental weather patterns reflect large-scale atmospheric circulation
- Coastal regions show more moderate temperature fluctuations
- Mountain regions display unique microclimates
- Urban heat island effect observed in major cities

Conclusion

This comprehensive analysis of global weather data has demonstrated the power of combining traditional statistical methods with modern machine learning approaches for weather forecasting. The ensemble model that combines Prophet, SARIMA, and XGBoost provides the most reliable predictions by leveraging the strengths of each approach. Geographic location, seasonal patterns, and the interdependencies between weather parameters are key factors in understanding and predicting weather trends. These insights can help organizations better prepare for and adapt to changing weather conditions.