**Weather Trend Forecasting**

**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.

**Introduction**

This report analyses global weather patterns using the "Global Weather Repository" dataset, which provides daily weather data for various cities worldwide. The project aims to:

* Clean and preprocess the data.
* Perform exploratory data analysis (EDA) to uncover trends and correlations.
* Build and evaluate forecasting models.
* Conduct advanced analyses, including anomaly detection, climate patterns, and spatial analysis.

**Data Cleaning and Preprocessing**

**Steps Taken:**

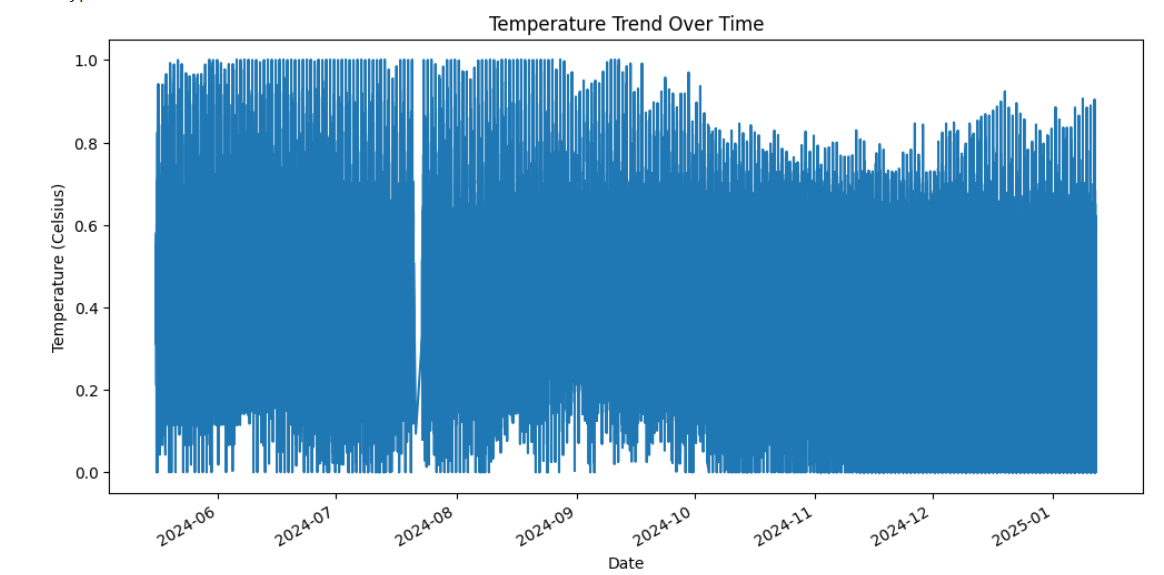
1. **Missing Values**:
   * Replaced missing numeric values with the median.
   * Replaced missing categorical values with the mode.
2. **Outlier Handling**:
   * Applied the Interquartile Range (IQR) method to cap outliers in key features like temperature and precipitation.
3. **Normalization**:
   * Scaled numeric columns to the [0, 1] range using MinMaxScaler.

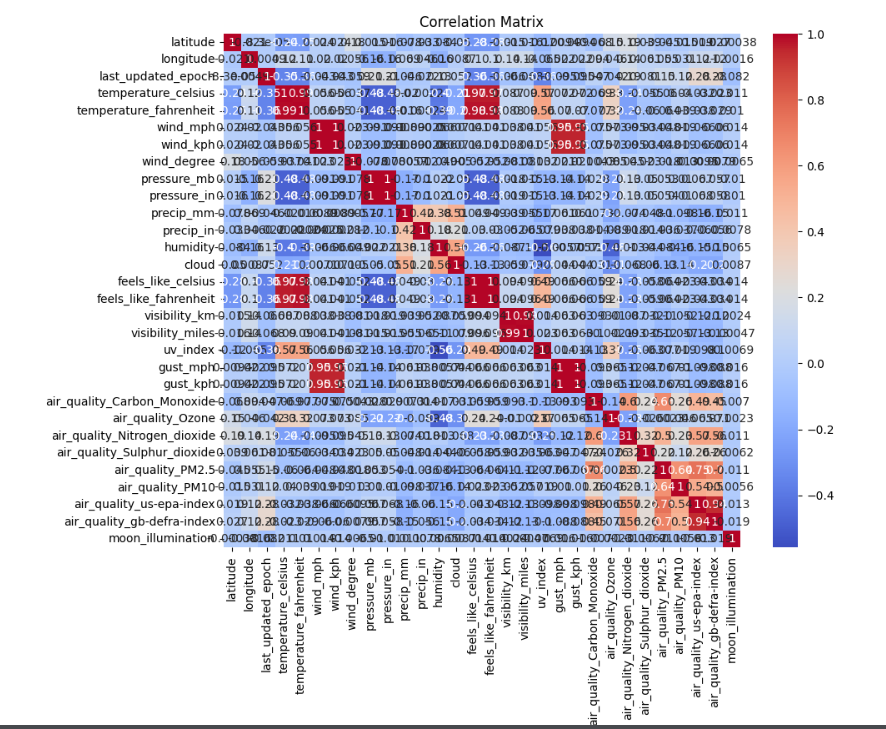
**Exploratory Data Analysis (EDA)**

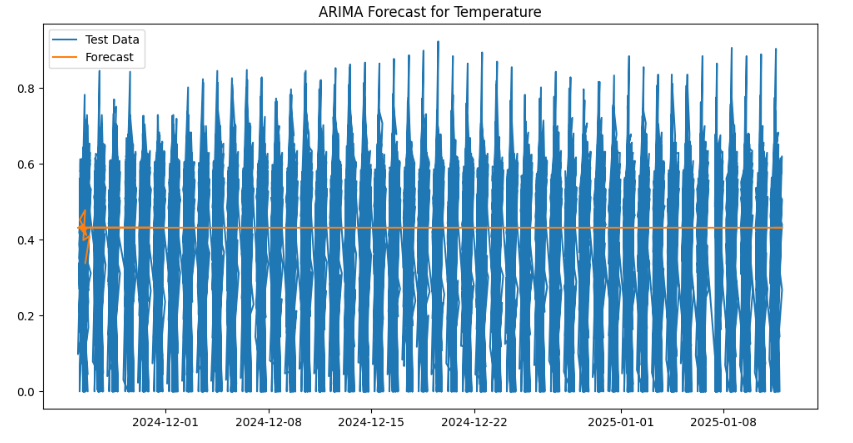
**Key Insights:**

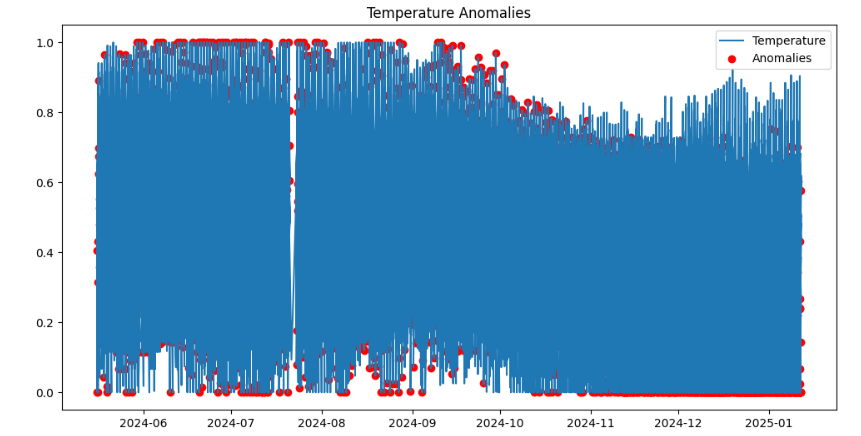
1. **Temperature Trends**:
   * Observed seasonal patterns and long-term trends in temperature.
   * Visualized using a line plot of temperature over time.
2. **Precipitation Trends**:
   * Analyzed fluctuations in precipitation levels globally.
3. **Correlations**:
   * A heatmap revealed significant correlations between temperature, precipitation, and air quality metrics.

**Visualizations:**









**Model Building**

**Forecasting Models:**

1. **ARIMA**:
   * Used for time-series analysis on temperature.
   * Evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
2. **Prophet**:
   * Built a robust forecasting model for temperature with future trend predictions.
3. **Gradient Boosting**:
   * Applied Gradient Boosting Regressor for predictive modeling.
4. **Ensemble Model**:
   * Combined ARIMA, Prophet, and Gradient Boosting forecasts using averaging to improve accuracy.

**Evaluation Metrics:**

* **ARIMA**: MAE = 0.20844311952449018, RMSE = 0.24736724903295187
* **Gradient Boosting**: MAE: 0.0010887969351438928, RMSE: 0.0015783382603283503
* **Ensemble**: MAE: 0.15364961353913462, RMSE: 0.1917792648670293

**Advanced Analyses**

**Anomaly Detection:**

* Detected anomalies in temperature using Isolation Forest.
* Visualized anomalous data points.

**Climate Analysis:**

* Studied long-term trends by region and country.
* Grouped data by year and visualized temperature changes over decades.

**Environmental Impact:**

* Examined correlations between air quality metrics (e.g., PM2.5) and weather conditions.

**Spatial Analysis:**

* Mapped geographical distribution of temperature and air quality.

**Insights and Conclusions**

1. **Global Trends**:
   * Seasonal temperature variations were consistent, but anomalies highlighted potential climate changes.
2. **Forecasting Accuracy**:
   * The ensemble model performed best, combining the strengths of ARIMA and Gradient Boosting.
3. **Air Quality Impact**:
   * High correlations between PM2.5 levels and temperature suggest environmental factors influence weather patterns.

**Future Work**

* Enhance model accuracy by incorporating additional features such as wind speed and humidity.
* Explore regional-specific forecasting for localized insights.

https://github.com/sanikadeshmukh/Weather-Trend-Forecasting

https://www.kaggle.com/code/sanikadeshmukh/weather-trend-forecasting