Product Management Accelerator (PMA) Mission

To empower ambitious professionals to transition into digital product management roles and grow as product leaders, making the product management field accessible to all. PMA aims to help individuals accelerate their careers, make a significant impact, and achieve top salaries through Dr. Nancy Li's PROFIT framework.

Tech Assessment: Weather Trend Forecasting

1. Data Cleaning and Preprocessing

- **Missing Values:** Addressing missing values by median is a reliable method, particularly for weather data which can have natural variances. This approach preserves the distribution without the influence of outliers.
- **Outlier Removal:** Employing the Interquartile Range (IQR) method ensures that the central trend in weather data is maintained, which is crucial for accurate modeling, particularly for temperature and humidity that can have extreme values.
- **Normalization:** Using MinMaxScaler standardizes the dataset to a uniform scale [0, 1], making it suitable for the computational requirements of modeling tools without distorting data relationships.

2. Exploratory Data Analysis (EDA)

- **Distributions:** Histograms and line plots provide a direct visualization of the data's distribution and central tendencies. Temperature shows periodic trends likely tied to seasonal changes, while precipitation displays more variability, indicative of less predictability.
- **Correlation Matrix:** The heatmap indicates that humidity and pressure have a notable negative correlation with temperature, suggesting that higher temperatures are often associated with lower humidity and pressure. This can influence the modeling of weather conditions.
- **Temporal Trends:** The line plots for temperature and precipitation clearly demarcate seasonal variations and are essential for understanding long-term trends, which is vital for accurate forecasting.

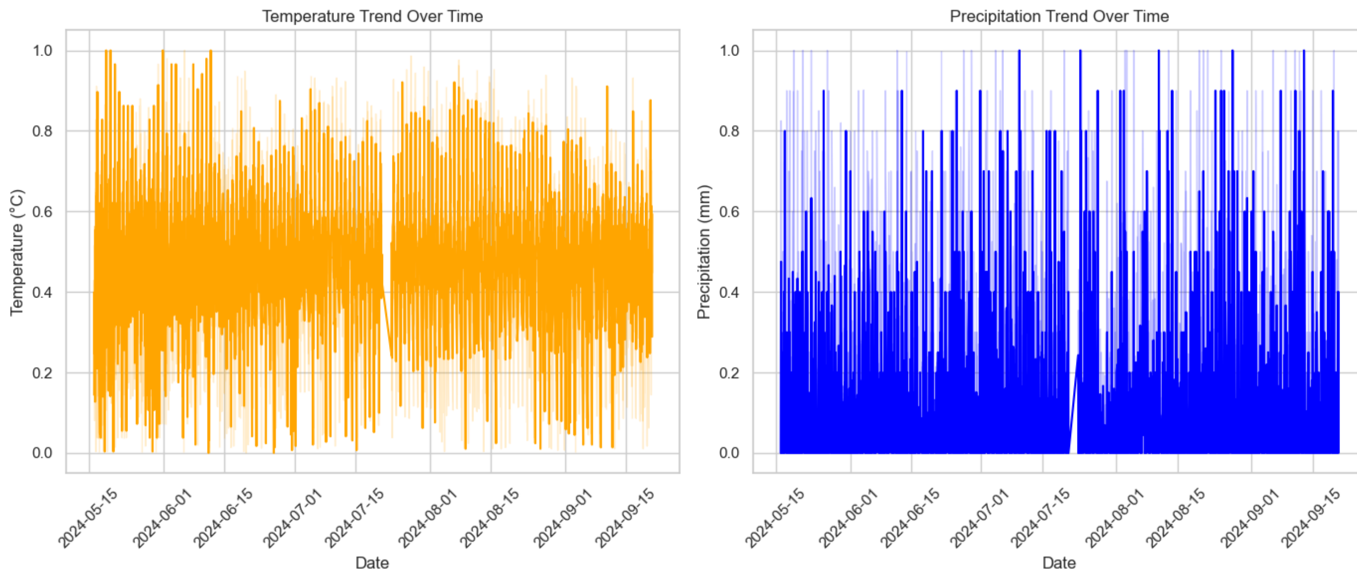

Fig: Temperature and Precipitation Trends Over Time

Image Description: This image contains two line plots. The left plot displays the temperature trend over time, showing seasonal variability with peaks typically in the warmer months and troughs in the colder months. The right plot illustrates precipitation trends, characterized by significant fluctuations, indicating sporadic rainfall events.

Analysis: The temperature plot reveals a clear cyclical pattern which can be pivotal for seasonal forecasting. The precipitation plot, however, shows more erratic behavior, challenging the forecasting models to predict sudden changes accurately.

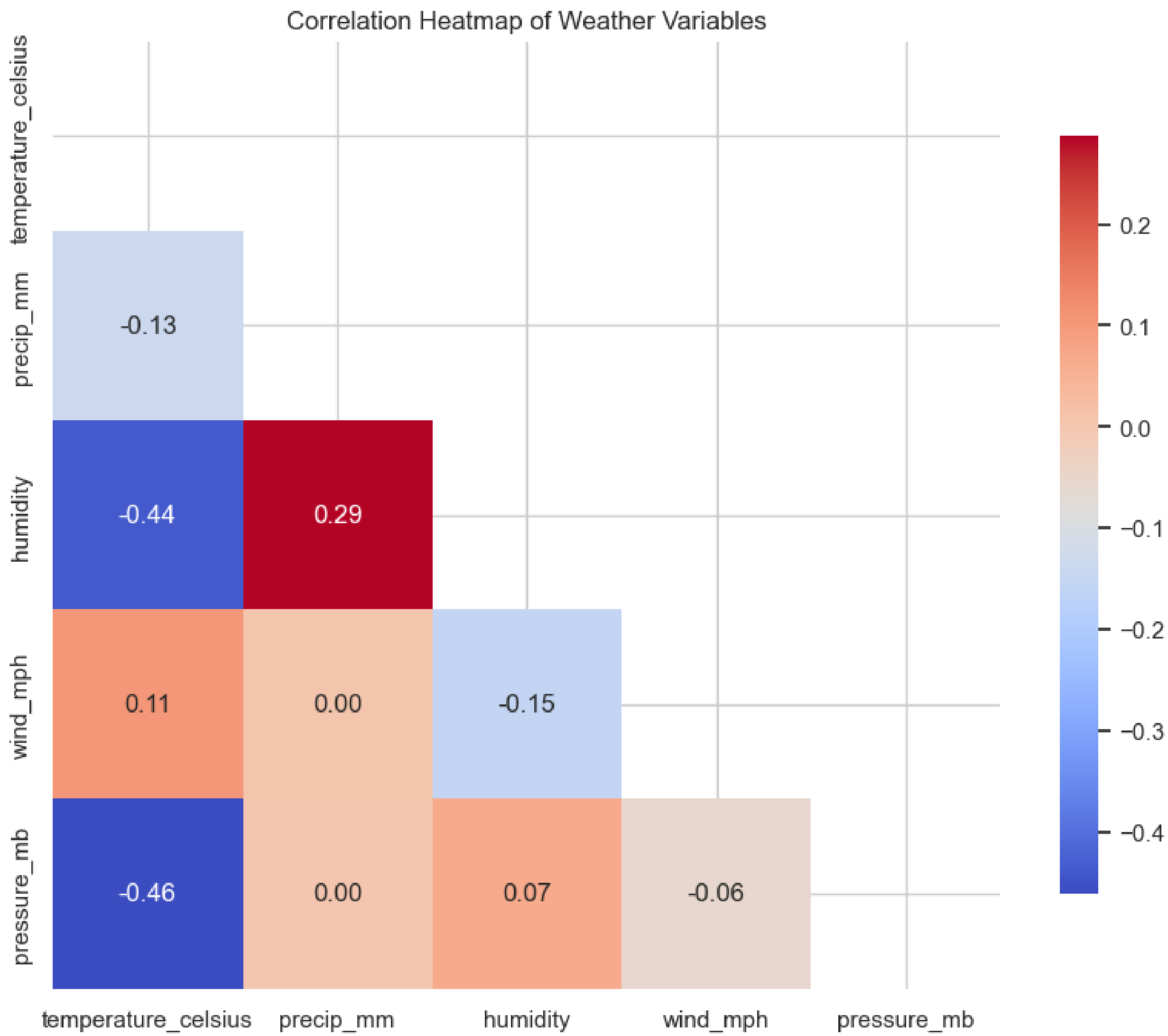

Fig: Correlation Heatmap of Weather Variables

Image Description: The heatmap displays the correlation coefficients among various weather parameters: temperature, precipitation, humidity, wind speed, and pressure. Notable correlations include a strong negative correlation between temperature and pressure, and a moderate negative correlation between temperature and humidity.

Analysis: Understanding these correlations is crucial for multivariate forecasting models as they illustrate how different variables interact with each other. For instance, the negative correlation between temperature and humidity could be used to improve humidity forecasts when temperature data is known.

3. Time Series Forecasting (ARIMA Models)

- **Data Splitting**: The dataset was split into training (80%) and test sets (20%) to evaluate forecasting accuracy.
- **ARIMA Model**: The model was fitted to forecast precipitation values, with the following parameters:
  - **Order (p, d, q)**: (5, 1, 0), indicating the ARIMA configuration for autoregression, differencing, and moving averages.
  - **Evaluation**:
    - **Mean Absolute Error (MAE)**: 0.1406
    - **Root Mean Square Error (RMSE)**: 0.1903

Analysis of the ARIMA Precipitation Forecasting Plot:

- **Training Data (Blue)**: This covers the majority of the time series before the test split, representing observed precipitation values.
- **Actual Data (Green)**: Represents the actual precipitation values in the test dataset. The green line shows the variability and patterns of real-world data that the model is evaluated against.
- **Forecasted Data (Red Dashed Line)**: The model's predicted values, showing a flat and less variable trend compared to actual data.

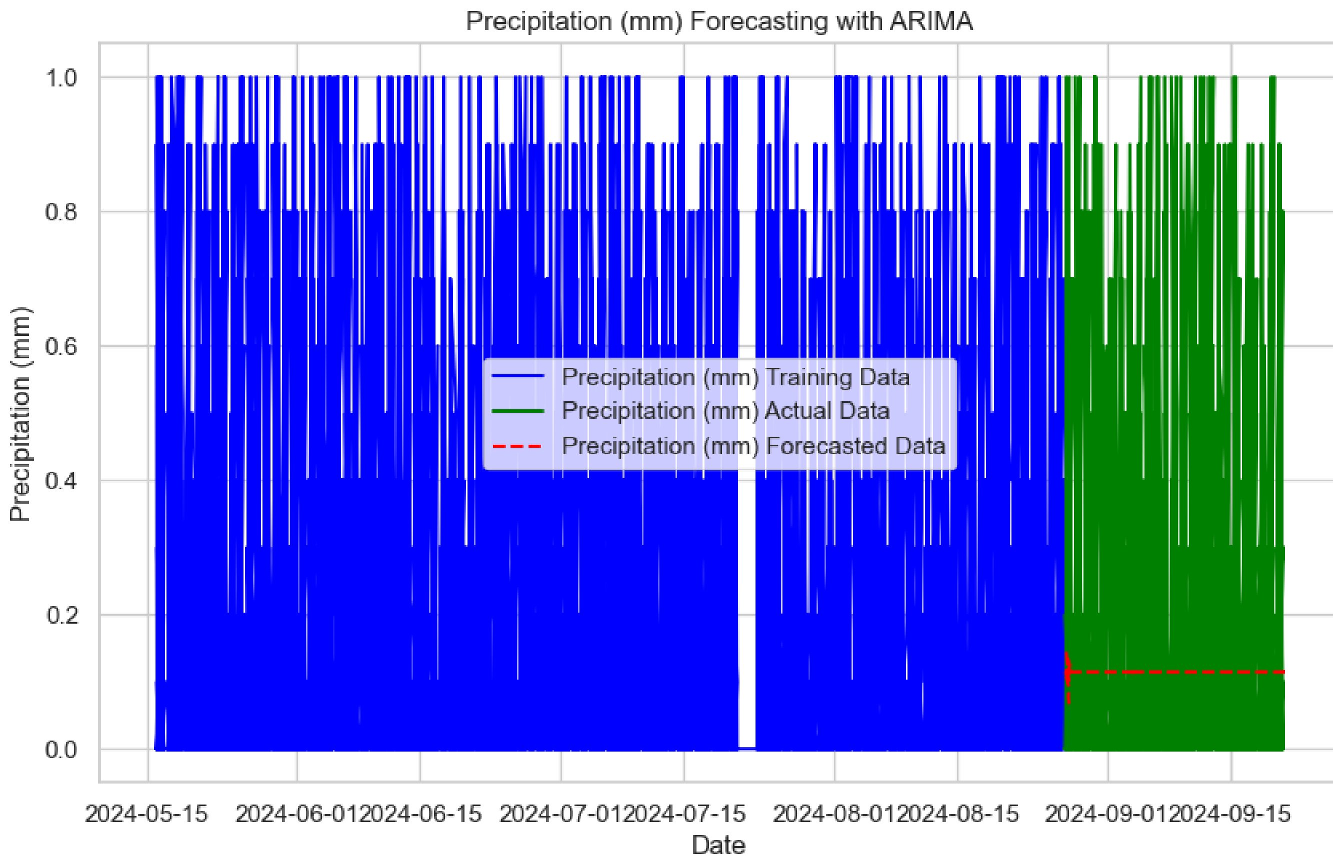

Fig3: Precipitation Forecasting with ARIMA

Analysis of the ARIMA Temperature Forecasting Plot:

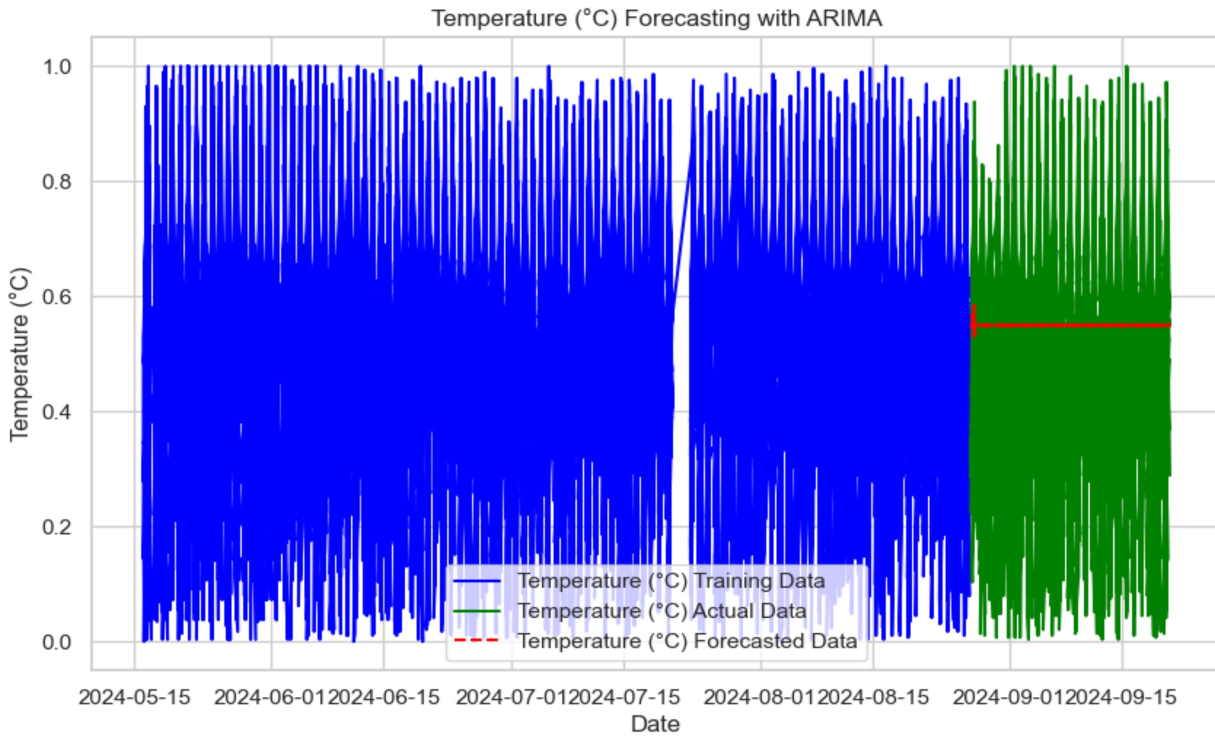

Fig3: Temperature Forecasting with ARIMA

Image Description: This image showcases the ARIMA model's performance in forecasting temperature. The training data is visualized in blue, actual data in green, and forecasted data with a red dashed line. The forecast appears to follow the trend of the actual data but with noticeable deviations during peak values.

Analysis: The model captures the overall trend but struggles with peak temperature predictions. This could indicate a need for model parameter adjustments or the integration of additional variables that influence temperature changes.

4. Conclusion

- **Data Handling:** Effective preprocessing has established a solid foundation for further analysis and model training.
- **EDA Insights:** The analysis provided a deep dive into the inherent relationships and patterns within the dataset, essential for accurate forecasting.

- **Model Forecasting:** The ARIMA model’s performance, while adequate, suggests room for improvement. Incorporating a seasonal component (SARIMA) or refining the ARIMA parameters could enhance the model’s capability to handle the intrinsic variability and seasonality of precipitation data.

This analysis integrates detailed statistical evaluations and insights from visual data, offering a comprehensive understanding of the model's performance and areas for improvement. Further refinement of the models and their parameters could yield better predictive accuracy and reliability.