SMITHI

MAHENDRAN

PM Accelerator

Technical Assessment Overview

“

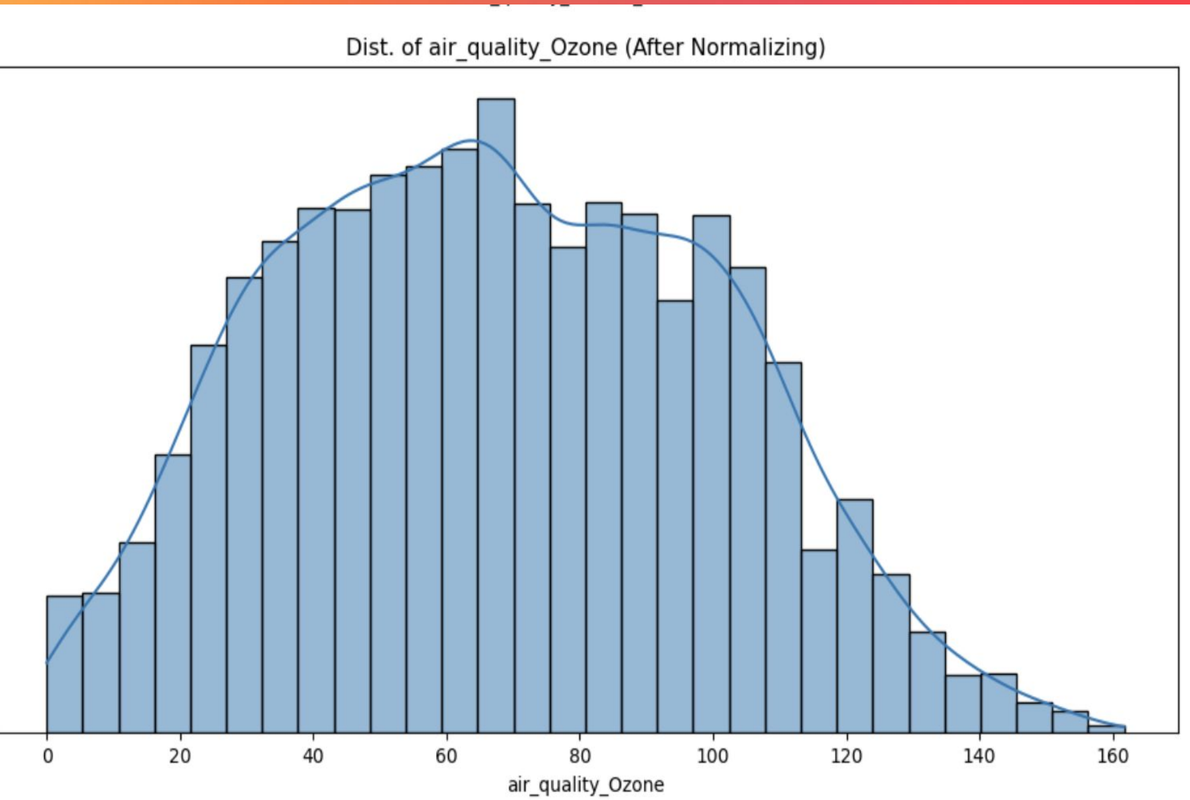
By making industry-leading tools and education available to individuals from all backgrounds, we level the playing field for future PM leaders. This is the PM Accelerator motto, as we grant aspiring and experienced PMs what they need most – Access. We introduce you to industry leaders, surround you with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

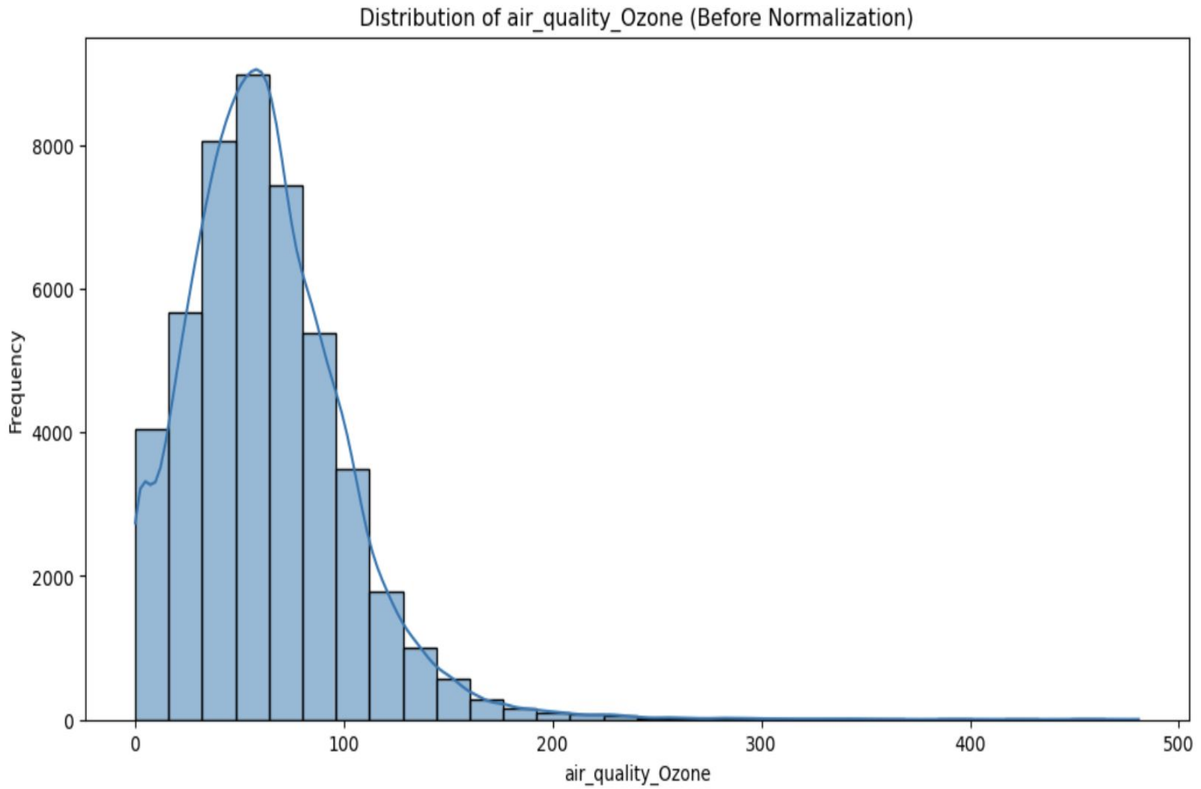
1. Data Cleaning + Preprocessing
2. Basic EDA
3. Model Building/Forecasting (Basic + Adv.)
4. Advanced EDA
5. Unique Analyses

01: Data Cleaning + Preprocessing

- Checked Missing Data: No missing values.
- Explored Data Types: Used info() to check column data types.
- Checked for Duplicates: No duplicate rows.
- Handled Outliers: Used IQR method to detect and remove outliers.
- No outliers remained after cleaning.
- Visualized Distributions: Plotted histograms with KDE before and after normalization.

01: Data Cleaning + Preprocessing (cont’d)

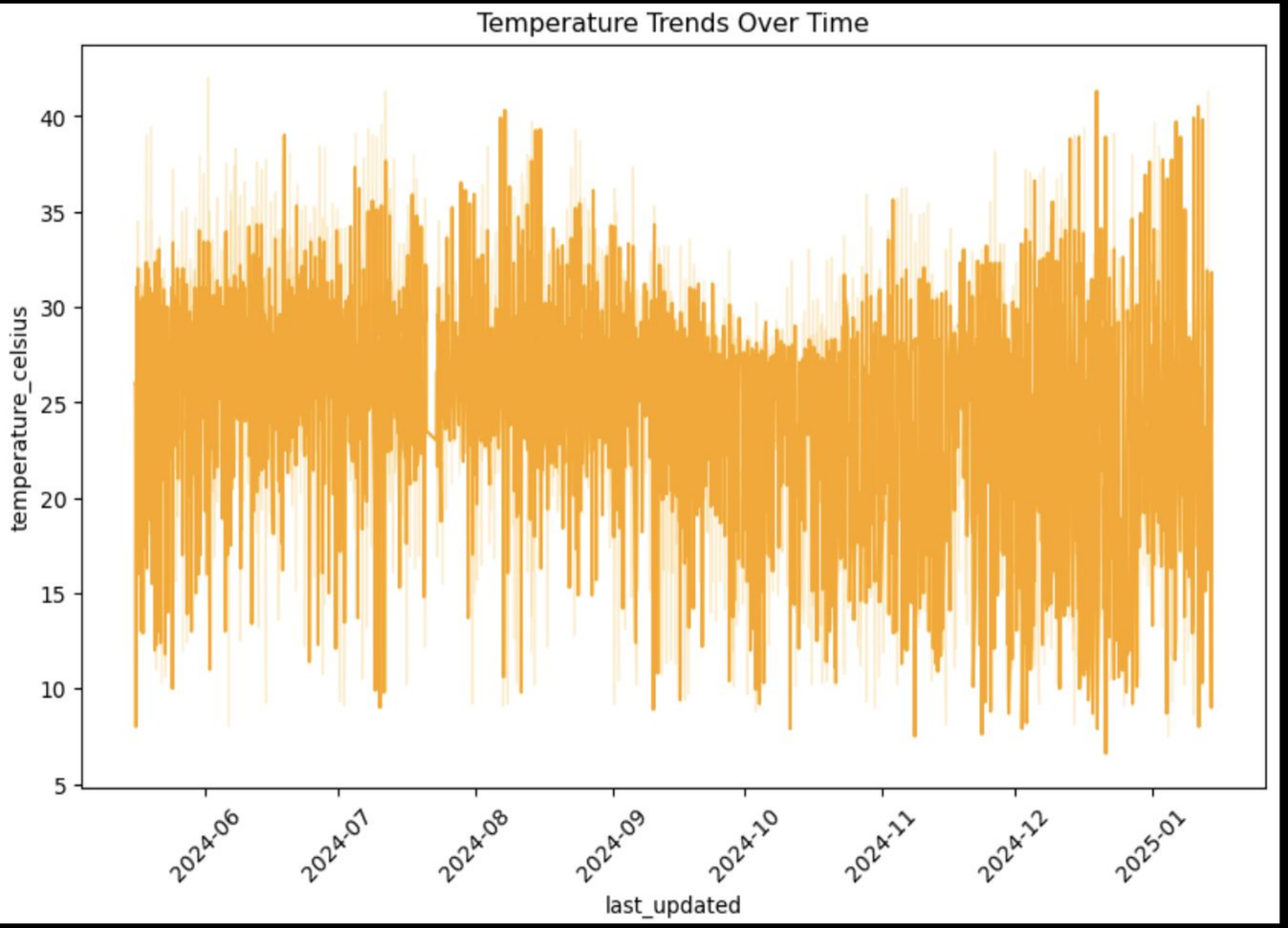


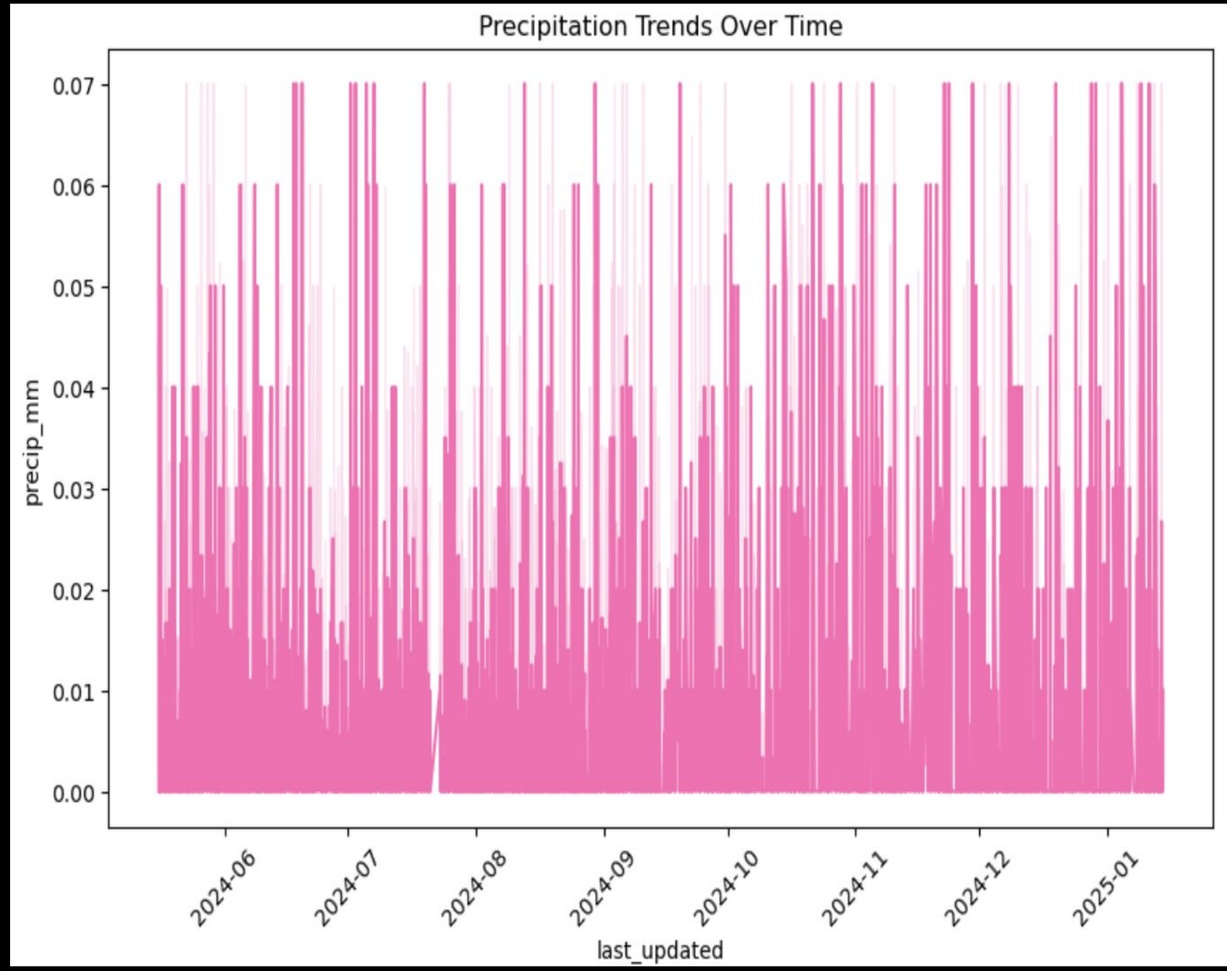


02:Basic EDA

- Temperature & Precipitation Trends: Line plots for temperature and precipitation over time.
- Precipitation by Country: Line plot for country-wise precipitation trends.
- Temperature by Country: Line plot for country-wise temperature trends.
- Wind Speed vs Gust Speed: Scatter plot shows positive correlation.
- Weather Variables Correlation: Heatmap of weather variables’ correlations.
- Air Quality Correlation: Heatmap shows strong PM2.5 & PM10 correlation.

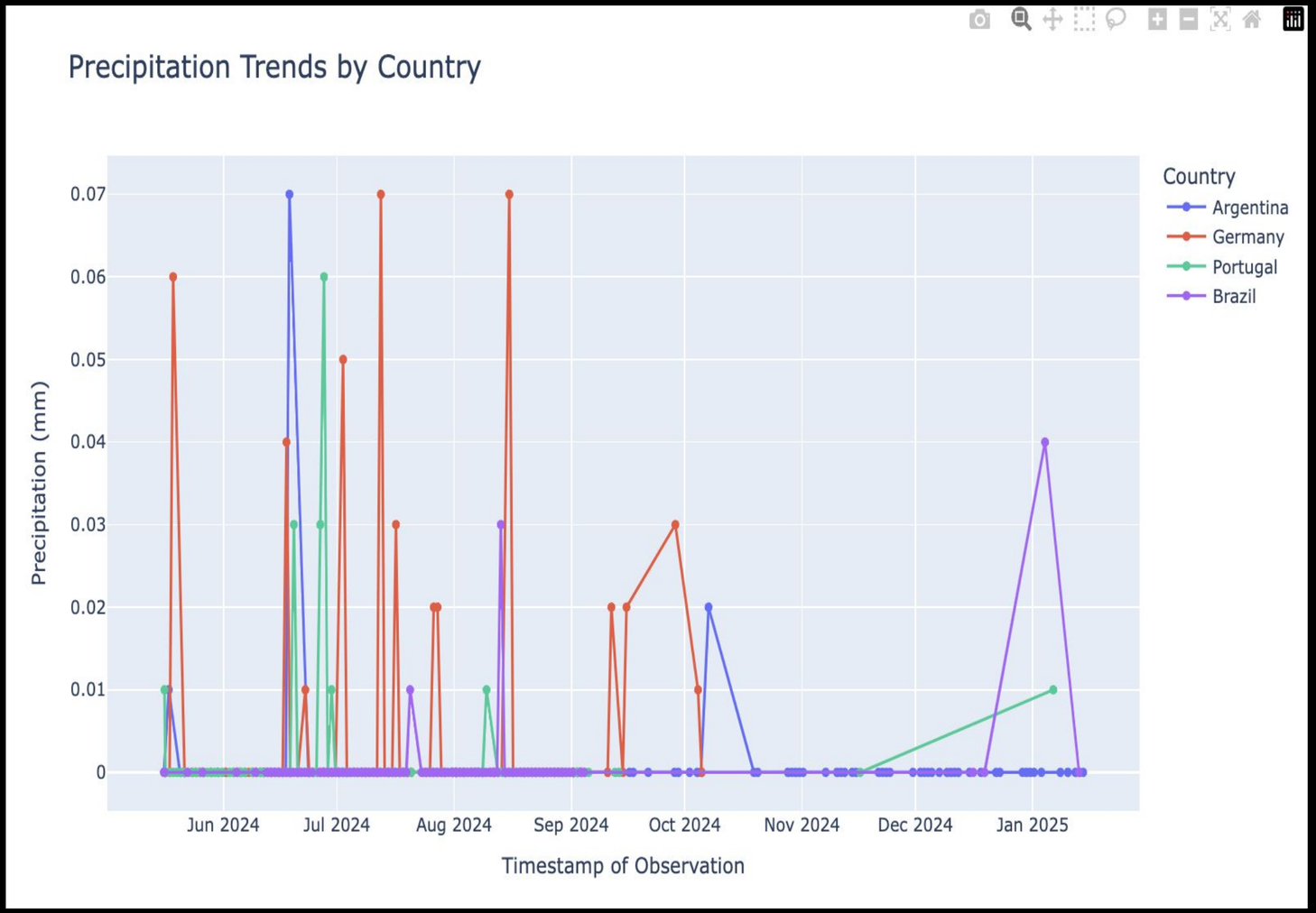
02: Basic EDA(cont’d)



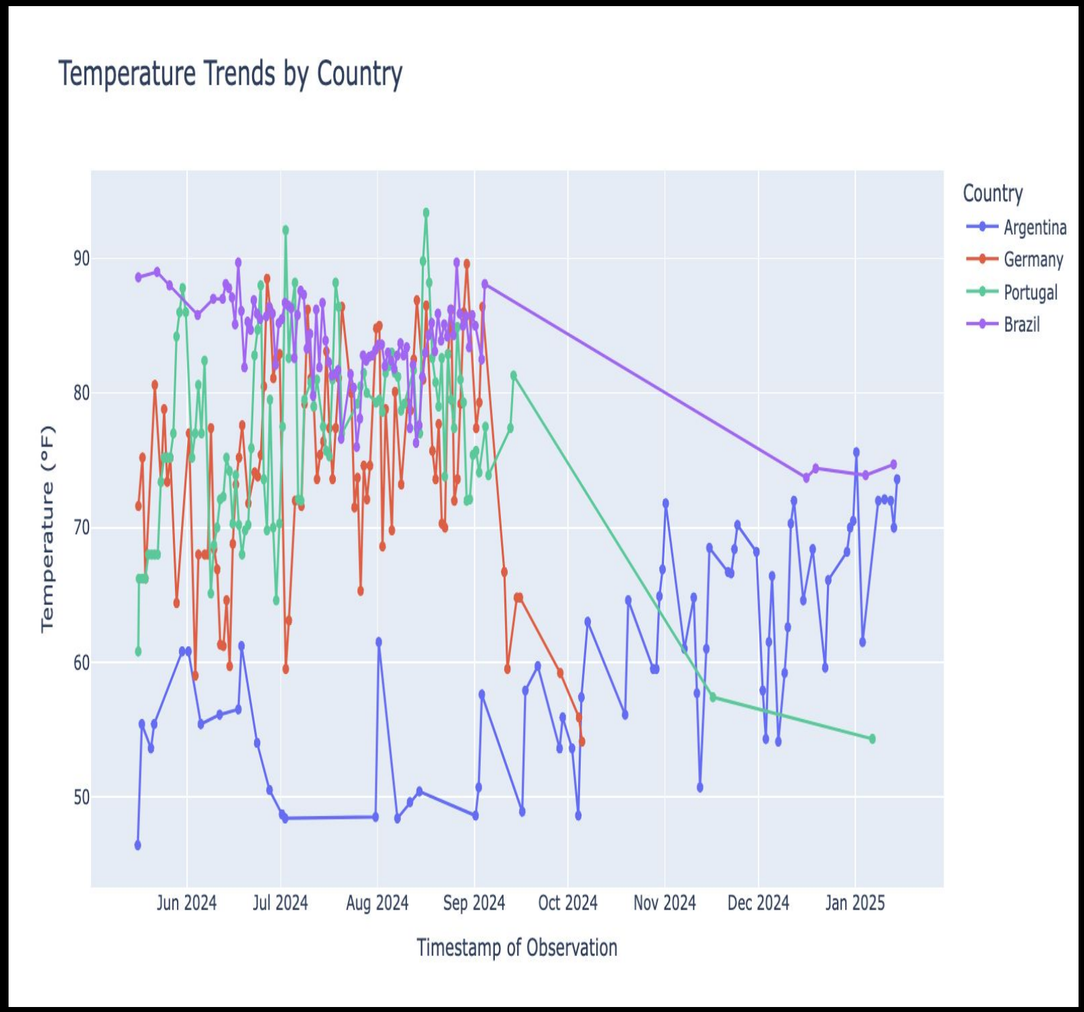


This code creates line plots to show the trends of temperature and precipitation over time using last_updated as the x-axis. The orange line graph shows the trends in temperature in celsius over time with temperature increasing in the summer and decreasing in the winter. The pink line graph shows spikes of rain followed by periods of dryness.

02: Basic EDA(cont’d)

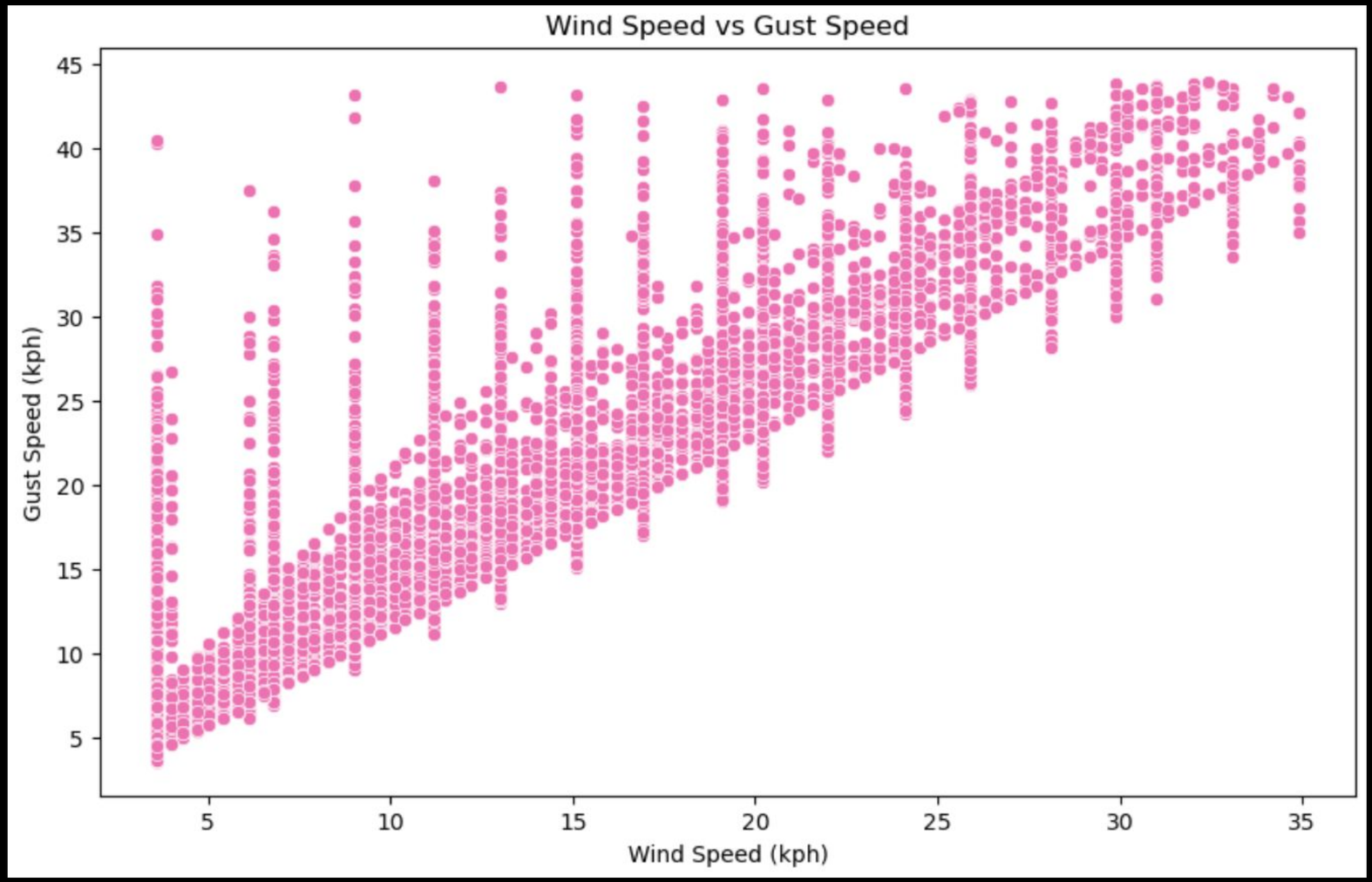


Argentina, Germany, and Portugal show relatively low precipitation levels, with sporadic high spikes in some months. Whereas Brazil shows a significant spike in precipitation around January 2025.



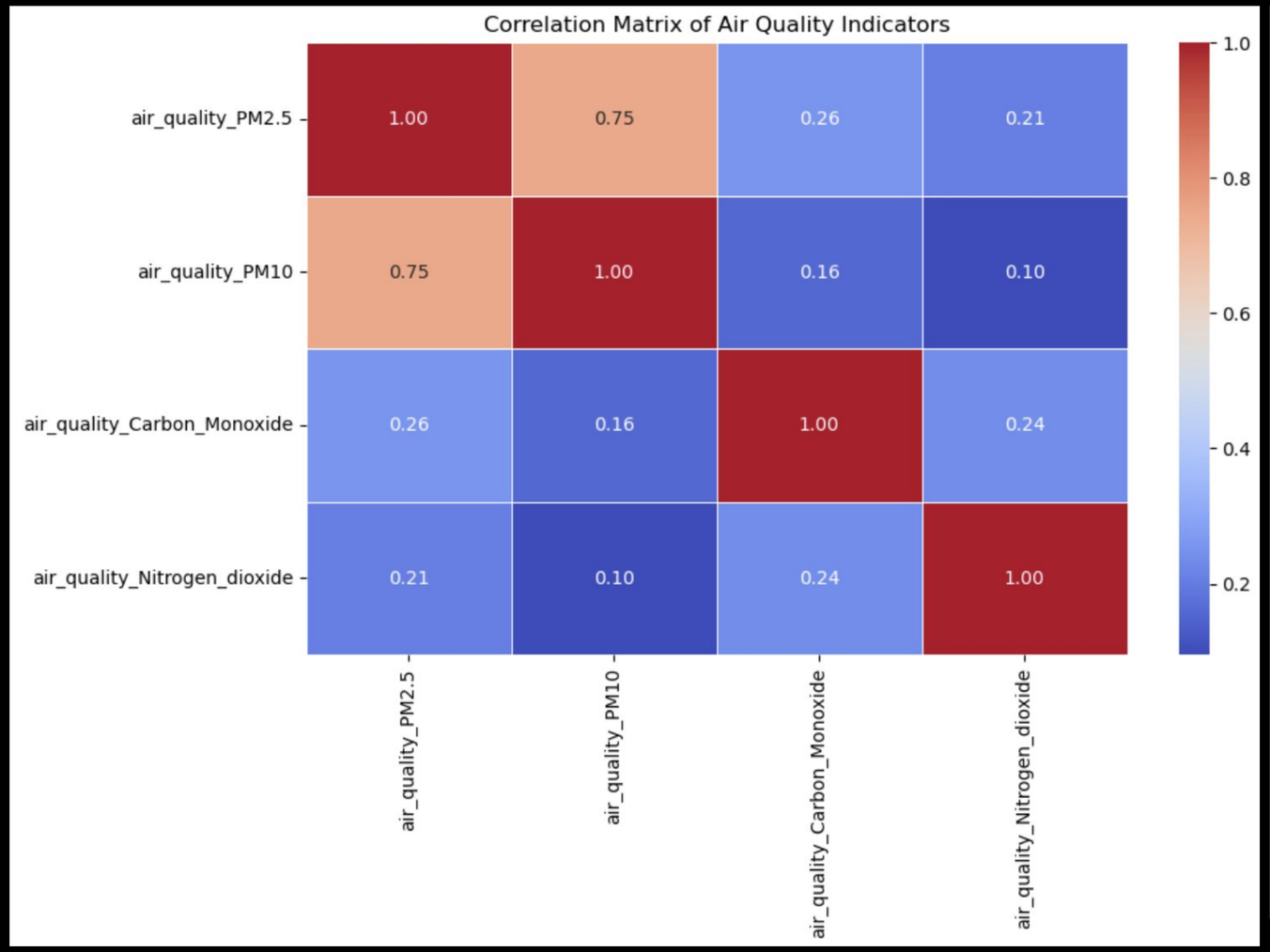
The line graph illustrates the temperature trends of Argentina, Germany, Portugal, and Brazil from June 2024 to January 2025, showing significant fluctuations in Argentina, Germany, and Portugal while Brazil shows a steady decrease in temperature over the same period.

02: Basic EDA(cont’d)

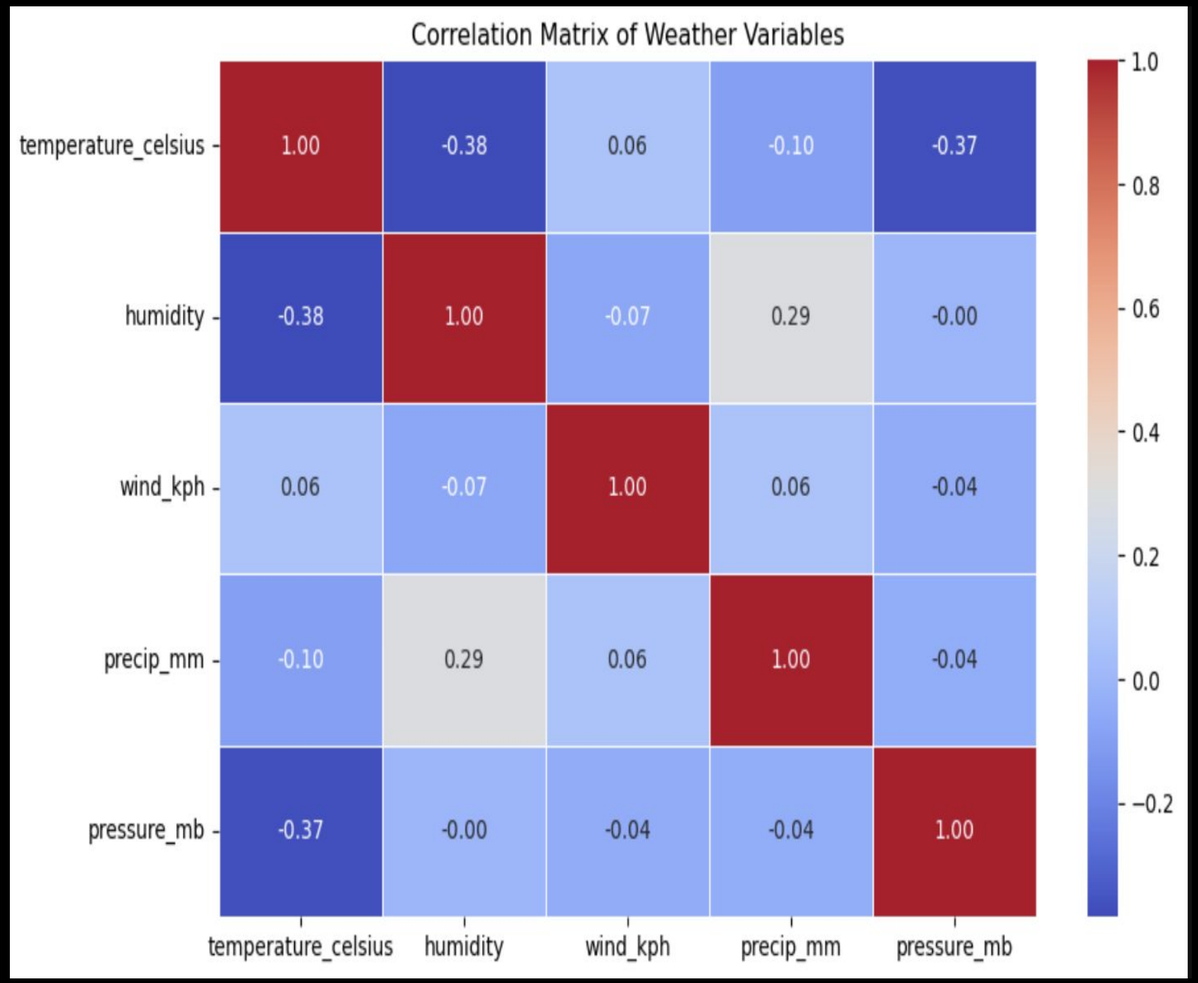


The scatter plot reveals a clear positive linear relationship between wind speed and gust speed, indicating that higher wind speeds are associated with stronger gusts. The close clustering of data points suggests a strong correlation, emphasizing wind speed as a significant factor in gust intensity.

02: Basic EDA(cont'd)



Air Quality PM2.5 and Air Quality PM10 have a strong positive correlation (0.75). This suggests that when PM2.5 levels are high, PM10 levels are also likely to be high.



Overall, the correlation matrix suggests that temperature and humidity as well as pressure and temperature have the strongest relationship amongst this group of features, and precipitation and humidity have a moderate positive relationship. The other variables do not appear to be strongly correlated with each other.

03:Model Building/Forecasting (Basic + Adv.)

- Model Building: Trained Gradient Boosting, Random Forest, and Linear Regression models to predict precipitation.
- Model Evaluation: Used MAE, RMSE, and R2 to evaluate model performance.
- Ensemble Model: Combined predictions from all models for improved accuracy.
- Performance Metrics: Displayed MAE, RMSE, and R2 scores for each model and the ensemble.

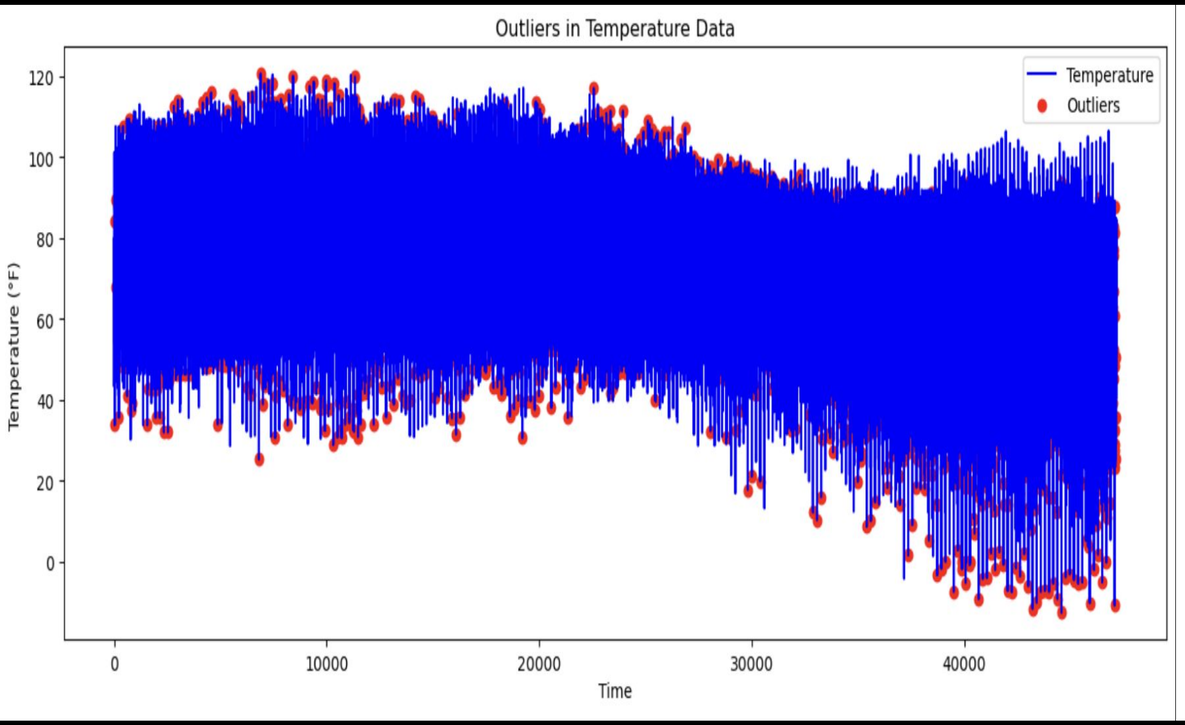
03:Model Building/Forecasting (Basic + Adv.), Cont’d

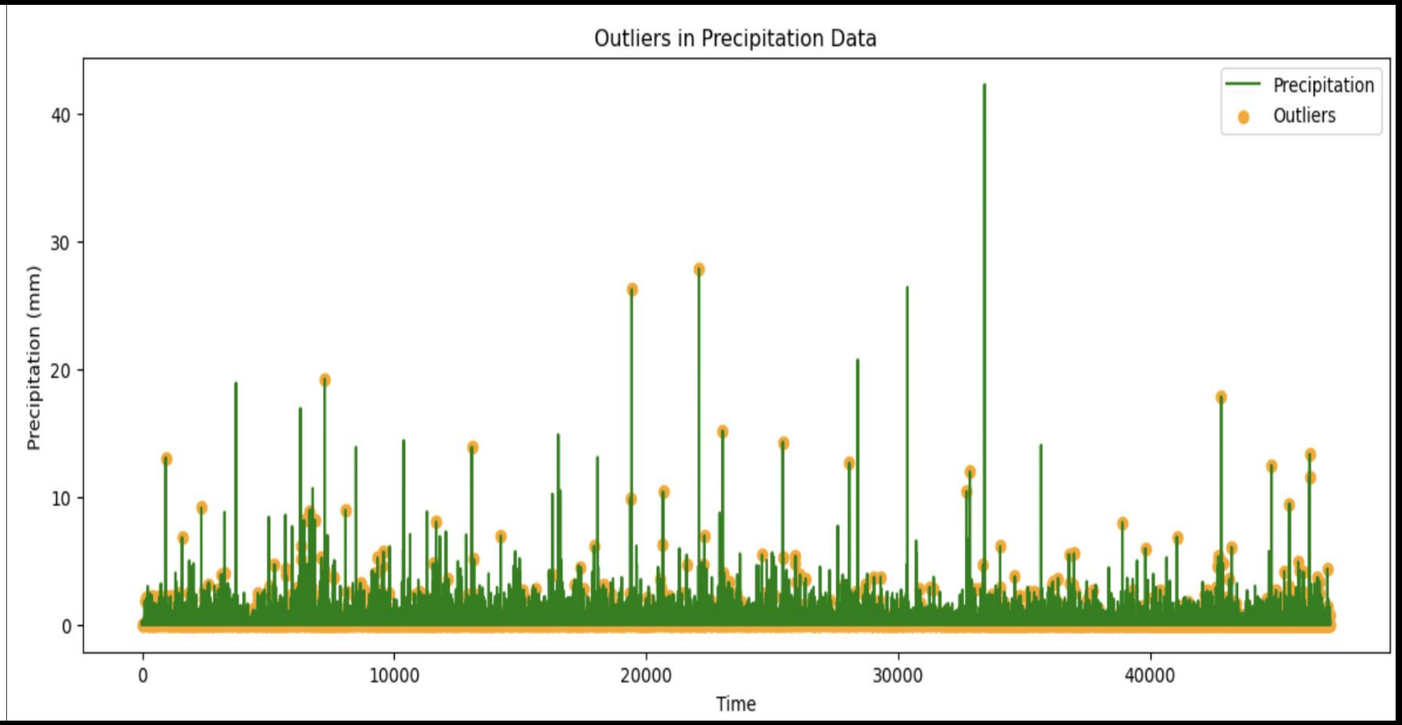
- Gradient Boosting: MAE=0.006, RMSE=0.012, R2 Score=0.398
- Random Forest: MAE=0.005, RMSE=0.011, R2 Score=0.411
- Linear Regression: MAE=0.020, RMSE=0.164, R2 Score=-120.266
- Ensemble: Ensemble: MAE=0.010, RMSE=0.056, R2 Score=-13.010

04:Advanced EDA

- Outlier Detection: Used Isolation Forest to detect outliers in temperature and precipitation data.
- Outlier Visualization: Plotted temperature and precipitation data with outliers marked in red and orange.
- Outlier Count: Detected 2,359 outliers in the dataset.

04: Advanced EDA(cont'd)

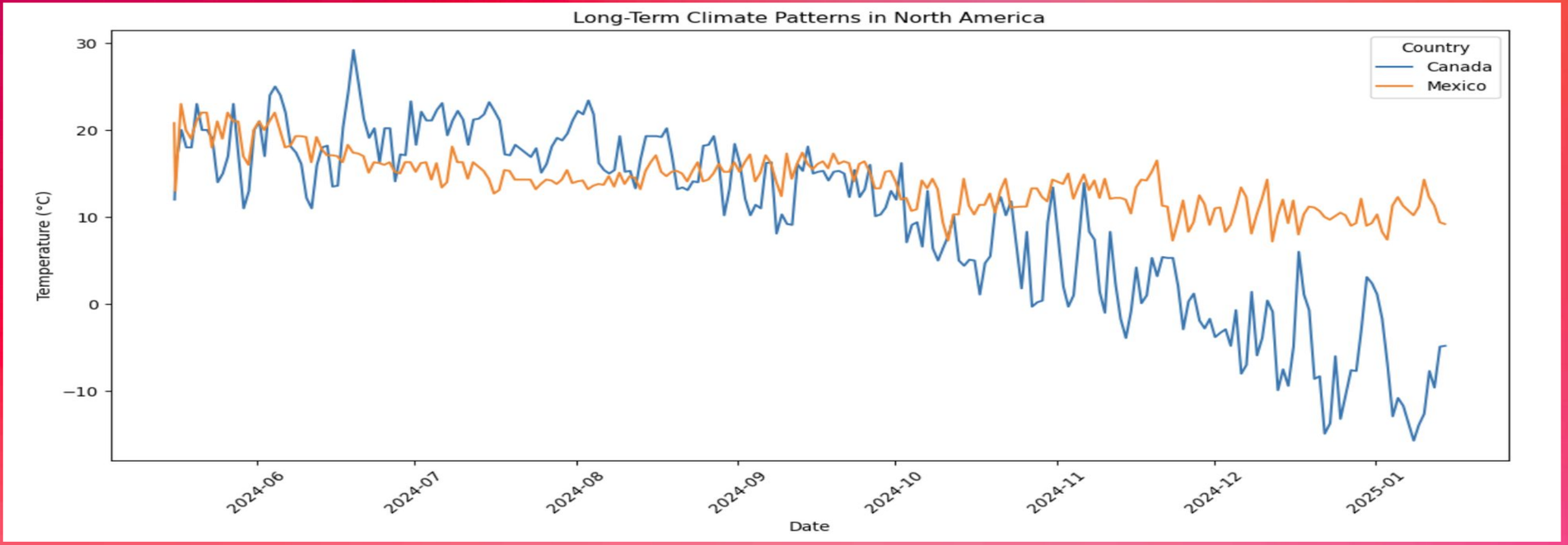




I applied the Isolation Forest method to find outliers in temperature and precipitation data. It then creates plots showing the data over time, marking the outliers in red for temperature and orange for precipitation.

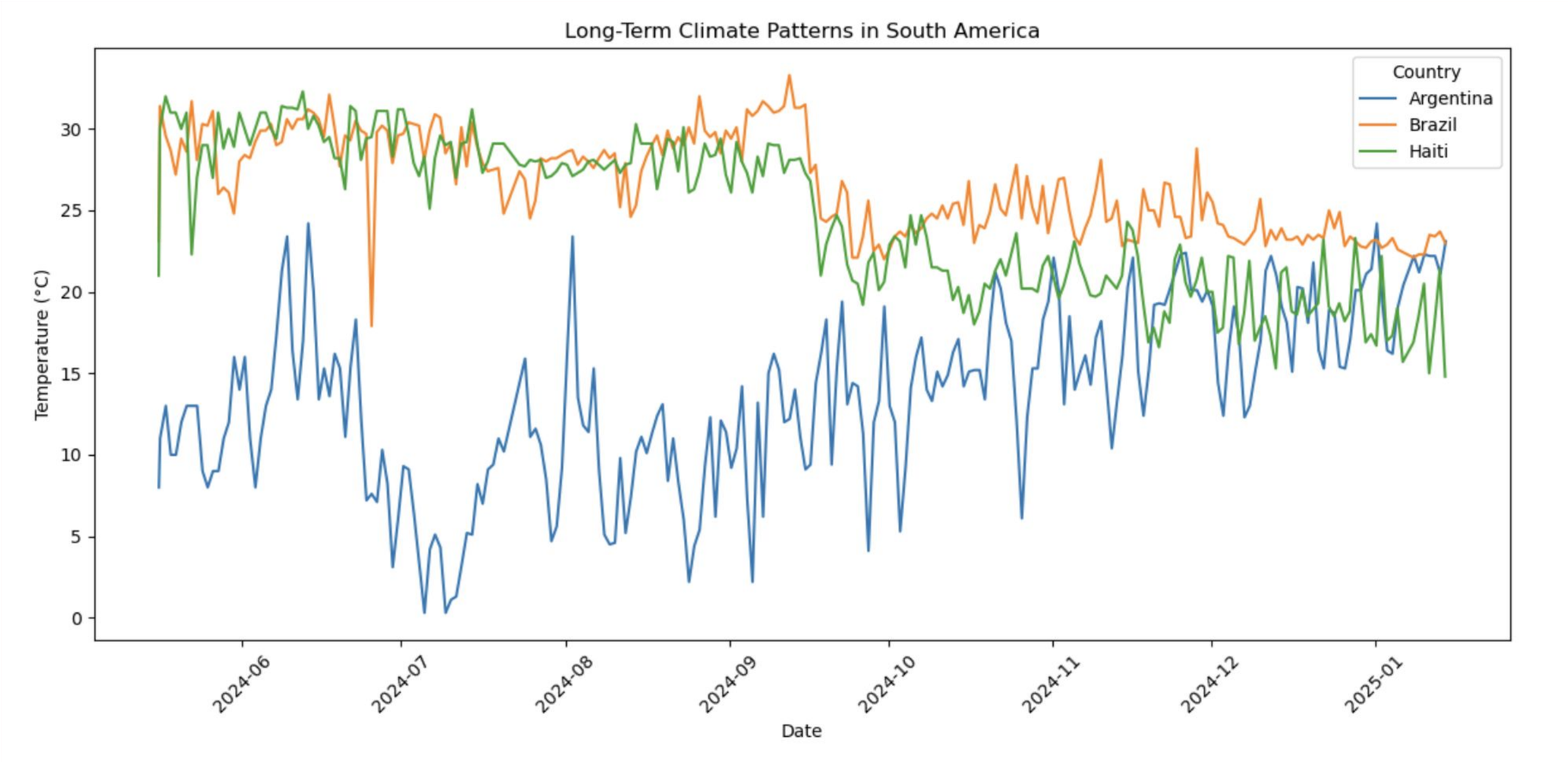
05:Climate Analysis

- I selected countries from each continent and plotted their temperature trends over time. Each plot shows how temperature has changed, with different colors for each country.



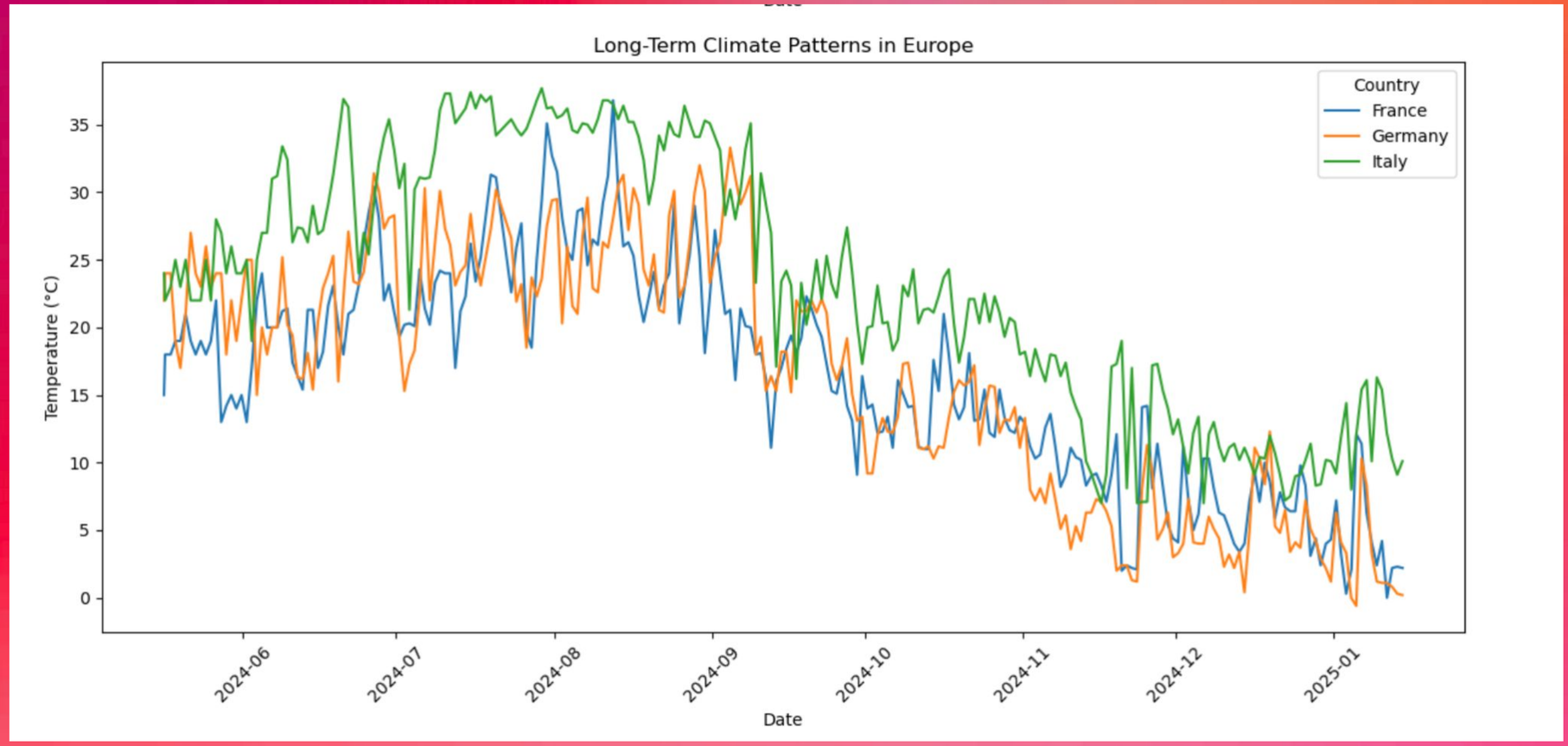
The trends depicted in the graph reveal that Canada experiences a significant decline in temperatures from June to January, showcasing pronounced seasonal fluctuations that lead to much colder winter months, while Mexico maintains a relatively stable and warmer temperature range throughout the same period, indicating less variability in its climate.

05:Climate Analysis (cont’d.)



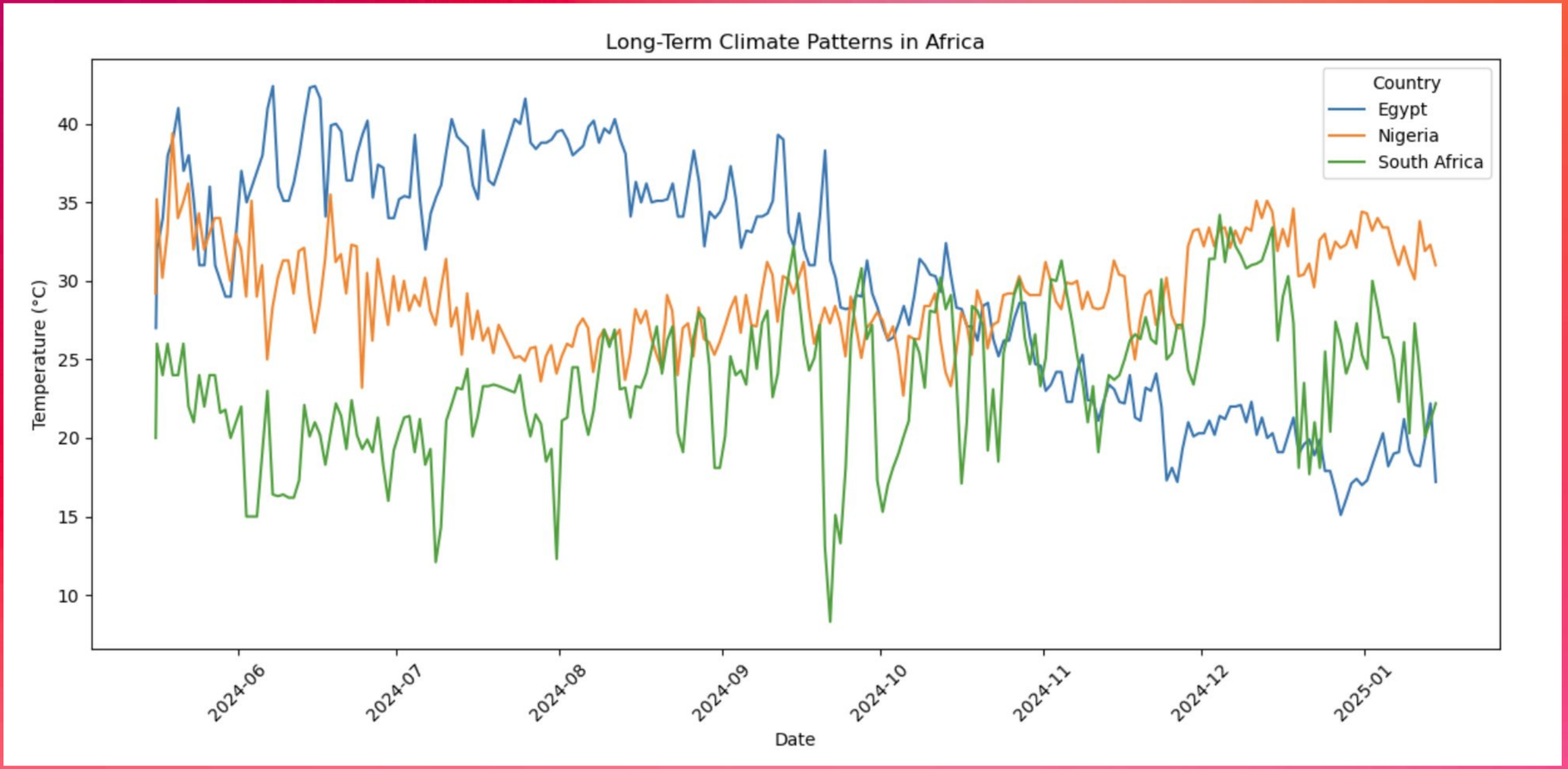
The graph shows that Argentina typically has a cooler climate than Brazil and Haiti. While all three countries experience seasonal temperature shifts(warmer in the summer and cooler in the winter) Argentina’s temperature changes are less extreme, showing that it has a more stable climate overall.

05:Climate Analysis (cont’d.)



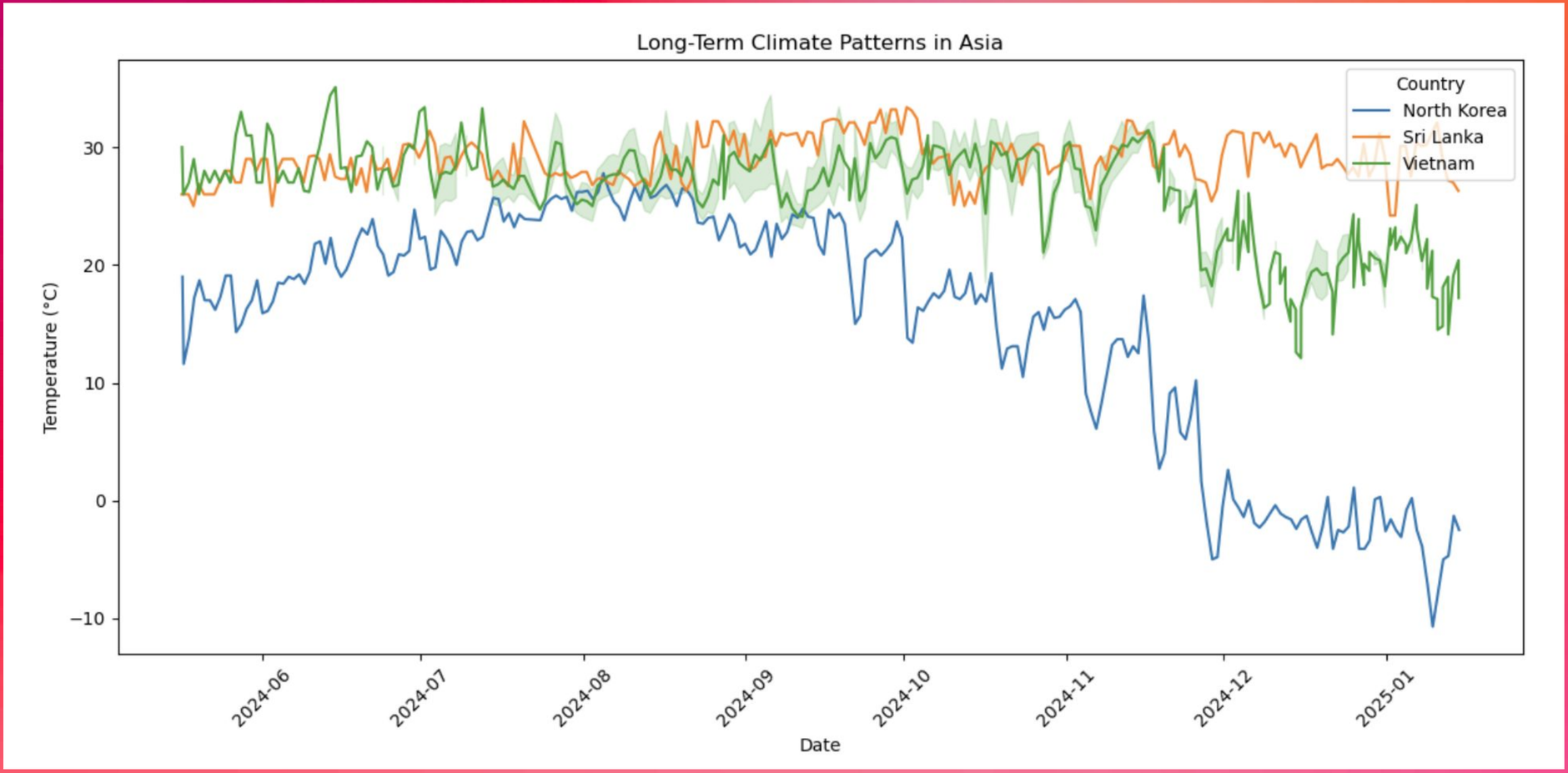
The graph shows a trend of decreasing temperatures across all three countries, France, Germany, and Italy, from June 2024 to January 2025. All three countries exhibit similar patterns with fluctuations, indicating the influence of shared climate factors in Europe.

05:Climate Analysis (cont’d.)



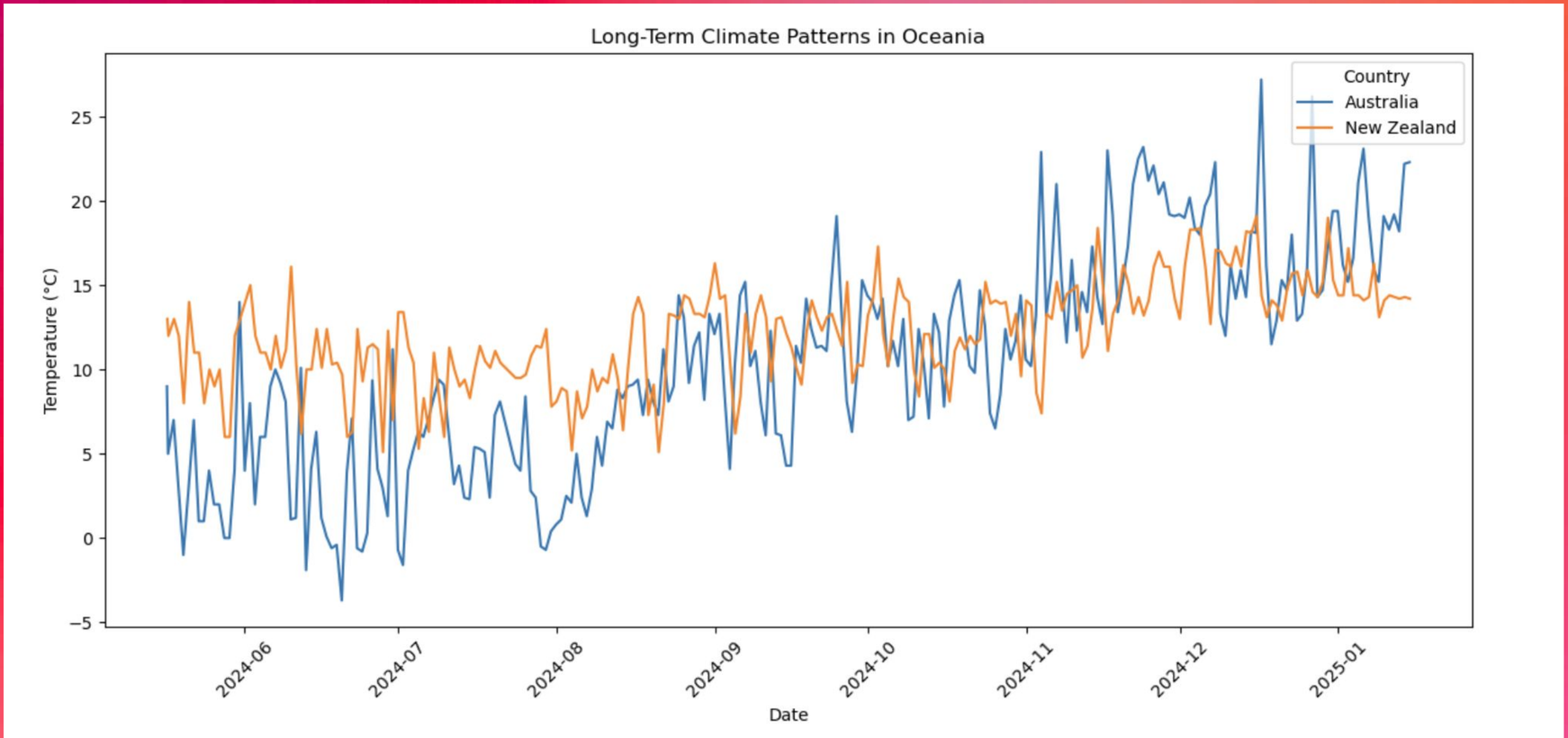
The line graph shows temperature trends over a half a year in Egypt, Nigeria, and South Africa. Egypt's temperatures are generally higher than Nigeria and South Africa, but all three countries experience temperature fluctuations throughout the year.

05:Climate Analysis (cont’d.)



The graph shows the long term climate patterns in North Korea, Sri Lanka, and Vietnam over a year, which reveals that North Korea experienced a colder year than the other two countries. Sri Lanka and Vietnam have similar average temperatures with greater fluctuations than North Korea.

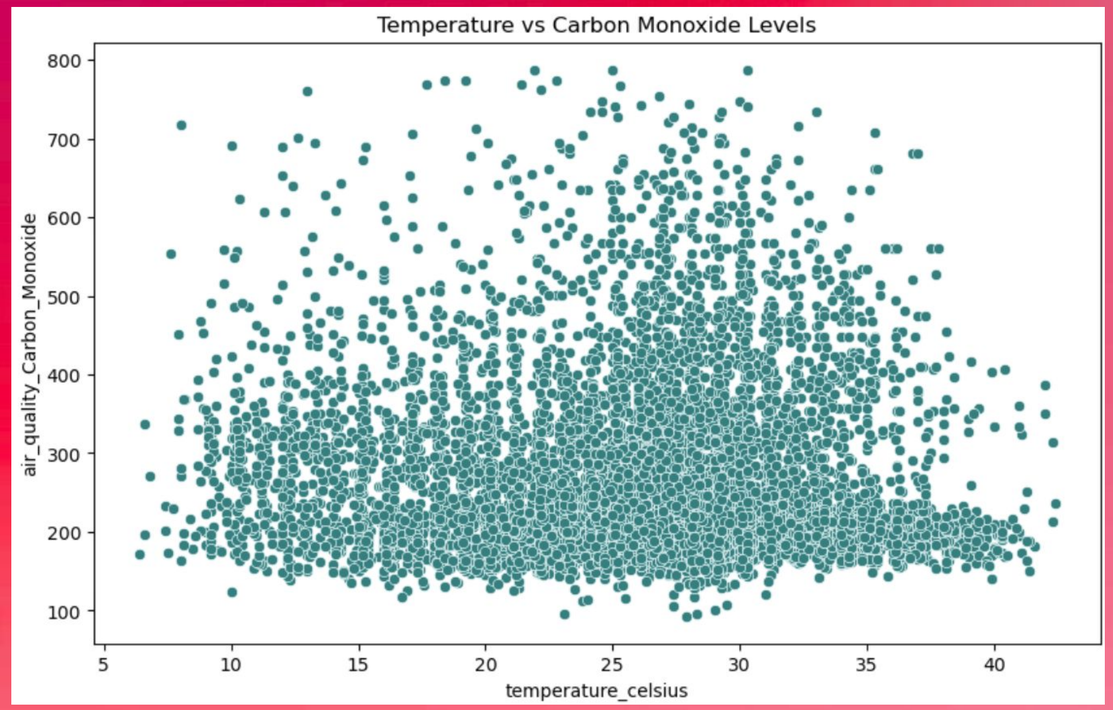
05:Climate Analysis (cont’d.)

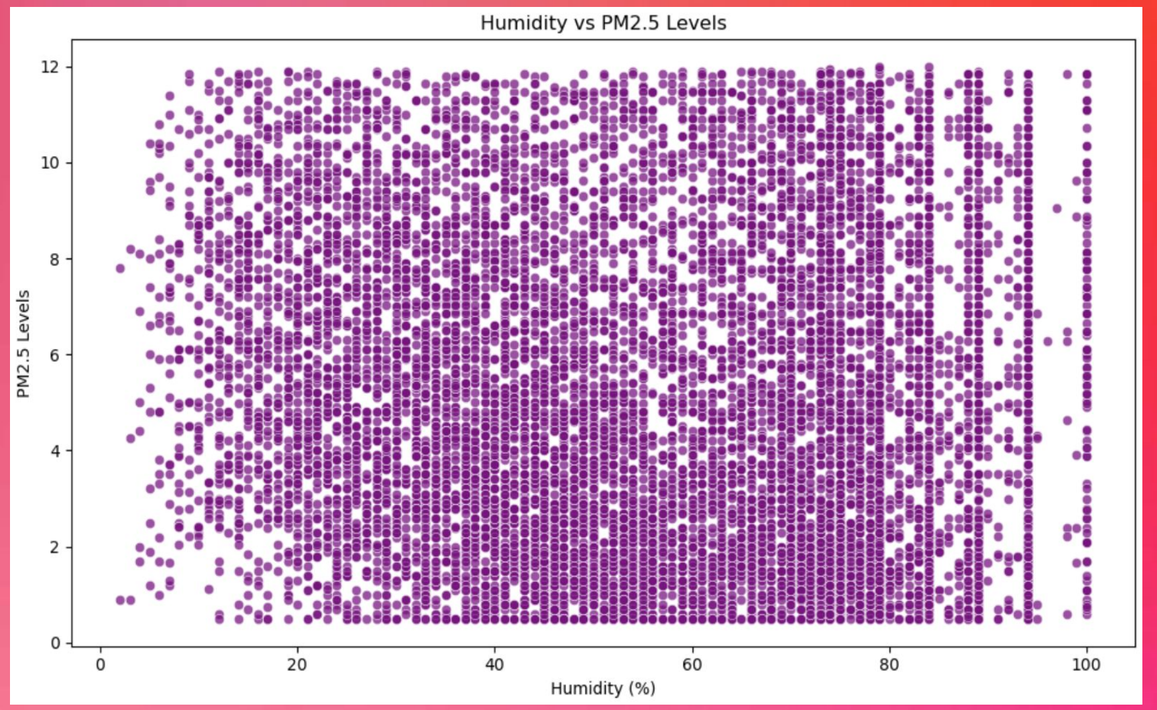


The data shows that the temperature in both Australia and New Zealand tend to fluctuate throughout the year, with both countries experiencing peaks and valleys in temperature. The overall trends suggest that New Zealand experiences a slightly warmer climate than Australia.

05:Environmental Impact Analysis

- I analyzed the relationship between air quality and weather parameters, such as temperature and humidity. The scatter plots show weak or no strong correlation between temperature and carbon monoxide levels, and humidity and PM2.5 levels.



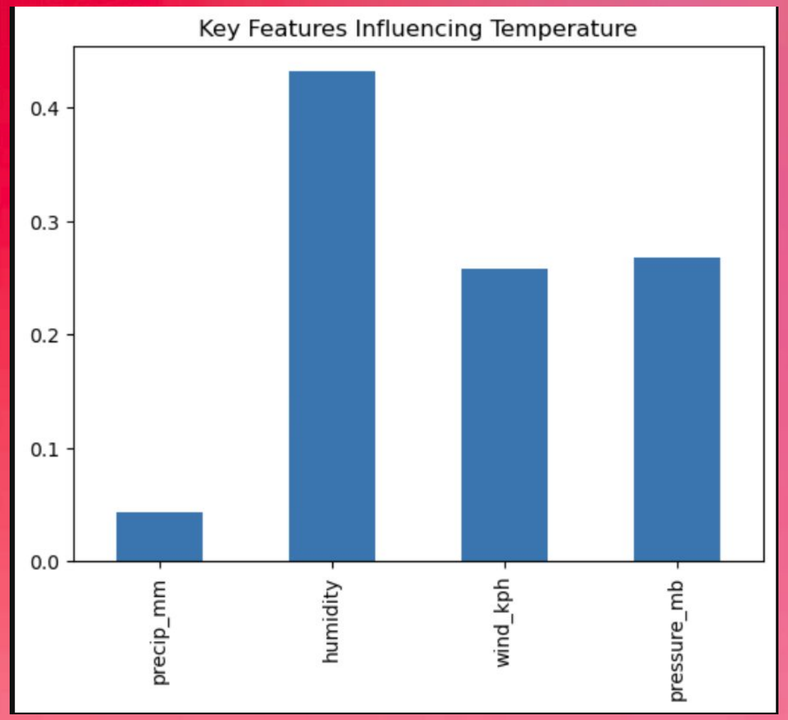


The data shows that the temperature in both Australia and New Zealand tend to fluctuate throughout the year, with both countries experiencing peaks and valleys in temperature. The overall trends suggest that New Zealand experiences a slightly warmer climate than Australia.

The data shows that the temperature in both Australia and New Zealand tend to fluctuate throughout the year, with both countries experiencing peaks and valleys in temperature. The overall trends suggest that New Zealand experiences a slightly warmer climate than Australia.

05:Advanced Feature Importance

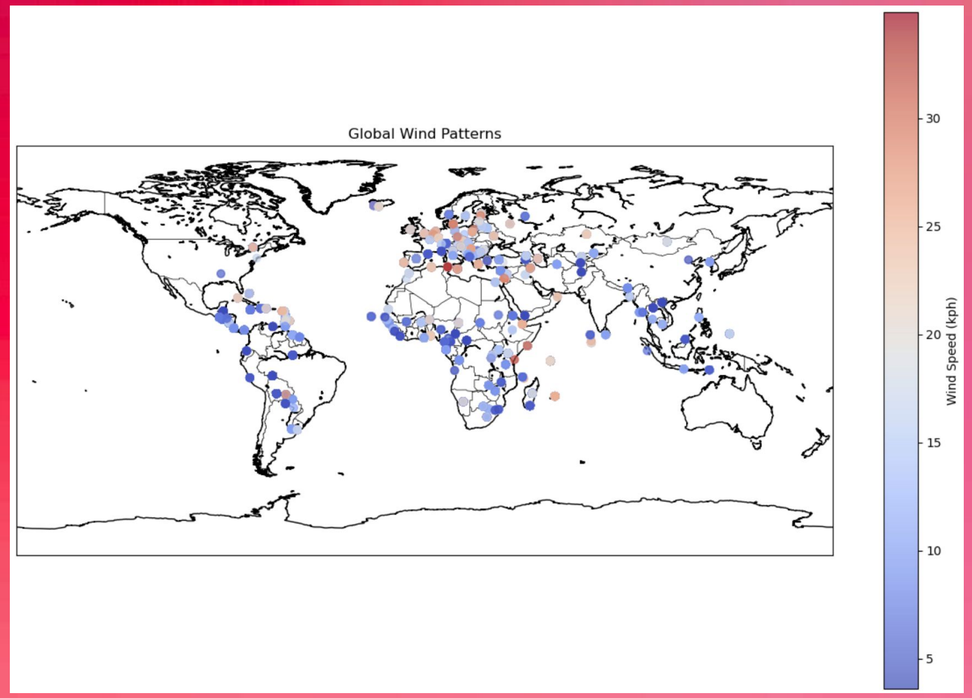
- I used a RandomForestRegressor to assess feature importance in predicting temperature. The bar plot shows how each feature (precipitation, humidity, wind speed, and pressure) contributes to the model's prediction of temperature



Humidity has the strongest influence on temperature. Wind speed and pressure have moderate influence. Precipitation has the weakest influence

05:Advanced Spatial Analysis

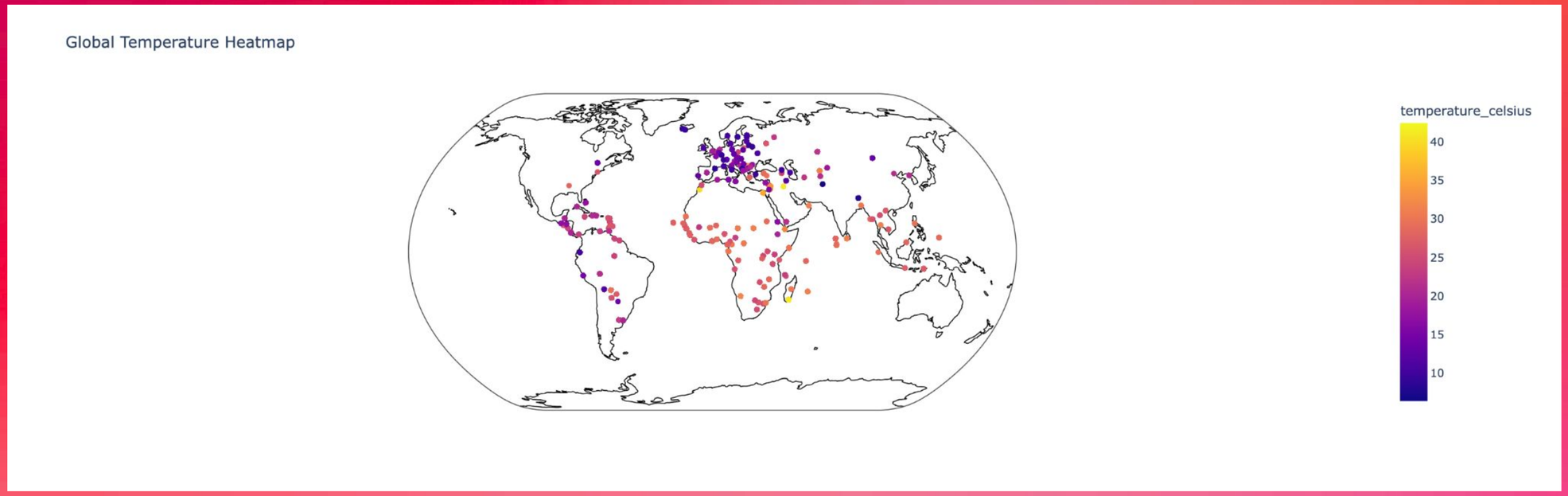
- I visualized global wind speed patterns by plotting wind speed data on a world map. I used latitude and longitude coordinates to place wind speed data points, with colors showing wind speed variations across different regions.



The map shows that wind speeds tend to be higher in the Northern Hemisphere, with some areas reaching over 30 kph. Whereas the Southern Hemisphere generally has lower wind speeds, mostly between 10 and 20 kph, showing differences in how air circulates between the two hemispheres.

05:Advanced Spatial Analysis

- I generated an interactive global temperature heatmap using plotly. It shows temperature variations across the world, with yellow representing the hottest areas and purple for the coldest regions, highlighting high temperatures in parts of Africa, the Middle East, and Australia.



The heatmap displays global temperatures, with yellow indicating the hottest areas primarily in the tropics, while purple represents the coldest regions. It is interesting to note the high temperatures in parts of Africa, the Middle East, and Australia.

Thank you!

I am looking forward to the possibility of applying my data science skills on the PM Accelerator team!