**WEATHER FORCASTING REPORT**

**PMA 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.

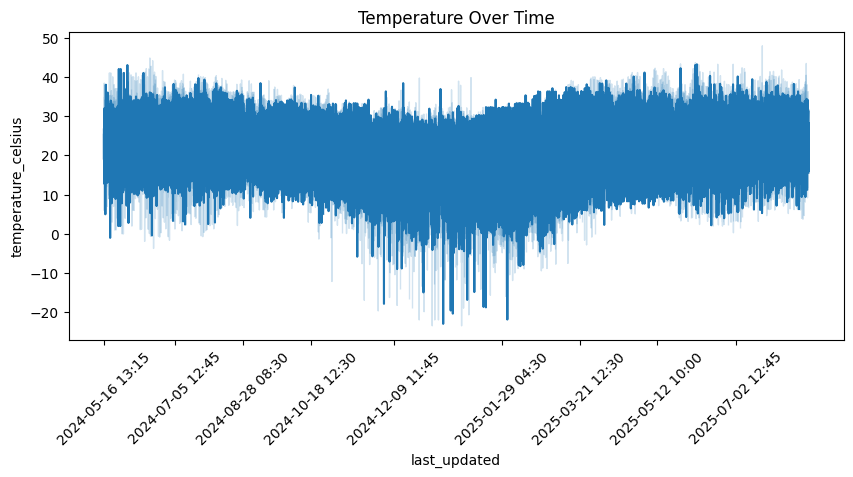
**Dataset Description: -**

The dataset I used comprises daily weather records for capital cities across the globe, beginning on August 29, 2023. It contains over 40 meteorological and environmental features—including temperature, wind speed, pressure, precipitation, humidity, visibility, and air quality—enabling comprehensive analysis of global weather patterns, climate trends, and correlations between various atmospheric parameters.

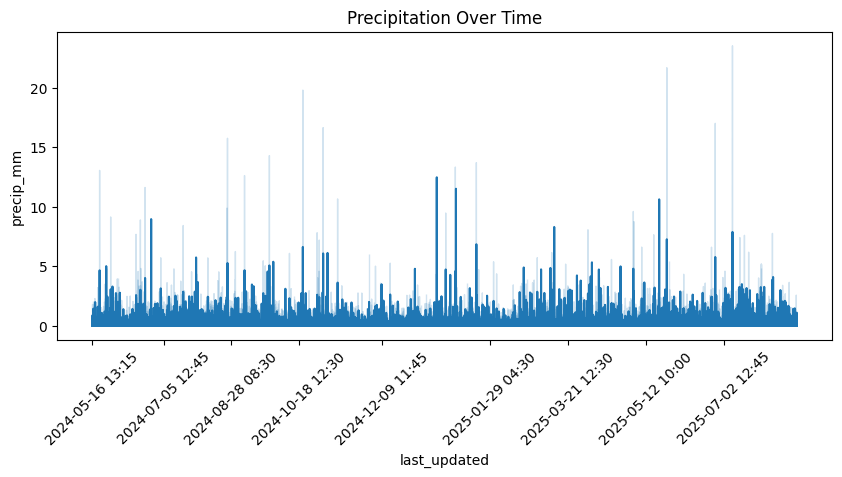
**Basic Visualisation and Model Training: -**

We will do Data cleaning, EDA Analysis and the train the model to check the predictions.

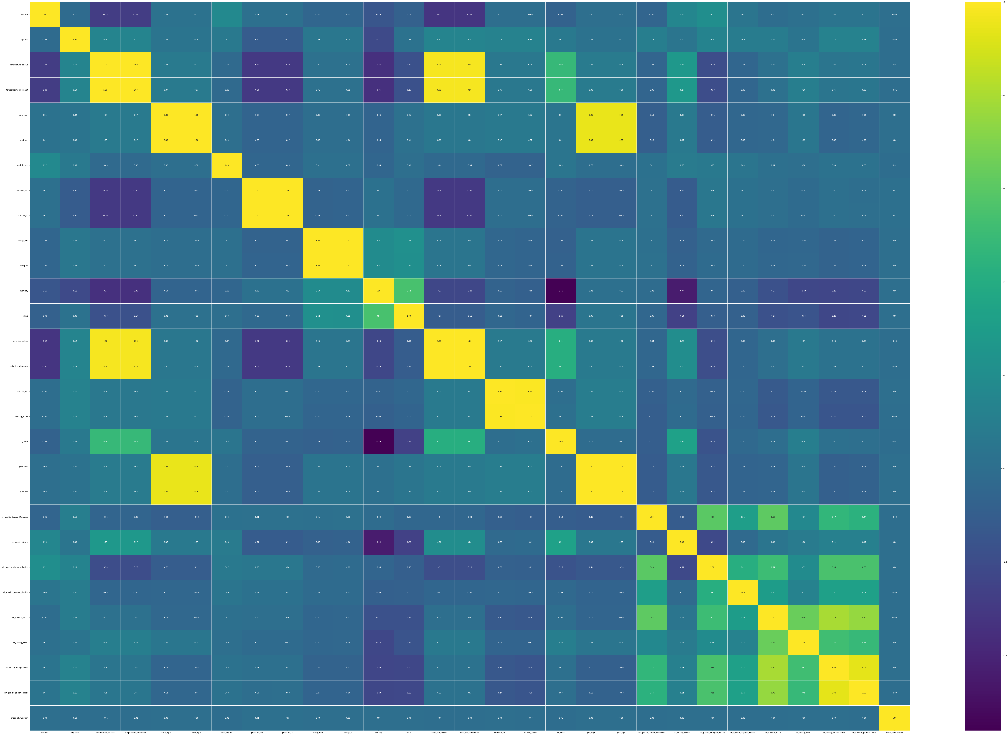
* **Missing Values Check**
  + The df.isnull().sum() output confirms no missing values in any column.
  + This indicates the dataset is clean and ready for analysis without imputation.
* **Data Visualisation**
  + This is the output of Temperature in Celsius with different date and time ranges



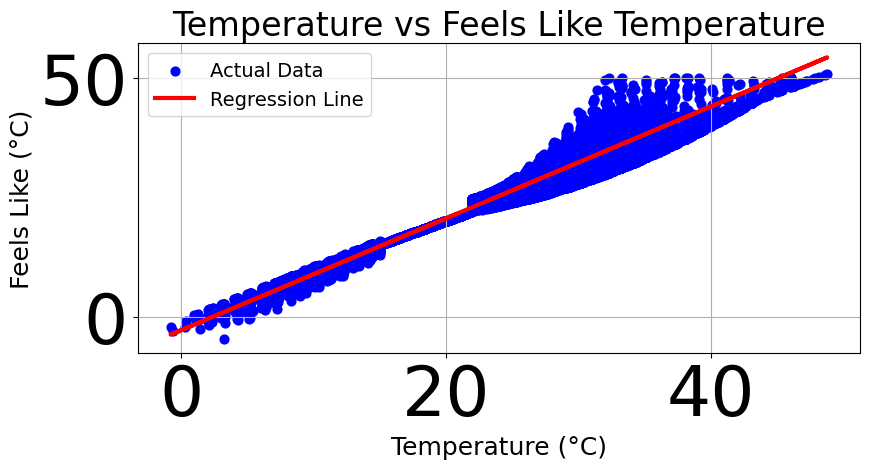
* + This is the output of Precipitation with different date and time ranges



* + This plot is a correlation heatmap of numerical features in your dataset.
    - Each square shows the correlation coefficient (from -1 to +1) between two variables.
    - Yellow = strong positive correlation (values move together), Dark purple = strong negative correlation (values move in opposite directions), and Green/Blue = weak or no correlation.
    - The diagonal is always yellow (correlation of each variable with itself = 1).



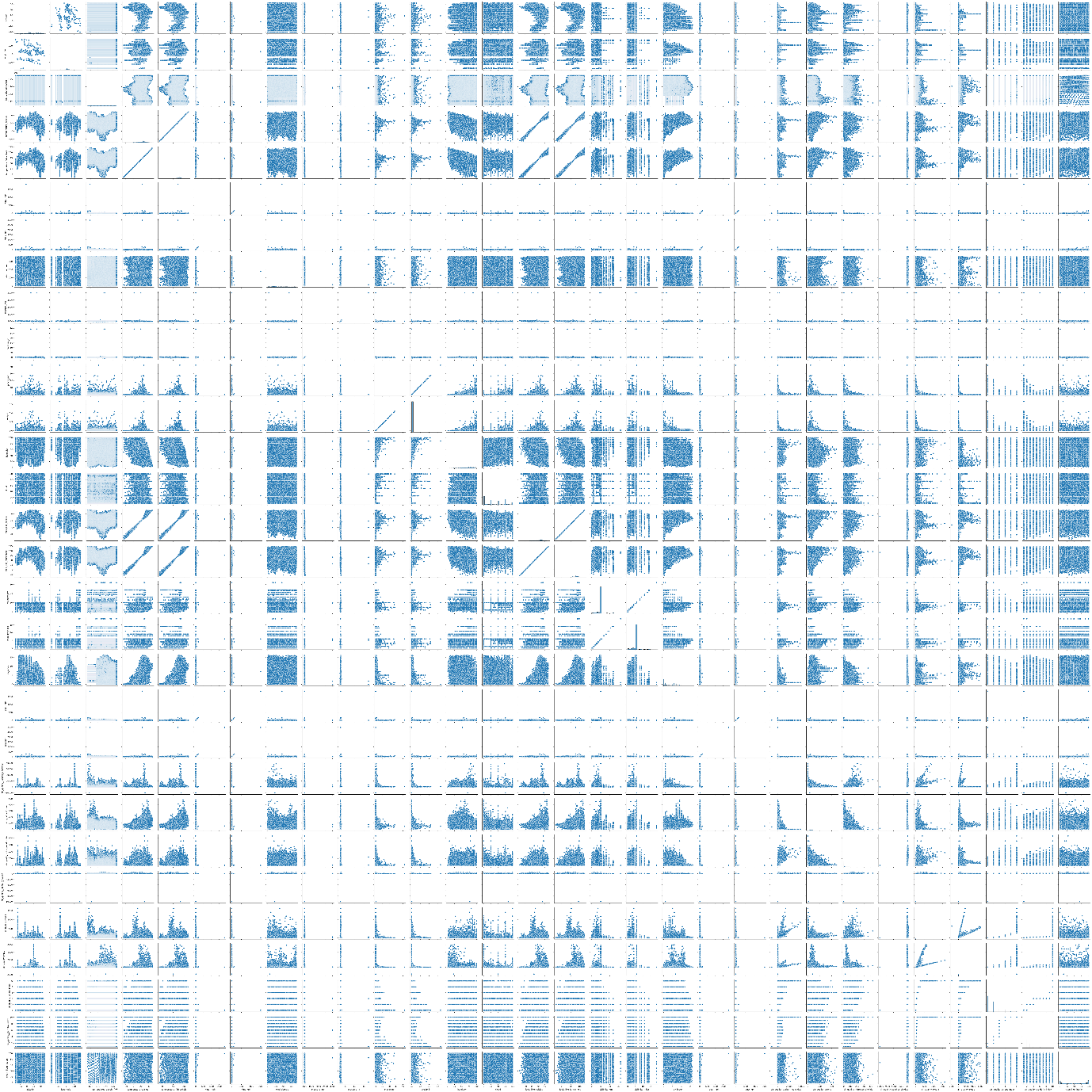
* **Train-Test Split & Model Training**
  + The dataset is split into training and testing sets for temperature prediction.
  + The model used is evaluated on unseen data to measure performance.
* **Model Performance Metrics**
  + Results show R² = 0.927, indicating the model explains ~93% of temperature variance.
  + MAE ≈ 1.65°C, meaning predictions are on average within ~1.6°C of actual values.
* **Performance Visualisation**
  + This graph shows the relationship between actual air temperature and "feels like" temperature (heat index), with blue dots representing real data points and a red regression line showing the overall trend. At lower temperatures the actual and feels-like temperatures are nearly identical, but as temperature increases above 20°C, the feels-like temperature becomes progressively higher than the actual temperature due to factors like humidity making hot weather feel even hotter.



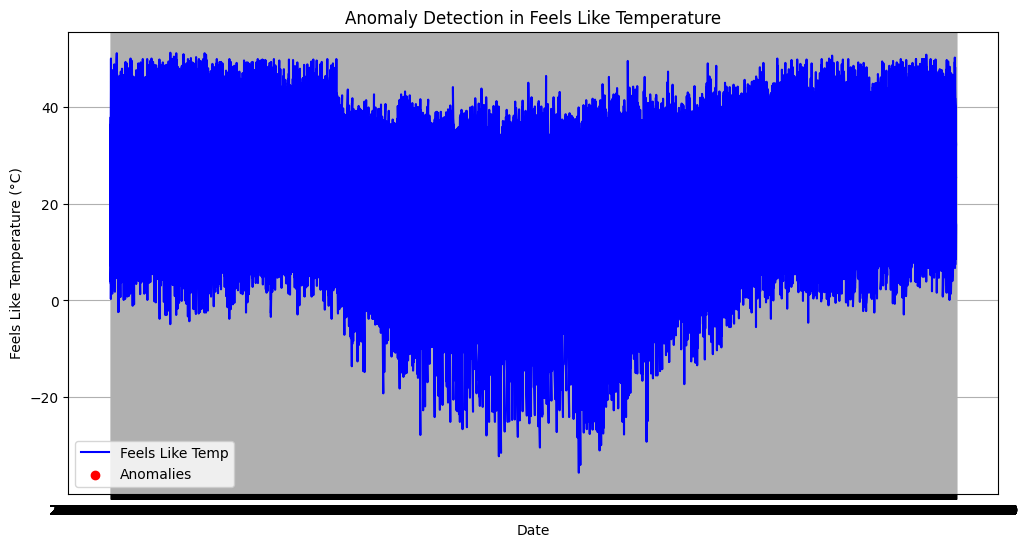
**Advance Visualisation and Training: -**

In this we did some advance EDA Analysis and train different models and evaluate them and then visualise the best model performance.

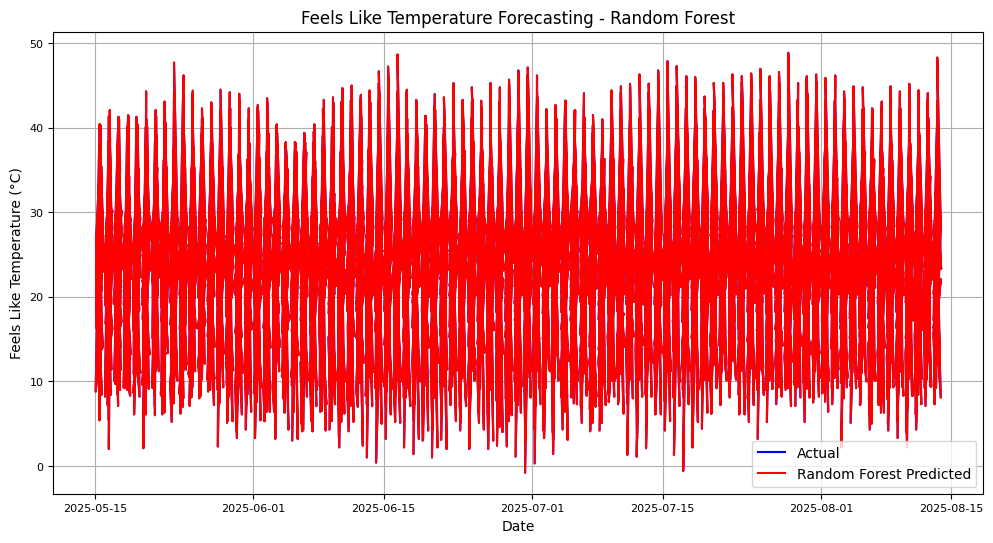
* **Comprehensive Correlation Matrix**
  + This is a comprehensive correlation matrix or pairs plot showing relationships between multiple variables in a dataset, with each small panel displaying either scatter plots (showing correlations between different variable pairs) or distribution plots (along the diagonal). The blue dots and patterns reveal the strength and nature of correlations across all possible variable combinations, allowing for quick identification of which variables are strongly related, weakly related, or uncorrelated with each other.



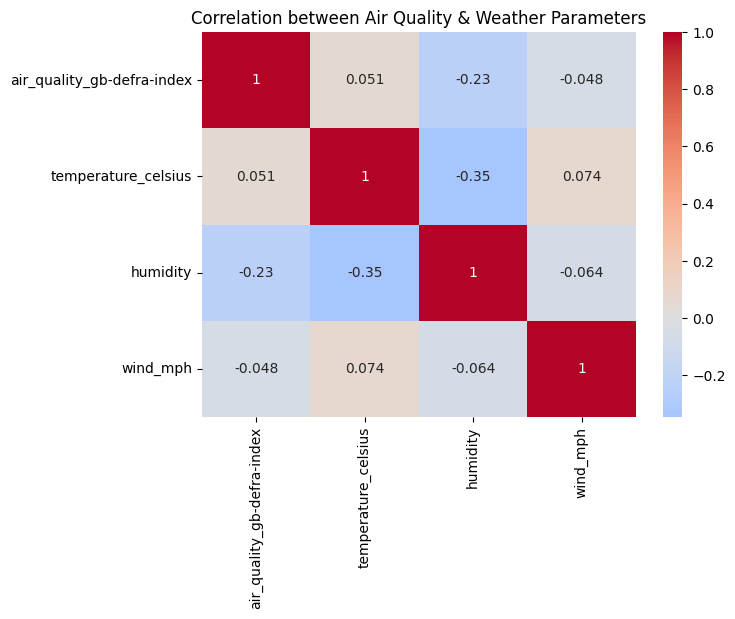
* **Anomalies Detection**
  + This graph shows anomaly detection in "feels like" temperature data over time, where the blue line represents the temperature values and red dots (if any) would indicate detected anomalies or outliers from normal temperature patterns.



* **Model Trainings**
  + We trained four different machine learning models (Random Forest, Linear Regression, Gradient Boosting, and XGBoost) to predict temperature data, with all models achieving exceptionally high accuracy (R² > 99.9%). Random Forest performed best with the lowest MAE (0.017) and RMSE (0.030), followed closely by Linear Regression, while XGBoost had the highest errors despite still maintaining very strong predictive performance.
  + This graph shows the Random Forest model's forecasting performance for "feels like" temperature from May to August 2025, where the red predicted values closely track the black actual values, demonstrating the model's high accuracy in temperature prediction.



* **Correlation between Air Quality and Weather Parameters**
  + This correlation heatmap shows that air quality has a moderate negative correlation with humidity (-0.23) and temperature (-0.35), suggesting that higher humidity and temperature are associated with poorer air quality, while correlations with wind speed are negligible.



* **Feature Importance of Air Quality Index**
  + This bar chart shows that temperature is the most important feature for predicting air quality (importance ~0.42), followed by wind speed and humidity which have roughly equal but lower importance (~0.28-0.30 each).

