# Global Weather Forecasting Project Report

## 1. Introduction

Project Overview:  
 This project focuses on forecasting global weather conditions by applying robust data cleaning, exploratory data analysis (EDA), and multiple forecasting techniques. The analyses cover both basic tasks—such as handling missing values and generating trend graphs—and advanced assessments, including anomaly detection, multi-model forecasting, and spatial analysis.

Company 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.”*

Purpose of the Report:  
 This report documents the entire workflow from data preprocessing and visualization to model building and evaluation. It also includes unique analyses like climate pattern studies, environmental impact assessment, and spatial distribution analysis. The report is structured to demonstrate technical competence and to align with the company’s mission of empowering future product management leaders.

## 2. Data Cleaning & Preprocessing

### 2.1 Data Acquisition and Initial Inspection

Data Source: GlobalWeatherRepository.csv

#### Key Steps:

* + Imported necessary libraries and installed dependencies.
  + Loaded the dataset and performed an initial inspection using functions like .info(), .describe(), and .head().
  + Converted date columns (e.g., last\_updated) into datetime format to facilitate time series analysis.

### Summary Table



## 2.2 Handling Missing Values and Outliers

### Missing Values:

* + Identified missing values with. isna(). sum() and detected dates with null entries.
  + Handled missing data by copying previous day’s values for specific dates.

### Outlier Handling:

* + Since every value contributed to the analysis, no aggressive outlier removal was applied.

## 2.3 Feature Engineering and Data Transformation

### Column Selection and Removal:

* + Removed redundant features (e.g., Fahrenheit values since Celsius was used as the target).
  + Selected essential features such as latitude, longitude, temperature (in Celsius), and various weather parameters.

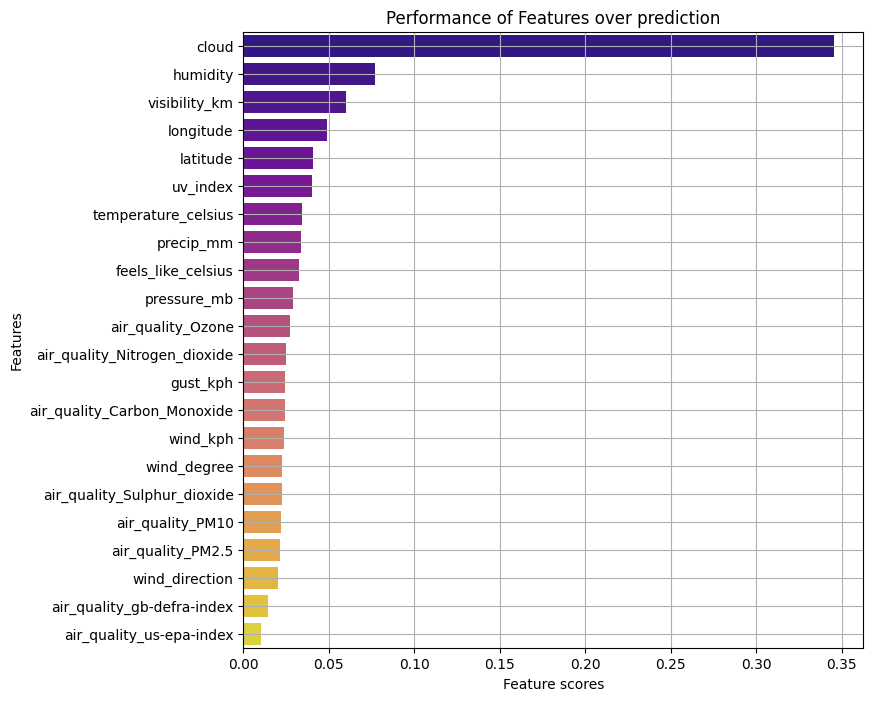
## Categorical Encoding:

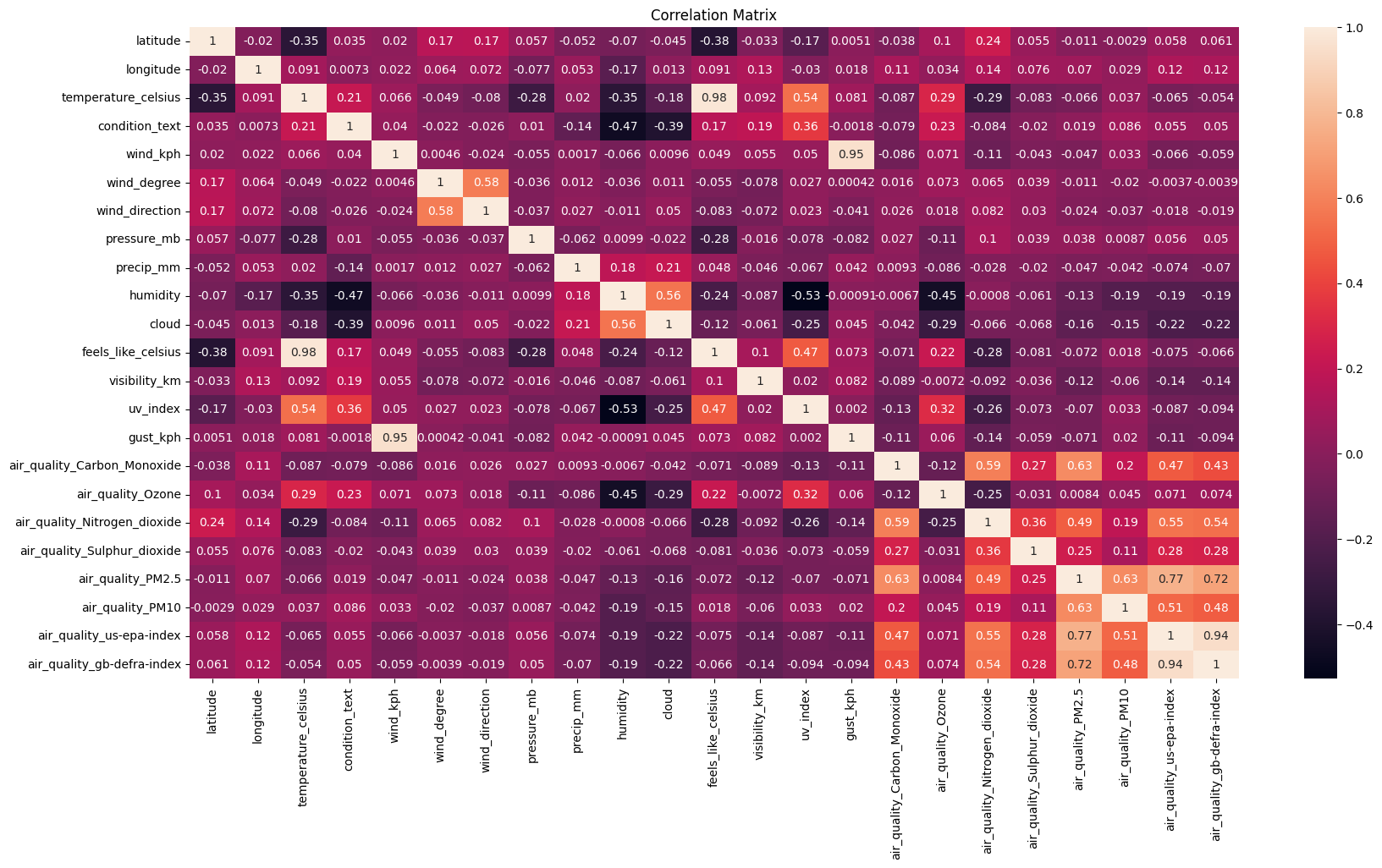
* + Applied label encoding to categorical features (e.g., condition\_text and wind direction) to convert them into numerical format.

## Normalization and Scaling:

* + Utilized scaling techniques (MinMaxScaler) to normalize numerical features as needed.

**Feature Performance Graph(Using ExtraTreesClassifier model)**



**Corelation with the target value (condition text)** 

## 3. Exploratory Data Analysis (EDA)

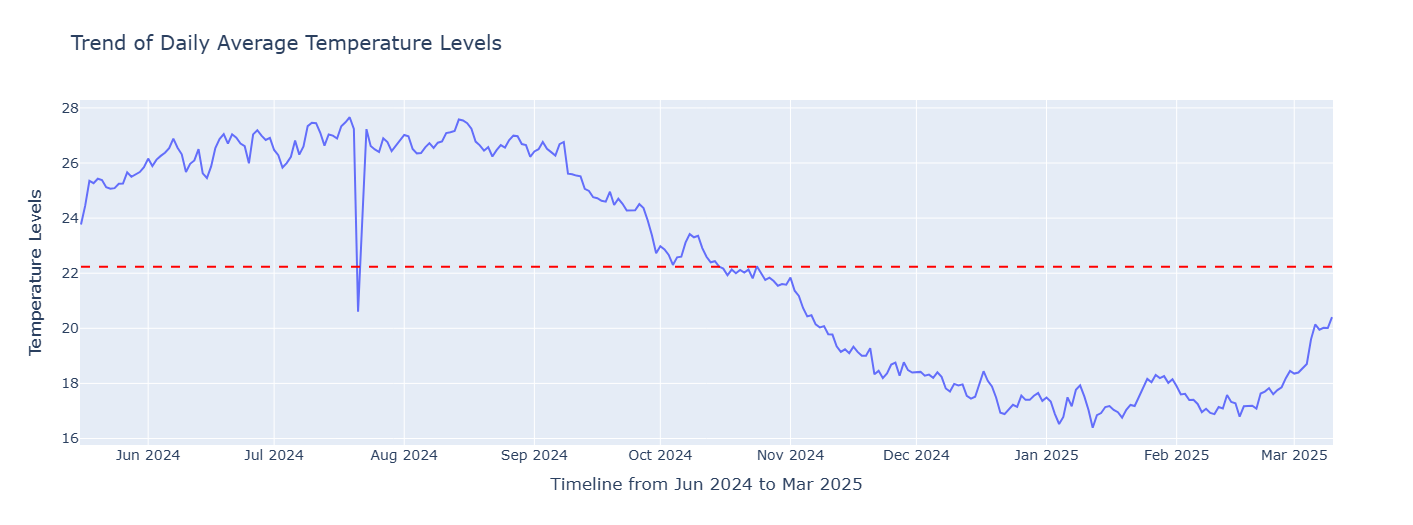
### 3.1 Trend Analysis of Weather Parameters

The EDA involved generating multiple visualizations to understand trends and patterns over time.

##### Temperature Trends:

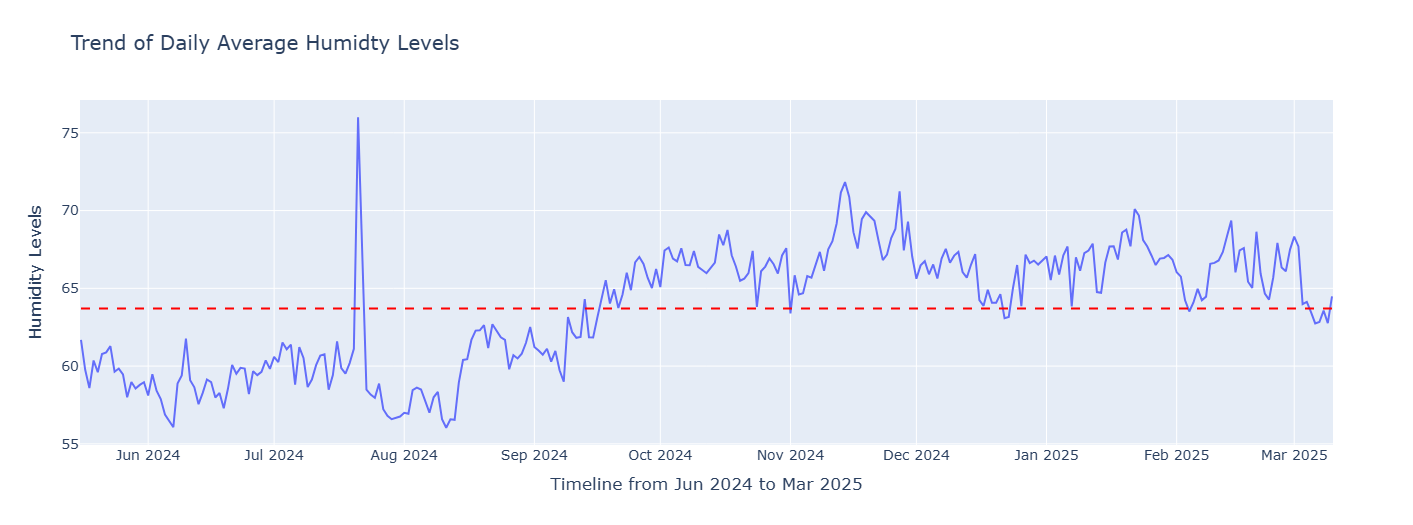
* + Daily averages of temperature were computed.
  + A Plotly line graph with a horizontal line representing the overall average temperature was produced.

**Graphical Representation (Refer colab for better visualization)**



##### Humidity Trends:

* + Similar analysis was performed on humidity levels with daily averages and overall mean.
* **Graphical Representation (Refer colab for better visualization)**

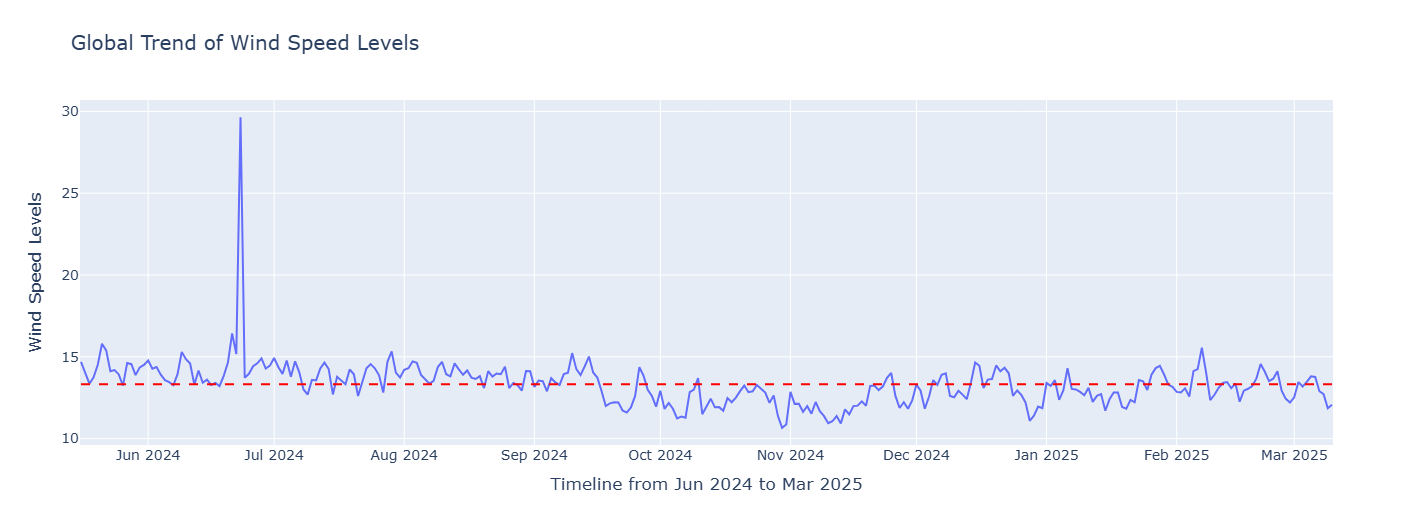


##### Wind Speed and Wind Direction:

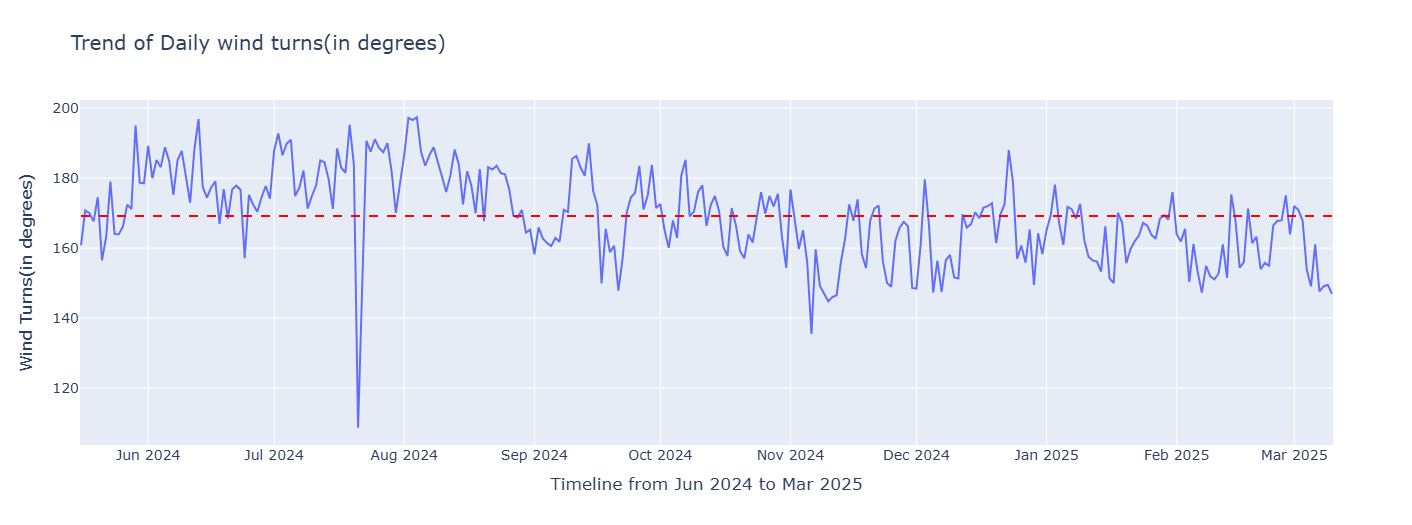
* + Trend analysis was conducted on wind speed (kph) and wind degree (direction) with appropriate visualizations.

**Graphical Representation (Refer colab for better visualization)**

**Wind Speed Trend Over Years:**

****

**Wind Degree Turn Trend Over Years:**

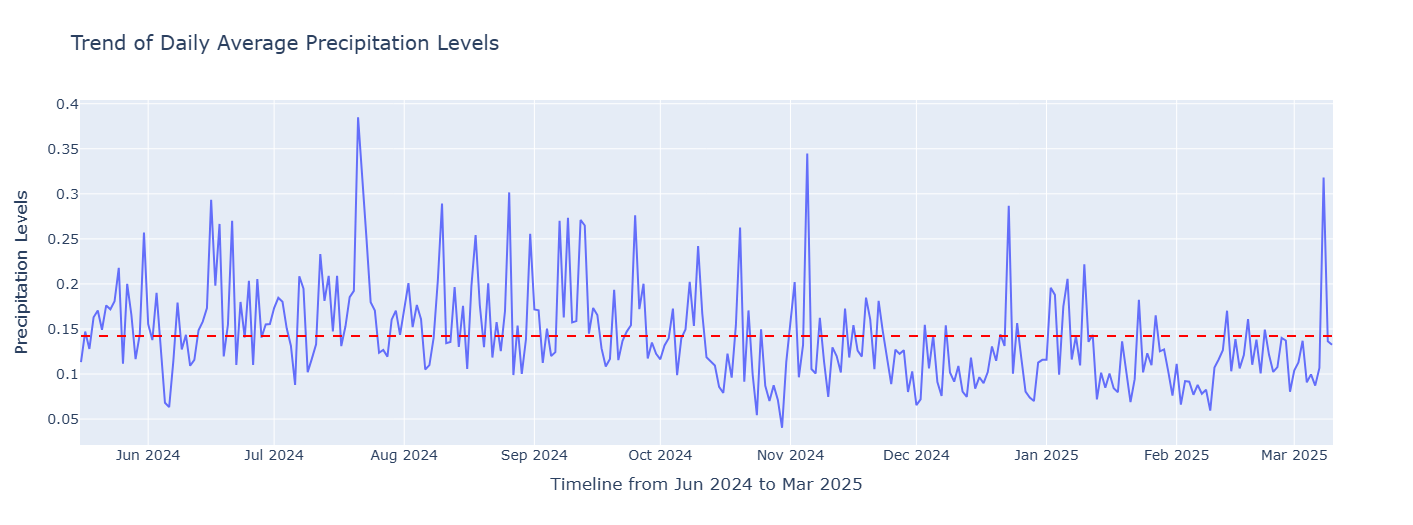
****

##### Precipitation Trends:

* + Daily precipitation levels were analyzed with annotated overall averages.

**Graphical Representation (Refer colab for better visualization)**

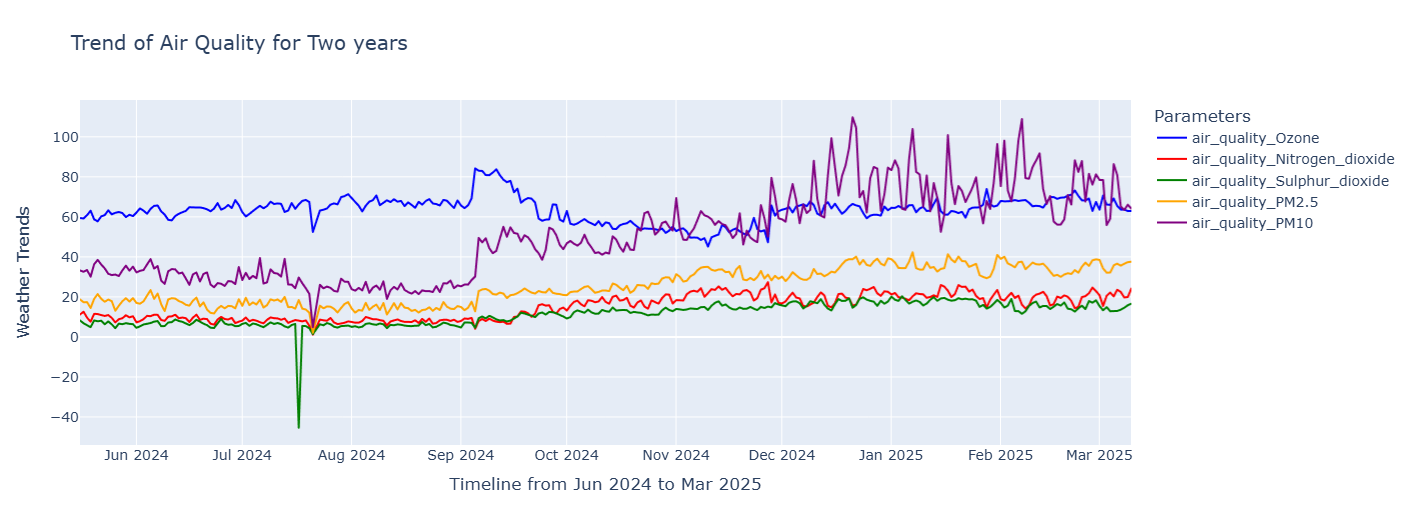
**Average Precipitation over years:**



##### Air Quality Trends:

* + Trends for various air quality indicators (Carbon Monoxide, Nitrogen Dioxide, Ozone, etc.) were examined.
* **Graphical Representation (Refer colab for better visualization)**

**Air Quality Over years:**

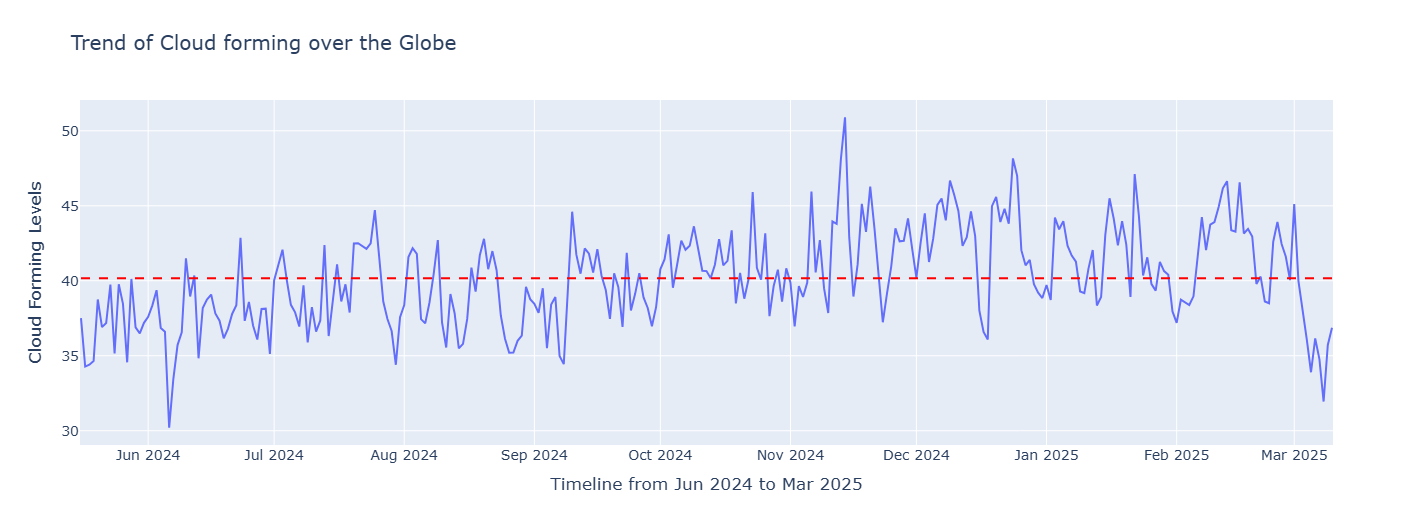


##### Cloud Formation Trends:

* + Analyzed daily average cloud levels and compared them to overall averages.

**Graphical Representation (Refer colab for better visualization)**

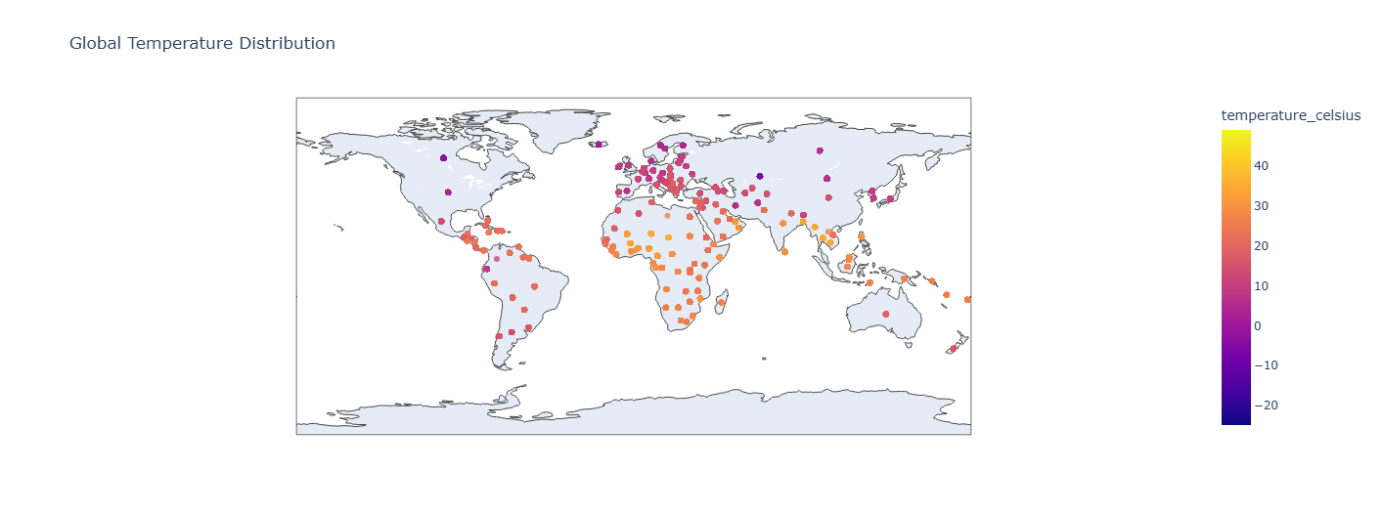
**Cloud Formation Trend Over Years:**

****

## 3.2 Spatial Analysis

### Global Distribution of Weather Conditions:

* + A scatter geo-plot was created using latitude and longitude to visualize global temperature distribution



## 4. Model Building and Forecasting

### 4.1 Basic Forecasting Model with XGBoost

##### Data Preparation:

* + Created a time-series dataset by resampling daily averages.
  + Split the dataset into training and testing sets.

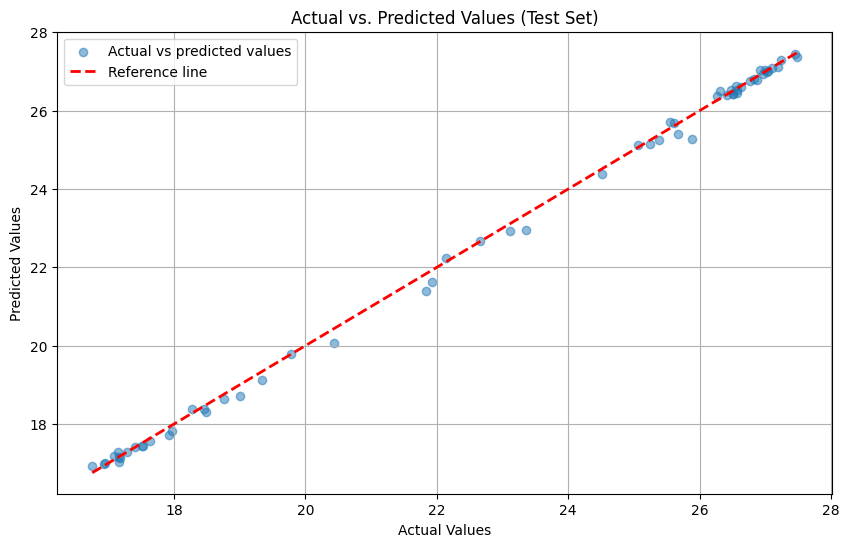
##### Model Training:

* + Built an XGBoost regression model to predict temperature\_celsius.
  + Evaluated model performance using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

##### Results Visualization:

* + A scatter plot was generated to compare actual versus predicted values.

**Actual Vs Prediction Performance Graph:**



## 4.2 Time Series Forecasting using ARIMA and SARIMA

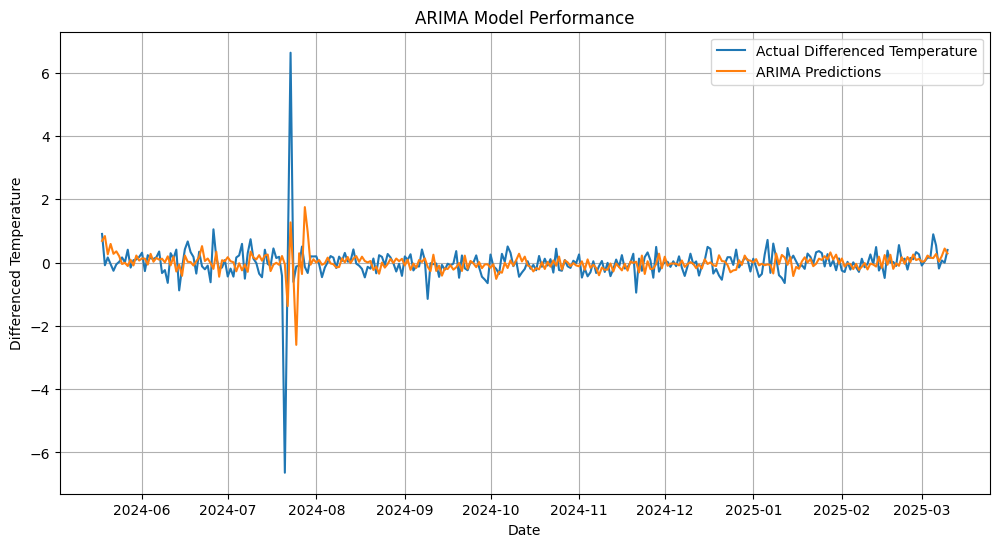
### Stationarity Testing:

* + Conducted the Dickey-Fuller test to evaluate stationarity.
  + Applied differencing to achieve a stationary time series.

### ARIMA Forecasting:

* + Developed and fitted an ARIMA model on the differenced data.
  + Computed performance metrics (MSE, RMSE) and visualized the model’s performance.

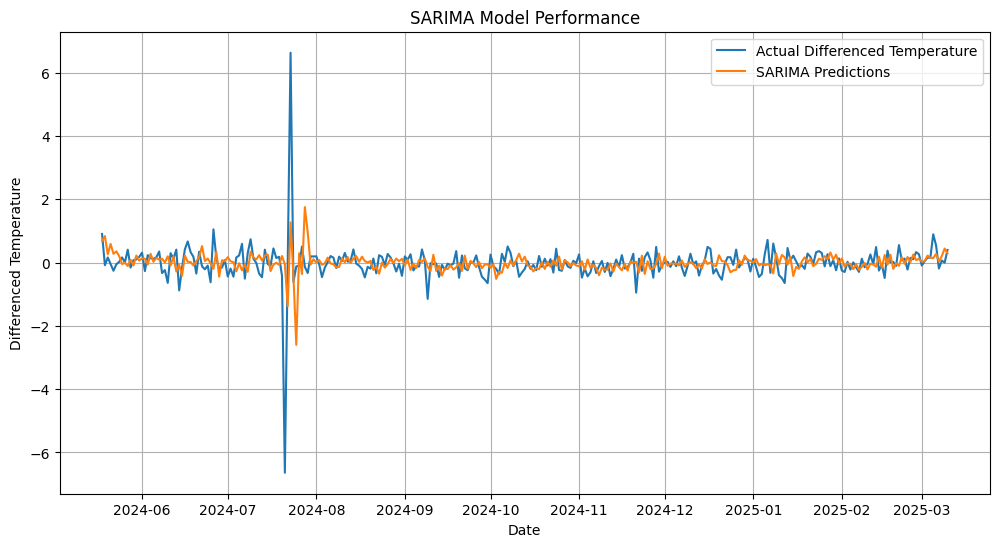
**ARIMA Model Performance (Actual Vs Predicted):**



### SARIMA Forecasting:

* + Built a SARIMA model with seasonal parameters to capture seasonal trends.
  + Evaluated and compared performance with the ARIMA model.

**SARIMA Model Performance (Actual Vs Predicted):**



## 4.3 Forecasting with Facebook Prophet

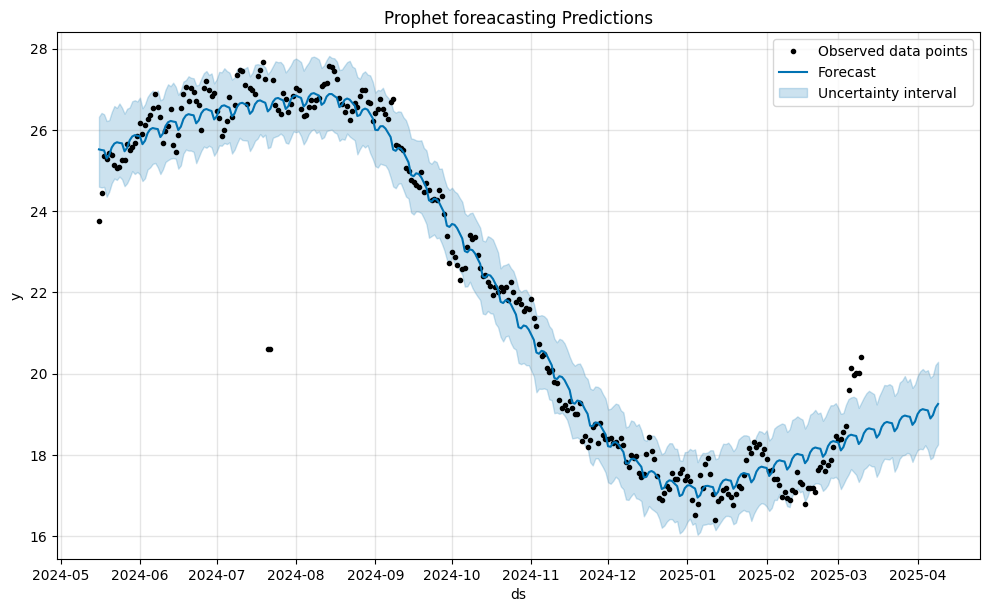
### Prophet Model Implementation:

* + Prepared the dataset according to Prophet’s requirements (renaming columns to ‘ds’ and ‘y’).
  + Fitted the model and generated future forecasts.

### Visualization:

* + Produced a forecast plot (including trend and seasonality components if available).

**Actual Vs Predicted weather forecasting using Facebook Prophet:**



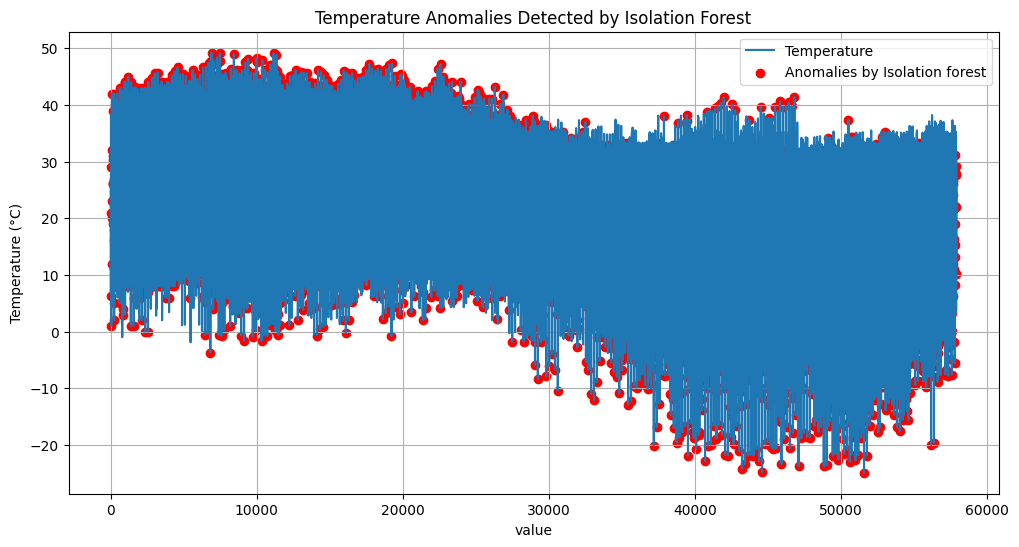
## 5. Advanced Analysis

### 5.1 Advanced Exploratory Data Analysis

##### Anomaly Detection:

* + Implemented an Isolation Forest algorithm to detect anomalies in temperature and related features.
  + Visualized anomalies on a time series plot where outliers were highlighted.

**Anomalies Detected on Temperature by Isolation Forest:**



## 5.2 Forecasting with Multiple Models and Ensemble Techniques

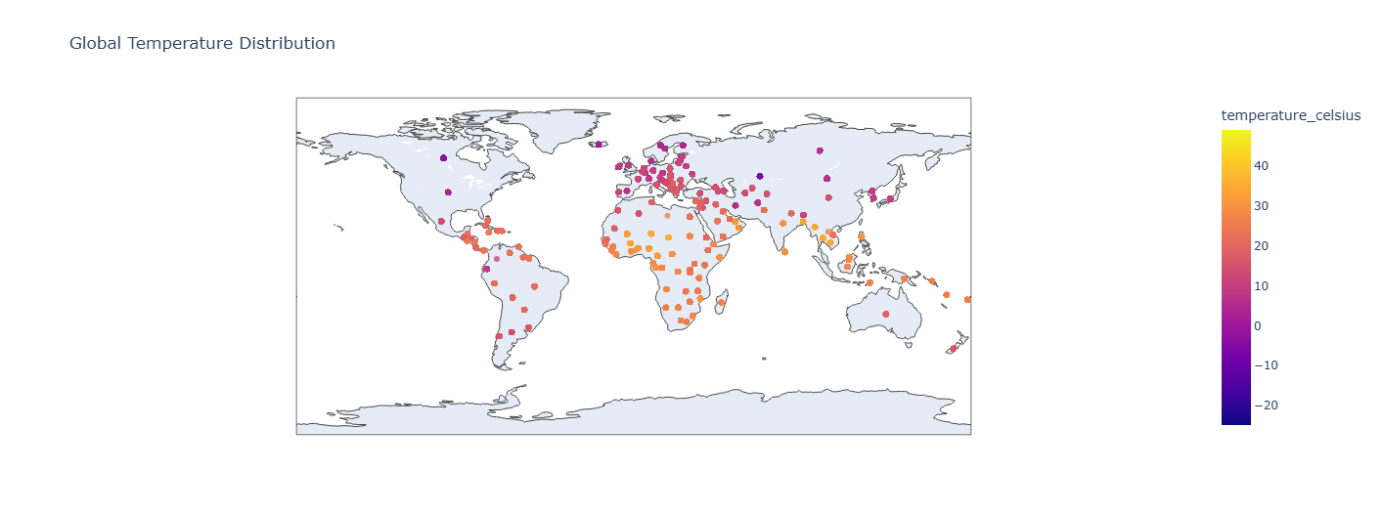
### Multi-Model Comparison:

* + Built multiple forecasting models (XGBoost, ARIMA, SARIMA, Prophet) and compared their performance.
  + Considered ensemble techniques to combine model predictions for improved accuracy.

## 5.3 Unique Analyses

### Climate Analysis:

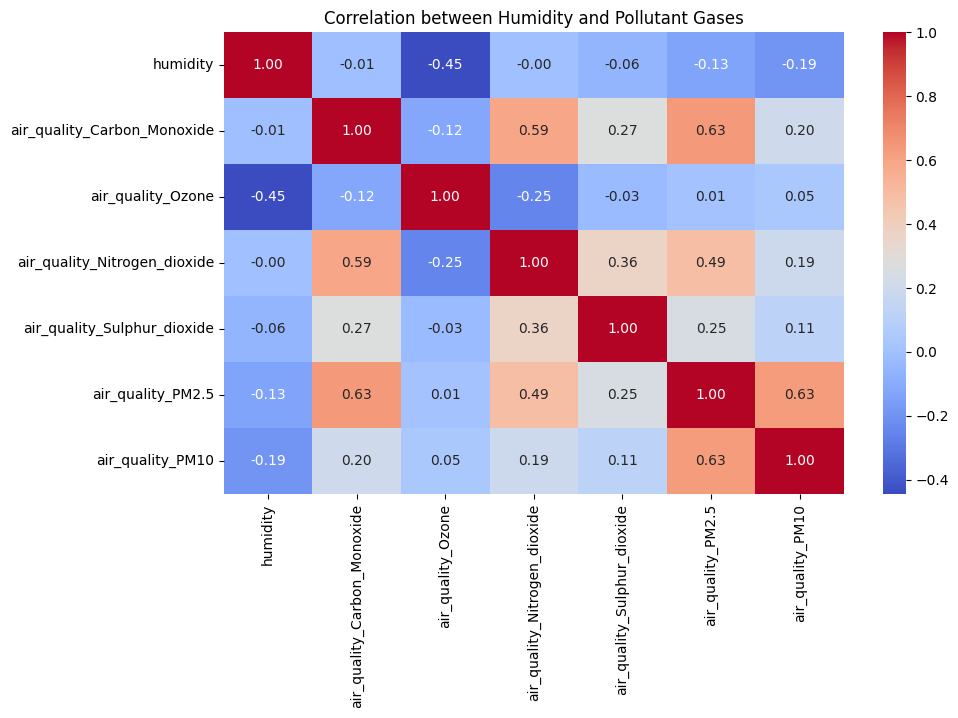
* + Explored long-term climate patterns and variations across different regions.



### Environmental Impact Assessment:

* + Studied the relationship between weather parameters and air quality, analyzing how environmental factors correlate.

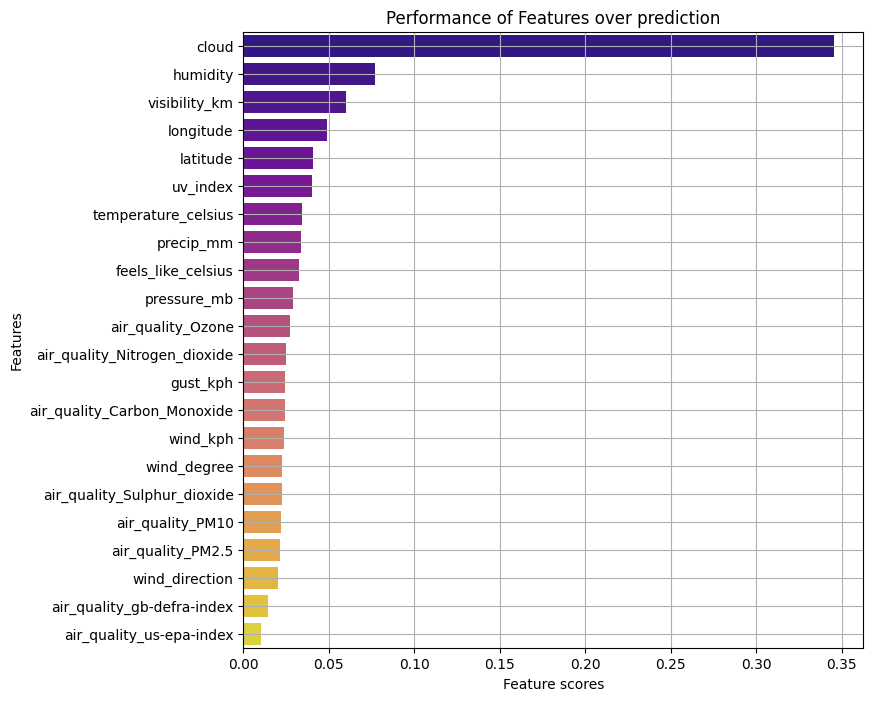
**Corelation between Humidity and Pollutant Gases:**



### Feature Importance Analysis:

* + Applied an ExtraTreesClassifier to assess the importance of various features.
  + Visualized the feature importance rankings to identify key predictors.

**Performance of features over prediction:**



## 6. Conclusion

### Key Findings:

* **Data Cleaning & Preprocessing:**  
   Successfully managed missing values and refined the dataset, ensuring robust input for subsequent analysis.
* **Exploratory Data Analysis:**  
  Identified significant trends in temperature, humidity, wind, and air quality which informed the choice of forecasting models.
* **Model Performance:**  
  Forecasting models (XGBoost, ARIMA, SARIMA, Prophet) produced varying degrees of accuracy, with ensemble methods showing potential for improved predictions.
* **Advanced Analyses:**  
  Anomaly detection and spatial analysis offered deeper insights into regional variations and environmental impacts.

**Final Note:**  
This project demonstrates a comprehensive workflow—from data acquisition to advanced forecasting and unique environmental analyses—in line with the company’s mission of providing industry-leading tools and education to empower future PM leaders.