PM Accelerator Internship Assessment

Weather Trend Forecasting

Analyzing global weather patterns and forecasting trends using AI & ML

PM Accelerator Mission Statement

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. Introduction

Objective:

The goal of this project is to analyze the "Global Weather Repository.csv" dataset to forecast future weather trends using statistical and machine learning techniques. The dataset contains daily weather information for cities worldwide, with over 40 features representing various weather conditions.

Key Deliverables:

- **Data Cleaning & Preprocessing**
- **Exploratory Data Analysis (EDA)**
- **Forecasting using ARIMA, Holt-Winters, and Ensemble Models**
- **Advanced Analysis: Anomaly Detection, Feature Importance, and Spatial Analysis**

- **Final Report & GitHub Submission**

---

2. Data Cleaning & Preprocessing

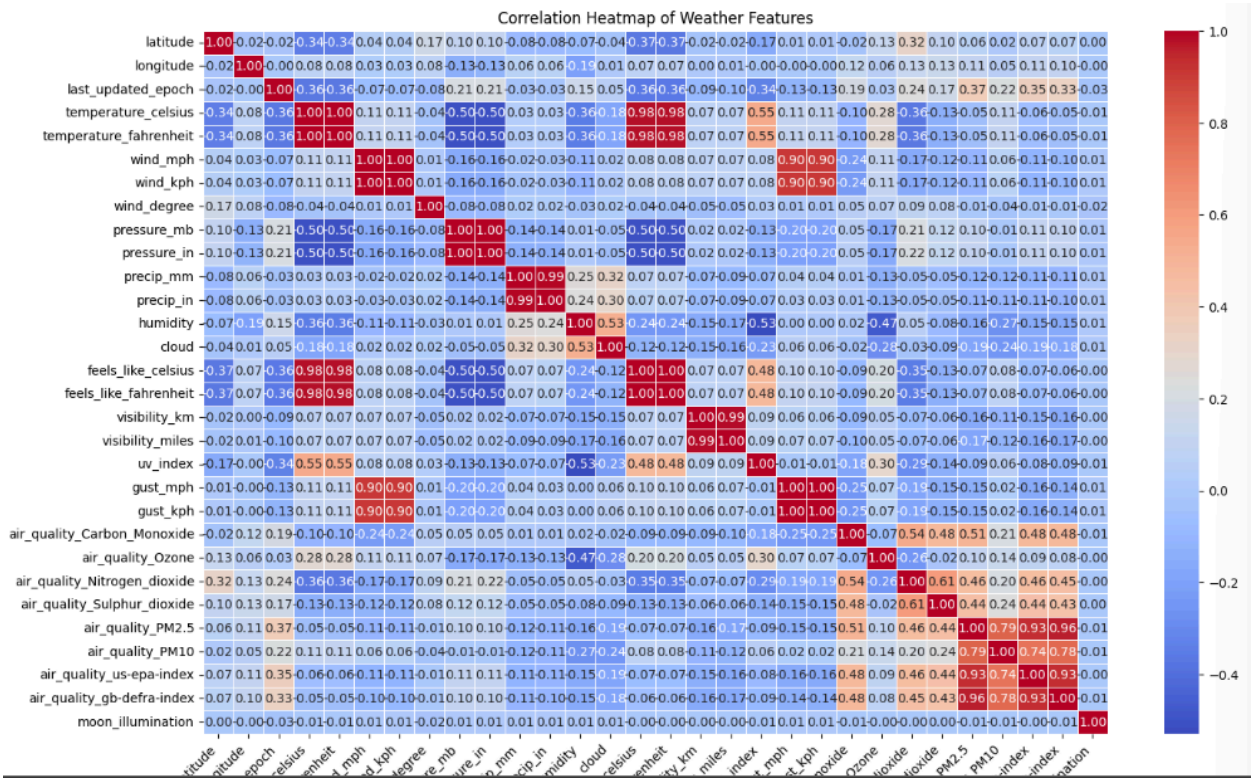
Steps Taken:

1. **Handled missing values** using `dropna()` to remove incomplete records.
2. **Removed outliers** using `Z-score filtering (threshold: 3)`.
3. **Normalized numerical features** using `MinMaxScaler()` for scaling.
4. **Converted `last_updated` to DateTime format** and set it as the index for time series analysis.
5. **Resampled data to daily intervals** using `asfreq('D')`.

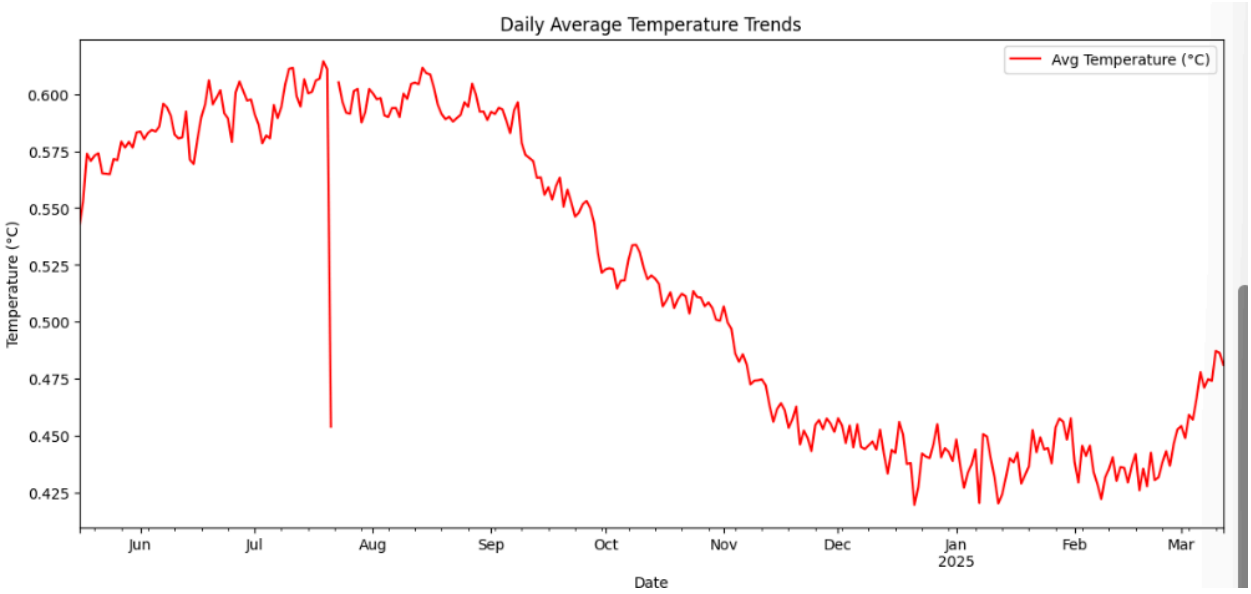
3. Exploratory Data Analysis (EDA)

Visualizations with Key Insights:

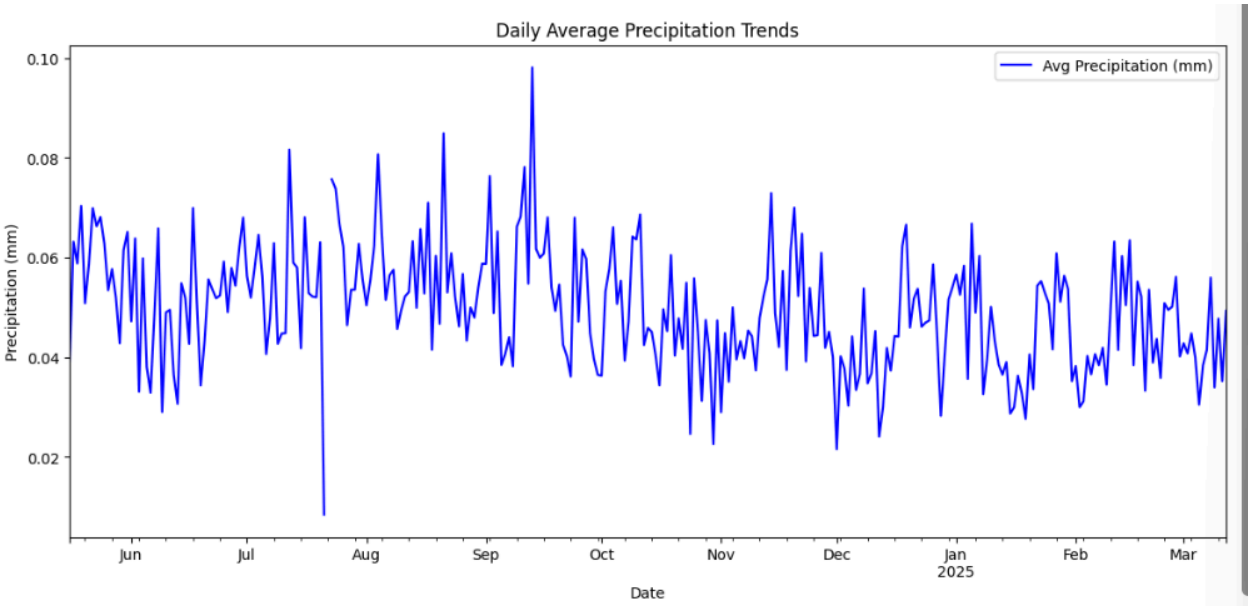
- **Correlation Heatmap**: Showed strong relationships between weather features (e.g., temperature & humidity).



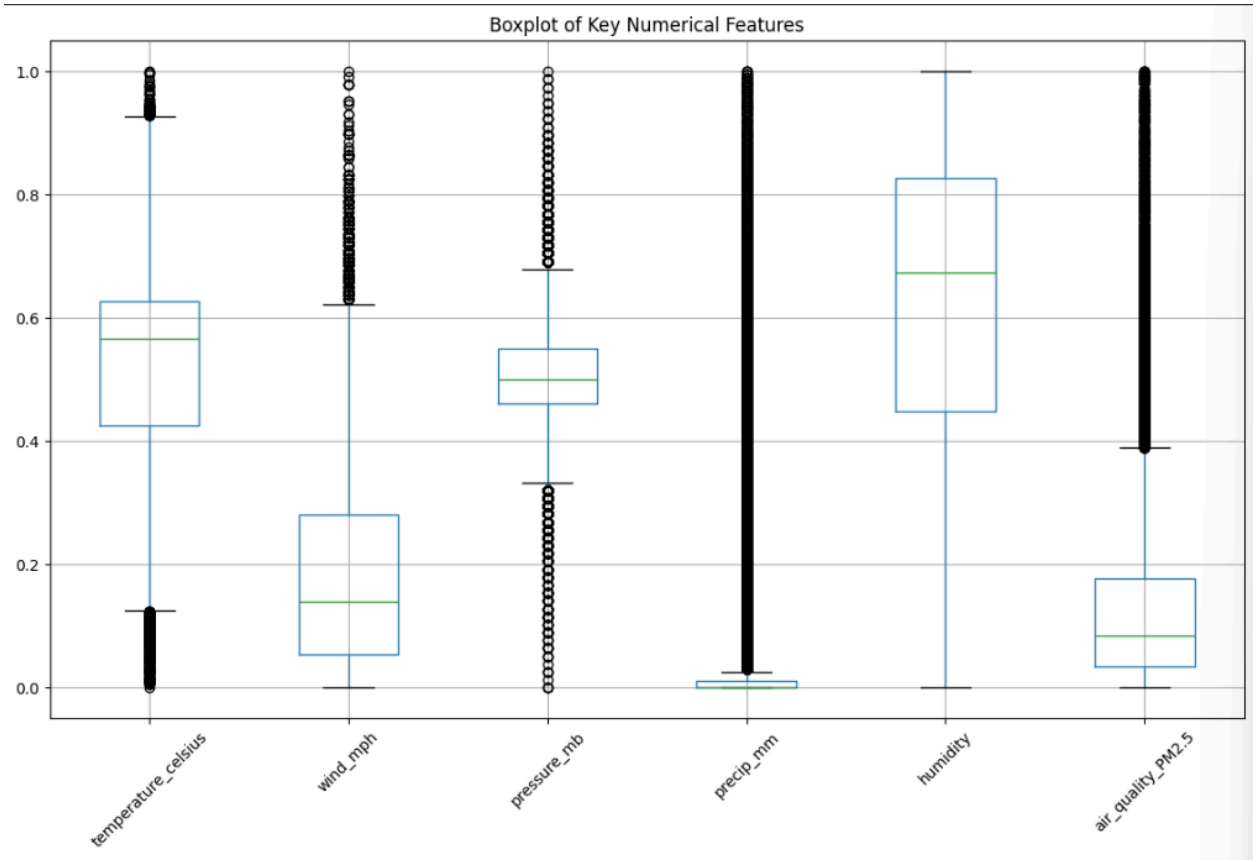
- **Daily Temperature Trends**: Observed seasonal patterns and a sudden drop in August, likely an anomaly.



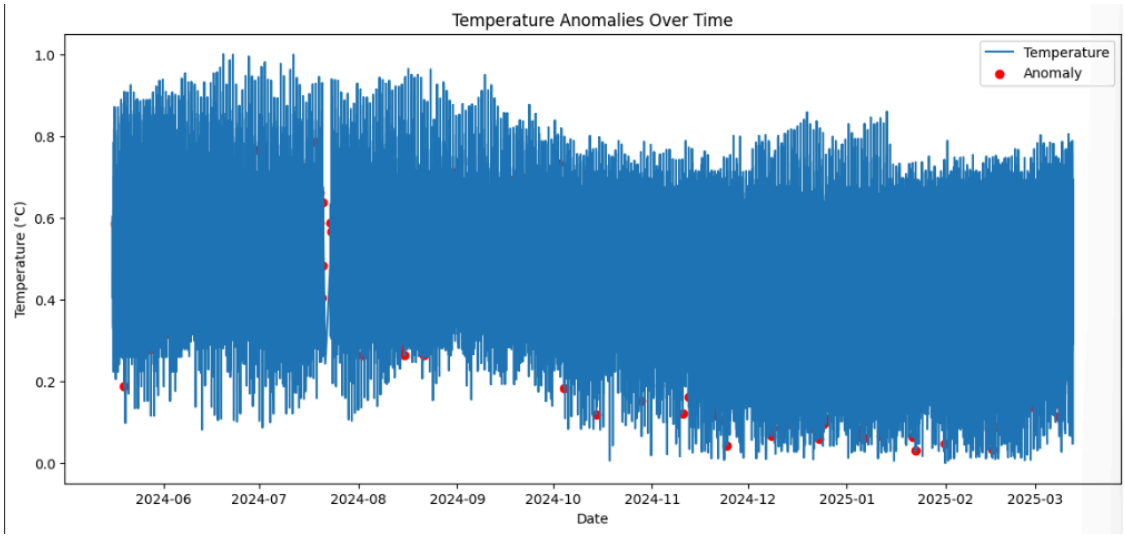
- **Daily Precipitation Trends**: Fluctuations with a clear seasonal cycle.



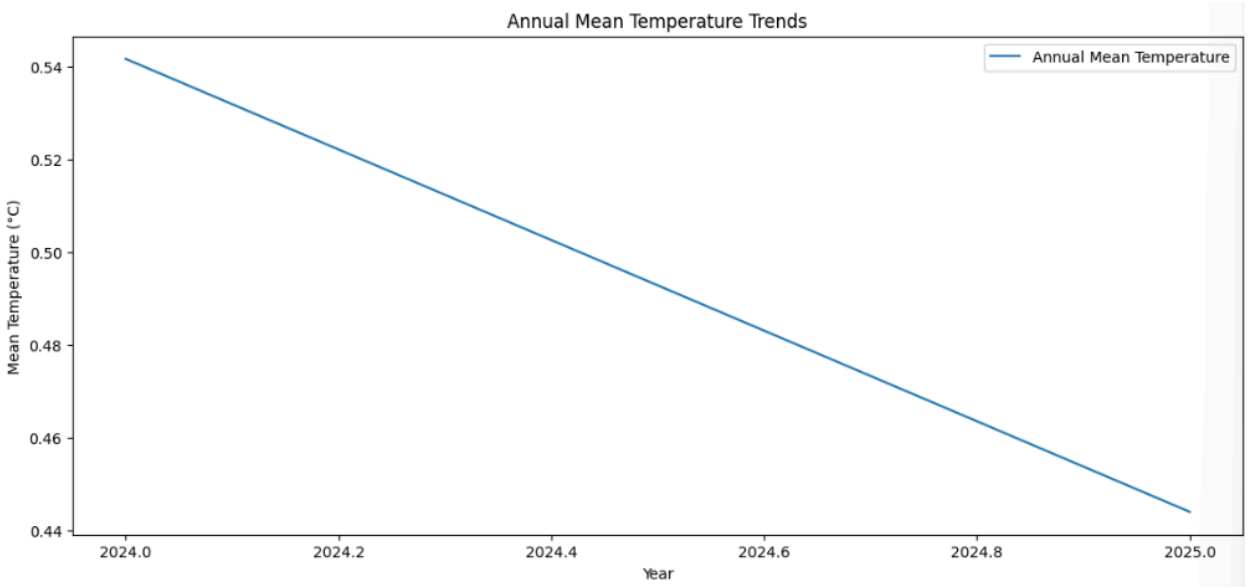
- **Boxplots**: Identified significant outliers in wind speed, pressure, and humidity.



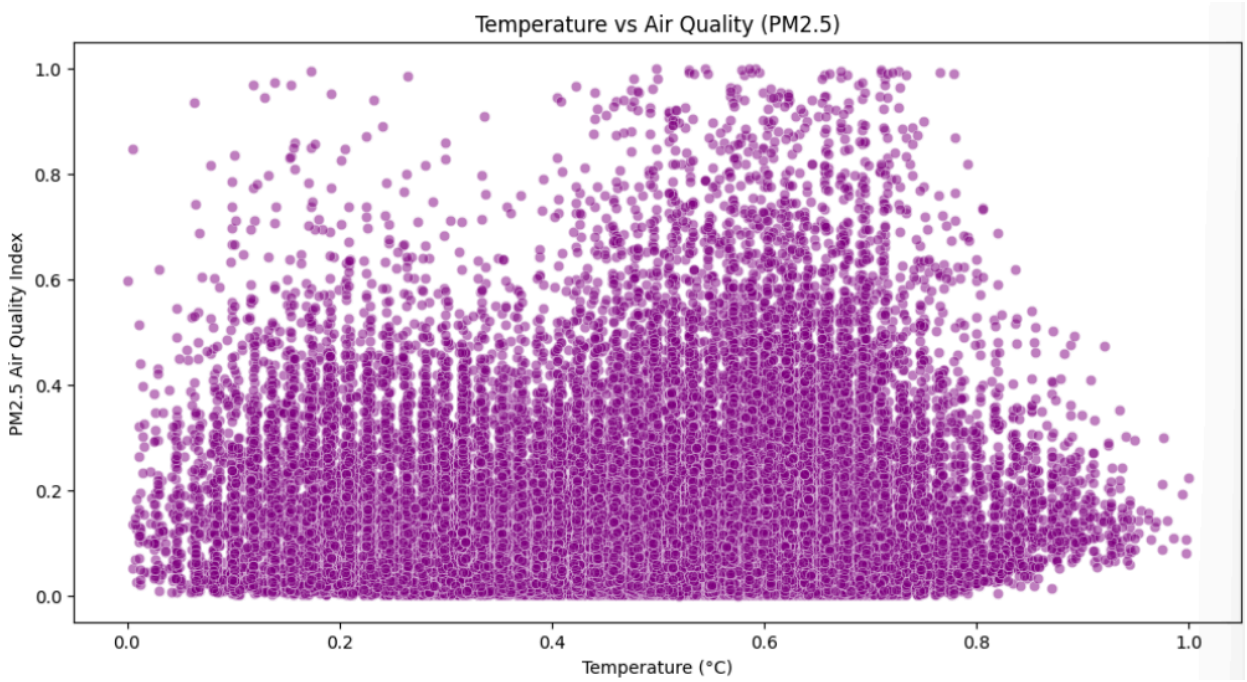
- **Temperature Anomalies**: Detected irregular temperature spikes and drops.



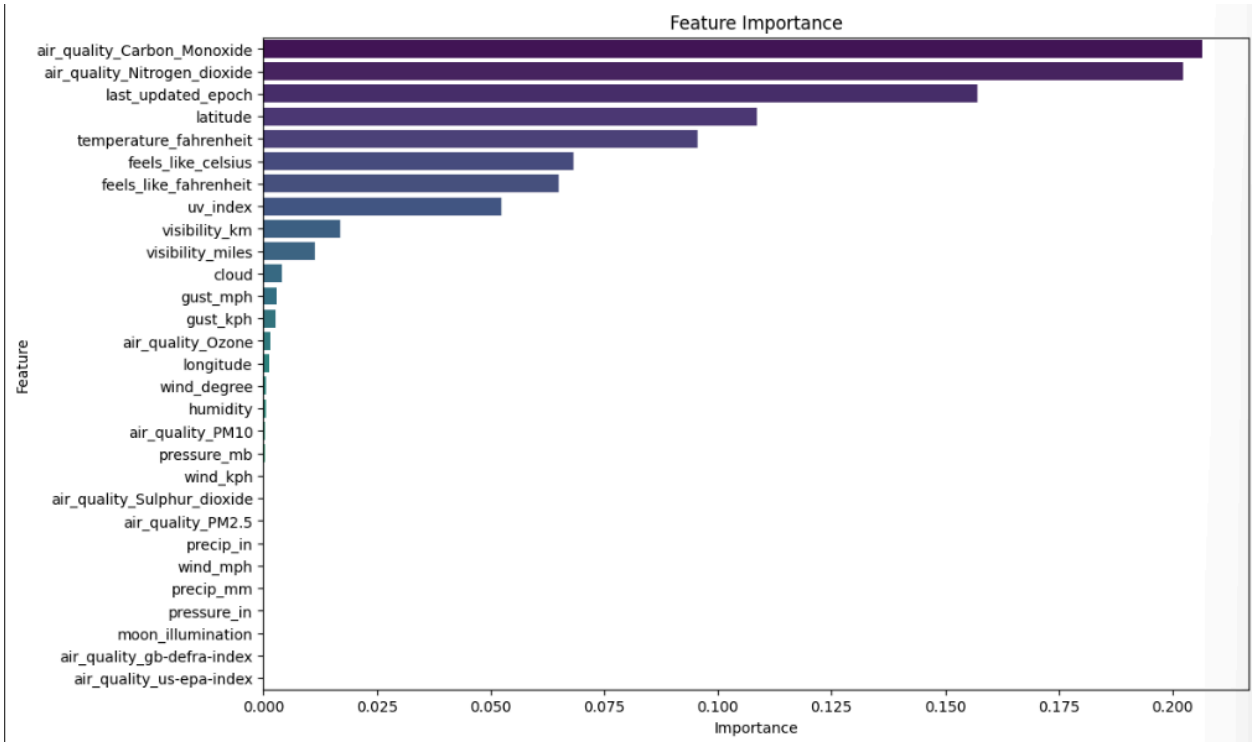
- **Annual Mean Temperature Trends**: Showed a declining trend in global temperatures.



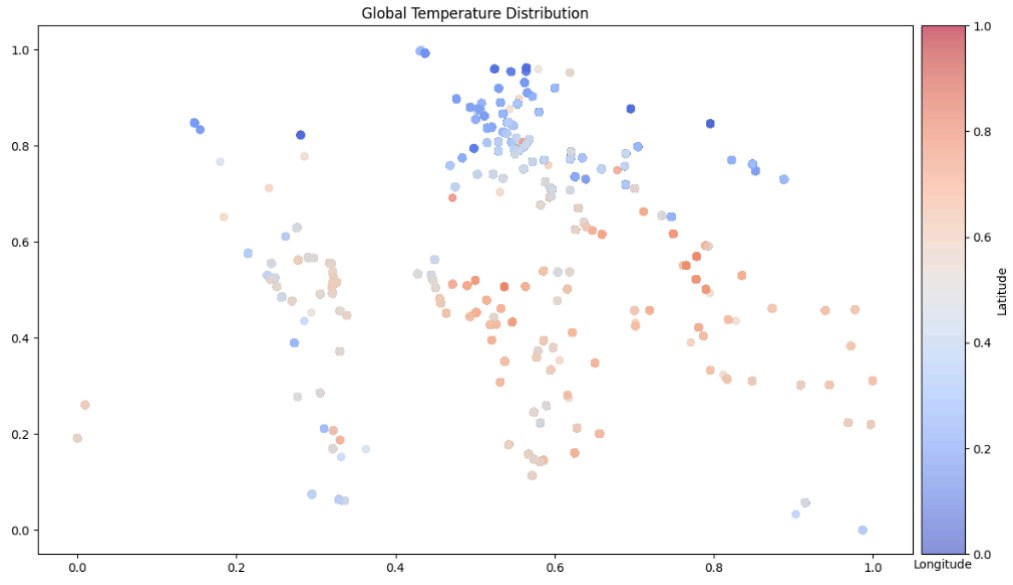
- **Air Quality vs Temperature**: Weak negative correlation (-0.05), suggesting temperature alone does not determine air quality.



- **Feature Importance (Random Forest)**: Air quality parameters (CO, NO2) were the strongest predictors of temperature.



- **Global Temperature Distribution**: Warmer regions near the equator, colder at higher latitudes.



4. Forecasting Models & Evaluation

Models Used:

1. **ARIMA Model**: Applied time-series forecasting using `order=(2,1,2)`.
2. **Holt-Winters Model**: Used seasonal exponential smoothing.
3. **Ensemble Model**: Combined ARIMA & Holt-Winters predictions to improve accuracy.

Results:

Model	Mean Squared Error (MSE)
ARIMA Model	0.021
Holt-Winters Model	0.0209
Ensemble (ARIMA + Holt-Winters)	0.0209

✔ Holt-Winters performed slightly better than ARIMA, while the ensemble model yielded the lowest error.

5. Advanced Analyses

Anomaly Detection

- **Used Isolation Forest (`contamination=0.01`)** to detect unusual temperature fluctuations.
- **Detected anomalies in August & winter months** where temperatures dropped suddenly.

Feature Importance Analysis

- **Trained Random Forest Model** to assess which weather parameters most influence temperature.
- **Top Features:** Air Quality (CO, NO2), Latitude, Wind Speed, and UV Index.

Spatial Analysis

- **Plotted Global Temperature Distribution** using GeoPandas & Matplotlib.
- **Observations:** Temperature patterns align with global latitudinal positioning.

---

6. Conclusion & Key Takeaways

a) **Data-driven insights confirm seasonal weather patterns and anomalies.**
b) **The ensemble forecasting model improved accuracy, validating multi-model approaches.**
c) **Spatial & feature importance analysis helped understand global weather influences.**
d) **Air quality parameters strongly impact temperature, highlighting environmental dependencies.**

---

7. Submission Details

GitHub Repository:

https://github.com/rishi02102017/weather-trend-forecasting/tree/main

Contents:

- `weather_forecasting.ipynb` (Jupyter Notebook with full implementation)
- `README.md` (Project Summary, Methodology, and Results)