Data Preparation & Initial Analysis

Infosys Springboard Internship Project: ClimateScope

Prepared by: Snehardha Ghosh

Project Overview

This project, conducted under the Infosys SpringBoard Internship, focuses on ClimateScope—a global weather analytics initiative aimed at analyzing worldwide climate behaviors, seasonal trends, and atmospheric patterns. The dataset includes daily weather observations, atmospheric conditions, air quality indicators, and astronomical data.

Objective

The objective of this phase was to clean, validate, and standardize the dataset to ensure data quality and prepare it for climate trend analysis, visualization, and anomaly detection.

Data Schema Overview

The dataset contains location details, timestamps, temperature metrics, atmospheric conditions like humidity and pressure, wind parameters, air quality indicators, and astronomical information such as sunrise, sunset, and moon phases.

Data Quality Checks

A detailed data quality assessment was performed, covering missing values, inconsistency detection, invalid ranges, incorrect formats, and duplicates. No missing values were found, and all inconsistencies were corrected.

Unit Conversion

Standardized unit conversions were applied for accuracy and consistency:
- Wind: mph to kph
- Pressure: inHg to mb
- Visibility: miles to km
- Precipitation: inches to mm
- Temperature: Fahrenheit to Celsius

Normalization

All numeric features were normalized using MinMaxScaler to scale values into a 0–1 range, enabling accurate comparison and improving readiness for data modeling.

Anomaly Detection

Outliers were identified using IQR and domain-based rules such as unrealistic temperatures, humidity outside 0–100%, negative rainfall, and incorrect wind speeds. Anomalies were logged for review.

Aggregation

Daily weather data was aggregated into monthly averages for all numeric columns, supporting seasonal and comparative climate analysis across regions.

Final Summary

The dataset is now fully cleaned, standardized, normalized, and aggregated. It is ready for visualization and advanced climate analytics as part of the ClimateScope project under Infosys SpringBoard.