Weather Data Analysis – Project Report

Time-Series Exploratory Data Analysis

Tools Used: Python, Pandas, Seaborn

Author: Suraj Singh

1. Project Overview

This project focuses on exploratory analysis of historical weather data using Python. The objective is to identify seasonal patterns, analyze temperature behavior over time, and understand relationships between key meteorological variables.

2. Dataset Overview

The dataset contains 96,453 observations and 12 original features. After preprocessing, 11 relevant features were retained. The data is time-series in nature, enabling temporal analysis.

3. Data Cleaning & Preparation

- Converted date column to proper datetime format.
- Removed Cloud Cover feature due to zero variance.
- Handled missing precipitation values by explicit labeling.
- Dataset prepared for time-series analysis.

4. Key Variables Analyzed

Temperature, Apparent Temperature, Humidity, Wind Speed, and Precipitation Type were analyzed.

5. Exploratory Data Analysis Findings

Mean temperature is approximately 11.9°C with realistic seasonal extremes (-21°C to 39°C). The distribution is approximately normal, reflecting genuine seasonal behavior.

A strong linear relationship exists between actual and apparent temperature, with apparent temperature typically lower due to wind chill effects.

6. Time-Series & Seasonal Analysis

Clear monthly seasonality is observed with repeating yearly cycles. Yearly average temperatures remain stable between 10°C and 12°C.

7. Weather Relationships

Temperature and humidity show an inverse relationship. Higher wind speeds increase the difference between actual and apparent temperature.

8. Key Insights

- Strong seasonal temperature patterns.
- Stable yearly climate behavior.
- Wind significantly impacts perceived temperature.

- Dataset is consistent and reliable.

9. Conclusion & Next Steps

The analysis demonstrates a structured EDA workflow on time-series data. The dataset is suitable for forecasting, regression modeling, and extended climate analysis. End-to-end analysis was automated using Python.