ASSIGNMENT - 4

SCHOOL OF ENGINEERING AND TECHNOLOGY

SUBMITTED BY : Khushbu Singh
SUBMITTED TO : Sameer Farooq
ROLL NO. : 2501730161
COURSE : B TECH CSE (AI&ML)
SECTION : B
SUBJECT : PYTHON

Report - Weather Data Analysis & Visualization

Introduction

In this assignment, I analyzed real weather data using Python to understand temperature, rainfall, and humidity patterns. The aim was to clean the dataset, calculate important statistics, create visualizations, and summarize the insights in a clear format.

Dataset Description

The dataset contains daily weather information including:

- Date
- Minimum and maximum temperature
- Temperature at 9am and 3pm
- Humidity at 9am and 3pm
- Rainfall
- Other columns such as wind information and pressure

For this assignment, I used the following main columns:

- Temp3pm → treated as temperature
- Humidity3pm → treated as humidity
- Rainfall → rainfall amount

- Date → for time-based analysis

These were renamed to make the analysis easier.

Data Cleaning

The dataset was cleaned by performing the following steps:

1. Converted the Date column into proper datetime format.
2. Filled missing temperature values with the mean temperature.
3. Filled missing rainfall values with 0.
4. Filled missing humidity values with the median.
5. Selected only the useful columns: Date, Temperature, Rainfall, Humidity.
6. Created a Month column for grouping and monthly analysis.

Statistical Summary

After cleaning the data, the following numerical summaries were calculated:

- Mean temperature
- Minimum temperature
- Maximum temperature
- Standard deviation of temperature
- Monthly rainfall totals

- Monthly average temperature

(The exact values can be taken from the program output.)

Visualizations

Three visualizations were created using Matplotlib:

1. **Daily Temperature Trend (Line Chart)**
   Shows how temperature changed over time and helps identify warm and cool periods.

2. **Monthly Rainfall (Bar Chart)**
   Displays total rainfall for each month and highlights wet and dry months.

3. **Temperature vs Humidity (Scatter Plot)**
   Shows the relationship between temperature and humidity, helping to identify how these two variables behave together.

Grouping and Monthly Analysis

Using grouped data, the following monthly summaries were calculated:

- Average temperature per month
- Total rainfall per month
- Average humidity per month

This helps in understanding seasonal changes and overall weather behaviour across the year.

Exported Results

The following output files were generated as part of the assignment:

- cleaned_weather.csv
- daily_temperature.png
- monthly_rainfall.png
- humidity_vs_temp.png
- summary_report.txt

Conclusion

This assignment helped me understand how to work with real datasets by cleaning, analyzing, and visualizing weather data. Using Pandas, NumPy, and Matplotlib, I was able to apply important data analysis skills and present insights in a clear manner. The project also helped in understanding the complete workflow from raw data to meaningful interpretation.