

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT : C.S.E.

PROGRAMME: B. Tech. C.S.E.(Specialization in AI&ML)

COURSE NAME : Problem Solving using Python

COURSE CODE : ETCCPP102

FACULTY : Dr. Sameer Farooq

ASSIGNMENT NUMBER : 04

STUDENT NAME : PRACHI MANWAL

ROLL NUMBER : 2501730365

SEMESTER : 1

PROJECT NAME : Weather Data Visualizer

1. Introduction

This project focuses on acquiring, cleaning, analyzing, and visualizing weather data. The primary variables studied include **temperature**, **humidity**, **rainfall**, and **date**.

The objective is to understand daily, monthly, and seasonal weather trends through structured data analysis and statistical techniques.

2. Dataset Description

The dataset consists of daily weather observations containing the following columns:

- **Date:** The recorded day of weather measurement
- **Temperature (°C):** Average temperature of the day
- **Humidity (%):** Daily humidity level
- **Rainfall (mm):** Total rainfall recorded
- *(Optional additional columns depending on dataset availability: wind speed, atmospheric pressure, cloud cover)*

These variables together provide a meaningful representation of climatic conditions.

3. Tools and technology used

The following tools and technologies were used in the project:

- **Python 3**
- **Pandas** – data loading, cleaning, filtering
- **NumPy** – statistical and numerical computation
- **Matplotlib** – weather trend visualizations
- **Visual Studio Code (VS Code)** – development environment
- **GitHub** – version control and project hosting

4. Data cleaning and preprocessing

The raw dataset underwent several preprocessing steps:

1. **Loading the dataset** using Pandas.
2. **Converting the Date column** to a proper datetime format.
3. **Handling missing values** using removal or suitable imputation methods.
4. **Selecting required columns** for analysis: Date, Temperature, Humidity, Rainfall.
5. **Creating additional time-based columns** such as Month and Year for better grouping and visualization.

These steps ensure the dataset is clean, consistent, and ready for meaningful analysis.

5. Statistical Analysis

NumPy and Pandas were used to compute key statistical metrics, including:

- **Mean, minimum, and maximum temperature**
- **Average humidity**
- **Total monthly and yearly rainfall**
- **Standard deviation** of temperature and humidity
- **Group-based trends** using month and year aggregations

This statistical breakdown helps identify seasonal trends, anomalies, and variability in the data.

6. Data visualizations

Matplotlib was used to create multiple plots that illustrate different aspects of the weather dataset:

- **Line chart** showing daily temperature trends
- **Bar chart** visualizing monthly rainfall totals
- **Scatter plot** showing the relationship between humidity and temperature
- **Combined multi-plot figure** to visually represent multiple weather factors together

These visualizations help interpret trends more clearly than tabular data alone.

7. Observations and insights

Key insights derived from the analysis include:

- **Temperature** displays clear seasonal variation, rising in summer and dropping in winter.
- **Rainfall** peaks during monsoon months, with minimal rainfall in the dry seasons.
- **Humidity** strongly increases during rainy periods, as seen in scatter plots.
- **Monthly aggregation** reveals consistent climatic cycles that could benefit agriculture, planning, and environmental research.

8. Exported Outputs

The project generated the following outputs:

- **Cleaned CSV dataset**

- **PNG images** of all visualizations
- **A summary report** (this document)
- **Python script (.py)** containing data processing and visualization code
- **GitHub repository** containing all project files

9. Conclusion

This project demonstrates a complete data analysis pipeline—from data acquisition to cleaning, analysis, visualization, and reporting.

The insights derived from the dataset help in understanding weather patterns and seasonal cycles.

The project also highlights the effectiveness of Python tools such as Pandas, NumPy, and Matplotlib in performing real-world data science tasks.

GitHub Repository link :

https://github.com/prachiii0418-star/weather_data.git