Weather Data Analysis

- Visualizing Temperature Trends, Rainfall, and Seasonal Patterns

Prepared by: ADITI VERMA

Branch: CSE (AI & ML)

Sec- A

University roll number : 202401100400014

Date: 10-03-2025

INTRODUCTION

This project aims to analyze and visualize weather data to understand temperature trends, rainfall patterns, and seasonal variations. By leveraging historical weather data, we can identify patterns and fluctuations over time. The dataset used for this analysis is obtained from Kaggle and contains temperature, humidity, precipitation type, and other weather-related attributes.

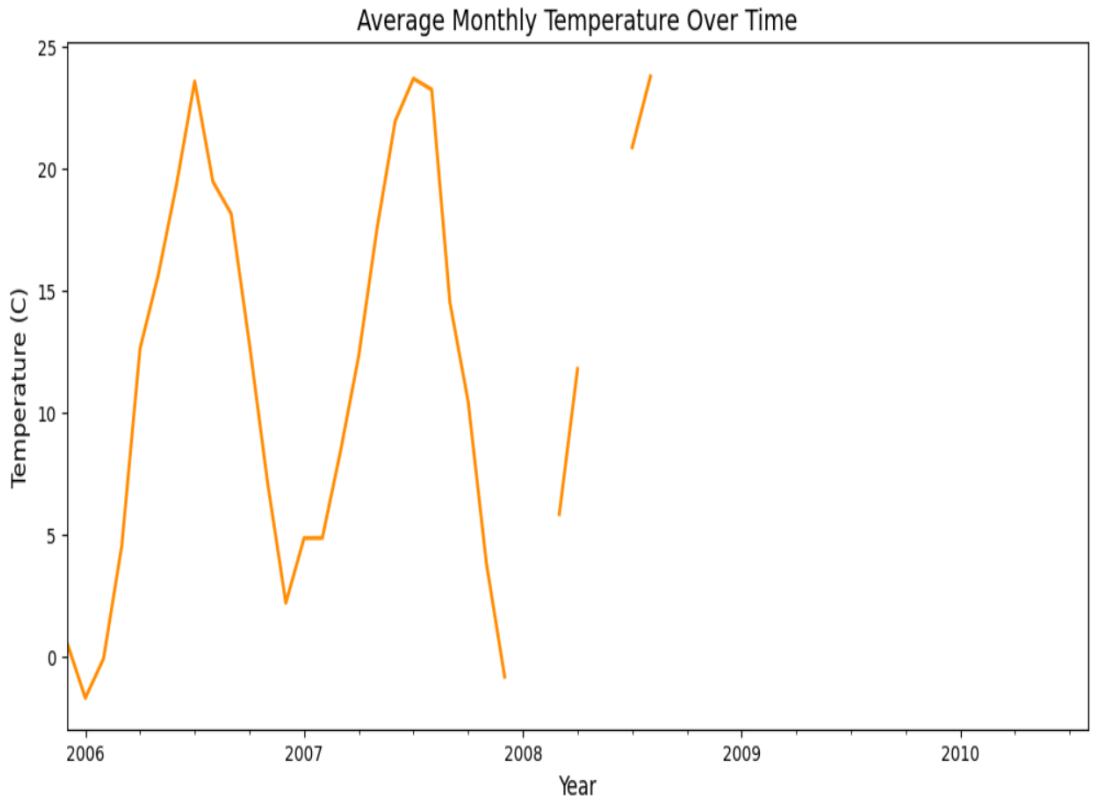
---

c. Methodology

1. **Data Cleaning**
    - **Removed missing values and inconsistencies.**
    - **Converted date columns into proper datetime format.**
    - **Extracted useful features like Year, Month, and Day.**
2. **Data Analysis & Visualization**
    - **Used Matplotlib and Seaborn for visualizations.**
    - **Temperature Trends: Line plot showing average monthly temperatures.**
    - **Rainfall Analysis: Bar chart indicating the number of rainy days per month.**
    - **Seasonal Patterns: Box plot to show temperature variations over different months.**

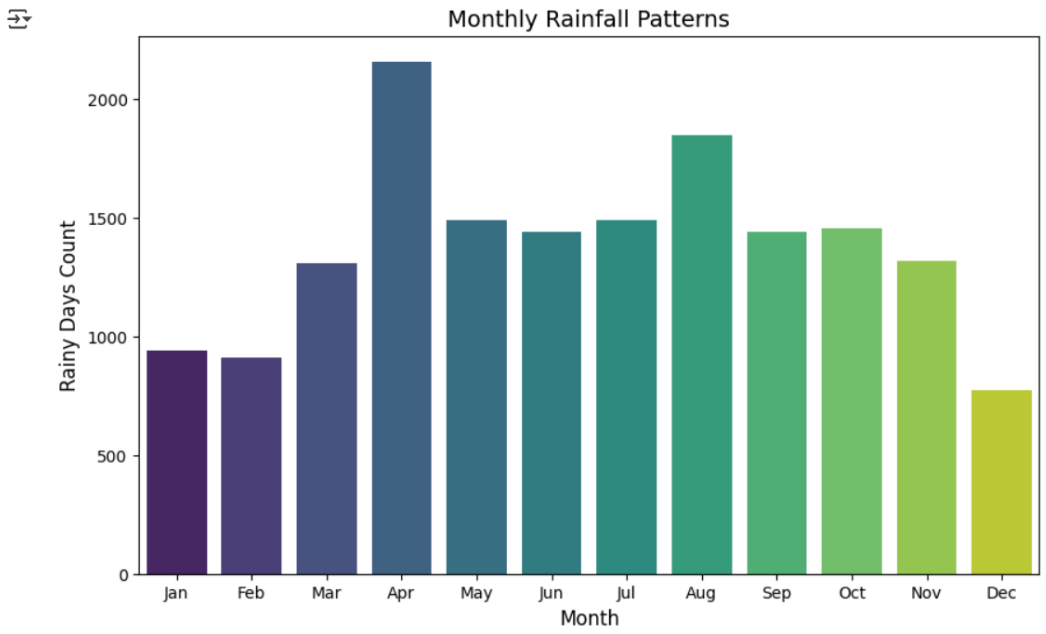
IMPORTING IMPORTANT LIBRARIES FOR DATA MANIPULATION , VISUALIZATIONS , NUMERCIAL COMPUTATION

```
import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import seaborn as sns

## Create a figure with a specified size

plt.figure(figsize=(12, 6))

##Resample the dataset to get the **monthly average temperature**

# 'M' represents monthly resampling, and mean() calculates the average temperature for each month

df.resample('M').mean(numeric_only=True)['Temperature (C)'].plot(color='darkorange', linewidth=2)

## # Add a title to the plot

plt.title('Average Monthly Temperature Over Time', fontsize=14)

### Label the x-axis as "Year"

plt.xlabel('Year', fontsize=12)

# Label the y-axis as "Temperature (C)"

plt.ylabel('Temperature (C)', fontsize=12)

# Display the plot

plt.show()
```

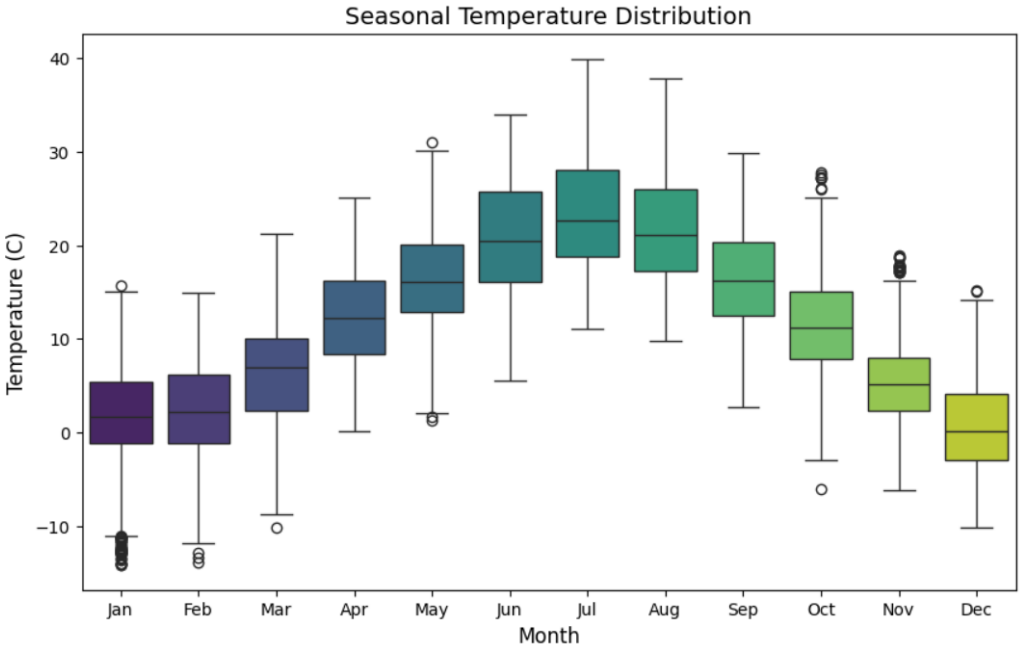



```
### Filter the dataset to include only rows where 'Precip Type' is 'rain'
# Group the data by 'Month' and count the number of rainy days in each month
monthly_rainfall = df[df['Precip Type'] == 'rain'].groupby("Month").size()
# Create a figure with a specified size
plt.figure(figsize=(10, 6))
custom_palette = "viridis"
###Create a bar plot showing the number of rainy days for each month
sns.barplot(x=monthly_rainfall.index, y=monthly_rainfall.values, palette=custom_palette)
## Add a title to the plot
plt.title("Monthly Rainfall Patterns", fontsize=14)
# Label the x-axis as "Month"
plt.xlabel("Month", fontsize=12)
# Label the y-axis as "Rainy Days Count"
plt.ylabel("Rainy Days Count", fontsize=12)
# Set custom month labels on the x-axis
```

```
plt.xticks(range(0, 12), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])
plt.show()
```




```
plt.figure(figsize=(10, 6))
# Create a box plot to show the distribution of temperature for each month
# The box plot helps visualize temperature variations, medians, and outliers
sns.boxplot(x=df['Month'], y=df['Temperature (C)'], palette=custom_palette)
# Add a title to the plot
plt.title("Seasonal Temperature Distribution", fontsize=14)
# Label the x-axis as "Month"
plt.xlabel("Month", fontsize=12)
# Label the y-axis as "Temperature (C)"
plt.ylabel("Temperature (C)", fontsize=12)
# Set custom labels for the x-axis months from January to December
plt.xticks(range(0, 12), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])
plt.show()
```



References/Credits

- **Dataset:** Kaggle Weather History Dataset
- **Libraries Used:** Pandas, Matplotlib, Seaborn, NumPy