

K.I.E.T. Group of Institutions Ghaziabad



Name: Arnav Bhardwaj

Branch: CSE(AI) - A

Roll No: 62

Date: 11/03/2025

Project Report On -. Weather Data Analysis

Introduction

This report presents an analysis of weather patterns observed over a specific period. It aims to examine key factors such as temperature fluctuations, precipitation levels, wind patterns, and humidity trends, providing insights into their impact on the environment. By understanding these elements, we can identify patterns, anticipate changes, and make informed decisions related to agriculture, energy consumption, and daily activities.

Methodology

- 1. Data Collection:** The dataset consists of Data, Temperature, Humidity, Rainfall,.
- 2. Data Processing:** The dataset is cleaned by handling missing values and converting Temperautre to numerical format.
- 3. Visualization Techniques:**
 - Histograms for Temperature.
 - Temperautre VS Rainfall graph.
 - Temperature VS Humidity.
 - Temperature , humidity , rainfall distribution, in comparision with the frequency.
 - Temperature Vs humidity bar graph.
- 4. Tools Used:** Python, Pandas, Matplotlib, and Seaborn.

Code

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read_csv('/content/weather_data.csv')

print(df.shape)

print(df.describe())

print(df.dtypes)

print(df.isnull().sum())

df.dropna(inplace=True)
value = 0
df.fillna(value, inplace=True)

sns.set_style("darkgrid")

plt.figure(figsize=(20, 4))
sns.histplot(df['Date'], bins=20, kde=True, color='blue')
plt.title("Temperature")
plt.xlabel("Rainfall")
plt.ylabel("humidity")
plt.show()


sns.histplot(df['Temperature'], kde=True, bins=30,
color='blue')
plt.title("Temperature Distribution")
plt.xlabel("Temperature (°C)")
plt.ylabel("Frequency")
plt.show()


sns.histplot(df['Humidity'], kde=True, bins=30,
color='blue')
plt.title('Humidity')
plt.xlabel('Humidity')
plt.ylabel('Frequency')
plt.show()
-----

sns.histplot(df['Rainfall'], kde=True, bins=30,
color='blue')
plt.title('Rainfall')
plt.xlabel('Rainfall')
plt.ylabel('Frequency')
plt.show()
```

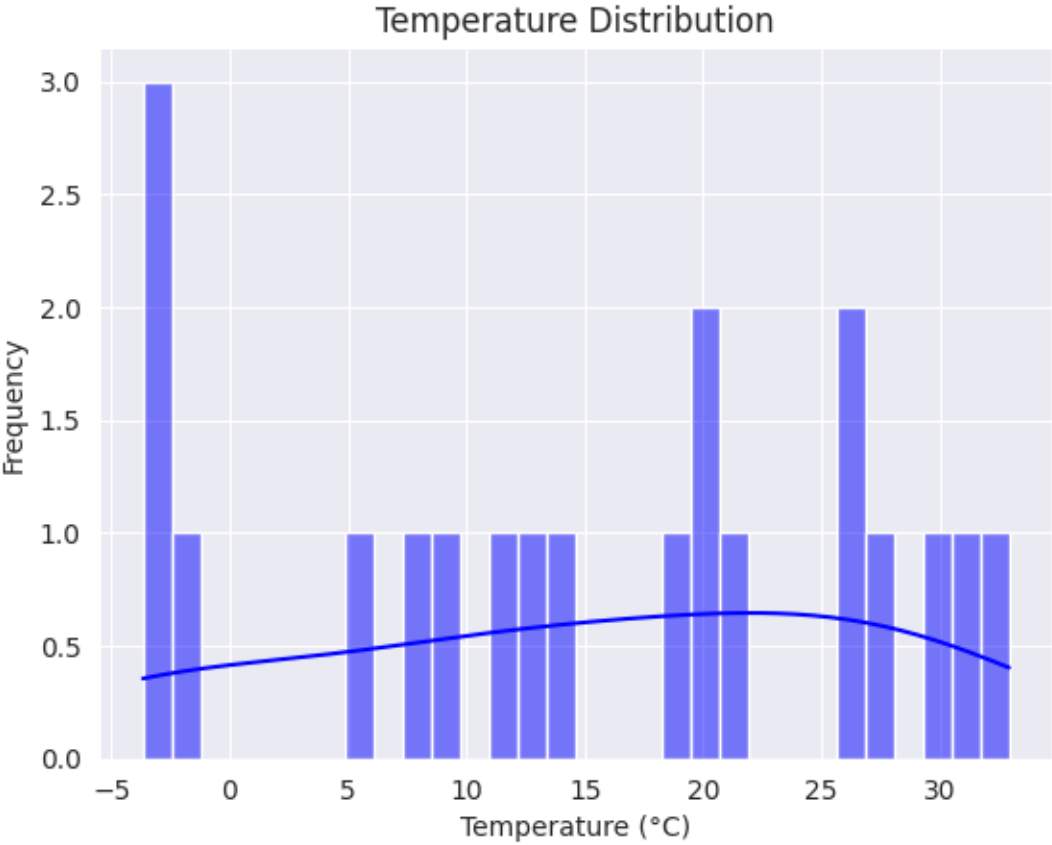
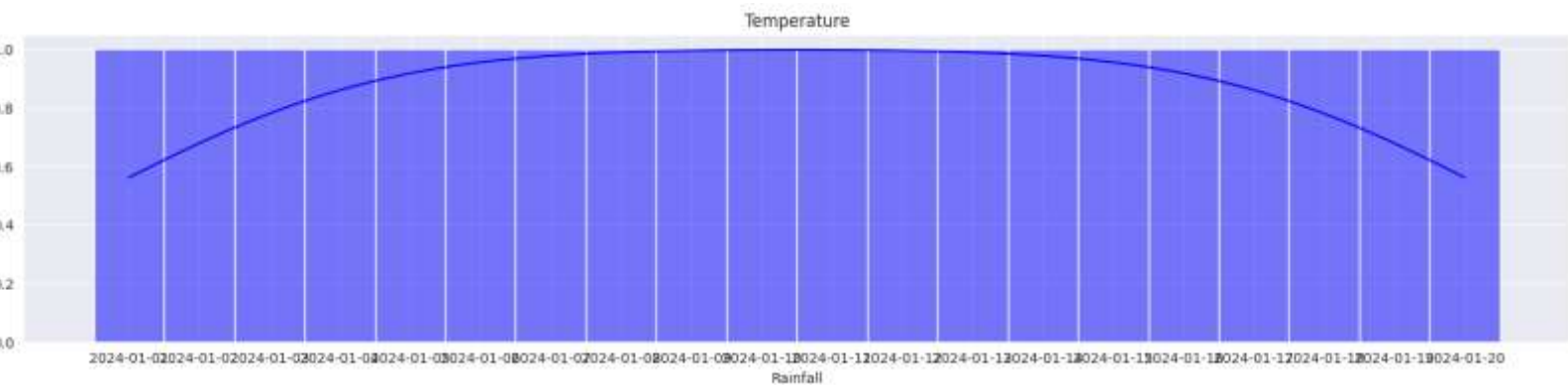
```
plt.hist(df['Temperature'], bins=10)
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.title('Temperature')
plt.show()
```

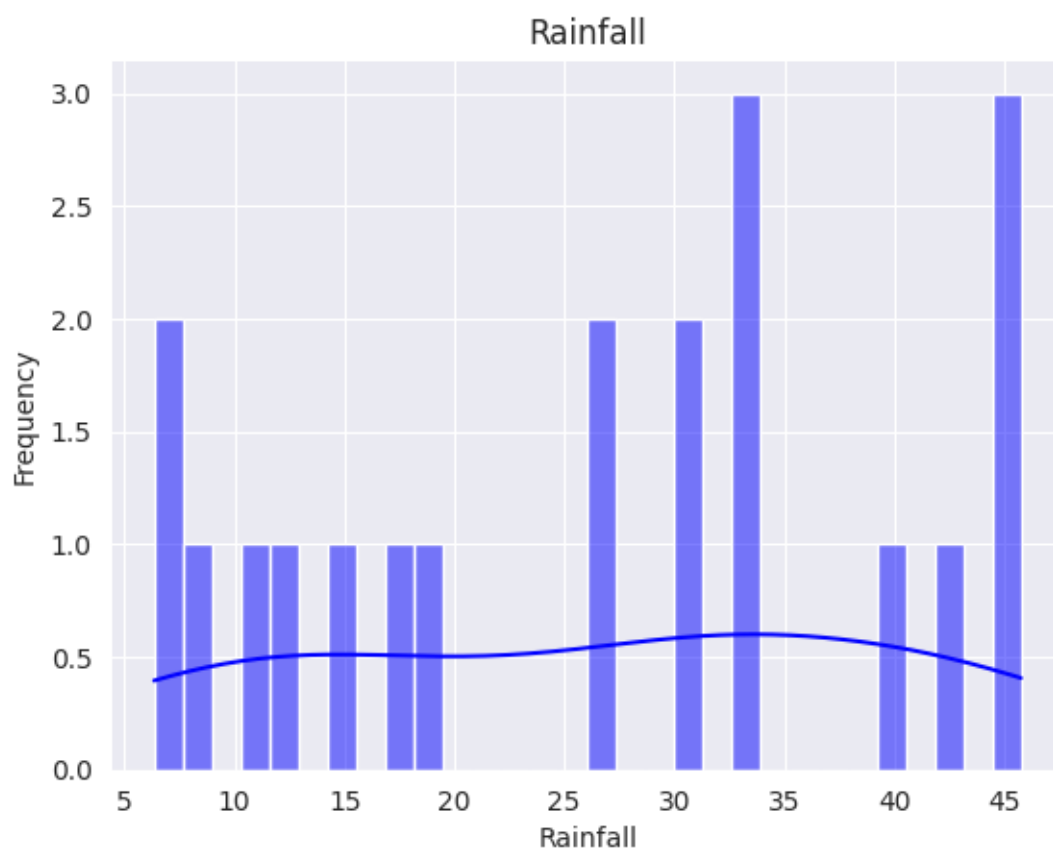
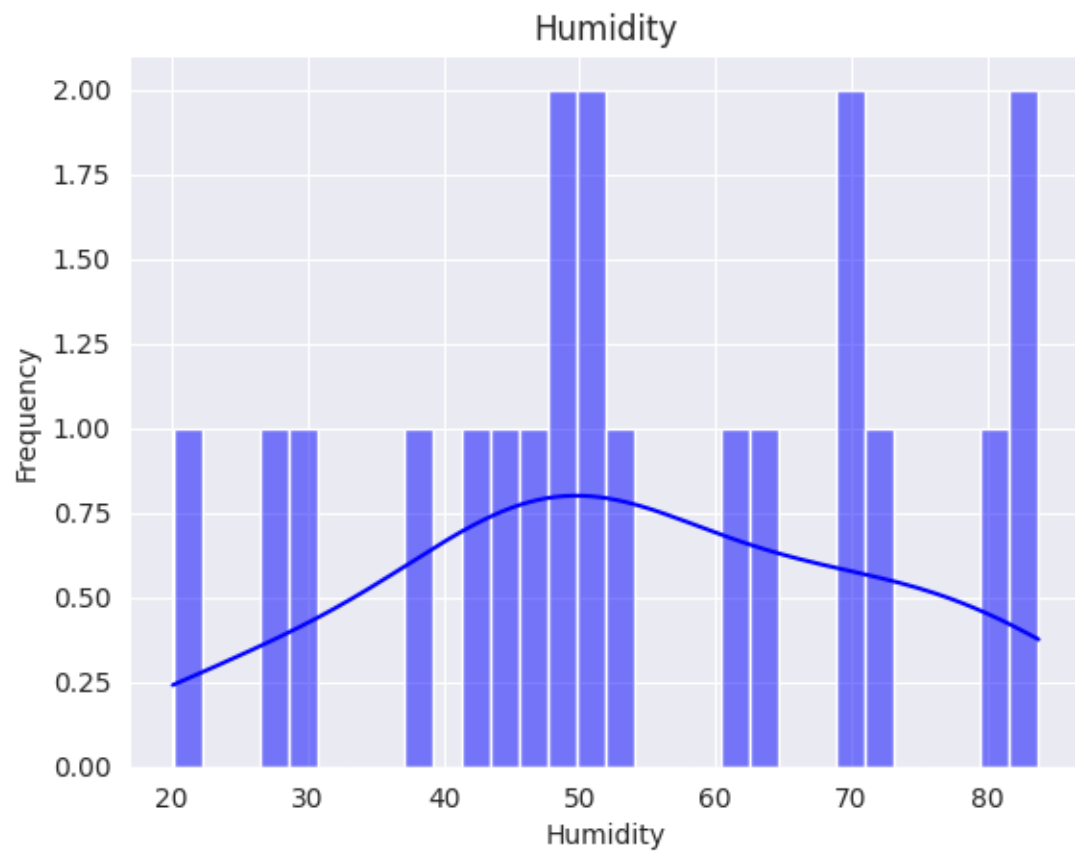
```
plt.scatter(df['Temperature'], df['Rainfall'])
plt.xlabel('Temperature')
plt.ylabel('Rainfall')
plt.title('Temperature vs Rainfall')
plt.show()
```

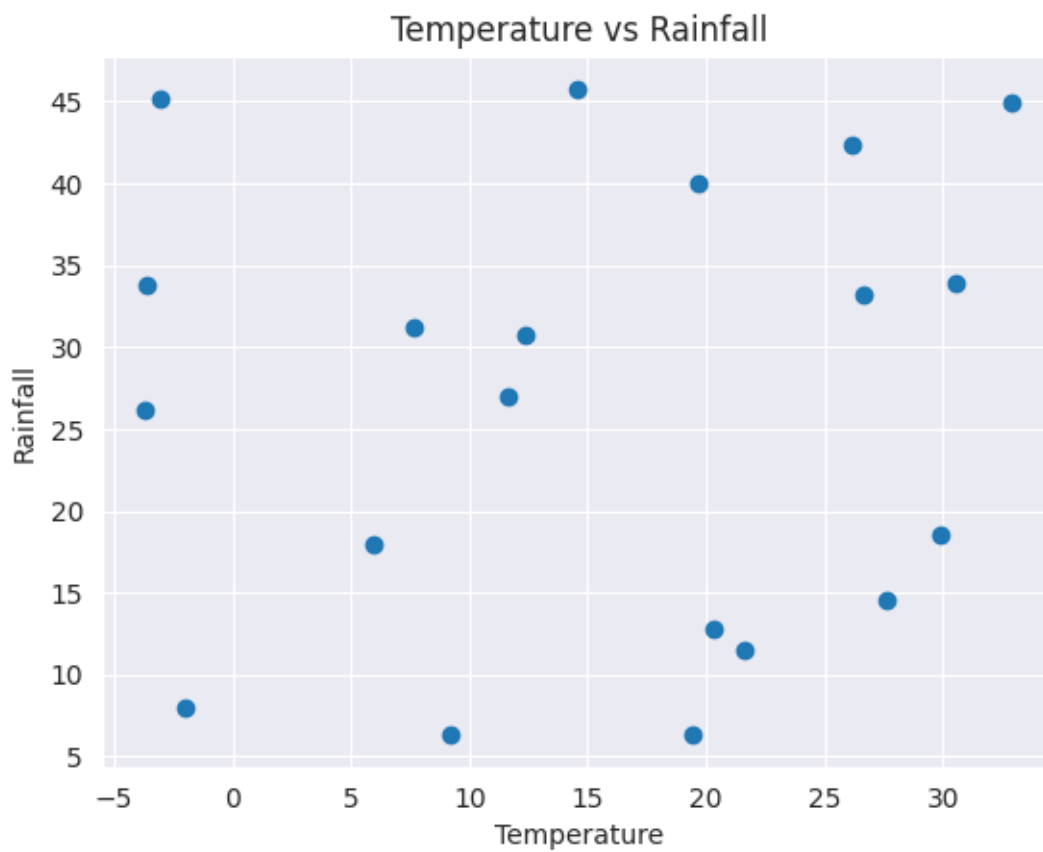
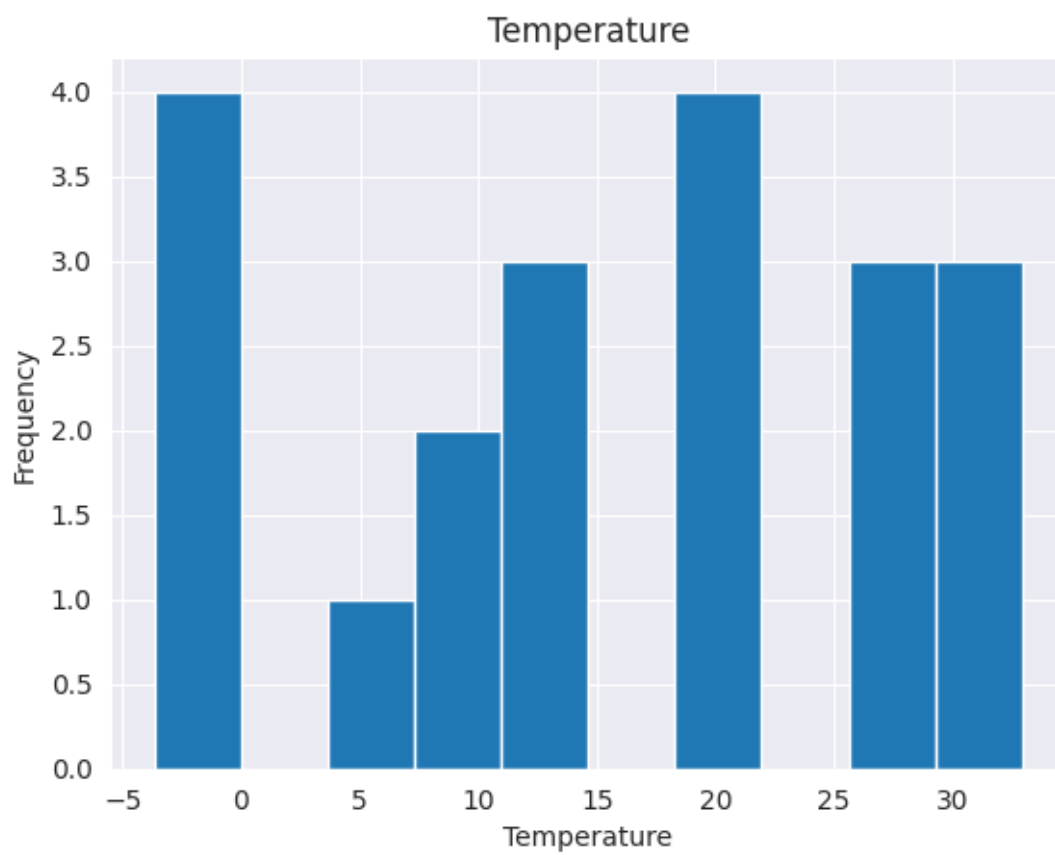
```
-----
sns.boxplot(x=df['Temperature'], y=df['Humidity'])
plt.xlabel('Temperature')
plt.ylabel('Humidity')
plt.title('Temperature vs Humidity')
plt.show()
```

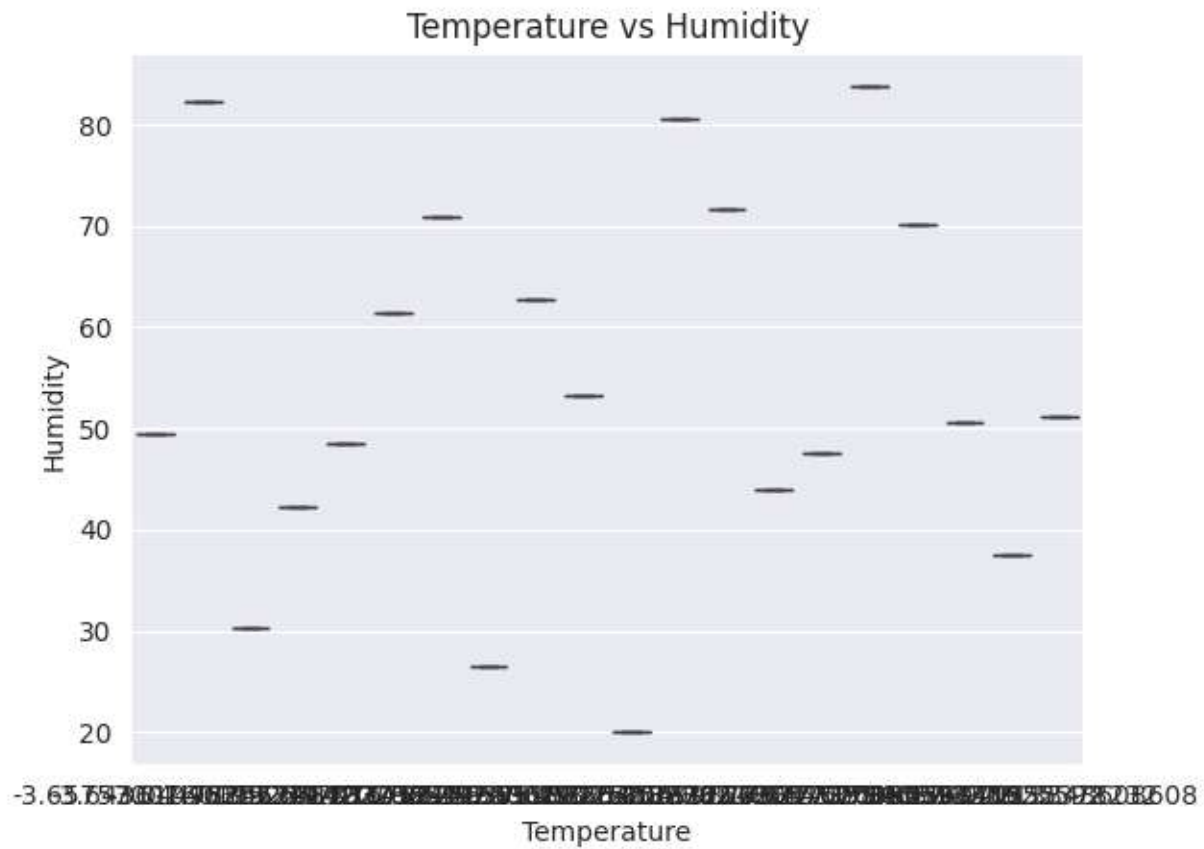
Output/Result

Observation from the analysis is given below:.









References/Credits:

- Dataset: [source: Kaggle]
- Libraries: Pandas, Matplotlib, Seaborn