**Weather Data Analysis Report**

**1. Title Page**

NAME – HARSH VARDHAN SHUKLA

PROGRAM – BTECH

BRANCH – CSE(AI)

ROLL NO – 202401100300116

SUBJECT – ARTIFICIAL INTELIGENCE

TOPIC – WEATHER DATA ANALYSIS

# 2. Introduction

This report provides an in-depth analysis of the given weather dataset, which includes temperature, rainfall, and humidity data recorded over a period of 20 days. Understanding weather patterns is essential for various applications such as agriculture, environmental monitoring, and urban planning.

The primary objectives of this analysis are:

* To identify trends and variations in temperature, rainfall, and humidity.
* To determine the correlation between different weather parameters.
* To present visual representations that aid in comprehending data distribution and relationships.
* To derive insights that can be useful for weather predictions and planning.

By employing statistical analysis and data visualization techniques, we aim to extract meaningful interpretations from the dataset, which can be beneficial for decision-making in weather-dependent sectors.

**3. Methodology**

**3.1 Data Collection**

**The dataset consists of 20 entries with 4 columns:**

* **Date: The date of recording.**
* **Temperature (°C): The temperature recorded on that day.**
* **Rainfall (mm): The amount of rainfall received.**
* **Humidity (%): The humidity percentage recorded.**

**3.2 Data Preprocessing and Cleaning**

* **The dataset was checked for missing values, and any null values were handled to ensure accuracy.**
* **Data types were verified to ensure that numerical computations could be performed correctly.**
* **Outliers were examined to identify any anomalies in the dataset.**

**3.3 Data Analysis Techniques**

**The following methods were employed to analyze the dataset:**

* **Descriptive Statistics: Summarizing key metrics such as mean, standard deviation, and range.**
* **Correlation Analysis: Measuring the relationships between temperature, rainfall, and humidity.**
* **Data Visualization: Utilizing histograms and scatter plots to represent distributions and relationships.**

**This approach ensures that we derive meaningful conclusions from the dataset while maintaining data integrity.**

**4. Code Implementation**

import pandas as pd

import matplotlib.pyplot as plt

# Load the weather data

# Replace 'weather\_data.csv' with the actual filename if different

df = pd.read\_csv('weather data.csv')

# Display basic information about the dataset

print("Dataset Overview:")

print(df.info())

# Display the first few rows to understand the structure

print("\nFirst 5 records:")

print(df.head())

# Basic statistics of the dataset

print("\nStatistical Summary:")

print(df.describe())

# Handling missing values (if any)

df.dropna(inplace=True)

# Exclude non-numeric columns before computing correlation

correlation\_matrix = df.select\_dtypes(include=['float64', 'int64']).corr()

# Visualizing temperature, rainfall, and humidity

def plot\_data():

    plt.figure(figsize=(12, 6))

    # Temperature distribution

    plt.subplot(1, 3, 1)

    plt.hist(df['Temperature'], bins=10, color='red', alpha=0.7)

    plt.xlabel('Temperature (°C)')

    plt.ylabel('Frequency')

    plt.title('Temperature Distribution')

    # Rainfall distribution

    plt.subplot(1, 3, 2)

    plt.hist(df['Rainfall'], bins=10, color='blue', alpha=0.7)

    plt.xlabel('Rainfall (mm)')

    plt.ylabel('Frequency')

    plt.title('Rainfall Distribution')

    # Humidity distribution

    plt.subplot(1, 3, 3)

    plt.hist(df['Humidity'], bins=10, color='green', alpha=0.7)

    plt.xlabel('Humidity (%)')

    plt.ylabel('Frequency')

    plt.title('Humidity Distribution')

    plt.tight\_layout()

    plt.show()

# Print the correlation matrix

print("\nCorrelation Matrix:")

print(correlation\_matrix)

# Plot the data distributions

plot\_data()

5. ScreenShots Output photo pasted



