TASK 3

Weather Data Analysis Project Report

This project involves analyzing a weather dataset to explore precipitation patterns and their variability. The analysis was conducted using Python and Jupyter notebooks, leveraging libraries such as pandas, matplotlib, and seaborn for data manipulation and visualization.

Dataset Description

The dataset used for this project is weather_data_extended.csv, which includes various weather-related variables such as temperature, humidity, and precipitation.

Analysis Objectives

The main objectives of this analysis are:

1. To understand the distribution of precipitation.
2. To explore the variability of precipitation across different times of the year (e.g., by month).
3. To investigate potential correlations between precipitation and other variables such as temperature and humidity.
4. To identify any notable trends or patterns in precipitation data over time.

Methodology

The analysis was structured as follows:

1. **Data Loading and Initial Exploration**
    - The dataset was loaded into a pandas Data Frame.
    - Basic information and summary statistics of the dataset were examined.

2. **Data Cleaning**
    - Missing values were handled appropriately.
    - Data types were checked and corrected if necessary.

3. **Time Series Analysis:**
    - If the dataset contained a 'Date' column, trends in precipitation over time were analyzed using line plots

4. **Exploratory Data Analysis (EDA)**
   - The distribution of precipitation was analyzed using histograms.
   - Precipitation variability by month was explored using box plots (if the dataset contained month data).
   - Scatter plots were used to explore correlations between precipitation and other variables like temperature and humidity.

Findings

1. Distribution of Precipitation

The histogram of the precipitation data showed the distribution, highlighting the most common precipitation values and the spread of the data.

```
plt.figure(figsize=(10, 6))
sns.histplot(weather_data['Precipitation'], kde=True)
plt.xlabel('Precipitation')
plt.title('Distribution of Precipitation')
plt.show()
```

2. Precipitation Variability by Month

If month data was available, a box plot was used to show how precipitation varied across different months.

```
plt.figure(figsize=(12, 6))
sns.boxplot(x='Month', y='Precipitation', data=weather_data)
plt.xlabel('Month')
plt.ylabel('Precipitation')
plt.title('Monthly Precipitation Variability')
plt.show()
```

3. Correlation with Temperature and Humidity

Scatter plots were used to examine the relationships between precipitation and other variables.

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Temperature', y='Precipitation', data=weather_data)
plt.xlabel('Temperature')
plt.ylabel('Precipitation')
plt.title('Temperature vs Precipitation')
plt.show()

plt.figure(figsize=(10, 6))
sns.scatterplot(x='Humidity', y='Precipitation', data=weather_data)
plt.xlabel('Humidity')
plt.ylabel('Precipitation')
plt.title('Humidity vs Precipitation')
plt.show()
```

4. Time Series Analysis

If the dataset included a 'Date' column, a line plot was used to visualize precipitation trends over time.

```
weather_data['Date'] = pd.to_datetime(weather_data['Date'],
errors='coerce')
plt.figure(figsize=(14, 7))
plt.plot(weather_data['Date'], weather_data['Precipitation'])
plt.xlabel('Date')
plt.ylabel('Precipitation')
plt.title('Precipitation Over Time')
plt.show()
```

Conclusions

- **Precipitation Distribution:** The distribution analysis provided insights into the typical precipitation levels and identified any outliers or extreme values.

- **Seasonal Variability:** The box plots revealed how precipitation varied by month, which could be useful for understanding seasonal patterns.

- **Correlation Analysis:** The scatter plots indicated whether there were any significant correlations between precipitation and other weather variables such as temperature and humidity.

- **Time Trends:** The time series analysis helped identify any trends or patterns in precipitation over the analyzed period.

Future Work

Future analyses could include:

1. More advanced time series analysis techniques, such as seasonal decomposition or ARIMA models, to better understand trends and seasonal effects.

2. Inclusion of additional weather variables or external data (e.g., geographical information) to enhance the analysis.

3. Application of machine learning models to predict future precipitation patterns based on historical data.

Code and Resources

The code and resources used for this analysis are available in the project's GitHub repository: https://github.com/guptaakshat2002