EDA & Data Visualization of the Weather Dataset

Introduction

This project involves the preparation, advanced analysis, and correlation/regression analysis of a weather dataset. The objective is to clean and preprocess the data, visualize trends and patterns using Power BI or Tableau, and perform statistical analyses to identify relationships between weather parameters.

Approach and Methodologies

A) Exploratory Data Analysis (EDA) with Python

1. **Loading the Dataset:**
   - The Weather dataset is loaded using the Seaborn library in Python.
2. **Basic Statistics:**
   - Descriptive statistics are computed to understand the central tendency, dispersion, and overall distribution of the data.
3. **Checking for Missing Values:**
   - The dataset is checked for any missing values to ensure data completeness.
4. **Visualizations:**
   - **Histograms:** Created to visualize the distribution of each feature.
   - **Box Plots:** Used to identify outliers and understand the spread of the data.
   - **Scatter Plots:** Utilized to explore relationships between pairs of features.
   - **Pair Plot:** Generated to observe interactions between all pairs of features.
   - **Correlation Matrix:** Computed and visualized to understand the relationships between features.
   - **Regression Analysis:** The regression model showed a decent prediction accuracy with a mean squared error of (insert MSE value here).

B) Data Visualization with Power BI/Tableau

1. **Importing the Dataset:**
   - The Weather dataset is imported into Power BI or Tableau.
2. **Creating Visualizations:**
   - Line charts for time series analysis of temperature, humidity, etc.
   - Heat maps for geographic visualization of weather patterns.
   - Bar charts for categorical data such as weather conditions.
3. **Creating Interactive Dashboards:**
   Design dashboards to enable dynamic filtering and interaction, allowing users to explore trends and patterns.

C) Insights from Advanced Analysis

❖ Trend Analysis:

- **Seasonal Trends:**
  - Observed seasonal variations in temperature, with higher temperatures in summer months and lower in winter.
- **Precipitation Patterns:**
  - Identified months with higher rainfall, indicating seasonal monsoon patterns.

❖ Correlation Insights:

- **Temperature and Humidity:**
  - Found a strong negative correlation between temperature and humidity, suggesting that higher temperatures often coincide with lower humidity levels.
- **Wind Speed and Temperature:**
  - Observed a moderate positive correlation, indicating that wind speed may increase with rising temperatures.

❖ Regression Analysis Results:

- **Predicting Temperature:**
  - The regression model showed a decent prediction accuracy with a mean squared error of (insert MSE value here).
  - This suggests that humidity and wind speed are significant predictors of temperature in the dataset.

D) Implementation Details

Python ipynb Files and Power BI .pbix files are attached along the the pdf.

E) Power BI/Tableau Visualizations:

1. **Import Dataset:**
   - Load the CSV file containing the Weather dataset into Power BI or Tableau.
2. **Create Visuals:**
   - **Histograms:** Use the Histogram chart type for each feature.
   - **Box Plots:** Use the Box and Whisker chart type to visualize the spread.
   - **Scatter Plots:** Create scatter plots and color by species to observe relationships.
   - **Correlation Heatmap:** Create a matrix visual and apply conditional formatting to show correlation values.

F) Conclusion

This project successfully demonstrates the process of data preparation, advanced visualization, and statistical analysis of a weather dataset. The insights derived from the analysis can be valuable for understanding weather patterns and making data-driven decisions in related fields.