Weather Data Analysis Report

Title: Analyzing Real-World Weather Data for Insights

Objective

The goal of this analysis is to clean, analyze, and visualize real-world weather data using Python libraries such as Pandas, Matplotlib, and Seaborn. The dataset contains historical weather information, and the insights generated will help in understanding temperature trends, precipitation distribution, and correlation between weather attributes.

Dataset Selection

For this analysis, the Raleigh-Durham (RDU) Weather History dataset was used. This dataset contains daily weather attributes, including temperature, precipitation, snowfall, and wind speed.

Data Attributes:

- Date
- Minimum Temperature (tmin)
- Maximum Temperature (tmax)
- Precipitation (prcp)
- Snowfall (snow)
- Snow Depth (snwd)
- Average Wind Speed (awnd)

Step 1: Data Cleaning

To ensure data quality, the following cleaning steps were performed:

1. **Handling Missing Values** – Forward fill method (`ffill()`) was used to fill missing values.
2. **Removing Duplicates** – All duplicate records were dropped.
3. **Standardizing Column Names** – Spaces were removed, and names were converted to lowercase.
4. **Data Type Conversion** – Ensured correct formats (e.g., dates to datetime, numerical conversions where necessary).

Step 2: Exploratory Data Analysis (EDA)

To understand the dataset, the following analyses were conducted:

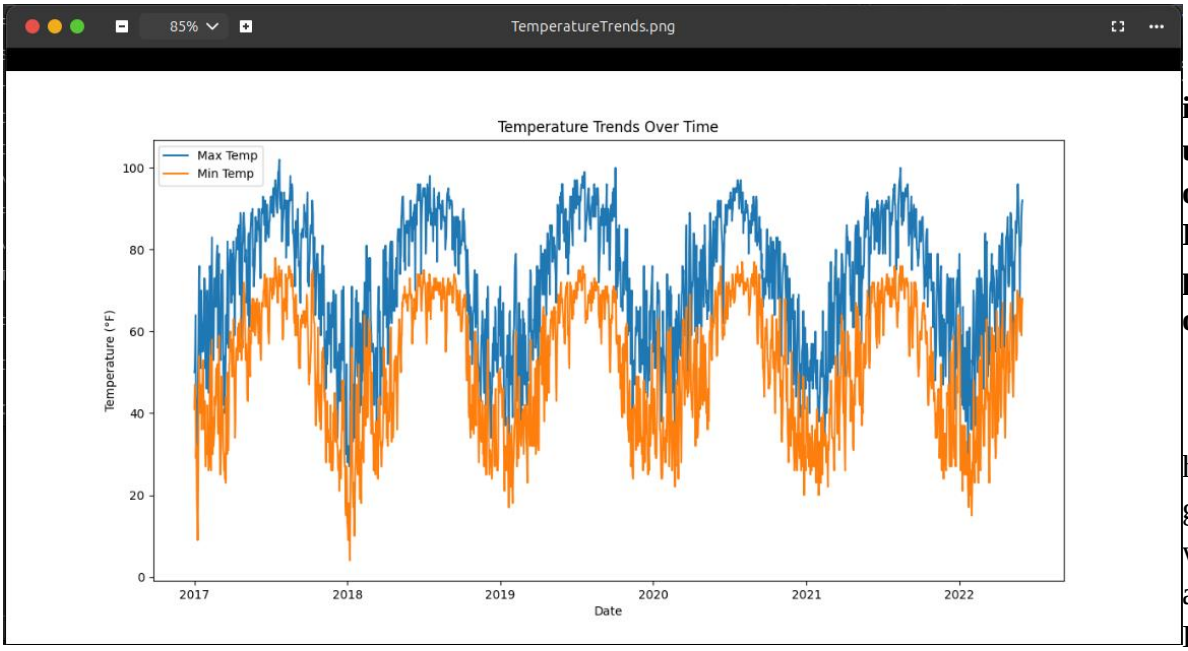
Summary Statistics

- Descriptive statistics such as mean, median, standard deviation, and percentiles were calculated to understand data distribution.

Visualizations

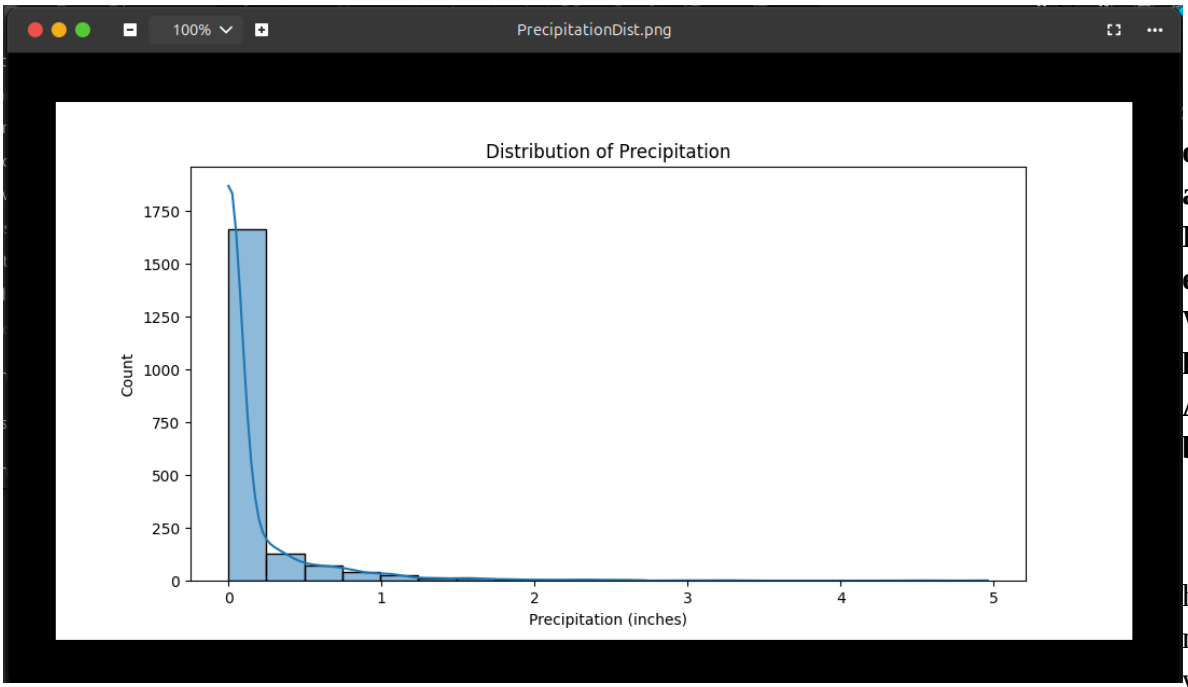
1. **Temperature Trends Over Time**

- A line plot was created to show how maximum and minimum temperatures varied over the years.



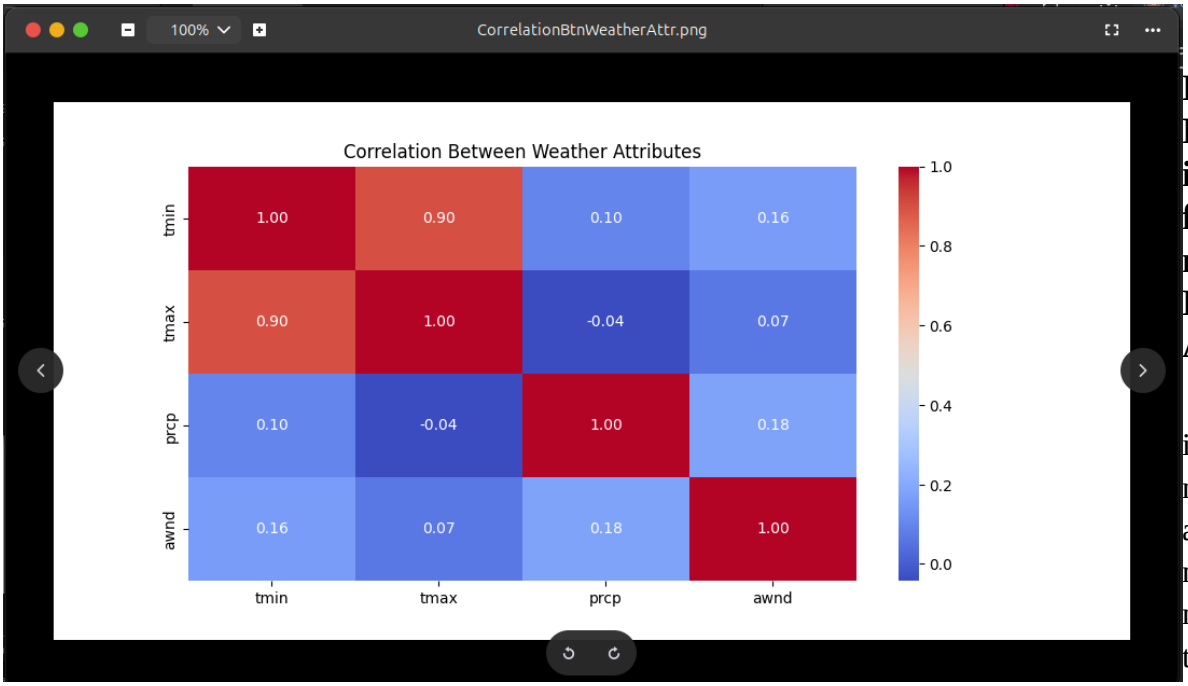
2. **Distribution of Precipitation**

- A histogram with a Kernel Density Estimate (KDE) overlay was plotted to show the distribution of daily precipitation levels.



3. **Correlation Between Weather Attributes**

- A heat map was generated to visualize the correlation between key numerical variables such as temperature, precipitation, and wind speed.



Key Findings from EDA:

- Minimum and maximum temperatures exhibit seasonal trends.
- Precipitation is mostly concentrated around lower values, indicating that heavy rainfall is relatively rare.
- There is a strong positive correlation between minimum and maximum temperatures.
- Weak correlation between precipitation and temperature suggests independent variations.

Step 3: Insights & Implications

From the analysis, the following actionable insights were identified:

1. **Temperature Trends:**
   - There are clear seasonal temperature variations, which can be useful for climate predictions and planning.
2. **Precipitation Patterns:**
   - Most precipitation events involve small amounts of rainfall, with occasional extreme events.
3. **Weather Correlations:**
   - Understanding the relationships between weather attributes can assist in forecasting models.

Conclusion

This project successfully analyzed historical weather data by cleaning, visualizing, and identifying key insights. The findings provide valuable information for meteorologists, researchers, and decision-makers interested in weather patterns and climate trends.

Deliverables

- **Jupyter Notebook (.ipynb)** – Contains Python code and visualizations.

- **Dataset (rdu-weather-history.csv)** – Used for the analysis.
- **PDF Report (this document)** – Summary of findings.

Prepared by: Meshack Kataboi BSE-01-0219/2024
Date: 22/3/2025
Course: Software Engineering, **Unit:** Data Science