Assignment 1 SOC Project ID : 111

Tanvi Sree Jaishankar, Snigdha Reddy

30 May 2025

Objective

Utilize the *Bike Sharing Dataset* from the UCI Machine Learning Repository to quantitatively assess how varying weather conditions influence daily bike rental counts. The analysis should encompass data preprocessing, statistical computation, visualization, and interpretation.This is to test your understanding of python, numpy , pandas and matplotlib.

Dataset

- **Source**: UCI Machine Learning Repository
- **File**: `day.csv`
- **Relevant Columns**:
  - `weathersit`: Categorical variable indicating weather situation:
    Rough Meanings of Numerical Values :
    - 1: Clear, Few clouds, Partly cloudy
    - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
    - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
    - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
  - `cnt`: Total count of bike rentals (casual + registered)

Tasks

1. **Data Preprocessing**
   - Load the `day.csv` dataset using pandas.
   - Check for and handle any missing values or anomalies.
   - Convert categorical variables into appropriate formats if necessary.
2. **Statistical Analysis**
   - Group the data by `weathersit` and calculate the average `cnt` for each category.
   - Compute the percentage change in average rentals between different weather conditions.
   - Use NumPy to calculate statistical measures (mean, median, standard deviation) of bike rentals under each weather condition.
3. **Data Visualization**
   - Create a bar chart to display average rentals per weather condition.
   - Annotate the chart with percentage changes between categories.
   - Generate scatter plots to visualize relationships between weather variables (e.g., temperature, humidity) and bike rentals.

4. **Interpretation**
    - Analyze the results to understand how different weather conditions affect bike rental counts.
    - Summarize key findings and provide actionable recommendations for the bike-sharing company based on the insights.

Deliverables

- Jupyter Notebook containing:
    - Data preprocessing steps
    - Statistical analysis computations
    - Visualizations with appropriate labels and annotations
    - Interpretation of results
- A brief report summarizing the methodology, findings, and recommendations.