

Topic Proposal

Team Members: Smitha Patil, Indulekha Siddabathuni, Mohiddin Bacha Shaik

Topic: Capital Bike Share Data Analysis

Capital Bike Share is a popular bike-sharing system in Washington D.C. that allows users to rent bicycles for short-term use. The system has grown rapidly since its launch in 2010 and now consists of over 5,000 bicycles and 600 stations across the city. The system generates a large amount of data on user behavior, including trip duration, start and end station locations, and user demographics.

Our data set is from Capital Bikeshare, which contains 20 columns and several rows from year 2010 to present. We are integrating Weather data, Holidays and Weekdays data from other datasets to support our smart questions.

Attributes:

1. *Start_at*: The date and time when the bike ride started.
2. *End_at*: The date and time when the bike ride ended.
3. *Start_station_number*: The ID number of the station where the bike ride started.
4. *Start_station*: The name of the station where the bike ride started.
5. *End_station_number*: The ID number of the station where the bike ride ended.
6. *End_station*: The name of the station where the bike ride ended.
7. *Bike_number*: The ID number of the bike that was used for the ride.
8. *Member_type*: The type of membership that the rider has.
9. *Temp*: The temperature is in degrees Fahrenheit.
10. *Dew*: The dew point in degrees Fahrenheit.
11. *Humidity*: The relative humidity as a percentage.
12. *Windspeed*: The wind speed is miles per hour.
13. *Holiday*: A binary variable indicating whether the date was a holiday or not.
14. *Weekday*: The day of the week (Monday-Sunday).
15. *TimeofDay*: The time of day (morning, afternoon, evening, or night).
16. *Season*: The season of the year (spring, summer, fall, or winter).
17. *HourOfDay*: The hour of the day (0-23).
18. *Month*: The month of the year (1-12).
19. *Day*: The day of the month (1-31).
20. *Year*: The year of the ride.

SMART questions:

1. What is the impact of seasonality on bike rental demand?
2. What are the key factors that influence bike rental demand, and how do they affect the overall bike-sharing system performance?
3. What are the most popular bike stations and routes, and how do they vary by time of day, season, and day of the week?
4. How does the weather (temperature, humidity, wind speed) impact bike usage patterns?
5. How does bike usage vary during holidays compared to regular days?

Models: Exploratory Data Analysis (EDA), Simple Linear Regression(SLR), Multi-Linear Regression(MLR), Logistic regression, Decision tree.

Dataset Source: [Capital Bike Share](#)

GitHub Repo Link: [Final project DATS6101 SIM](#)