# Final Report - Capital Bikeshare Analysis

# Final Report: Capital Bikeshare Analysis (Feb–June 2023)  
  
## 🧩 Introduction  
  
This project focuses on understanding the relationship between bike usage (pickups and drop-offs) and weather conditions using Capital Bikeshare data in Washington, DC from February to June 2023. The primary goal is to predict bike usage patterns through multiple linear regression models and identify optimal predictors.  
  
## 📁 Data Overview  
  
- \*\*Trip Data\*\*: Includes start and end timestamps, station details, and trip durations.  
- \*\*Weather Data\*\*: Includes daily weather parameters like temperature, humidity, precipitation, snow, and windspeed.  
- Data was merged, cleaned, and filtered to contain only the period from February to June 2023.  
  
## 🛠️ Methodology  
  
1. \*\*Data Cleaning\*\*  
 - Converted timestamps and filtered the target months.  
 - Grouped trip counts by date.  
 - Joined trip data with weather data.  
  
2. \*\*Model Development\*\*  
 - Built multiple linear regression models using combinations of weather variables.  
 - Trained models separately for predicting pickups and drop-offs.  
 - Evaluated using train/test split and Mean Squared Error (MSE).  
  
## 📊 Results  
  
| Model | Train MSE (Dropoffs) | Test MSE (Dropoffs) | Train MSE (Pickups) | Test MSE (Pickups) |  
|-------|----------------------|----------------------|----------------------|---------------------|  
| Model 3 | ~81.87 | ~86.97 | - | - |  
| Model 5 | ~81.46 | ~81.12 | - | - |  
  
- Model 3 with features (`temp`, `humidity`, `precip`) performed best for \*\*pickups\*\*.  
- Model 5 with additional weather features showed lowest MSE for \*\*drop-offs\*\*.  
  
## 📌 Conclusion  
  
Weather conditions significantly influence bike usage. Linear regression models can reasonably predict ride counts based on weather data. This insight can be useful for planning resources and optimizing station logistics in bikeshare systems.