# README - Capital Bikeshare Analysis

# Capital Bikeshare Analysis (Feb–June 2023)  
  
This project analyzes Capital Bikeshare data from February 2023 to June 2023 to explore the relationship between weather factors and bike ride patterns in Washington, DC.  
  
## 📊 Objective  
  
To predict the number of bike pickups and drop-offs using weather data (temperature, humidity, precipitation, windspeed, etc.) and identify the most effective linear regression models.  
  
## 📁 Dataset  
  
The data was collected from the official Capital Bikeshare system data repository:  
[https://ride.capitalbikeshare.com/system-data](https://ride.capitalbikeshare.com/system-data)  
  
## 🔧 Tools & Libraries  
  
- Python  
- Pandas, NumPy  
- Matplotlib, Seaborn  
- scikit-learn (LinearRegression, Ridge, Lasso, CV methods)  
  
## 📌 Methodology  
  
1. \*\*Data Cleaning & Preprocessing\*\*  
 - Merged multiple CSV files.  
 - Converted timestamps and filtered date range (Feb–June 2023).  
 - Joined weather data with trip data.  
   
2. \*\*Model Building\*\*  
 - Built 7 different linear regression models with different combinations of weather features.  
 - Applied Linear Regression, Ridge, and Lasso models.  
 - Evaluated models using Mean Squared Error (MSE).  
  
3. \*\*Evaluation\*\*  
 - Compared train and test MSE for all models to identify best-performing ones.  
  
## ✅ Key Findings  
  
- \*\*Best Pickup Model\*\*: Model 3 (`temp`, `humidity`, `precip`) with test MSE ≈ 86.97  
- \*\*Best Dropoff Model\*\*: Model 5 (`temp`, `humidity`, `precip`, `snow`, `windspeed`) with test MSE ≈ 81.12  
  
## 📂 How to Run  
  
Open the notebook `Final Submitted.ipynb` in Jupyter Notebook and run all cells in order after ensuring all required libraries are installed.  
  
---