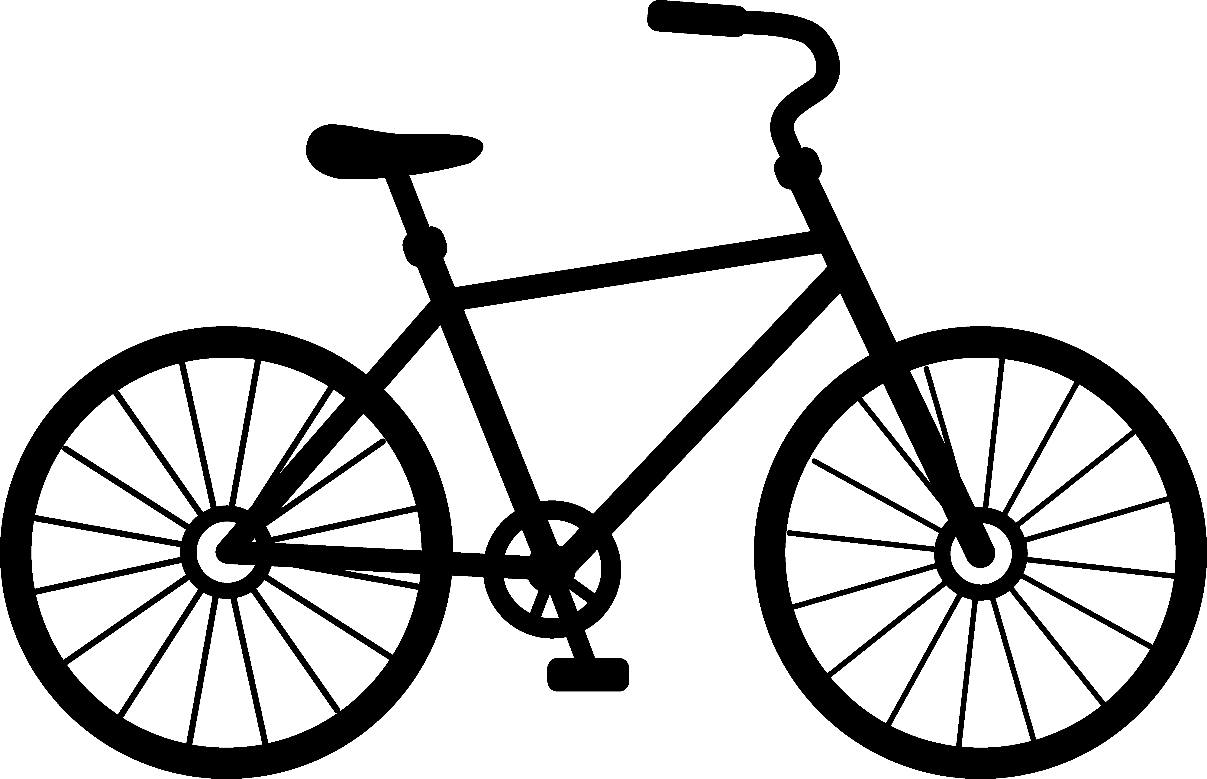
**Analysis of Casual vs. Registered Users of a Bike Share System in Washington DC**

April 13, 2023

UBC STAT 306

A report by Lab Group 2B1

|  |  |
| --- | --- |
| **Name** | **Student Number** |
| Alexander Proskiw | 27194166 |
| Youjung Kim | 38762639 |
| Mengfei Chen | 79921177 |
| Timothy Gao | 62490230 |



# Introduction

## Motivation

Bike sharing is where customers can rent a bike and return it to different places across a city, often meant to reduce vehicle dependency and extend transit capability in an urban area. Bike share services having been growing in popularity around Vancouver, with companies such as Mobi Bike Share, HOPR Bike Share, and Lime Electric Bike Sharing establishing themselves in recent years. With several cyclists in our group, we are accustomed to cycling all days of the week and in a variety of weather conditions in order to commute to school and/or work. In contrast, bicycle sharing services are often advertised as a fun and convenient way to enjoy a weekend while the weather is nice. But is this stereotype unfounded? In this report, we aim to explore the factors that influence bike share users.

## Data

The data being used for this project is [bike rental data](https://www.kaggle.com/datasets/prepinstaprime/bike-rental-data)[[1]](#footnote-1) found on Kaggle. The dataset contains information on bicycle rentals associated with bike-sharing systems in Washington, DC from 2011 to 2012. The Kaggle dataset doesn’t say where or how the data was obtained, although further research indicates it is likely from the Capital Bikeshare system.

## Variables

Response variables include:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| casual | The number of non-registered users renting bikes at that time |
| registered | The number of registered users renting bikes at that time |

Explanatory variables include:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| year | Categorical variable (2011, 2012) |
| time | Categorical variable representing 1-hour segments within a day (0, 1, 2, …, 23) |
| season | Categorical variable (January to March, April to June, July to September, October to December) |
| holiday | Categorical variable (not a holiday, holiday) |
| workingday | Categorical variable (weekend, weekday) |
| weather | Categorical variable (clear, mist/clouds, light rain/snow, heavy rain/snow) |
| atemp | Continuous variable representing the normalized feeling temperature (°C) |
| humidity | Continuous variable representing the humidity (normalized by 100) |
| windspeed | Continuous variable representing the wind speed (miles per hour normalized by 67) |

Several alterations were made to the raw dataset. Specifically:

* The raw dataset included a variable for datetime, representing a year-month-day-hour. We separated this into its components, and only kept year and hour as categorical variables. Month was dropped as the dataset already included a variable for season (representing a 3-month period). Day was dropped as the dataset already included a variable for weekday/weekends, which seems more applicable than whether the day is the 5th of the month vs the 17th of the month.
* The raw dataset included variables for both “temperature” and “apparent temperature”. These two variables were highly correlated (R2 = 0.985). As such, temperature was discarded in favour of apparent temperature, with the rationale being that the temperature a person “feels” would likely have a greater influence on their decision to bike.
* The raw dataset included variables for the casual, registered, and total number of users. The total number of users was not considered as our research questions focus on examining the differences between casual and registered users.

## Research Questions

One might expect that registered users use bike rentals more routinely, perhaps for commuting. In contrast, casual users might use bike rentals more regularly on weekends when the weather is nice.

We aim to use bike sharing data to investigate the difference in usage patterns between casual and registered users by exploring the following questions:

* Do casual users rent more bikes on weekends than weekdays?
* Do registered users rent more bikes at typical commuting times (such as 7-9am and 4-6pm)?
* Are the number of users that rent bikes influenced by the weather (rain, temperature, humidity)?

To answer these research questions, we will perform a combination of visualization and linear regression modelling.

# Analysis

## Visualization

To gain an understanding of the influence of each explanatory variable on the response variables, plots were made depicting the number of casual/registered users vs each explanatory variable, not accounting for any additions/interactions from other variables.

Figure 1 on the next page includes plots for the number of casual bike users, while Figure 2 on the following page includes plots for the number of registered bike users.

Figure 1: Effect of Different Explanatory Variables on the Number of Casual Bike Users

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Figure 2: Effect of Different Explanatory Variables on the Number of Registered Bike Users

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

## Summary

Discuss observations from the above graphs, differences between casual and registered.

## Models

Fit different models of varying numbers of parameters

Highlight which parameters are most important. For example, the 3 parameter models include:

* For casual: Weekday, temperature, humidity
* For registered: 8am, 5pm, 6pm

Fit full models, show plots

Discuss fitting other types of models (quadratic, log, etc.)

Fit Poisson models, show improvements in plots

# Conclusion

## Key Findings

Use data to answer the research questions

* Do casual users rent more bikes on weekends than weekdays?
* Do registered users rent more bikes at typical commuting times (such as 7-9am and 4-6pm)?
* Are the number of users that rent bikes influenced by the weather (rain, temperature, humidity)?

## Limitations and Future Improvements

Methods of improvement include:

* Use month instead of season (e.g., 12 months vs 4 seasons)
* use day of week (Mon, Tues, etc.) rather than just weekday vs weekend

1. Dataset accessed March 15, 2023 from <https://www.kaggle.com/datasets/prepinstaprime/bike-rental-data> [↑](#footnote-ref-1)