

Forecasting Bikesharing Usage for DC's Capital Bikeshare System

Table of Contents

I. Introduction 1. Brief Overview of the Project

II. Current Status of the Project A. Data Cleaning and Preprocessing B. Exploratory Data Analysis/Visualizations C. Key Findings/Adjustments D. Challenges Encountered

III. Ongoing Work A. Intermediate Data Preparation B. Preliminary Modeling C. Expected Milestones and Deliverables

IV. Explanation and Background Information A. Provide Background on the Statistical Models Used B. Explain the Methodology and Approach C. Justify the Selection of Models and Techniques

V. Literature Review Summary

VI. Works Cited

Introduction

Current Status of the Project

Data Cleaning and Preprocessing

The dataset has required minimal cleaning. We had to convert several variables into factor variables (season, holiday, weekday, workingday, weather). Additionally we noted that the data key for our dataset mislabeled the season factor variable, which was trivial to correct.

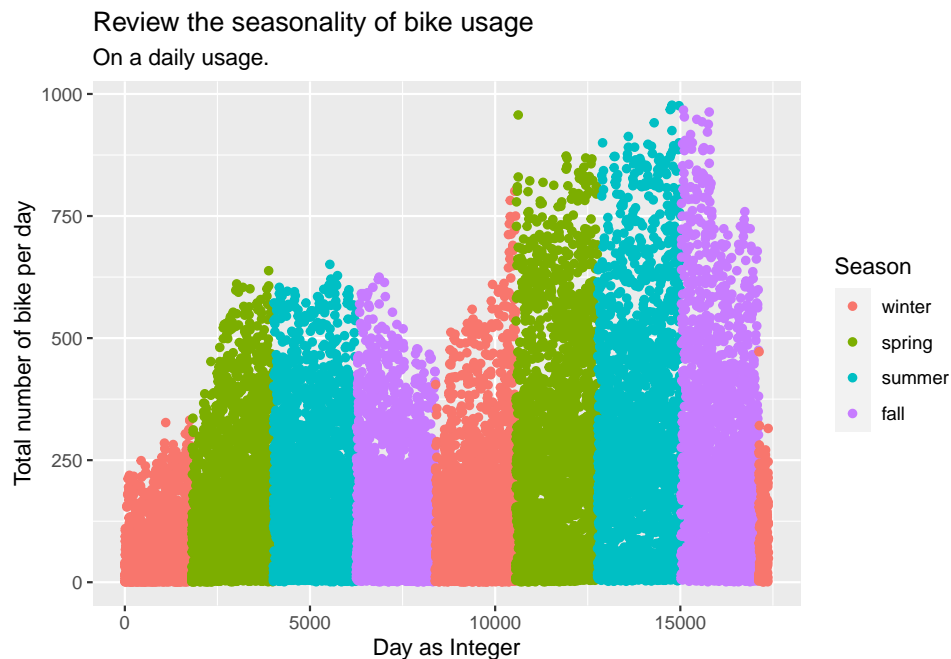
EDA / Visualization

Not all of our predictors or response variables are distributed normally. In particular, humidity exhibits leftward skew, and windspeed exhibits rightward skew.

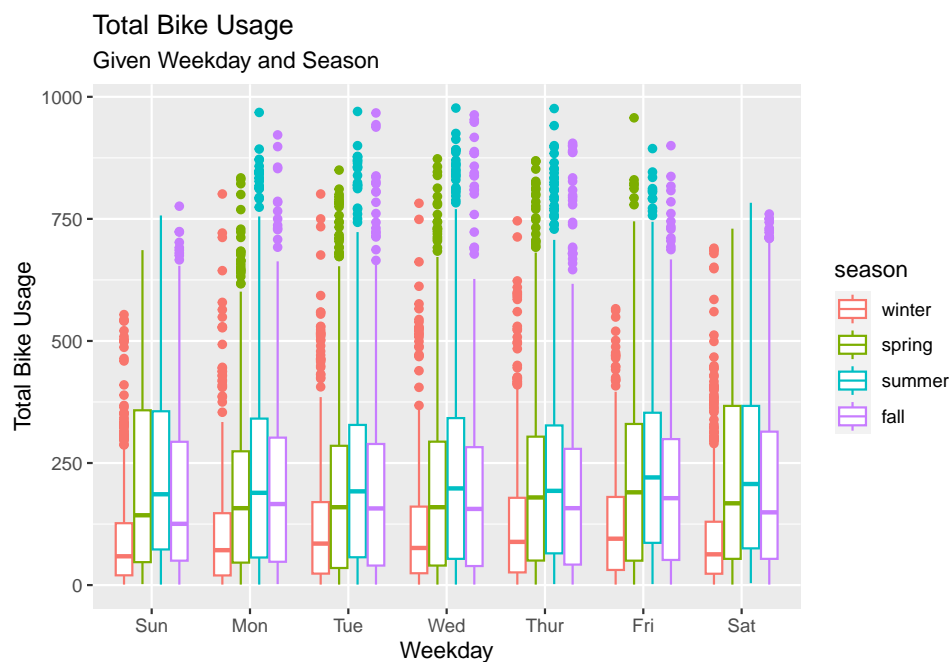
Casual users exhibit rightward skew, while registered users are not too far from the normal distribution.

Preliminary EDA reveals strong seasonality - overall usage on a daily basis is much higher in spring/summer than fall and much more than winter.

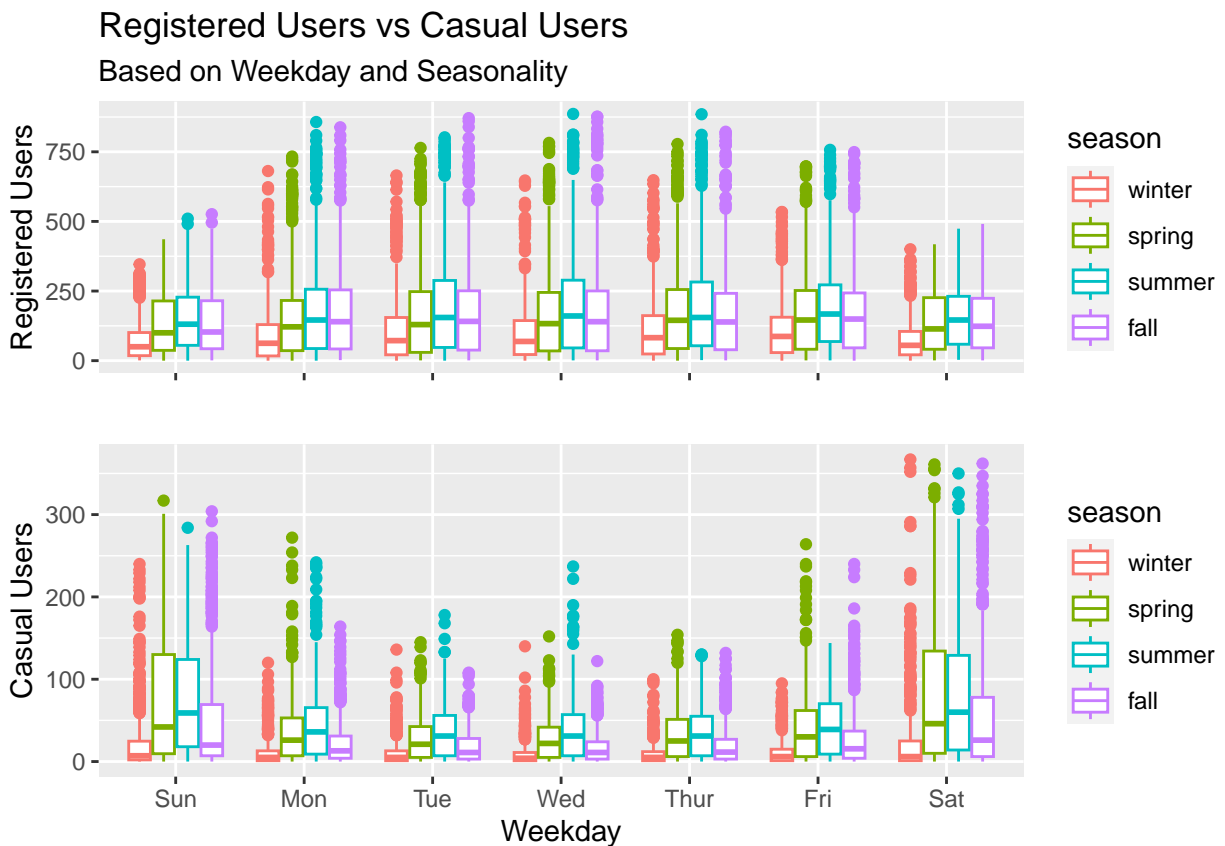
We wanted to look at the overall bike usage as it relates to seasonality. We had the hypothesis that we would have more users in the summer than in the winter. Looking at the plot below, we can see that not only do we have the max bike usage during the summer, but we also seem to have an increase to bike usage as the service becomes more mature.



The next thing we wanted to take a look at is see whether the day of the week had a large impact on the total bike usage. This is important to see whether a specific day will have the majority of usage or if they are relatively evenly spread out. Ideally we want to see the data spread out. This would allow us to increase on the capitalization of renting our bikes every day.



We can see by the above given boxplot, that the data appears to be evenly spread out between each weekday and across all the seasons. This shows that in general each day will yield approximately the same number of riders, which will help increase our profit. The question though is do we see a difference between casual and registered users.



We can see above that the majority of the **casual** user's usage is on the weekend, and the registered user uses them consistently throughout the week. This indicates that we might need to look at different models depending on the user type. We might see a little more non-linearity with casual users than with the registered users.

Ongoing Work

Anticipated challenges:

- Distinguishing between the influences of weather and the influence of the season, particularly on casual usage. DC has many visitors in Spring and Summer who would show up under the casual response variable.
- Accounting for the overall increase in usage over the two years spanning our dataset. The dataset was collected towards the beginning of the Capital Bikeshare program, so findings we draw from the overall increasing usage would not necessarily translate directly to mature systems.

Explanation and Background Information

Literature Review Summary

Existing literature around bikeshare usage generally emphasizes the following:

- Time of day is typically the most important predictor, but different days of the week have different trends based on time of day
- Specifically, usage is often bimodal on weekdays reflecting commuter patterns
- Usage is not bimodal on weekends, typically with the highest value in mid-afternoon.
- Usage increases as temperature increases, then starts to decrease as temperatures go into the 90s (Fahrenheit), which can be too hot
- Precipitation of any amount discourages cycling
- High humidity has a negative effect on cycling
- High winds can have a negative effect on cycling
- Usage is often higher in spring and summer, and lowest in winter

Works Cited

Bean, R., Pojani, D., & Corcoran, J. (2021). How does weather affect bikeshare use? A comparative analysis of forty cities across climate zones. *Journal of Transport Geography*, 95. <https://doi.org/10.1016/j.jtrangeo.2021.103155>.

Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, 54. <https://doi.org/10.1016/j.scs.2019.101882>

Ashgar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, 7(2), 261-268. <https://doi.org/10.1016/j.cstp.2019.02.011>