

Cyclistic Analytical Case Study Report

By: Mahir Tahmid

Contact: mahirtahmid@gmail.com

1. Introduction

This report documents the analysis process of the Cyclistic bike-share case study provided by the course, [Google Data Analytics Capstone: Complete a Case Study](#) in the [Google Data Analytics Professional Certificate](#) program. In the case study, I was guided by a few questions and one key task to complete the analysis where I had to follow the steps taught to me in the program.

2. Business Task

Finding the difference in usage of bikes between casual users and annual members of the fictional bike-sharing company, Cyclistic.

3. Prepare

Cyclistic trip data was downloaded from [Index of bucket "divvy-tripdata"](#). For this project, data from the second quarter of 2019 to the first quarter of 2020, time spanning one year, was used.

The data was created in and provided by the company itself and so, the data holds integrity. The data had enough information to determine how the casual riders differed in usage from the annual members.

3.1. Limitations

Because of data-privacy issues, the riders' personally identifiable information could not be used. So, past purchases could not be connected to credit card numbers to determine information such as the riders' living area or whether the casual riders have purchased multiple single purchases. As a result, the rides could not be connected to the users. There was no way to know how many rides a single user took.

The dataset did not provide any details on the pricing of the rides. So, no insight can be drawn in that regard for distinguishing between the types of riders.

The data had some issues. They are listed below.

- The column names were inconsistent.
- The ride_id and the rideable_type columns had different types of data for the different datasets.
- Some fields (lat, long, birthyear and gender) were exempted from the datasets from 2020.
- There is no field for the length of the rides.
- There are some entries when the bikes were taken out of docks and checked for quality. In these cases, the calculated ride length appeared negative.

4. Process

R programming language was used for this project. The data was too big to be compiled in a spreadsheet and SQL does not offer visualization tools. These problems are nullified if R/ Python is used.

The problems with data stated in the Prepare step were dissolved through R programming. The documentation about the cleaning process can be found [in this html file](#) created using R Markdown. The original data was divided into four parts in accordance with the four quarters. The data was combined into one big dataframe to make the task easier.

[A dashboard](#) was created with Tableau to further visualize the data.

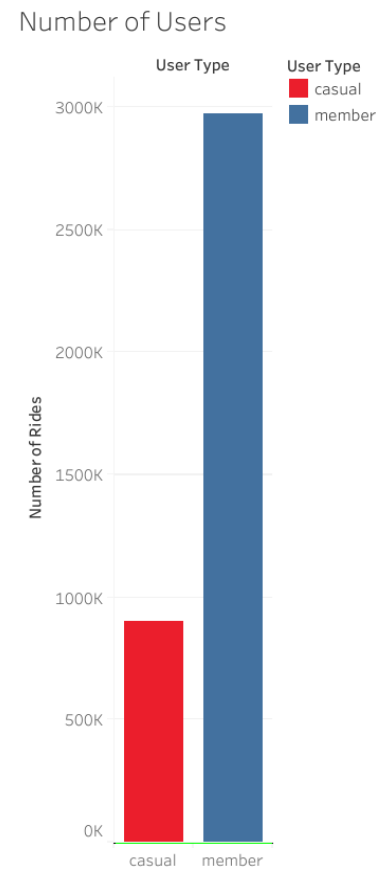
5. Analyze

There were a total of 3876042 rides recorded in Cyclistic's database during the time period of one year spanning from the second quarter of 2019 to the first quarter of 2020. The number of rides by annual members was more than three times that of the casual riders.

The average and median for casual riders was much higher than those of the annual members. The maximum ride length for both types of users was very high.

	Trip Duration (second)
Minimum	1
Median	712
Average	1479
Maximum	9386024

Table: Overall Trip Duration Statistics

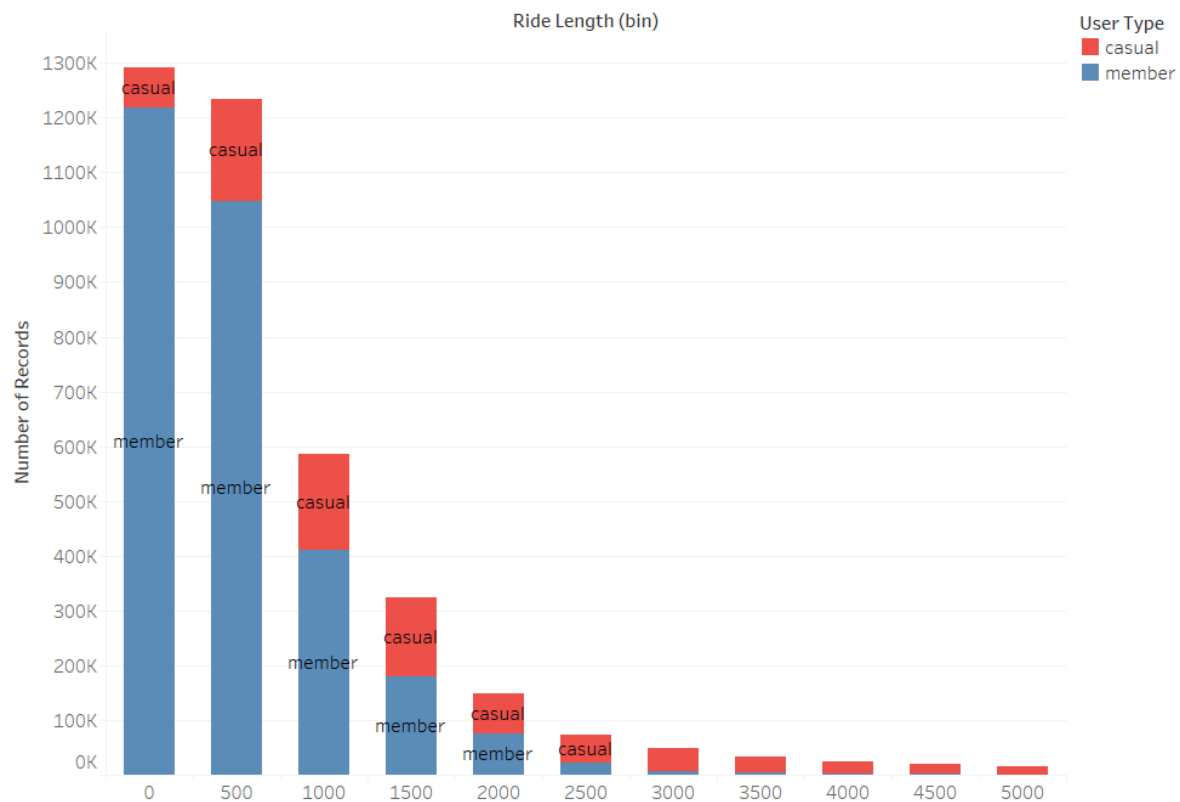


	Trip Duration (second)	
	Casual	Member
Minimum	2	1
Median	1546	589
Average	3553	850
Maximum	9387024	9056634

Table: Trip duration comparison between casual and member customers

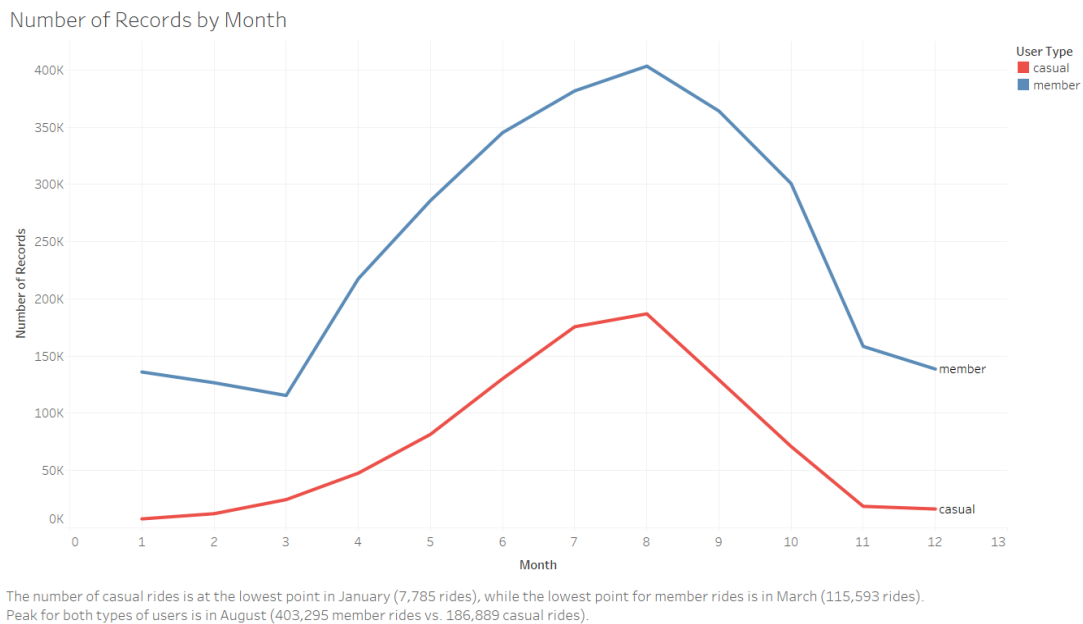
Annual members mostly use Cyclistic's services for short durations. They use the company's services more frequently compared to the casual riders. Most of the high ride lengths were recorded by casual riders as shown in the visual below. The trend continues if the filter for the ride-length is removed.

Ride-Length's Distribution



X-axis shows ride length in seconds. As ride length increases, the ratio moves towards casual riders.

No definitive insights or differences between the different types of riders could be extracted by categorizing the records by month.



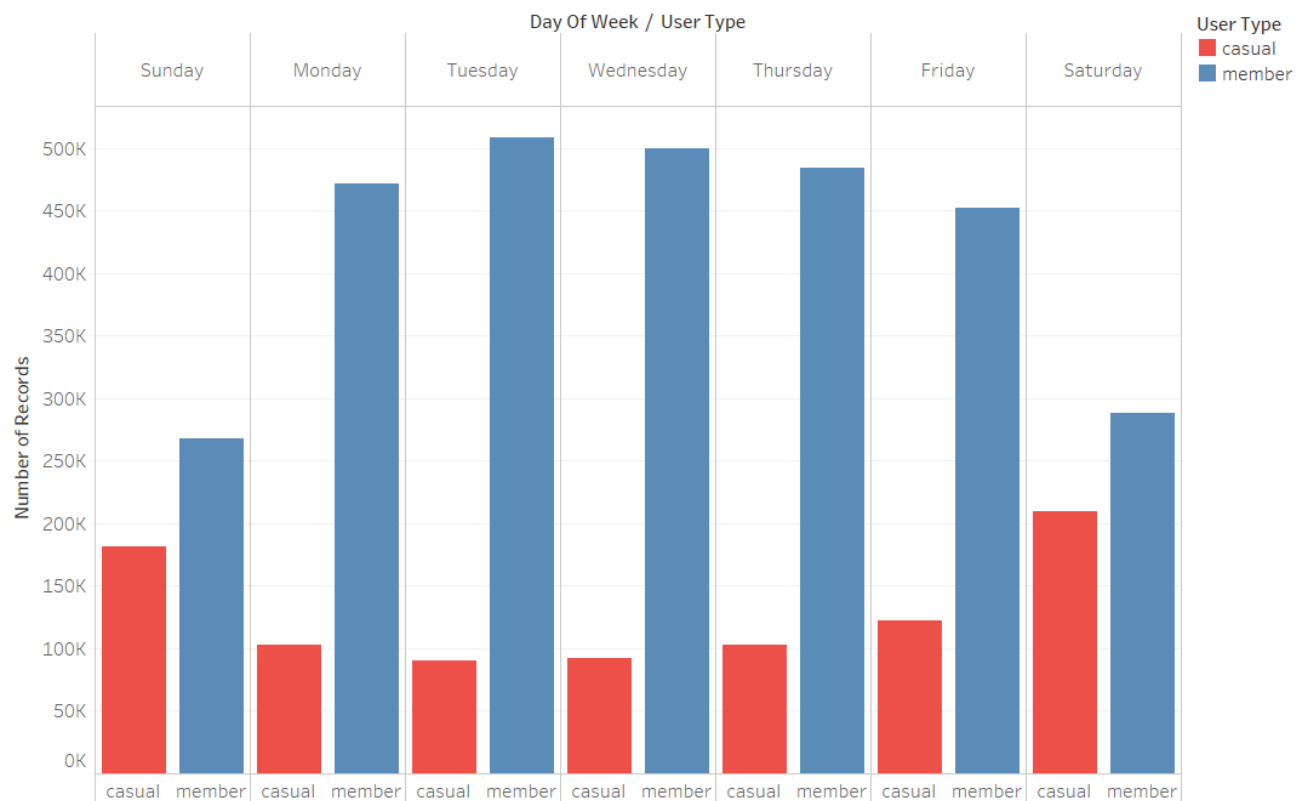
The number of rides for annual members was higher on all weekdays, but one weekends, the number of members fell and the number of casual riders rose. The average ride-length was similar on all weekdays. The table and visuals below support this.

	Day	Average	Number of rides
Casual	Sunday	3581	181293
Member	Sunday	920	267965
Casual	Monday	3372	103296
Member	Monday	843	472196
Casual	Tuesday	3596	90510
Member	Tuesday	826	508445
Casual	Wednesday	3719	92457
Member	Wednesday	824	500329
Casual	Thursday	3683	102679
Member	Thursday	824	484177

Casual	Friday	3774	122404
Member	Friday	825	452790
Casual	Saturday	3332	209543
Member	Saturday	969	287958

Table: Number of Rides And Average trip duration by day

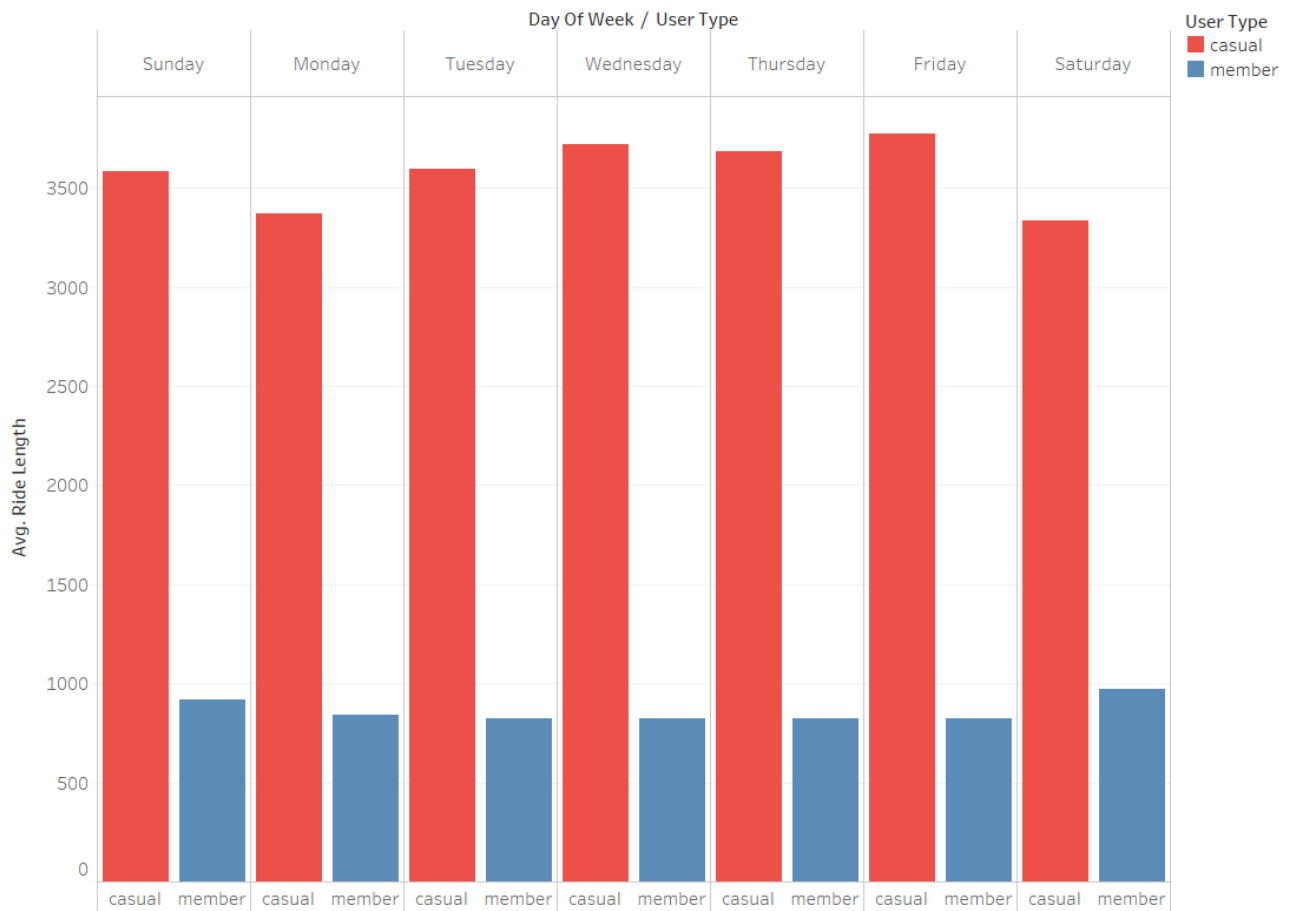
Number of Rides by Weekday



Number of casual rides is higher on all weekdays.

In weekends, number of members increase and inversely, number of casuals decrease. But the number of casual rides remains higher than the number of member rides.

Average Ride Length by Weekday



Average of Ride Length for each User Type broken down by Day Of Week. Color shows details about User Type.

5.1. Summary

Based on the information extracted, the following insights could be inferred.

- I. Annual members use Cyclistic's services a lot more frequently than casual riders, but the average ride length of casual riders is much higher than that of annual members.
- II. Most of the high ride lengths are recorded by casual riders.
- III. On weekends, the number of casual rides increases while the number of member rides decreases.

6. Recommendations

Based on the analysis, here are the top three recommendations.

- I. Make longer rides less expensive exponentially for annual members. On the other hand, the rides can be made more expensive as the length gets bigger for casual riders.
- II. Make a package for annual members where a few rides (3-4) per month will be charged less irrespective of the ride length. The riders can select which rides will have less charge.
- III. Make rides cheaper for members on weekends.