

CYCLISTIC BIKE-SHARE ANALYSIS

Sylvain Boutros

Table of Contents

Introduction	2
Ask.....	3
Prepare.....	4
Process	5
Cleaning.....	5
Analyze and Share	6
Count of Riders.....	6
Average Ride Duration by Rider Types.....	7
Number of Rides by Rider Types and Weekday	7
Average Ride Duration by Rider Types and Weekday	8
Summary Table of Day of Week Usage and Average	8
Act	9
Conclusion.....	9
Recommendations	9

Introduction

This project is part of the [Google Data Analytics Professional Certificate](#). As a junior data analyst in Cyclistic's marketing analyst team, I am tasked with analyzing the bike usage patterns of annual members and casual riders. Cyclistic, a bike-share company based in Chicago, launched its program in 2016 and has experienced remarkable growth. With a fleet of 5,8214 geotracked bicycles distributed across 692 stations in the city, the program offers riders the flexibility to pick up and return bikes at any station within the system.

The finance analysts at Cyclistic have determined that annual members contribute higher profits compared to casual riders. Recognizing the importance of annual memberships for future growth, the director of marketing has assigned me the responsibility of understanding the differences in how annual members and casual riders utilize the bike-share service. By conducting an in-depth analysis of the historical bike trip data, my goal is to provide a comprehensive summary, including supporting visualizations and key findings, along with my top three recommendations based on the analysis.

This analysis will focus on identifying distinct usage patterns, including ride duration, frequency, popular routes and specific times and days of the week when the service is utilized the most. To align with the marketing strategy development timeline, I will complete this analysis within one week. The outcome of this project will play a crucial role in driving the future success of Cyclistic's bike-share program.

To complete this project, I followed the six steps of the data analysis process:

- **Ask**
- **Prepare**
- **Process**
- **Analyze**
- **Share**
- **Act**

Ask

1. **How do annual members and casual riders use Cyclistic bikes differently?**
2. **The key stakeholders are:**
 - a. **Lily Moreno:** The director of marketing and my manager
 - b. **Cyclistic executive team:** The executive team will decide whether to approve the recommended marketing program.

Prepare

- Obtained the historical trip data from **Cyclistic** for analysis by downloading the previous 12 months of [Cyclistic trip data](#) (May 2022 to April 2023).
- The data is organized in a structured way, in 12 different comma-separated value (CSV) files for each of the months.
- There is no issue of bias or credibility. **For the purpose of this case study** the data seems to be Reliable, Original, Comprehensive, Current, and Cited.
- The data is made available by **Motivate International Inc.** under the specified [license](#).
- The data does not contain any personally identifiable information (PII) of riders.
- Data integrity is performed using data cleaning, validation, and verification to ensure accuracy and integrity of the data. These steps may include checking for missing values, outliers, and inconsistencies within the dataset.

Notes:

There are some limitations with the dataset provided such as prices for membership, demographic information about the riders that could help in covering a more in-depth analysis. But as stated prior, for the purpose of this exercise the data is perfectly fine.

Process

Excel and R were used to process the data. Excel is great at quickly glancing at the individual files before using R. After the initial glance, R was used for cleaning, analyzing, and visualizing the data.

Cleaning

- Handled missing data in “**start_station_name**” and “**end_station_name**” by looking up exact match in the “**start_lat**”, “**start_lng**”, “**end_lat**”, “**end_lng**”, “**start_station_id**”, or “**end_station_id**”
- Removed **missing data** when it could not be added by the lookup of similar data
- Removed **duplicates** from files
- Removed the following from the dataset: “**start_lat**”, “**start_lng**”, “**end_lat**”, “**end_lng**”, “**start_station_id**”, and “**end_station_id**”
- Created “**day_of_week**” which represent the day of the week for each observation, “**ride_length**” as the duration (**in minutes**) based on start and end times, and “**date**” as the full date of the observation.
- Removed observations that had “**ride_length**” **less than or equal to zero**.
- Removed observations that had “**ride_length**” **outliers**.
- Formatted “**day_of_week**” and “**ride_length**” as numeric, and later formatted “**day_of_week**” to a string.

These steps transformed the data from its raw format to a usable format suitable for analysis.

Analyze and Share

After cleaning the data in R, I ran some descriptive analysis on the it. From May 2022 to April 2023, Cyclistic has serviced over 5, 000, 000 customers. I calculated the count of annual members and casual riders as can be seen in Figure 1. There was a total of 5,322,080 riders that used Cyclistic bike-share. Out of those 2,004,701 were casual riders and 3,317,379 were annual members. The days where we saw most of the activity were on Thursday for annual members and Saturday for casual riders. The average ride duration was approximately 13 minutes for the casual riders and approximately 11 minutes for the annual members. The daily average ride duration was also calculated and can be found in Table 1.

Count of Riders

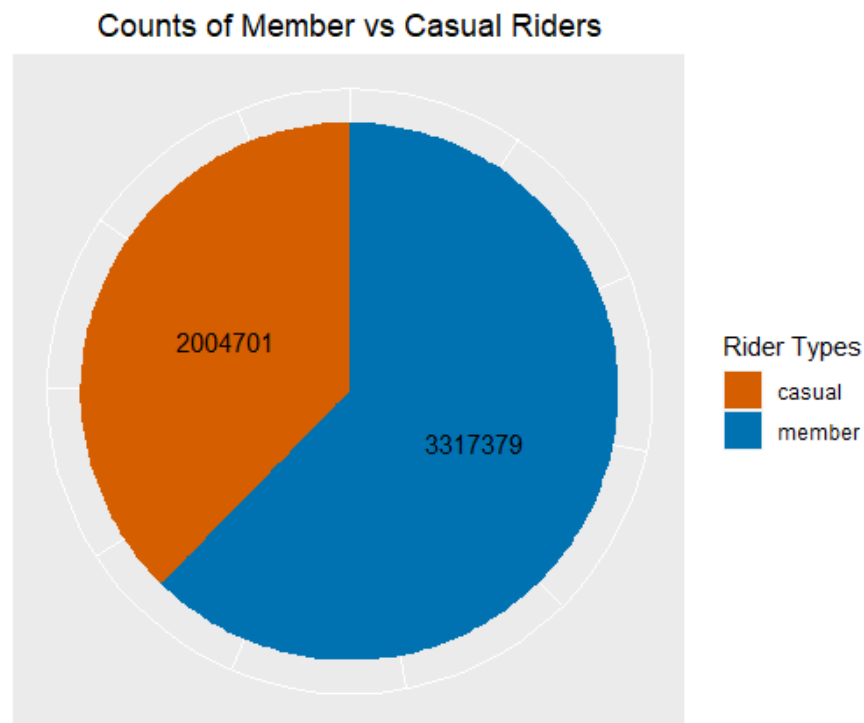


Figure 1. There was a total of 5,322,080 riders over the last twelve months, 2,004,701 were casual rider and 3,317,379 were annual members

Average Ride Duration by Rider Types

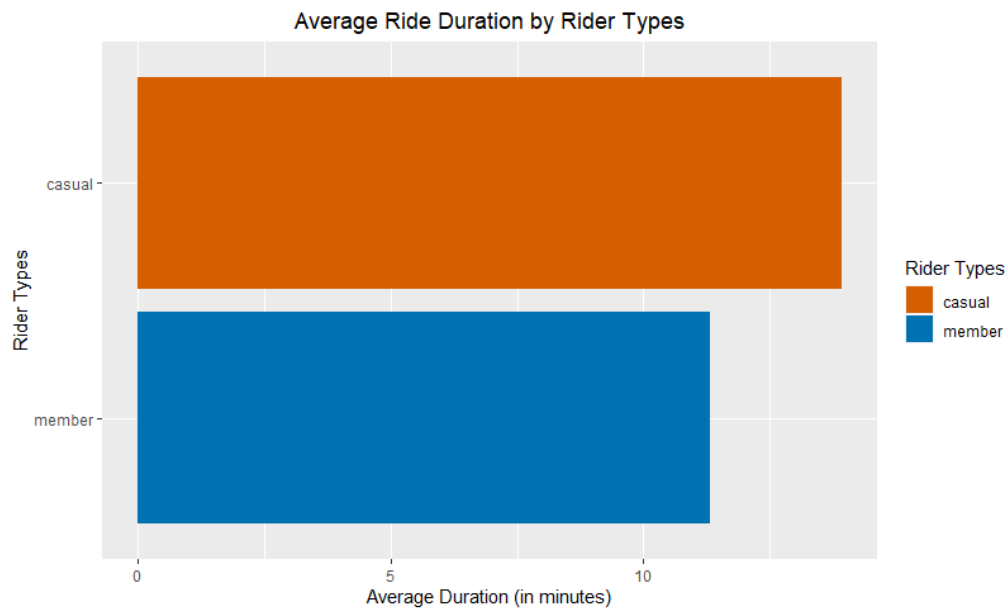


Figure 2. The average duration of the ride by rider type. Casual riders had a 12.88 minute ride on average whilst the annual members had a 10.53 minute ride

Number of Rides by Rider Types and Weekday

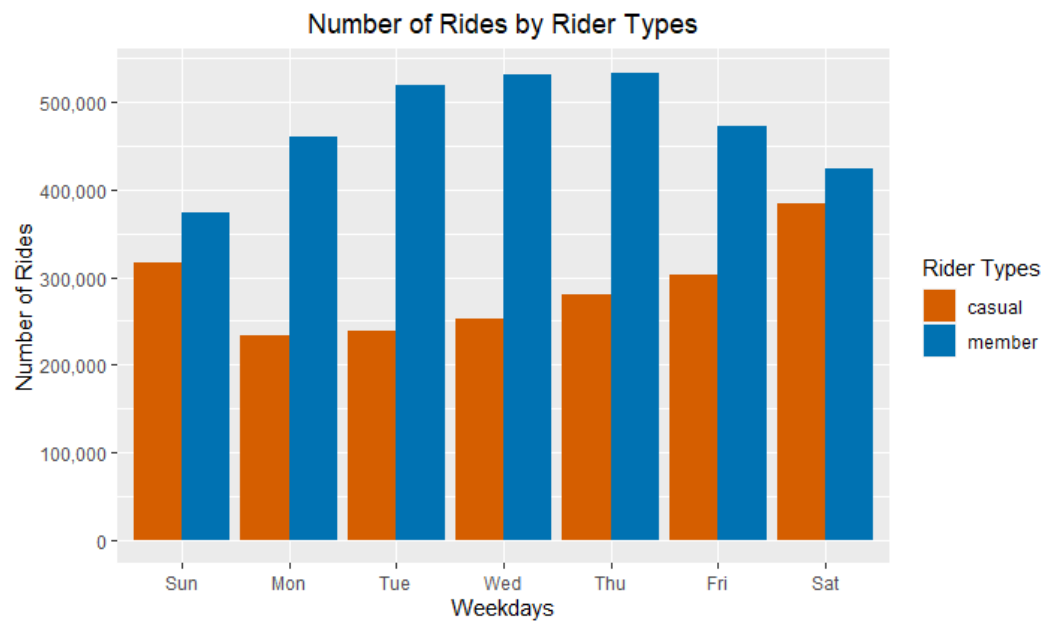


Figure 3. Number of rides by rider type and weekday

Average Ride Duration by Rider Types and Weekday

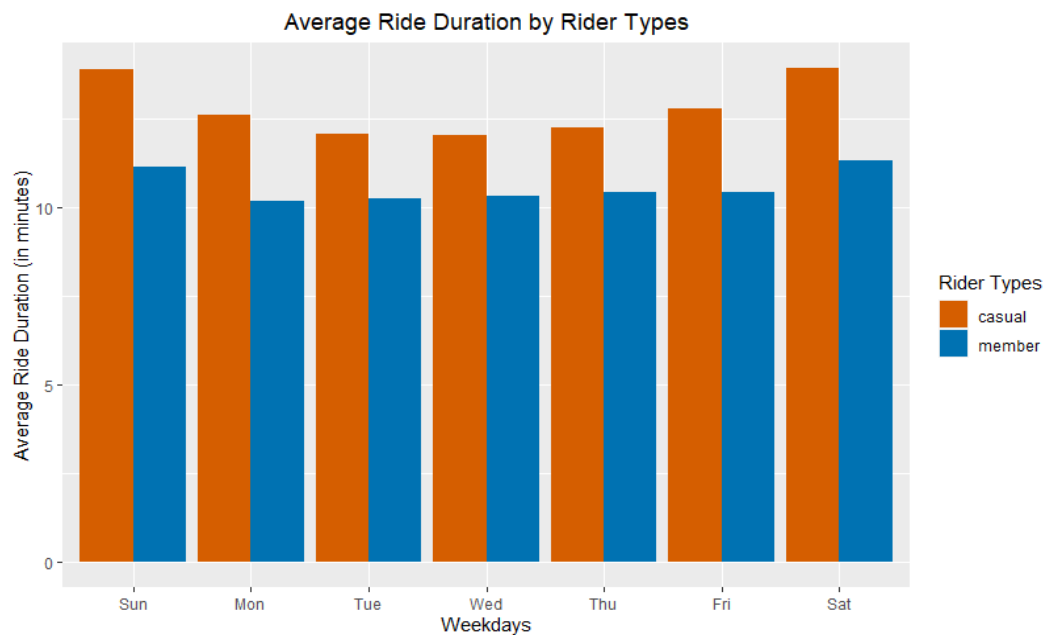


Figure 4. Average of ride duration by rider type and weekday

Summary Table of Day of Week Usage and Average

Rider Types	Day of Week	Number of Riders	Average Duration
Casual rider	Sunday	316151	13.86361
Casual rider	Monday	232515	12.59493
Casual rider	Tuesday	238470	12.04986
Casual rider	Wednesday	251981	12.03749
Casual rider	Thursday	279628	12.24219
Casual rider	Friday	301916	12.75971
Casual rider	Saturday	384040	13.92121
Annual member	Sunday	374652	11.12097
Annual member	Monday	460965	10.16491
Annual member	Tuesday	519641	10.24115
Annual member	Wednesday	531549	10.31205
Annual member	Thursday	534221	10.40817
Annual member	Friday	472210	10.43187
Annual member	Saturday	424141	11.30499

Act

Conclusion

My conclusion is that Cyclistic's customers used the bike-sharing services differently. Annual members used the services more frequently than the casual riders over Monday-Saturday and had a peak usage on Thursdays. Casual riders used the services more frequently on Saturday. Based on the average ride duration and consistency of user, we can conclude that annual member commute daily, whilst casual riders use the services for leisure.

Recommendations

Creating different memberships to attract casual riders based on their need, such as a weekend membership. Additionally, creating a loyalty program for annual members might help in increasing the numbers. Lastly, expanding the network to cover a bigger territory would help grow the services.