Google Data Analytics Capstone

Christopher Robles

2023-03-01

Cyclist Bike Share Case Study

Biography

Hello and thank you for taking a look at my project. Currently, I am a Business Analyst and since my time at University I have had a big interest in data. I enrolled in the Google Data Analytics Professional Certificate and as part of the course curriculum I have worked on a capstone project that encompasses skills learned throughout the course.

If you would like to contact me you can find me on:

- https://www.linkedin.com/in/christopher-robles-0512/

or if view any other projects I have worked on:

- https://christopherrobles.github.io/

Introduction

A bike sharing company named Cyclistic that has been operating since 2016 in Chicago. Since its inception, they have added 5,824 bicycles and 692 stations to its network.

Currently, Cyclistic's marketing strategy relies on building general awareness and appealing to broad consumer segments. The company has flexible pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members. However, the marketing team believes that maximizing the number of annual members will be key to future growth as annual members are much more profitable than casual riders. Rather than creating a marketing campaign that targets all-new customers, the Cyclistic team believes there is a very good chance to convert casual riders into members.

Problem

In this scenario, I have been tasked with analyzing Cyclistic's customer data to understand how annual members and casual riders use bikes differently and ultimately find the most effective marketing strategy of converting casual riders into annual members.

Process

In order to look at Cyclistic's data, I obtained a dataset from a 3rd party, which was made available by Motivate International Inc. With this public data, I was able to look at monthly data for 2022. I used R for data cleaning and analysis as well as Tableau for data visualization.

Load Libraries

The libraries that were used:

- tidyverse
- lubridate
- ggplot2

Data Cleaning

Before cleaning the data, it was important to get familiar with it. Summary and Column names were just a few examples I took to inspect the data to ensure the I knew what variables were being observed and if it made any sense. A few approaches I took was making sure the member_casual variable was properly recorded by verifying that it only had to options: member or casual. I took the same approach with rideable_type to make sure nothing unusual showed up in the data. In addition to this, any observation with missing data was omitted from the dataset along side data that did not make sense, for example having a negative time in a trip.

Creating New Columns

- Ride Length - which was obtained by subtracting the end time of a ride by its start time
- Day of Week
- Month
- Time - which looked start time of ride (converted in HH:MM:SS format)

Link to View Full Code

The full code can be found here: Github (Click here)

Findings

From the data it can be seen that members accounted for about 60% of total bike rides in the year 2022. If we dig a little deeper by looking at the frequency of rides by rider throughout the year, we can also see a bell-shaped like curve. The summer months reach their highest levels of users by both members and casual riders, the frequency of riders peaked in August for members and July for casual riders. It is also worth noting that members used Cyclistic bikes more frequently than casual riders in every month. When looking at the data by weekday, it can be seen that ridership is greater between Monday and Friday for members. This could signal that members may use the bikes to get to and from work. As for casual riders, ridership is highest on the weekends and particularly on Saturday. Lastly, viewing the data by the time showed a notably difference between members and casual riders. Member ridership peaked in the morning between 8 and 9 am and peaked again in the evening at 5 pm. Meanwhile, the number of casual riders increased gradually throughout the day and peaking around 5 pm.
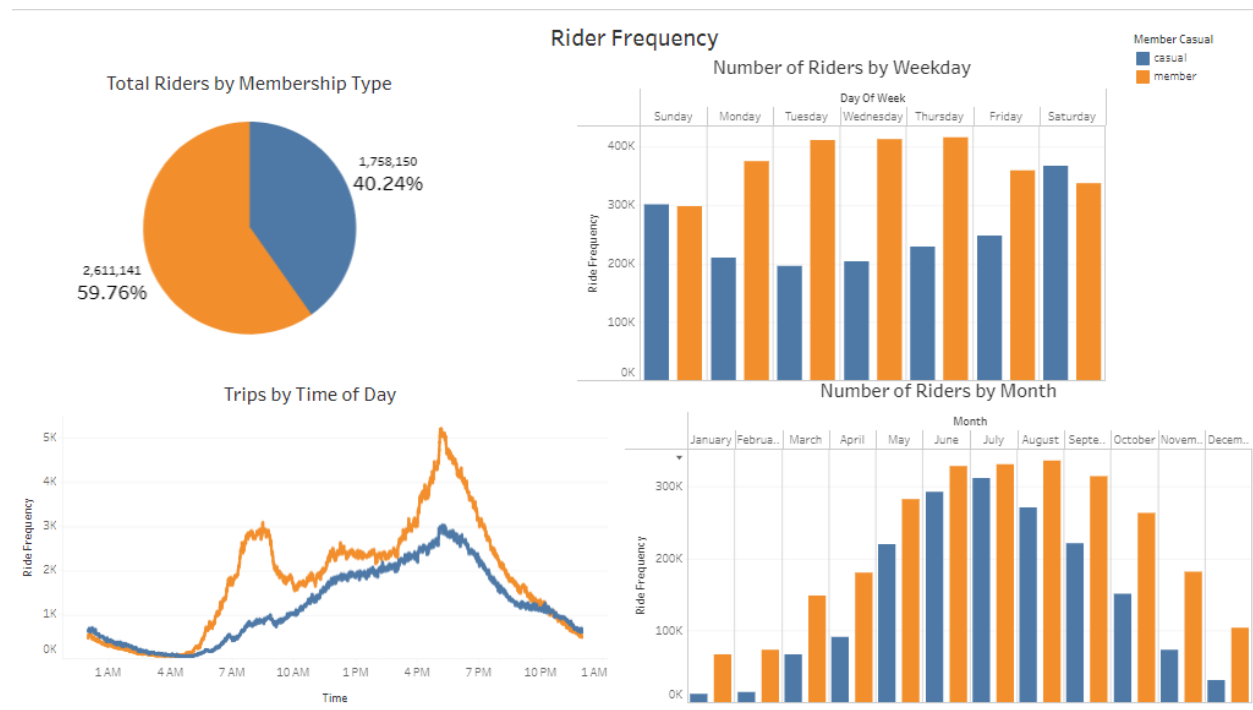

Figure 1: Rider Frequency

Another interesting point found was that casual riders on average rode for 23.99 minutes, while members rode for an average of 12.45 minutes. When viewing the data throughout the year, we can see that on average casual riders spent more time riding in every month than members. The bar chart show that the amount of time spent on the bikes dropped significantly in the second half of the year for casual riders, but it still remained higher than members. Meanwhile, the average bike ride for members remained relatively the same throughout the year.When looking by days of the week we can see that the average ride time for casual riders was higher every day and was at its highest on the weekends. Finally, when looking at the time of day members were between 10 and 13 minutes throughout the whole day. On the other hand, casual riders had a trip duration that would vary throughout the day.The average length of a trip was greatest between 10 am and 2 pm and the average length stayed above 20 minutes, except between 4 am and 8 am.
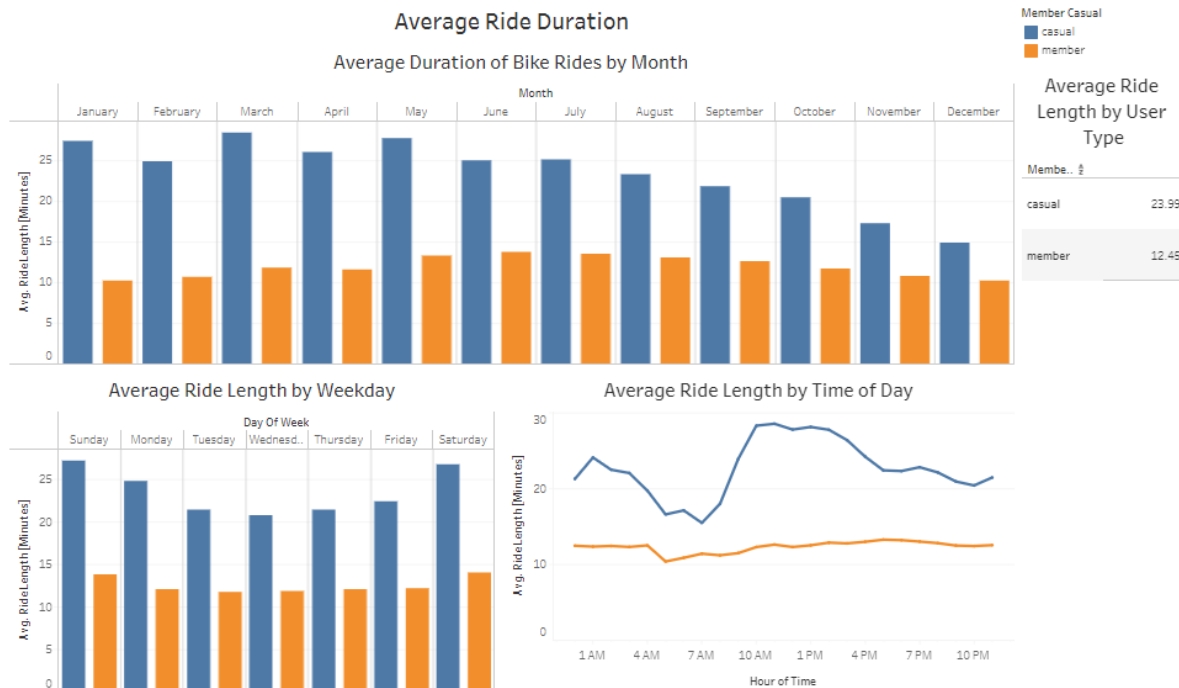

Figure 2: Average Duration of Rides

For interactive Tableau Dashboard: Tableau (Click here)

Summary of Data

Overall, it can be seen how differently members and casual riders behave. The data showed that members used Cyclistic bikes frequently compared to casual riders, but rider frequency was at its highest during the summer months for both groups. In addition, when we look further into the data we can see that members tend to use bikes more during the week, this could be due to members using bikes more often to get to and from work. This point is also supported by the fact that members are more likely to use bikes in the morning, 8 am, and evening, 5 pm, which is often synonymous with the times people go to work. However, it should be noted that further research should be conducted on this idea. As for trip duration, casual riders on average were more likely to ride longer than members as shown by the charts. In each chart members were between 10 and 13 minute for ride duration, while casual riders had a greater duration especially during the afternoon, weekends, and first half of the year.

Business Suggestions

My suggestions for the marketing team to convert casual riders to annual members:

1. Creating a Spring/Summer membership, as casual riders frequently used the bikes when approaching the summer months. The marketing for this should be done in the Winter and Spring months.

2. Introducing a weekend pass that would cater to casual riders that prefer to ride on the weekends, as casual riders were more frequent during the weekend.

3. Alter the existing payment structure to make it more appealing for casual riders to consider switching to a membership. This can be done by making single-use rides more costly or making the membership option more appealing for riders that have a trip duration greater than 20 minutes.

Sources and Links

Dataset:

- https://divvy-tripdata.s3.amazonaws.com/index.html

Tableau Dashboard:

- https://public.tableau.com/app/profile/christopher.robles/viz/GoogleCapstoneProject__16773758728230/ DataAnalyticsCapstone

R Code:

- https://github.com/ChristopherRobles/Google-Data-Analytics-Capstone-Project/blob/main/ Capstone__Code.txt