

Cyclist Market Analysis Report

Marcelo Duarte da Silva

04/08/2021

Introduction

This report contains the market analysis of the customers of **Cyclist** ride share company. This company is fictitious, and this report is part of the capstone project of the Google Data Analytics Professional Certificate.

Business Task

This report explores the differences in bike use between annual members and casual riders of the **Cyclist** bike sharing app.

Objectives

- Compare bike use behavior between annual members and casual riders, in order to identify trends.
- Elect and compare metrics, in order to better understand the differences in behavior of the two groups of customers, if they exist.

Preparing the data

The analysis used data from **Divvy Bikes** as proxy data. **Divvy Bikes** is a bike sharing company from Chicago and their data is licensed as public domain.

Datasets

The datasets were obtained from this download page, in 2021-08-4. The raw data is comprised of 33 datasets available for download as of the date mentioned. I restricted the analysis to the last 5 years, even though there are older datasets available. I made this decision to keep the analysis current and relevant.

Process

I used **R language** and **RStudio** to do the entire analysis of this project. **R** is considered to be a very flexible and capable instrument for data analysis, and contains all the tools needed to fulfill the listed objectives.

The first step in the data processing was to combine these datasets in a single dataset containing all the data. Then, we saved the data into a single dataset.

While combining the data, we had to rename and reformat values in the columns of the datasets from 2019 and older. These datasets have a different format compared to those from 2020 and earlier. We used the structure from the newer datasets as a baseline, because it contained some new information not present in the older ones. The most relevant addition is the geographical coordinates of the start and end of the ride.

Analysis

In order to start our analysis, I elected some metrics to compare between annual members and casual riders, based on the data available. The metrics are the following:

- Ride time;
- Ride frequency;
- Ride amount.

I chose to create an aggregated metric called *Ride Amount* calculated as the simple product of the average *ride time* and the daily *ride frequency*.

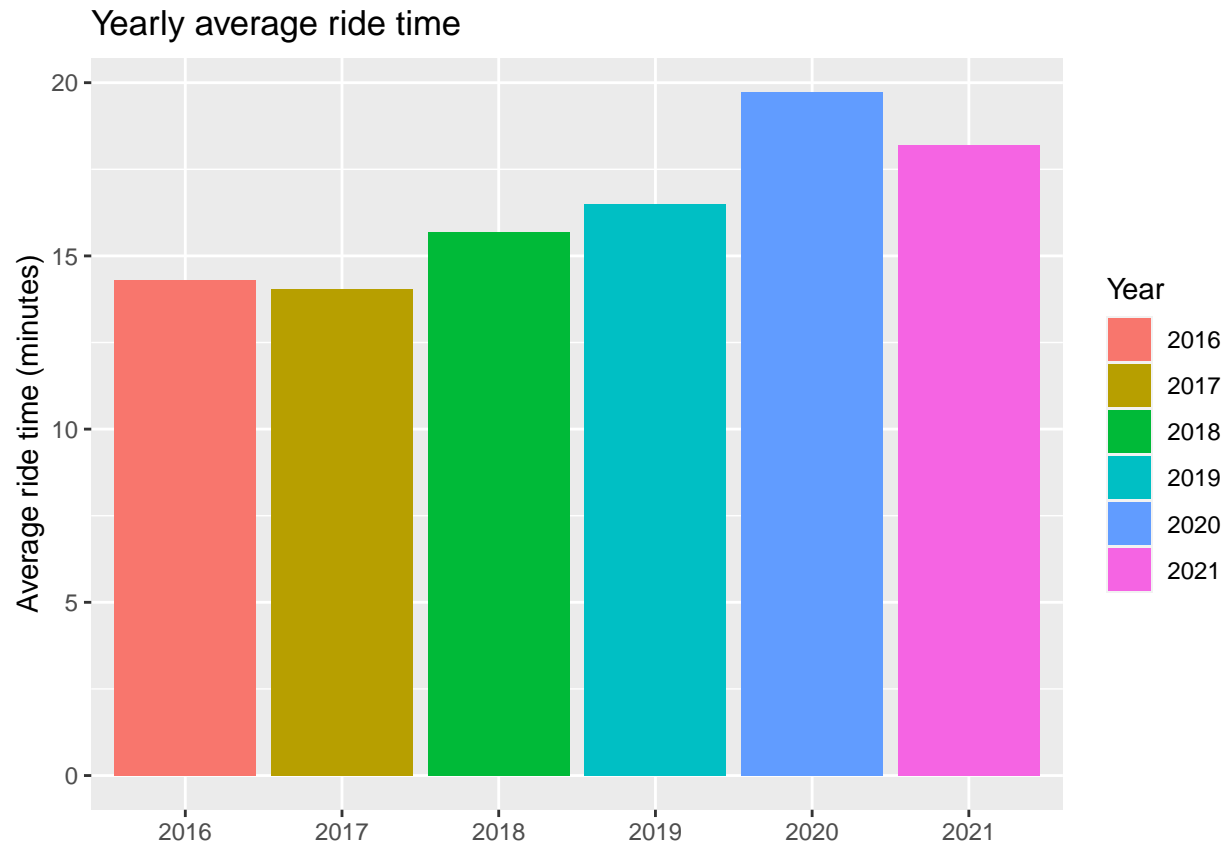
These metrics were analyzed and checked for seasonal trends in biker behavior, comparing differences based on period of the day (morning, afternoon and night), and differences during the year.

Initially I tried to calculate the ride length, by using the geographical coordinates. But, examining the data it turned out the these coordinates could not be used for this purpose. There was a very low correlation between ride time and ride length, with an $R = 0.23$. The probable explanation is that riders sometimes don't ride linearly between the starting point (or station) and the ending point. A rider, for instance, could start a trip, go somewhere and then return to end the trip at the same station. For this reason we will not consider the geographical coordinates.

We filtered out trips with missing start or end times. We also filtered out trips in which the start time was greater than the end time and trips with more than 12 hour duration, because they are not a good indicator of ride times since the bikes most likely were not being used by the customer for all that time.

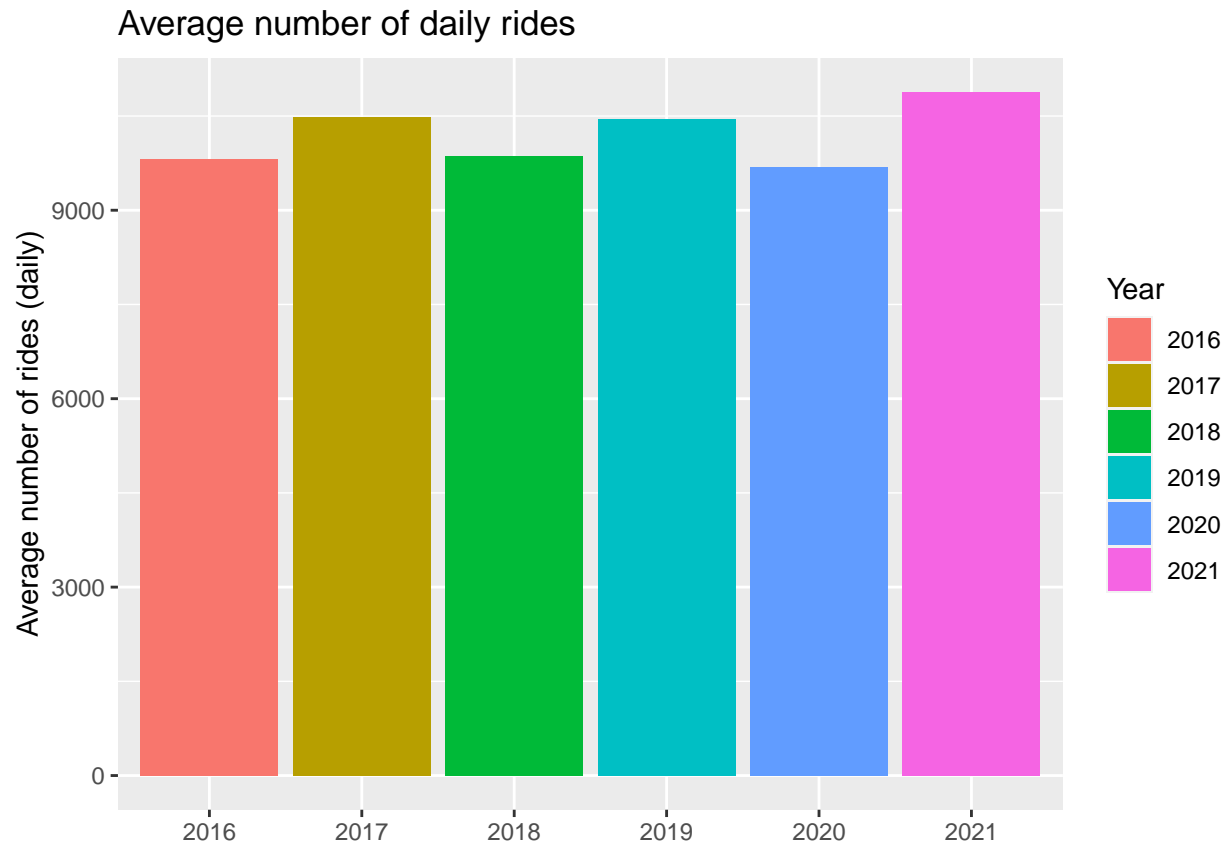
Results

First I established a baseline of values by finding the average ride time and the frequency of daily rides.



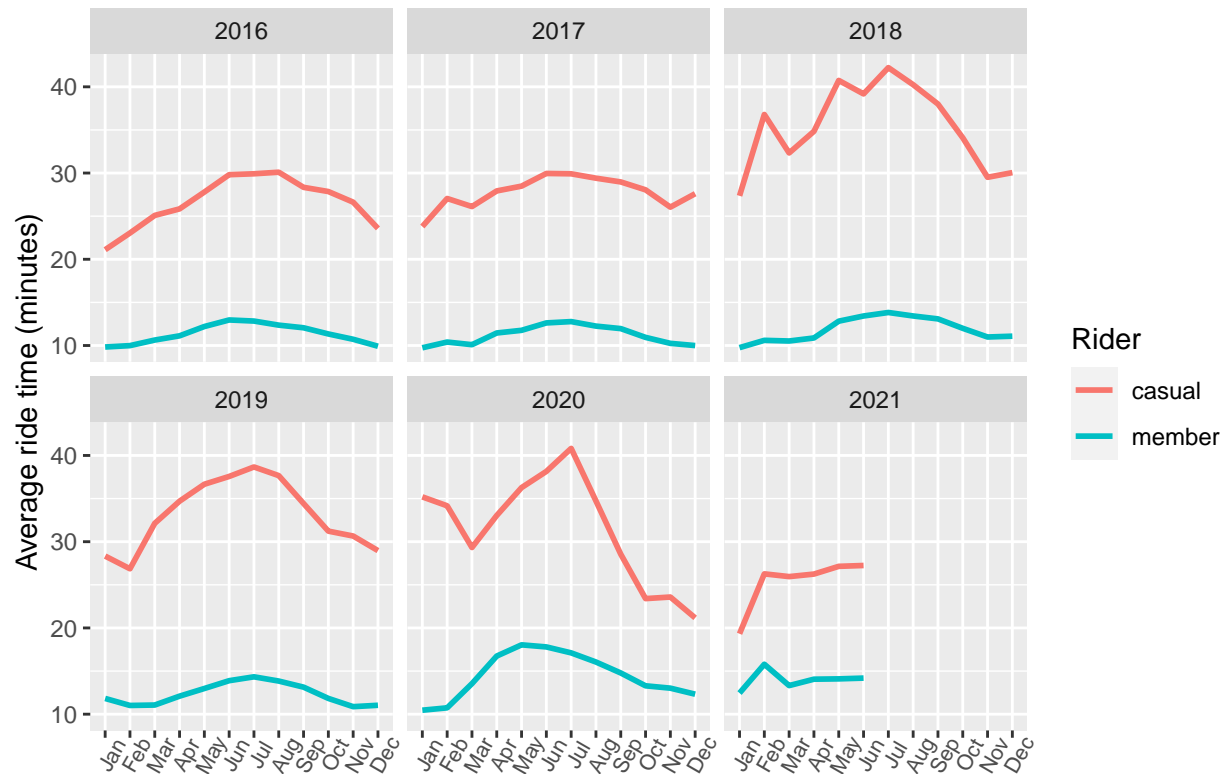
The average ride time across the five years of the analysis is 21.3 minutes. It varies from just under 15 minutes in 2016 to a little more than 23 minutes in 2021. There was a dip to 20 minutes in 2020, no doubt because of the pandemic, but the ascending trend is clear from 2018 onward.

There was a very sharp time increase in the average ride time in 2018, compared to 2017. The available data is not enough to provide a reason for this. It is necessary further exploration to better understand this effect.



The next step is start the exploration of the metrics in individual customer groups.

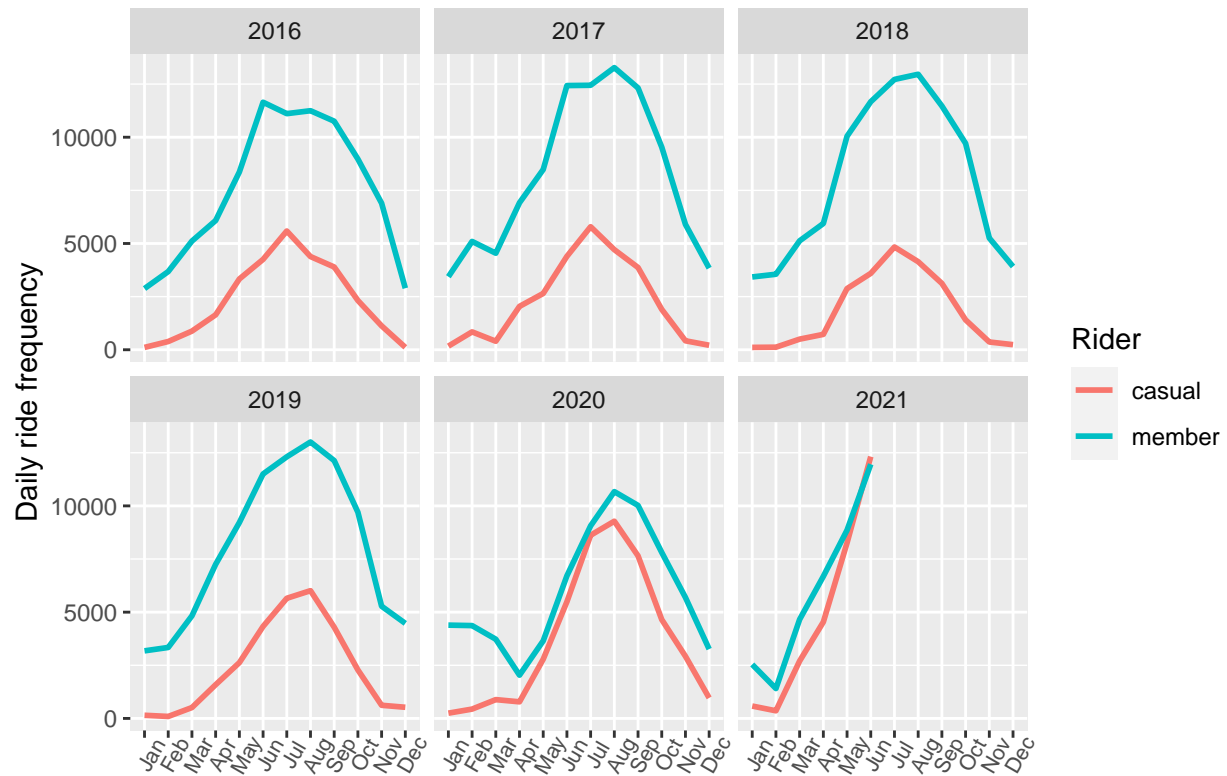
Average ride time by year



The data shows that casual riders have an average ride time greater than members, on all years of the analysis. The data shows a trend of increased ride time during the summer. This increase is more pronounced in casual riders, compared to members.

The year 2020 showed an increase in ride times from members, and there was a general increase in ride time from casual riders between 2018 and 2020. It is not possible to posit a reason for this behavior using only the available date.

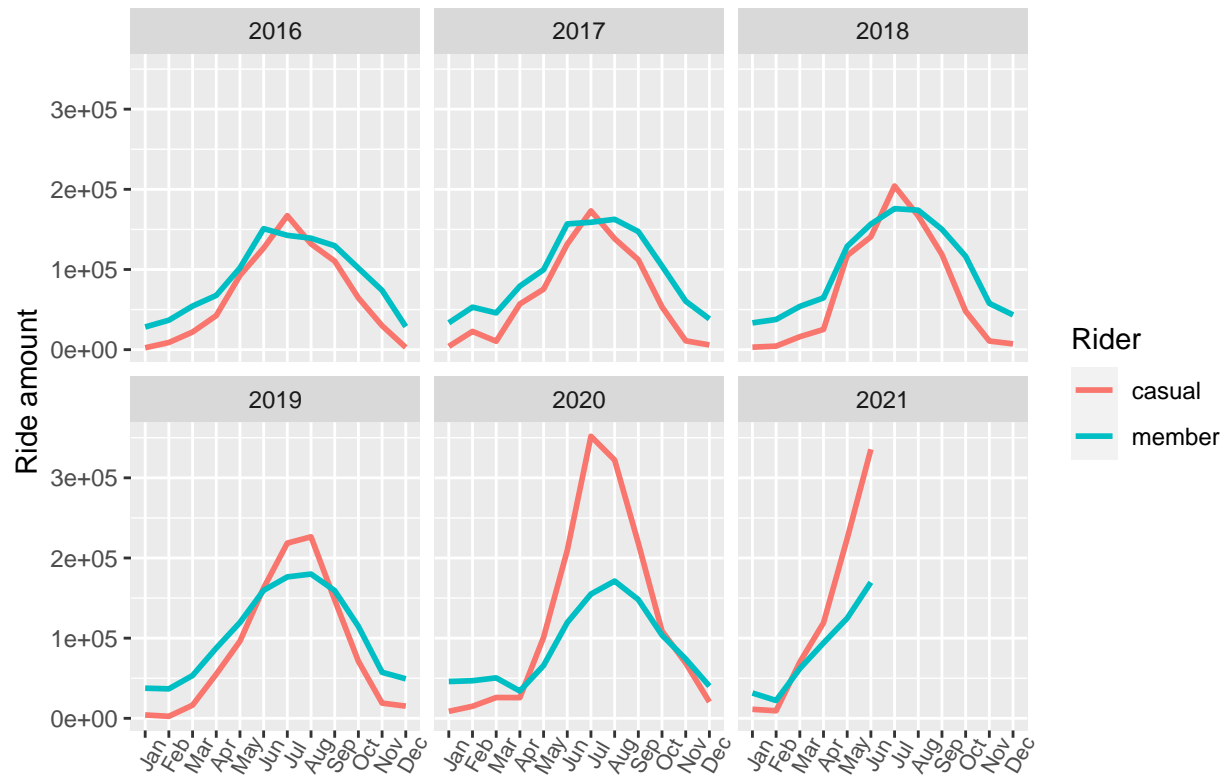
Average daily ride frequency by year



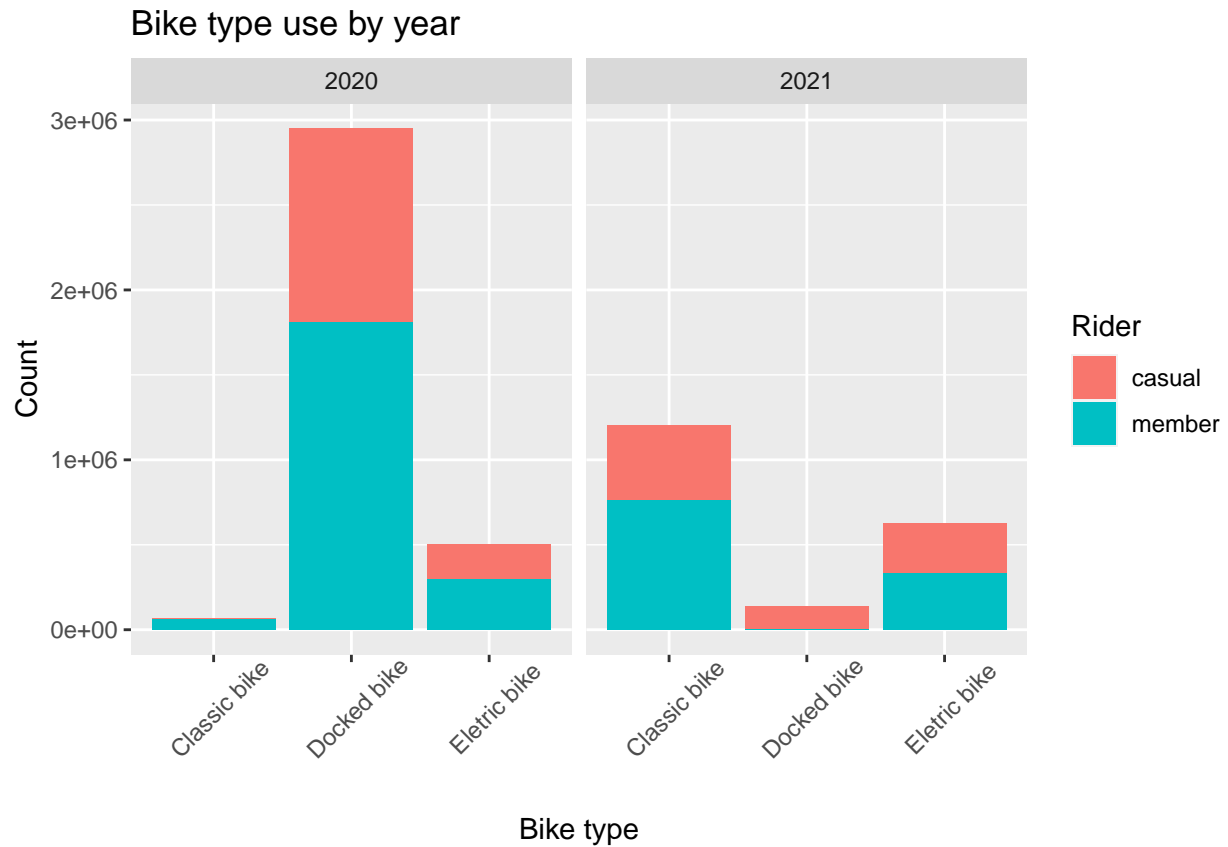
It is possible to see that the data shows a different trend in regards to average daily ride frequency. Members, on average, have a much higher average number of rides than non-members. It is also possible to verify that the number of rides increases during the summer.

In 2020, the trend changed a bit. The frequency of rides of casual customers increased and the number of members decreased, to the point of being very similar. This trend repeated on the first half of 2021.

Ride amount by year



When analyzing the *ride amount* aggregated metric, we can see that in most of the years of the analysis, casual riders and members showed similar ride amounts. With a slight increase from casual riders in 2019 during summer and early autumn. This difference became bigger in 2020 and, so far, 2021 is replicating the previous year tendency.



Bike type data is only available starting in 2020. By analyzing the available data regarding the types of bikes, It is not possible to discern any trend between casual riders and members behaviors. It is possible to see that classic bike use diminished, and docked bike use increased between 2020 and 2021.

Recomendations

1. Develop a service tier in order to monetize the increased ride time of casual users.
2. Further investigate the reason for ride time increase in 2018.