

Case Study: How Does a Bike-Share Navigate Speedy Success?

Alaa Ali

12/1/2021

Summary

This case study was completed by Alaa Ali in December 2021 as part of the Google Data Analytics Professional Certificate capstone. RStudio and Tableau will be used to complete this case study.

Scenario

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Moreno (The director of marketing and your manager) has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

Approach

In order to differentiate between the casual and the member riders we must study both of the riders thoroughly then we will be able to capture the patterns and trends of both riders.

Data cleaning

First we must inspect the data provided by the company and collect the most suitable data to analyze.

The link provided for the historical data is here

After viewing the data i decided in this project i will use the most current data for more accurate results (the 2021 data only) .

Now we have to open RStudio and import the data for analysis, in order to that we have to load some packages that will help us to import the data after downloading it then analyze it and here are the steps:

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

library(dplyr)
library(janitor)

##
## Attaching package: 'janitor'
## The following objects are masked from 'package:stats':
##
## chisq.test, fisher.test

library(here)

## here() starts at C:/Users/ME/Desktop/data courses/case study

library(skimr)
library(lubridate)

##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
## date, intersect, setdiff, union

library(ggplot2)
library(readr)
```

Now we import the selected data for analysis The data is stored monthly in separate spreadsheets so we import each dataset individually

```
X202101_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202101-divvy-tripdata.csv")
View(X202101_divvy_tripdata)

X202102_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202102-divvy-tripdata.csv")
View(X202102_divvy_tripdata)

X202103_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202103-divvy-tripdata.csv")
View(X202103_divvy_tripdata)

X202104_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202104-divvy-tripdata.csv")
View(X202104_divvy_tripdata)

X202105_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202105-divvy-tripdata.csv")
View(X202105_divvy_tripdata)

X202106_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202106-divvy-tripdata.csv")
View(X202106_divvy_tripdata)

X202107_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202107-divvy-tripdata.csv")
View(X202107_divvy_tripdata)

X202108_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202108-divvy-tripdata.csv")
View(X202108_divvy_tripdata)
```

```
X202109_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202109-divvy-tripdata.csv")
View(X202109_divvy_tripdata)
```

```
X202110_divvy_tripdata <- read_csv("C:/Users/ME/Desktop/data courses/2021 bike/202110-divvy-tripdata.csv")
View(X202110_divvy_tripdata)
```

Then we check the data details using `str()` and/or `col.names()` to make sure if the data is eligible for combination.

```
str(X202101_divvy_tripdata, X202102_divvy_tripdata, X202103_divvy_tripdata, X202104_divvy_tripdata, X202105_divvy_tripdata)
```

After confirming that the data is ready. Now we combine all the tables into one data set and we will name it “x2021” then open it to view it

```
X2021 <- bind_rows(X202101_divvy_tripdata, X202102_divvy_tripdata, X202103_divvy_tripdata, X202104_divvy_tripdata, X202105_divvy_tripdata)
View(X2021)
```

After scrolling through the data to understand it better and check what are the most important variables and eliminating the variables with numerous nulls and/or irrelevant to the task in hand.

I decided that the (start_lat, start_lng, end_lat, end_lng) columns are useless for my analysis and (start_station_name, start_station_id, end_station_name, end_station_id) columns had too much nulls and were irrelevant to the approach i am going with in my analysis

```
X2021 <- X2021 %>%
select(-c(start_lat, start_lng, end_lat, end_lng, start_station_name, start_station_id, end_station_name, end_station_id))
```

Now we will want to add column with the duration of each trip so it will help us greatly in our analysis.

to do that we will have to subtract the end time “ended_at” and the start time “started_at” for each trip to get the ride length and we will name the column “ride_len”

```
X2021 <- X2021 %>%
mutate(ride_len = ended_at - started_at)
```

Then we will need to change it to numeric value so we can analyze it

```
X2021$ride_len <- as.numeric(as.character(X2021$ride_len))
```

After scrolling through the new column we can see there are hundreds of columns with negative value so we remove them for accurate results.

```
X2021 <- X2021 %>%
filter(!(ride_len < 0))
```

For analysis purposes we will need columns with the exact date, time, day, week and month these trips were made so we can compare the rides more thoroughly

```
X2021$date <- as.Date(X2021$started_at)
X2021$year <- format(X2021$started_at, "%Y")
X2021$month <- format(X2021$started_at, "%m")
X2021$week <- format(X2021$started_at, "%w")
X2021$day <- format(X2021$started_at, "%d")
X2021$day_of_week <- format(X2021$started_at, "%A")
X2021$time <- format(X2021$started_at, "%H:%M:%S")
```

Now we conduct a descriptive analysis to our data so we can look for trends and relationships in the data to answer the business task in hand, we can get a summary of the data by getting the mean, median, max and min ride length of each rider type by aggregating the ‘ride_len’ & ‘member_casual’ columns

```

aggregate(X2021$ride_len ~ X2021$member_casual, FUN = mean)

aggregate(X2021$ride_len ~ X2021$member_casual, FUN = median)

aggregate(X2021$ride_len ~ X2021$member_casual, FUN = max)

aggregate(X2021$ride_len ~ X2021$member_casual, FUN = min)

```

we can dig even deeper for insights like to check the number of rides and the average ride length for each rider type through out the week

```

X2021 %>%
mutate(weekday = wday(started_at, label = TRUE)) %>%
group_by(member_casual, weekday) %>%
summarise(number_of_rides = n(), average_duration = mean(ride_len)) %>%
arrange(member_casual, weekday)

```

Now we have a clear view of how both type of riders use the company bicycles, now we will need to visualize such our findings for more clear presentation.

Data Analysis & Visualization

- Although RStudio offers wide variety of plots and graphs using the 'ggplot2' package but i decided to use Tableau because it offer even more options that Rstudio and it's both easier and faster to manipulate and work with.

The main question here was how do annual members and casual riders use Cyclistic bikes differently?

so we must have a clear view of how they acted during this year in the past 10 months (only data available this year)

Rider Type vs Rider Activities

First we want to have an overall view of both type of riders activities during this year in terms of both ride length per hour and number of rides

Members vs Casuals during 2021

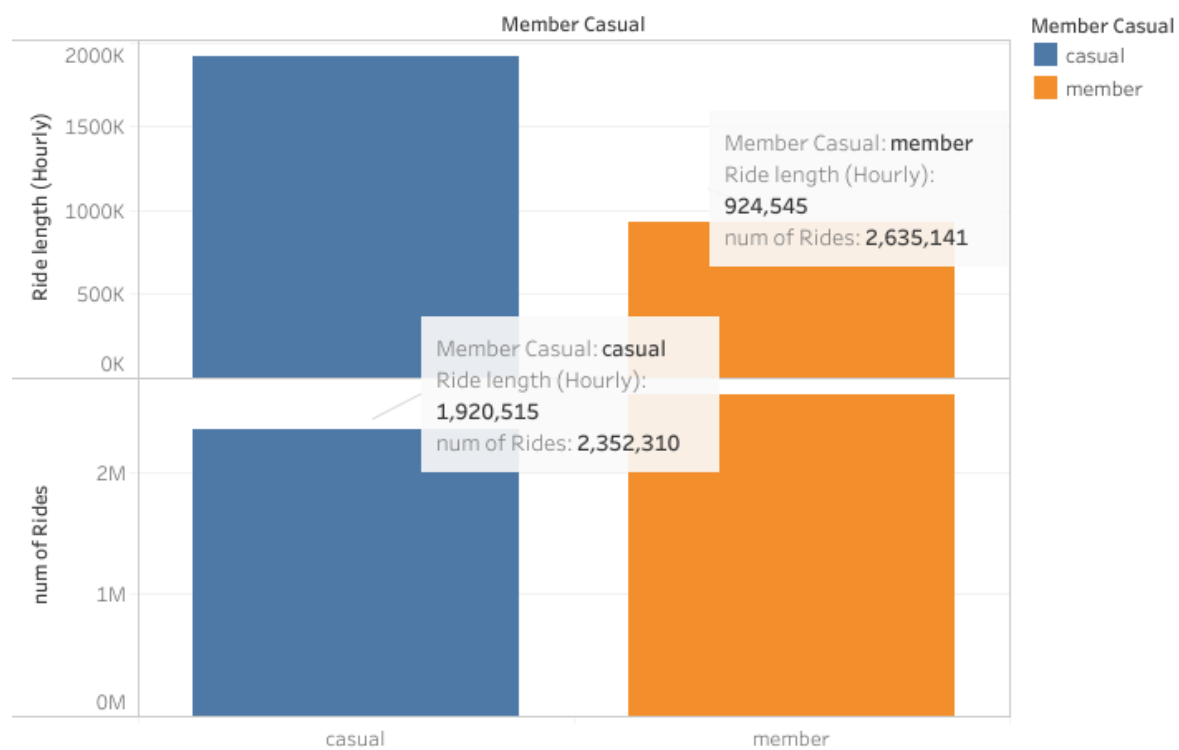
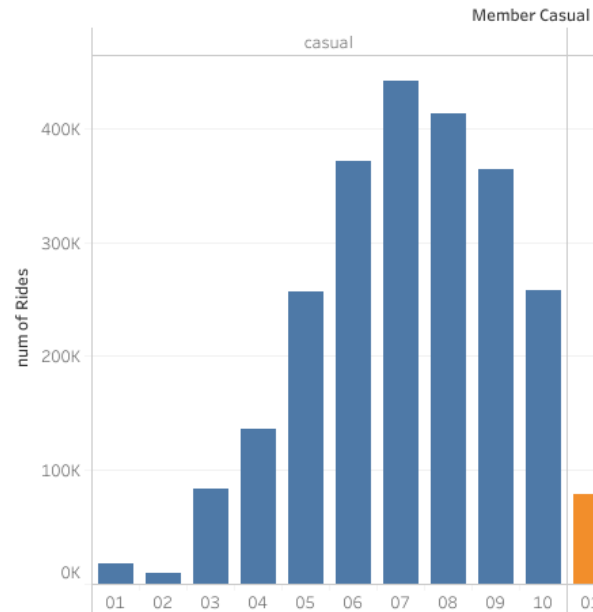


Figure 1: We can see that the casual rider had fewer number of rides but had more ride length than member rider

Rider type vs Number of Rides Monthly.

Members vs Casuals during 2021



In this graph we can see the number of rides per month for each rider type.

We can see that there is a huge spike in number of rides from June to September in both type of riders clearly it's because of the hot weather during the summer to insure that i checked Chicago temperature history during this year via weatherspark.com

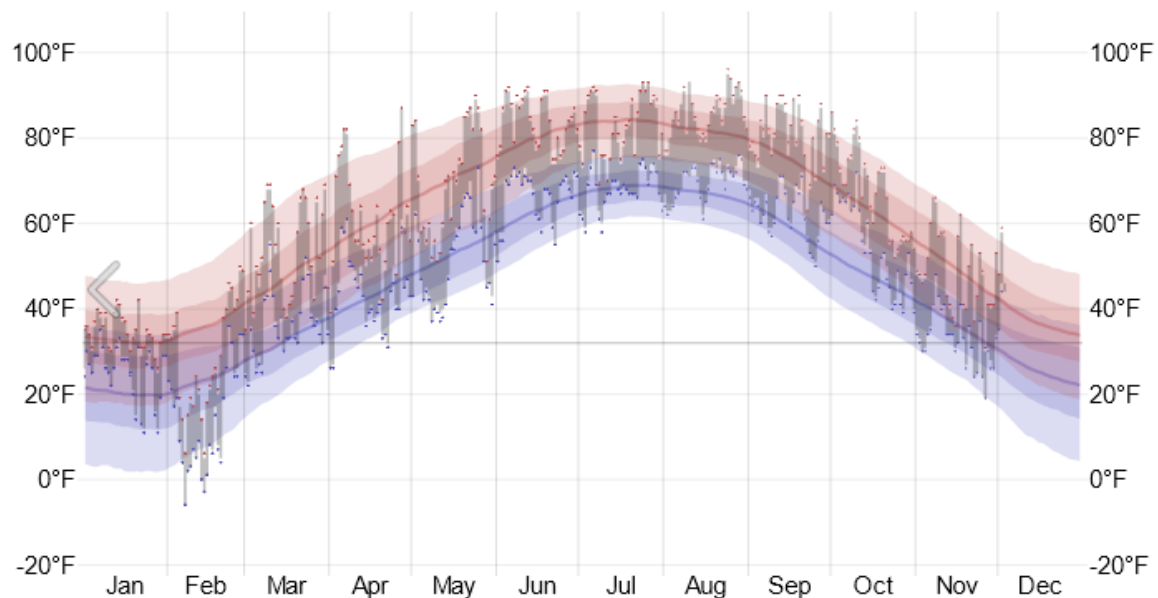


Figure 2: Temperature history in 2021 in Chicago

This graph validates the assumption.

Rider type vs Number of Rides through out the week.

Now we in this graph we will see each rider type activity during the weekdays

Members vs Casuals during 2021

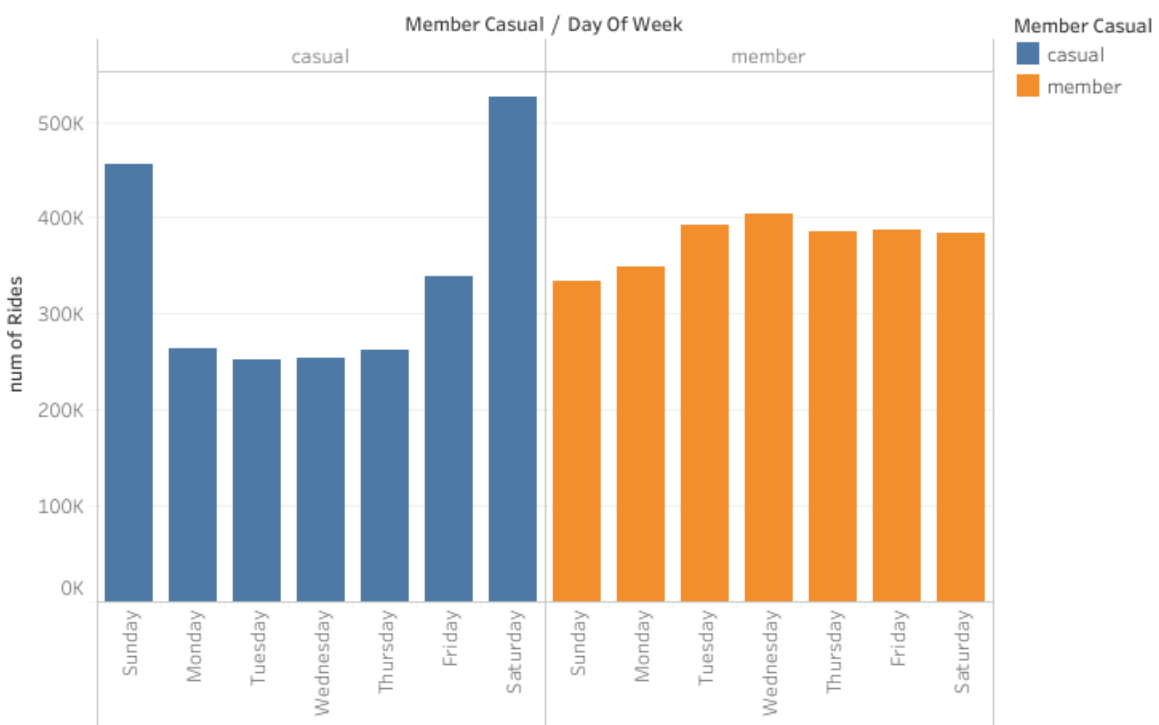


Figure 3: We can see that the member type riders activity is almost stable through out the week with a little spike in the middle of the week, in the other hand Casual riders activity during the weekend is much higher then their activity during the weekdays

Rider type vs Number of Rides through out the day.

Now we will see each rider type activity during the day

Members vs Casuals during 2021



Figure 4: We can see that between 00:00 and 05:00 am that both type of rider activity is low and also almost similar while between 05:00 am and 11:00 am their activity varies because we can see that the member riders activity gets clearly higher while the casual rider activity also get high but with a very small rate, between 11:00 am and 19:00 pm both activities get really higher especially between 17:00 pm and 19:00 pm (during and after sunset), then between 19:00pm and 00:00 it gets much lower for both riders

Rider type vs bike type

Cyclistic offers three (3) type of bikes for its customers now we will dig deeper to understand its customers preferences

in this graph we will compare the number of ride for each rider type compared to the bike type

Ride Length vs bike type

Now we will see how each type of riders use the 3 type of bikes differently

Conclusion

After analyzing this data set thoroughly there are a few takeaways that will help us strategize our market campaigns

Members vs Casuals during 2021

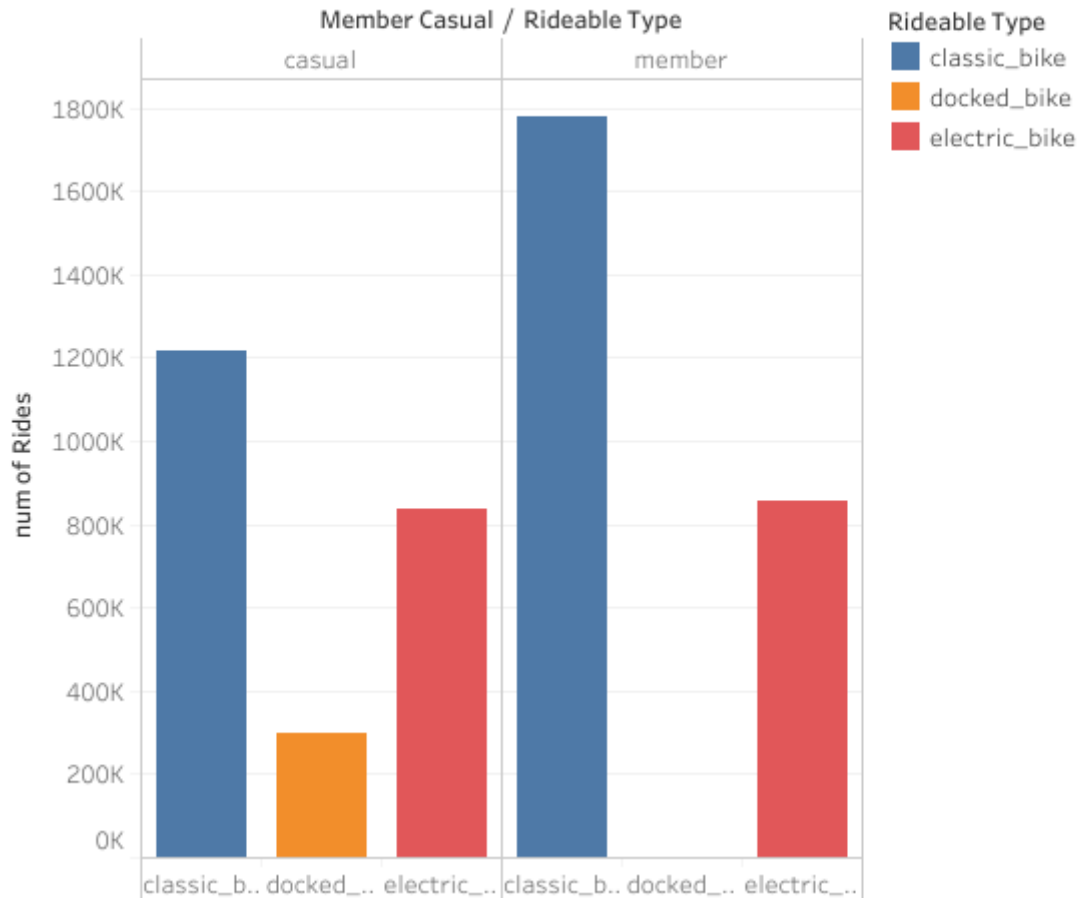


Figure 5: We can see in this graph that when it comes to the electric bikes both type of riders have almost the same number of rides, we can also see that the classic type bike is the most used for both type of riders with a clear edge for the member type riders with more over than 550,000 rides than the casual riders, lastly the docked bikes here we can see that has been only one ride made by a member rider using a docked bike but when it comes to casual riders they have used docked bikes 299.800 times in the past 10 months

Members vs Casuals during 2021

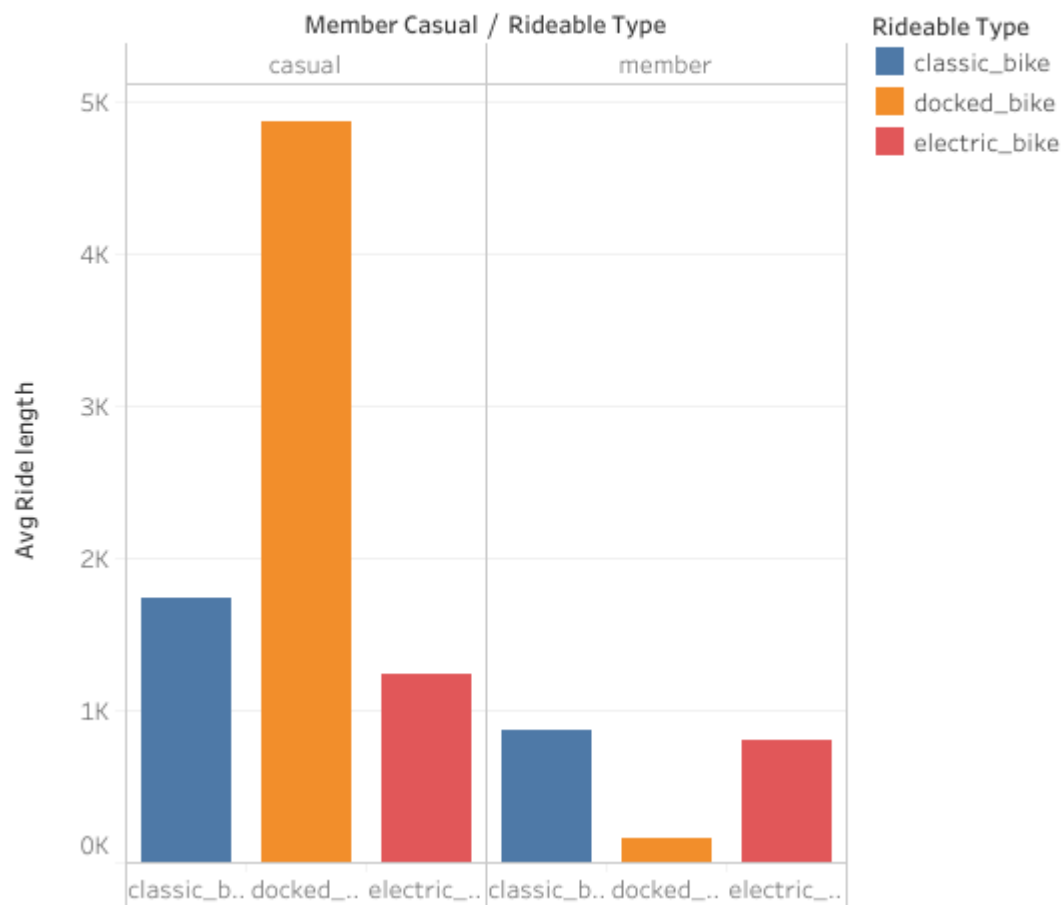


Figure 6: This graph shows that the average ride length (per seconds) for casual riders is clearly more than the member riders in every bike category, also shows how casual riders prefer docked bikes when it comes to long rides compared to the classic and electric type

1. The seasonal campaign

Since casual riders use our services more during the summer and the heat wave we should offer them an attractive package of offers during that time of the year that will lead to decrease the amount of money required to obtain an annual membership compared to the total fees required to buy a day pass/ single trip pass during the summer, we should show how buying an annual membership that comes with great offers during the summer is cheaper than buying rides passes daily during the summer

2. The weekly & daily trends

Since casual riders use Cyclistic bikes more often during the weekend we should increase the fees of the daily passes during the weekend in a way that makes the annual membership deal seems reasonable while also showing that the amount of money spend on total trips through out the month/year with the new prices will be much more compared to buying an annual memberships, we also should post more about how great it is to drive a bike during the weekend to view the sunset and after it with pictures of both views since this is the best time to ride a bike for casual riders according to our analysis

3. The bike type & The ride length

According to our analysis the average ride length of the casual rider is more than the average ride length of the member rider in all the bike classifications also that both riders use classic bikes over the electric bikes and the docked bikes also that the casual rider will prefer to ride a docked bike when it comes to long rides which makes me propose two ideas, firstly we change how we price the bikes to casual riders so we charge more for classic bikes while showing also that it will be cheaper if the casual riders bought the membership by showing them the difference of the two offers, also we could change the way we price docked bikes to an Uber-like pricing strategy that we charge more for the amount of time that the bike had been used while stating that won't happen if they purchase the annual membership while showing how good it will be to ride freely without worrying about the cost of the ride