

# Cyclistic Case Study | Final Report

Date: 07/14/2022

Report by James Nesbitt

## **About Cyclistic**

The program has over 5,800 bikes and 600 dock stations. The company's marketing director is planning to maximize the number of Cyclistic Members by creating marketing strategies to convert Casual riders to annual Member riders. Casual riders are customers that purchase single-ride or full-day passes. Cyclistic Members are customers that purchase annual Memberships. As a junior data analyst for Cyclistic, I was tasked to make data-driven recommendations for the marketing campaign.

## **Business Task**

Design marketing strategies aimed at converting Casual riders into annual Members.

## **Business Questions**

- How do annual Members and Casual riders use Cyclistic bikes differently?
- Why would Casual riders buy Cyclistic annual Memberships?
- How can Cyclistic use digital media to influence Casual riders to become Members?

## **Data Sources Used**

Data was collected by the customers activity on Cyclistic but customers were kept anonymous and were issued a unique ID.

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

Provided by Motivate Int. under [license](#).

The provided data files cover a one-year period from June 2021 to May 2022.

## Data Cleaning Process

I chose to use the R programming language for this project as my main tool. Here is what I did with the 12 downloaded CSV files, which I merged.

First, I loaded my libraries:

```
library(dplyr)
library(readr)
library(tidyverse) #helps wrangle data
library(lubridate) #helps wrangle date attributes
library(ggplot2) #helps visualize data
library(hms) # for time
library(data.table) # for exporting the data frame
```

I imported the raw data file (CSV) into R-Studio:

```
jun_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/jun_2021.csv')
jul_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/jul_2021.csv')
aug_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/aug_2021.csv')
sep_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/sep_2021.csv')
oct_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/oct_2021.csv')
nov_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/nov_2021.csv')
dec_2021 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/dec_2021.csv')
jan_2022 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/jan_2022.csv')
feb_2022 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/feb_2022.csv')
mar_2022 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/mar_2022.csv')
```

```
apr_2022 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/apr_2022.csv')
may_2022 = pd.read_csv('C:/Users/jimmy/Downloads/Google Certified
Data Analyst Course/Case Studies/Cyclistic_v2/Data/may_2022.csv')

cyclix_cleaned <- rbind (jun_2021, jul_2021, aug_2021, sep_2021,
oct_2021, nov_2021, dec_2021, jan_2022, feb_2022, mar_2022,
apr_2022, may_2022)
```

Deleted fragmented dataframes to free up memory:

```
remove(jun_2021, jul_2021, aug_2021, sep_2021, oct_2021,
nov_2021, dec_2021, jan_2022, feb_2022, mar_2022, apr_2022,
may_2022)
```

Then I created and formatted a day of week, month, day, year, time, hour column:

```
cyclix_cleaned$date <- as.Date(cyclix_cleaned$started_at) # is
yyyy-mm-dd
cyclix_cleaned$day_of_week <- wday(cyclix_cleaned$started_at) #day
of the week
cyclix_cleaned$day_of_week <- format(as.Date(cyclix_cleaned$date),
"%A")
cyclix_cleaned$month <- format(as.Date(cyclix_cleaned$date), "%m")
cyclix_cleaned$day <- format(as.Date(cyclix_cleaned$date), "%d")
cyclix_cleaned$year <- format(as.Date(cyclix_cleaned$date), "%Y")
cyclix_cleaned$time <- format(as.Date(cyclix_cleaned$date),
"%H:%M:%S")
cyclix_cleaned$time <- as_hms((cyclix_cleaned$started_at))
cyclix_cleaned$hour <- hour(cyclix_cleaned$time)
```

Then added a new ride\_length column:

```
cyclix_cleaned$ride_length <-
difftime(cyclix_cleaned$ended_at, cyclix_cleaned$started_at
units="mins")
```

I deleted all negative values from the ride\_length column:

```
cyclix_cleaned <- filter(cyclix_cleaned, ride_length > 0)
```

I removed blank rows and missing values, as well as checking for duplicates:

```
cyclix_cleaned <- na.omit(cyclix_cleaned)
cyclix_cleaned <- distinct(cyclix_cleaned)
```

Created a new column called ride\_length, which is essentially the difference

between started\_at and ended\_at columns.

```
cyclix_cleaned$started_at <-  
as.POSIXlt(cyclix_cleaned$started_at, format = "%m/%d/%Y %H:%M",  
tz="EST")  
cyclix_cleaned$ended_at <- as.POSIXlt(cyclix_cleaned$ended_at,  
format = "%m/%d/%Y %H:%M", tz="EST")  
cyclix_cleaned$ride_length <- difftime(cyclix_cleaned$ended_at,  
cyclix_cleaned$started_at)
```

Also, did some data cleaning for negative values again and other NAs.

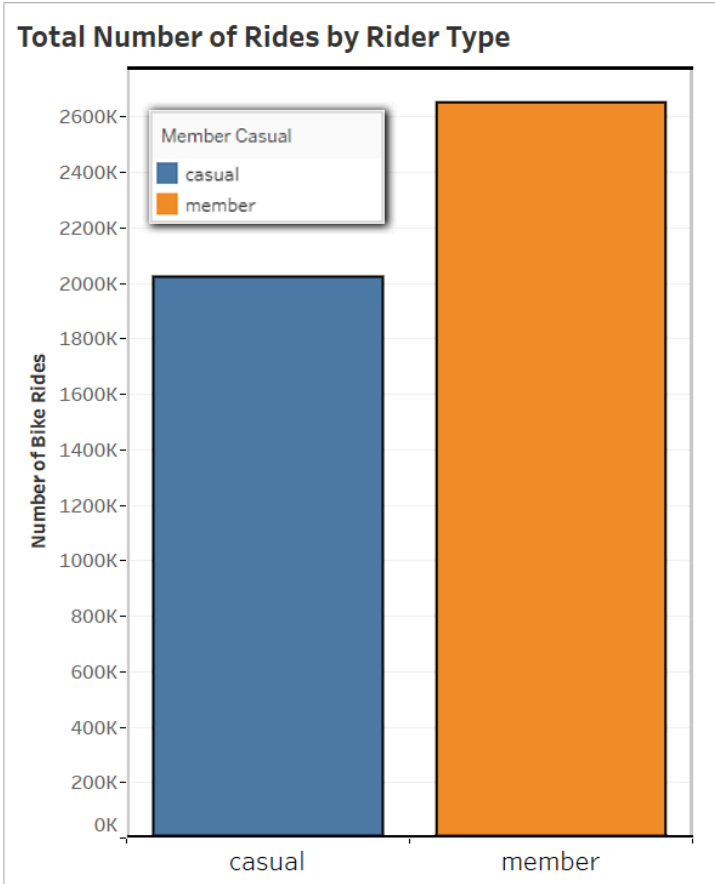
```
cyclix_cleaned<- na.omit(cyclix_cleaned) #remove rows with NA  
values  
cyclix_cleaned<- distinct(cyclix_cleaned) #remove duplicate  
rows  
cyclix_cleaned<- cyclix_cleaned[!(cyclix_cleaned$ride_length  
<=0),]
```

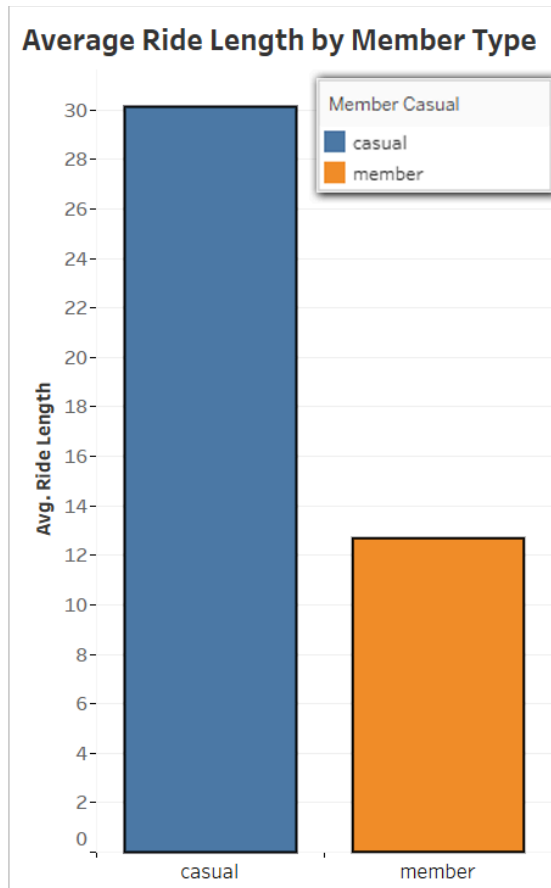
Exported CSV to Tableau for Analysis.

Final Record Count After Cleaning: **4,667,073** records

## Analysis

<u>Total Rides &amp; Avg Ride Length by Rider Type</u>		
<u>Rider Type</u>	<u>Ride Count</u>	<u>Avg Ride Length</u>
Casual	2019136	30.10 mins
Member	2647937	12.67 mins



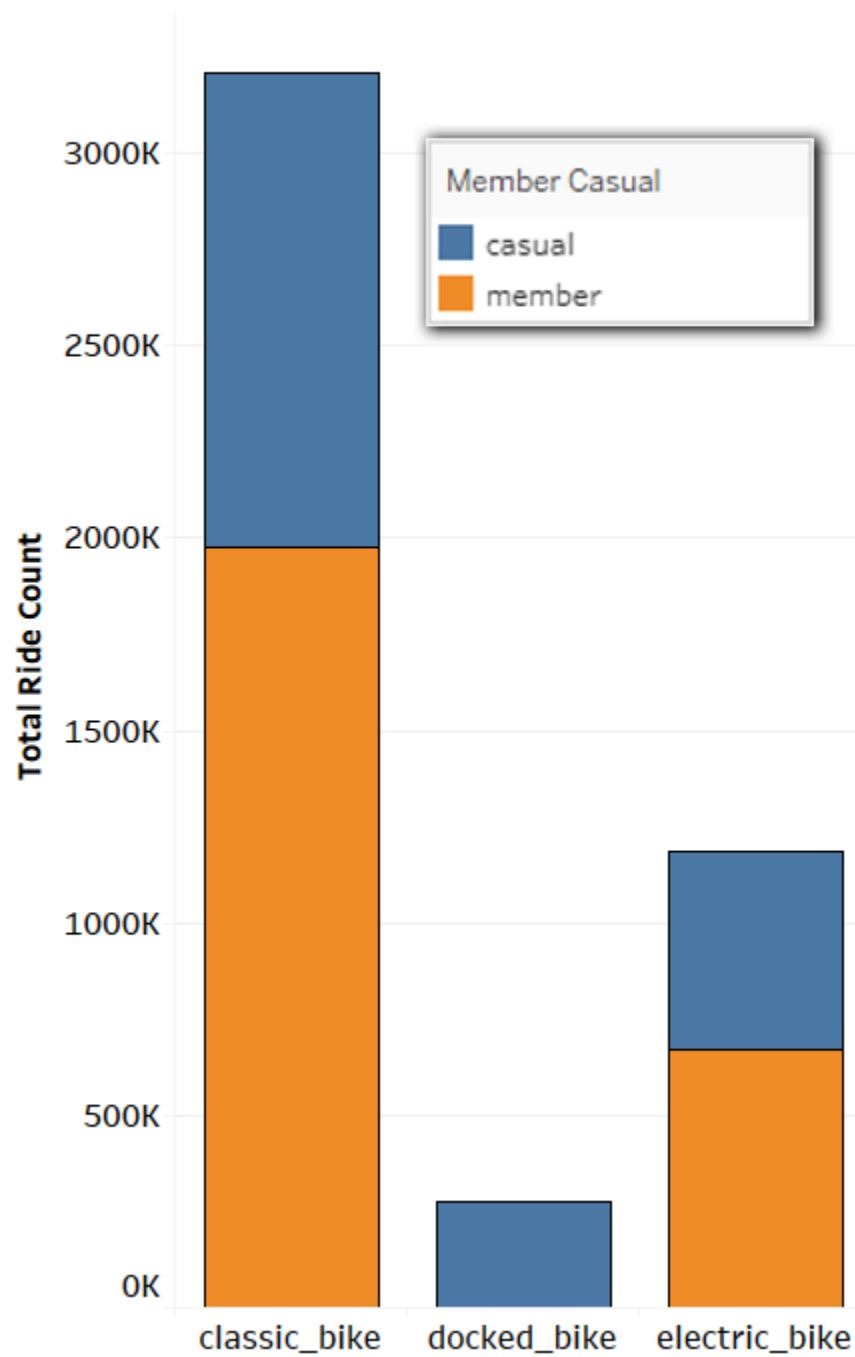


- Members had more bike rides overall (**2647937**), but Casual riders average ride length was almost three times longer (**30.10 mins**) than the Member's average ride length.

<u>Total Rides by Bike &amp; Rider Type</u>		
<u>Rider Type</u>	<u>Bike Type</u>	<u>Ride Count</u>
Casual	classic_bike	1231669
Member	classic_bike	1977831
Casual	docked_bike	273550
Member	electric_bike	670106

Casual	electric_bike	513917
--------	---------------	--------

Total Rides by Bike Type & Rider Type



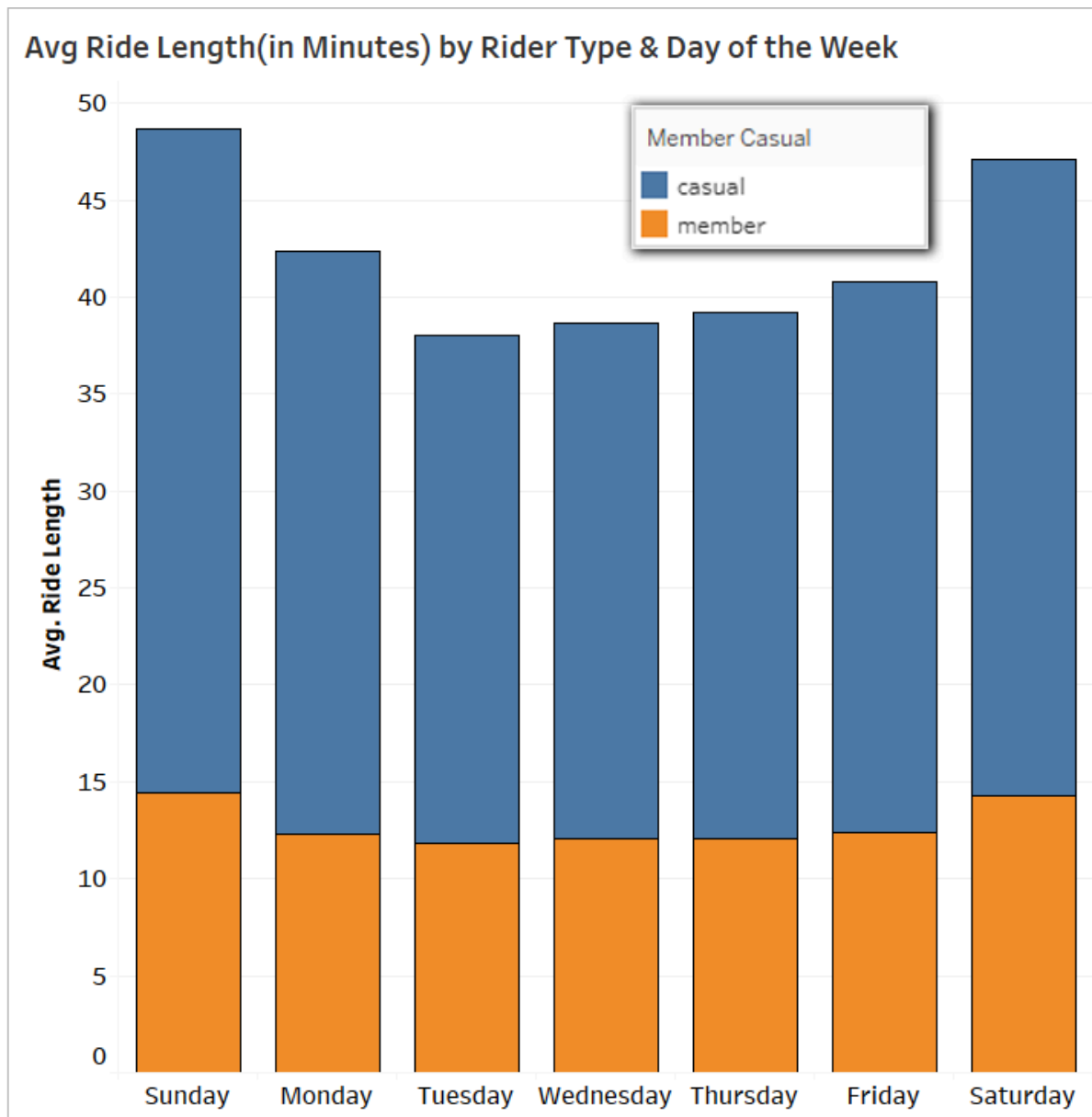
- Members and Casual riders both used the Classic bike the most (**1977831** and **1231669**) followed by the Electric Bike (**670106** and **513917**) and Docked bike (**273550**).

<b>Total Rides And Avg Ride Length by Hour of the Day &amp; Rider Type</b>							
<u>Rider Type</u>	<u>Hour</u>	<u>Ride Count</u>	<u>Avg Ride Length</u> <u>Casual Riders</u>	<u>Avg Ride Length</u> <u>Members</u>	<u>Ride Count</u>	<u>Hour</u>	<u>Rider Type</u>
Casual	12:00:00AM	41045	32.75 mins	12.61 mins	26162	12:00:00AM	Member
Casual	1:00:00 AM	29401	37.23 mins	13.41 mins	16977	1:00:00 AM	Member
Casual	2:00:00 AM	19606	38.73 mins	13.45 mins	9554	2:00:00 AM	Member
Casual	3:00:00 AM	10311	39.90 mins	13.26 mins	5555	3:00:00 AM	Member
Casual	4:00:00 AM	6673	47.75 mins	11.98 mins	6464	4:00:00 AM	Member
Casual	5:00:00 AM	9185	21.44 mins	10.37 mins	27669	5:00:00 AM	Member
Casual	6:00:00 AM	20386	20.10 mins	11.11 mins	76330	6:00:00 AM	Member
Casual	7:00:00 AM	38351	19.61 mins	11.46 mins	142852	7:00:00 AM	Member
Casual	8:00:00 AM	52335	22.73 mins	11.34 mins	167157	8:00:00 AM	Member
Casual	9:00:00 AM	61658	28.06 mins	11.76 mins	114685	9:00:00 AM	Member
Casual	10:00:00AM	84693	32.66 mins	12.72 mins	108068	10:00:00AM	Member
Casual	11:00:00AM	111316	32.88 mins	12.95 mins	129720	11:00:00AM	Member
Casual	12:00:00 PM	129762	32.37 mins	12.63 mins	148377	12:00:00 PM	Member
Casual	1:00:00 PM	137460	32.93 mins	12.83 mins	144731	1:00:00 PM	Member
Casual	2:00:00 PM	140881	33.18 mins	13.12 mins	142249	2:00:00 PM	Member



Casual	3:00:00 PM	149492	31.97 mins	13.05 mins	168252	3:00:00 PM	Member
Casual	4:00:00 PM	164471	28.95 mins	13.11 mins	227668	4:00:00 PM	Member
Casual	5:00:00 PM	192057	27.27 mins	13.29 mins	286455	5:00:00 PM	Member
Casual	6:00:00 PM	173390	27.56 mins	13.23 mins	234070	6:00:00 PM	Member
Casual	7:00:00 PM	133544	29.01 mins	13.08 mins	165641	7:00:00 PM	Member
Casual	8:00:00 PM	96612	28.84 mins	12.93 mins	110879	8:00:00 PM	Member
Casual	9:00:00 PM	82271	29.53 mins	12.57 mins	83453	9:00:00 PM	Member
Casual	10:00:00 PM	76242	30.15 mins	12.52 mins	62945	10:00:00 PM	Member
Casual	11:00:00 PM	57994	31.04 mins	12.51 mins	42024	11:00:00 PM	Member

- For Casual riders, 4AM was the quietest hour with the lowest amount of rides happening during that hour (**6673**). For Members, it was 3AM (**5555**).
- The busiest hour for both Casual (**192057**) and Member (**286455**) riders was 5PM
- In terms of Avg Ride Length, Casual Riders rode the longest during the hour of 4AM with an Average Ride Length of **47.75** mins. Member riders longest ride took place during the hour of 2AM with an Average Ride Length of **13.45** mins.
- The shortest Average Ride Length for Casual Riders was at 7AM with **19.61** mins, and for Member Riders it was 5AM with an Average Ride Length of **10.37** mins.



Avg Ride Length by Rider Type & Day of Week		
Rider Type	Day of Week	Avg Ride Length
Casual	Monday	30.11 mins
Member	Monday	12.28 mins

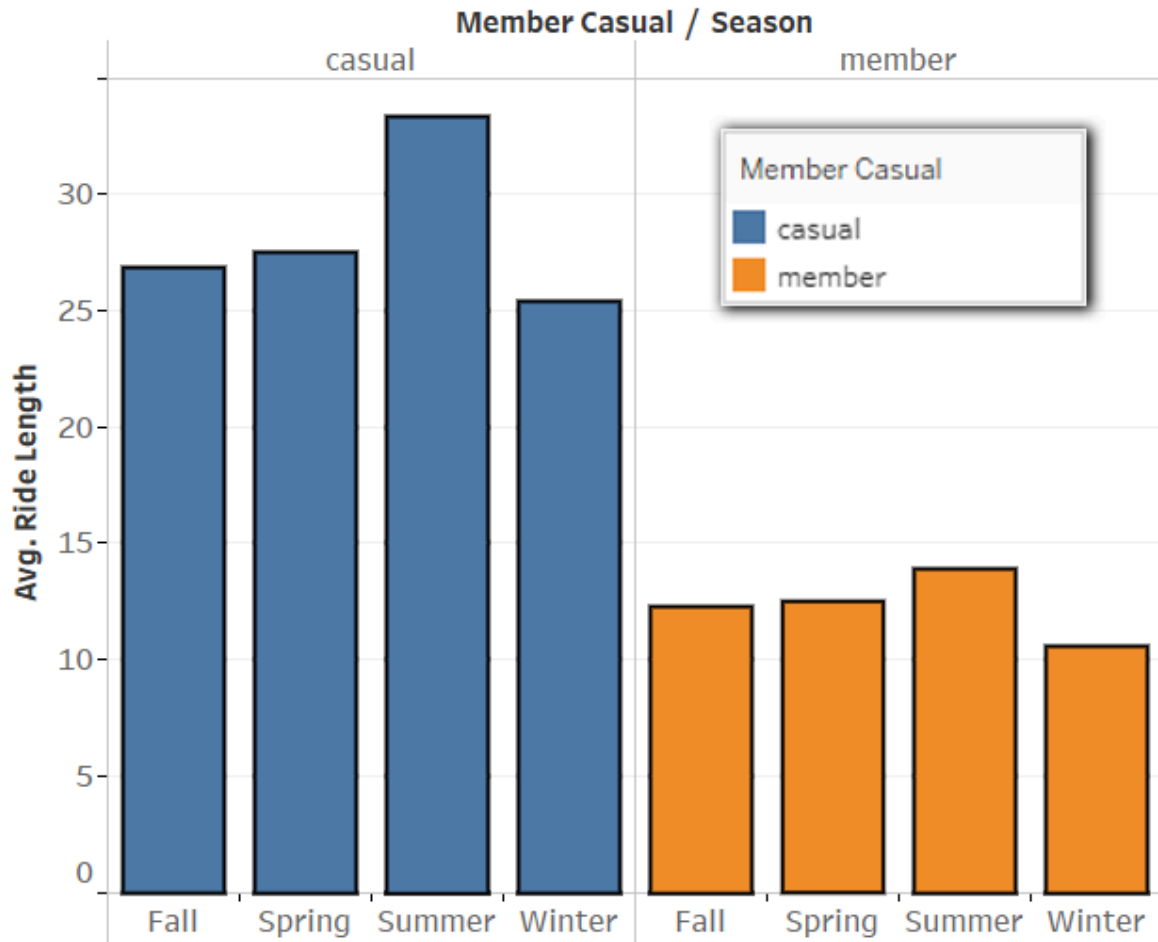
Casual	Tuesday	26.17 mins
Member	Tuesday	11.86 mins
Casual	Wednesday	26.62 mins
Member	Wednesday	12.03 mins
Casual	Thursday	27.14 mins
Member	Thursday	12.07 mins
Casual	Friday	28.38 mins
Member	Friday	12.37 mins
Casual	Saturday	32.80 mins
Member	Saturday	14.28 mins
Casual	Sunday	34.23 mins
Member	Sunday	14.45 mins

- For the average ride length per weekday both casual riders and members had an increase in the average ride length on the weekends.

<b><u>Total Rides by Rider Type &amp; Day of Week</u></b>		
<b><u>Rider Type</u></b>	<b><u>Day of Week</u></b>	<b><u>Ride Count</u></b>
Casual	Saturday	448040
Member	Saturday	350989
Casual	Sunday	384610
Member	Sunday	314098
Casual	Monday	234497

<b>Member</b>	<b>Monday</b>	<b>375438</b>
<b>Casual</b>	<b>Tuesday</b>	<b>218148</b>
<b>Member</b>	<b>Tuesday</b>	<b>424906</b>
<b>Casual</b>	<b>Wednesday</b>	<b>217431</b>
<b>Member</b>	<b>Wednesday</b>	<b>415253</b>
<b>Casual</b>	<b>Thursday</b>	<b>236560</b>
<b>Member</b>	<b>Thursday</b>	<b>403059</b>
<b>Casual</b>	<b>Friday</b>	<b>279850</b>
<b>Member</b>	<b>Friday</b>	<b>364194</b>

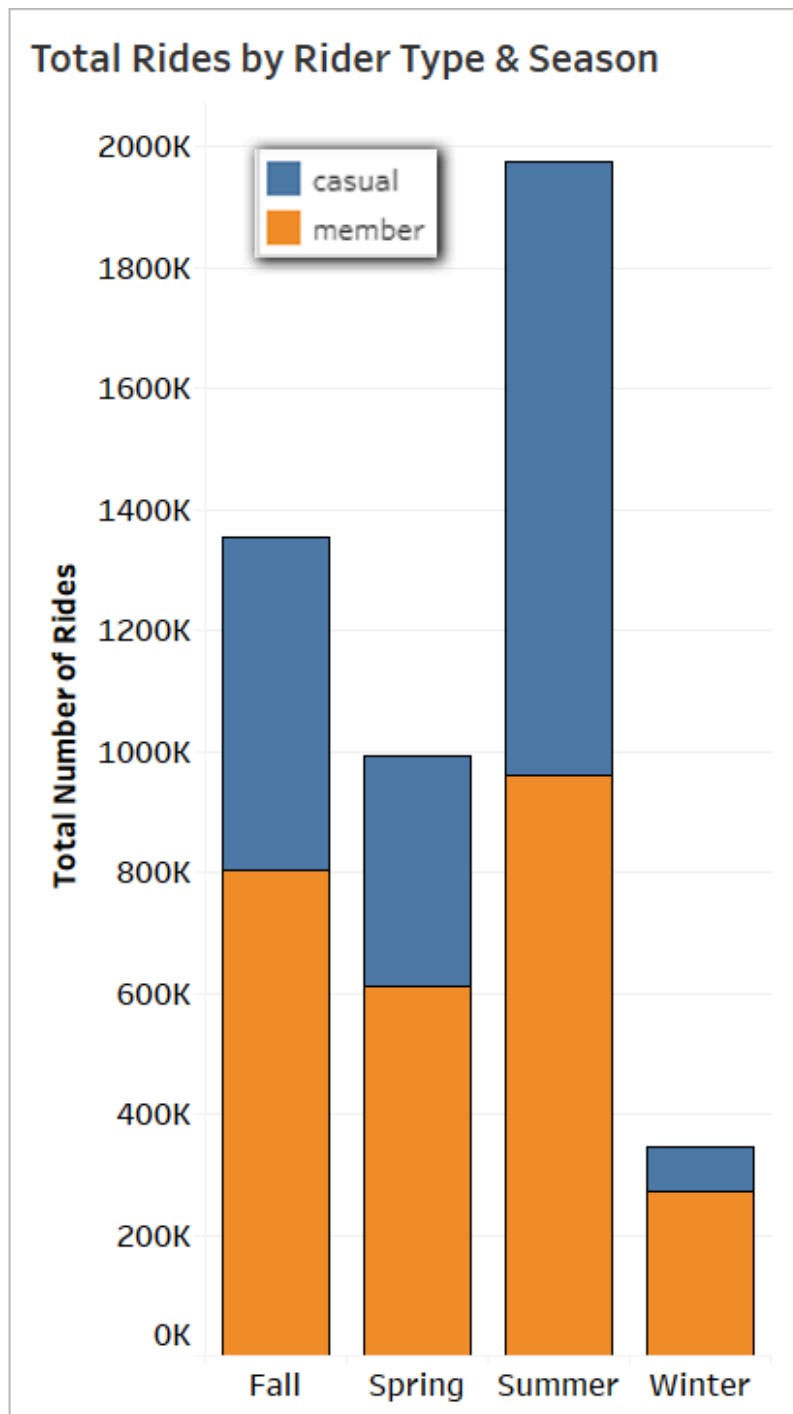
- For the total rides per weekday Tues, Wed, Thur were the top three days for members. For casual users, the weekends were busier.



<b><u>Avg Ride Length by Rider Type &amp; Season</u></b>		
<b><u>Rider Type</u></b>	<b><u>Season</u></b>	<b><u>Avg Ride Lgth</u></b>
Casual	Fall	26.78 mins
Member	Fall	12.22 mins
Casual	Spring	27.41 mins
Member	Spring	12.44 mins
Casual	Summer	33.27 mins

<b>Member</b>	<b>Summer</b>	<b>13.82 mins</b>
<b>Casual</b>	<b>Winter</b>	<b>25.27 mins</b>
<b>Member</b>	<b>Winter</b>	<b>10.52 mins</b>

- Both Rider groups show an obvious spike in the summertime. Ride length in Fall and Spring is pretty consistent across both Members and Casual riders.



**Total Rides by Rider Type & Season**

<u>Rider Type</u>	<u>Season</u>	<u>Ride Count</u>
Casual	Fall	551993
Member	Fall	802943
Casual	Spring	379271
Member	Spring	611762
Casual	Summer	1015048
Member	Summer	960385
Casual	Winter	72824
Member	Winter	272847

## Final Key Findings

- The busiest season is by far the Summer.
- Peak time for both riders would be 5:00pm.
- The busiest weekday is Saturday, and Member riders are the top consumer.
- Total amount of rides for Casual riders was **2019136 with an Average Ride Length of 30.10 mins.**
- Total amount of rides for Member riders was **2647937 with an Average Ride Length of 12.67 mins.**
- **Members use their bikes to go back and forth routinely (probably to work) during the week. This is proven by the data.**

## Recommendations

1. Start a digital media campaign targeted at Casual Riders that is strictly focused on people who are out of shape. It is a Challenge to see who can "Ride Your Bike To Work for 90-Days" IronMan Challenge. Any permutation



of this will suffice. Doing this during the peak times and summer season would be best. Contestants have to upload a testimonial of themselves recording their commitment and video of their ride every week, etc. This can be spun a million different ways.

2. Start a Member Rewards program with some real tangible benefits that people talk about online on popular venues such as Twitter, YouTube, etc.
3. Offer discounts for Membership on the weekends to Casual riders. This is kind of a no-brainer, but it should work especially when combined with other recommendations.