Bicycle Report

Wyatt Morrison

2023-01-09

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

Introduction

Welcome to the Cyclistic bike-share analysis case study! In this case study, you will perform many real-world tasks of a junior data analyst. You will work for a fictional company, Cyclistic, and meet different characters and team members. In order to answer the key business questions, you will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path. By the end of this lesson, you will have a portfolio-ready case study. Download the packet and reference the details of this case study anytime. Then, when you begin your job hunt, your case study will be a tangible way to demonstrate your knowledge and skills to potential employers.

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.

Characters and teams

- Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.
- Lily Moreno: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.
- Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about

Cyclistic's mission and business goals — as well as how you, as a junior data analyst, can help Cyclistic achieve them.

- Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

About the company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its 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. Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs. Moreno 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.

Ask

Three questions will guide the future marketing program: * 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? Moreno has assigned you the first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

Loading Packages and Tranforming Tables

```
library("here")
library("skimr")
library("janitor")
library("plyr")
library("dplyr")
library("tidyverse")
library("stringr")
library("rmarkdown")
```

```r
library("readxl")
library("knitr")

colnames(bicycle_data) [colnames(bicycle_data) == 'started_at'] <- 'Month'
colnames(bicycle_data) [colnames(bicycle_data) == 'ended_at'] <- 'Day'
```

The two codes above changed the format of ride_length, started_at, and ended_at columns and also changed the column names of started_at & ended_at to Month & Day so that the graphs and models will be more clear.

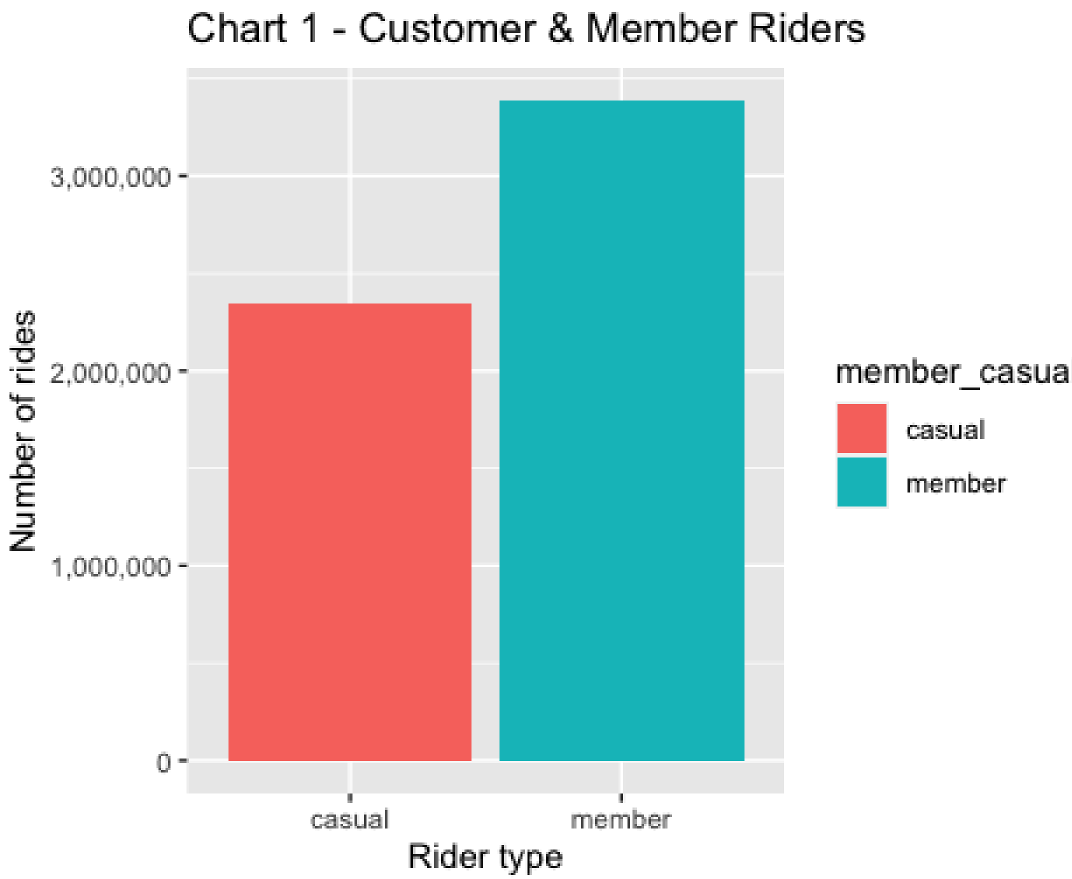
Analysis

Number of Casual and Member Riders

```r
bicycle_data %>% group_by(member_casual) %>% select(member_casual) %>%
  count()

## # A tibble: 2 × 2
## # Groups:   member_casual [2]
##   member_casual       n
##   <chr>           <int>
## 1 casual        2346876
## 2 member        3386575

bicycle_data %>% group_by(member_casual) %>% dplyr::summarise(number_of_rides
= n()) %>%
  ggplot(aes(x= member_casual, y= number_of_rides, fill= member_casual)) +
  geom_col(position = "dodge") + scale_y_continuous(labels= scales::comma) +
  labs(title = "Chart 1 - Customer & Member Riders", y= "Number of rides", x=
"Rider type")
```



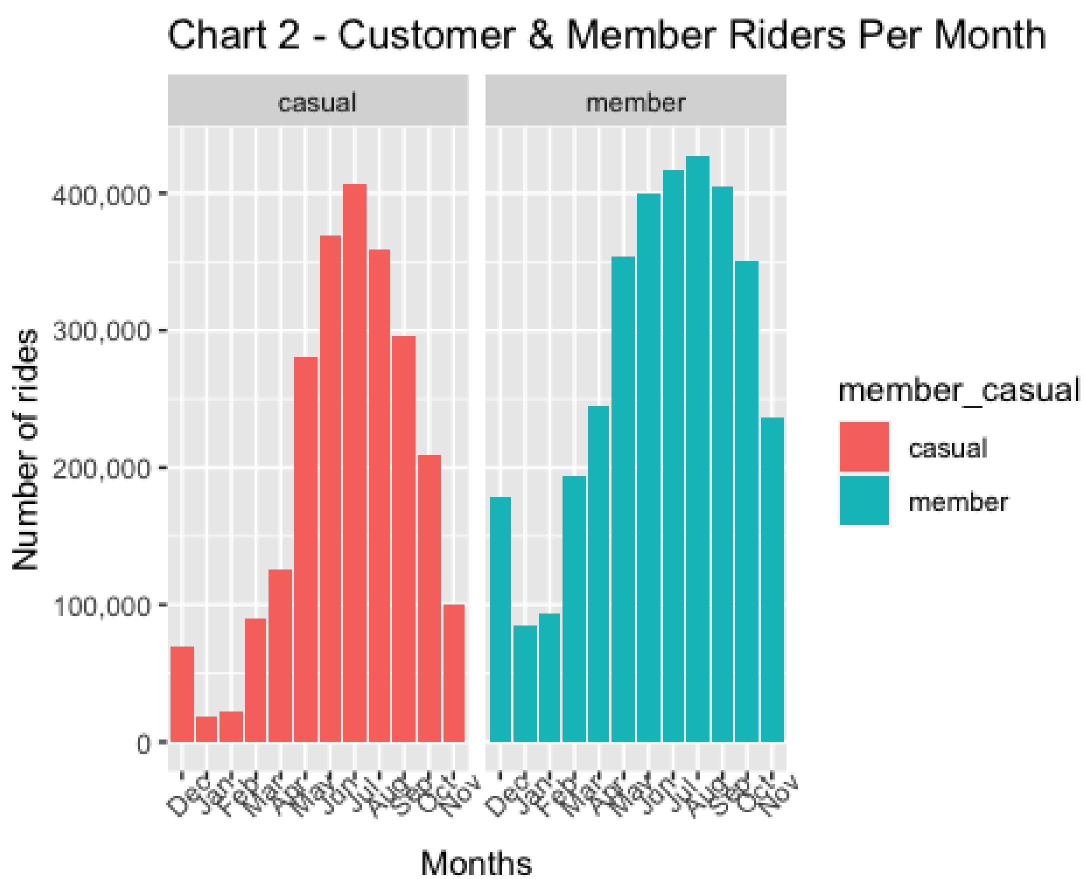
- There is currently 3,386,575 members compared to only 2,346,876 casual riders

Number of Riders Per Month

```
bicycle_data %>% group_by(Month, member_casual) %>% select(Month,
member_casual) %>%
  count()

## # A tibble: 24 × 3
## # Groups:   Month, member_casual [24]
##    Month member_casual      n
##    <chr> <chr>          <int>
##  1 Apr   casual        126417
##  2 Apr   member        244832
##  3 Aug   casual        358924
##  4 Aug   member        427008
##  5 Dec   casual         69738
##  6 Dec   member        177802
##  7 Feb   casual         21416
##  8 Feb   member         94193
##  9 Jan   casual         18520
## 10 Jan   member         85250
## # … with 14 more rows
```

```
bicycle_data$Month <- ordered(bicycle_data$Month, levels=c("Dec", "Jan",
"Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov"))
bicycle_data %>% group_by(member_casual, Month) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  ggplot(aes(x= Month,y= number_of_rides, fill = member_casual )) +
  geom_col(position = "dodge") +
  facet_wrap(~ member_casual) +
  scale_y_continuous(labels= scales::comma) + theme(axis.text.x =
element_text(angle = 45)) +
  labs(title="Chart 2 - Customer & Member Riders Per Month", y= "Number of
rides", x= "Months")
```



- Before graphing this data, I ordered the months from December to November due to the data starting in December, 2021.
- From the chart it is clear that Q2 and Q3 are the most popular riding times for both casual and member riders.

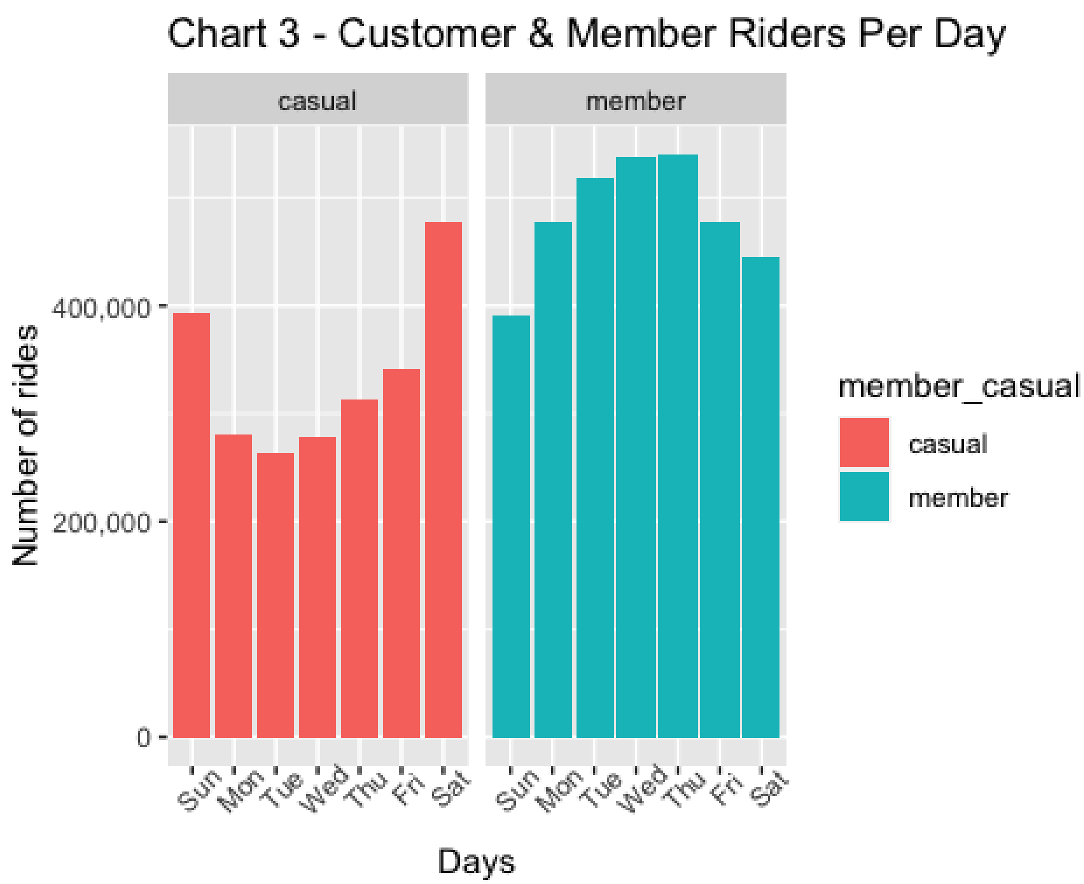
Riders Per Day

```
bicycle_data %>% group_by(day_name, member_casual) %>% select(day_name,
member_casual) %>% count()

## # A tibble: 14 × 3
## # Groups:   day_name, member_casual [14]
```

```
##    day_name member_casual      n
##    <chr>    <chr>          <int>
##  1 Fri      casual        340499
##  2 Fri      member        476908
##  3 Mon      casual        280472
##  4 Mon      member        476935
##  5 Sat      casual        476588
##  6 Sat      member        445473
##  7 Sun      casual        392130
##  8 Sun      member        390502
##  9 Thu      casual        313739
## 10 Thu      member        540347
## 11 Tue      casual        264068
## 12 Tue      member        518665
## 13 Wed      casual        279380
## 14 Wed      member        537745
```

```
bicycle_data$day_name <- ordered(bicycle_data$day_name, levels=c("Sun",
"Mon", "Tue", "Wed", "Thu", "Fri", "Sat"))
bicycle_data %>% group_by(member_casual, day_name) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  ggplot(aes(x= day_name,y= number_of_rides, fill = member_casual )) +
  geom_col(position = "dodge") +
  facet_wrap(~ member_casual) +
  scale_y_continuous(labels= scales::comma) + theme(axis.text.x =
element_text(angle = 45)) +
  labs(title="Chart 3 - Customer & Member Riders Per Day", y= "Number of
rides", x= "Days")
```



- These charts show that casual riders use the bikes more heavily on the weekends where as the member use them more during the week. This can be an early indication that casual riders are using the bikes more for leisure and activities where members would be using them more for transportation for school/work.

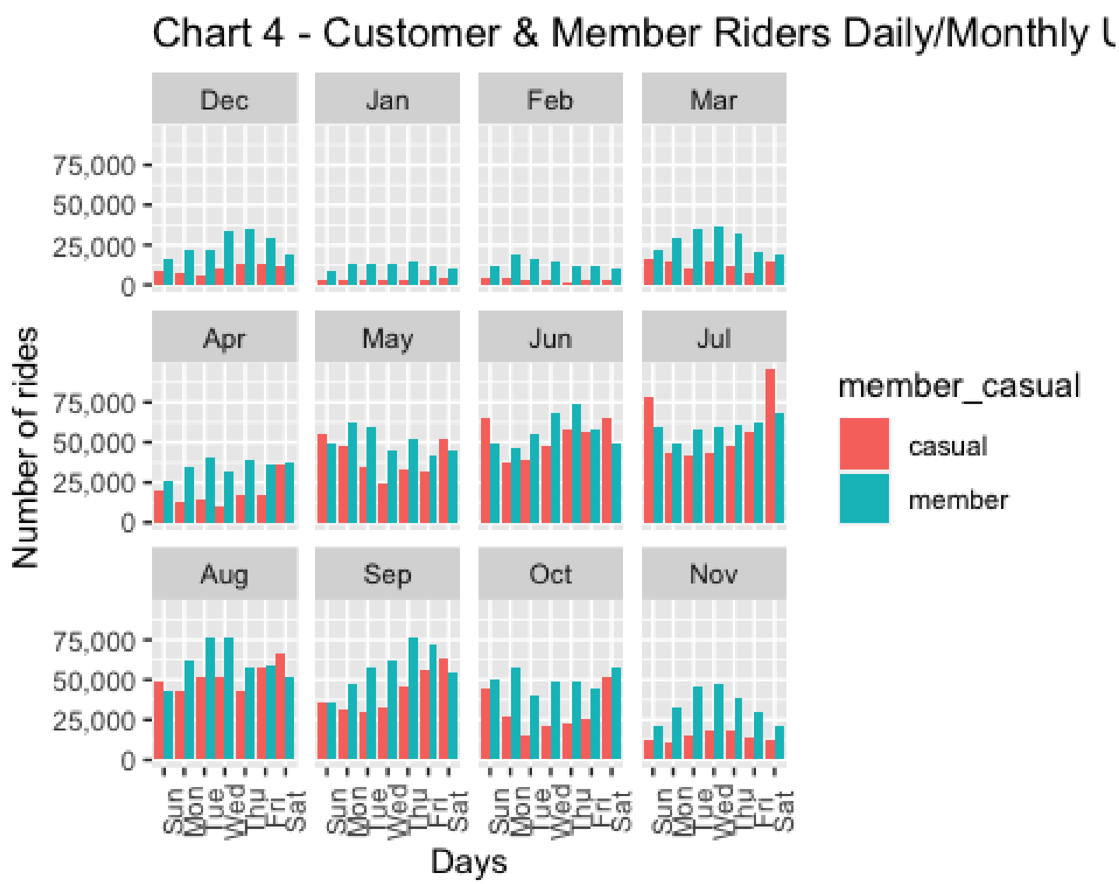
Riders Per Month & Day

```
bicycle_data %>% group_by(member_casual, day_name, Month) %>%
select(member_casual, day_name, Month) %>% count()

## # A tibble: 168 × 4
## # Groups:   member_casual, day_name, Month [168]
##    member_casual day_name Month     n
##    <chr>         <ord>    <ord> <int>
##  1 casual        Sun      Dec    8437
##  2 casual        Sun      Jan    2515
##  3 casual        Sun      Feb    4206
##  4 casual        Sun      Mar   16575
##  5 casual        Sun      Apr   19388
##  6 casual        Sun      May   55321
##  7 casual        Sun      Jun   65851
##  8 casual        Sun      Jul   78251
##  9 casual        Sun      Aug   48154
```

```
## 10 casual         Sun       Sep    36254
## # … with 158 more rows
```

```
bicycle_data %>% group_by(member_casual, day_name, Month) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  ggplot(aes(x= day_name,y= number_of_rides, fill = member_casual )) +
  geom_col(position = "dodge") +
  facet_wrap(~ Month) +
  scale_y_continuous(labels= scales::comma) + theme(axis.text.x =
element_text(angle = 90)) +
  labs(title="Chart 4 - Customer & Member Riders Daily/Monthly Use", y=
"Number of rides", x= "Days")
```



- This chart really puts into perspective how busy it is in Q2 and Q3 in comparison to Q1 and Q4.

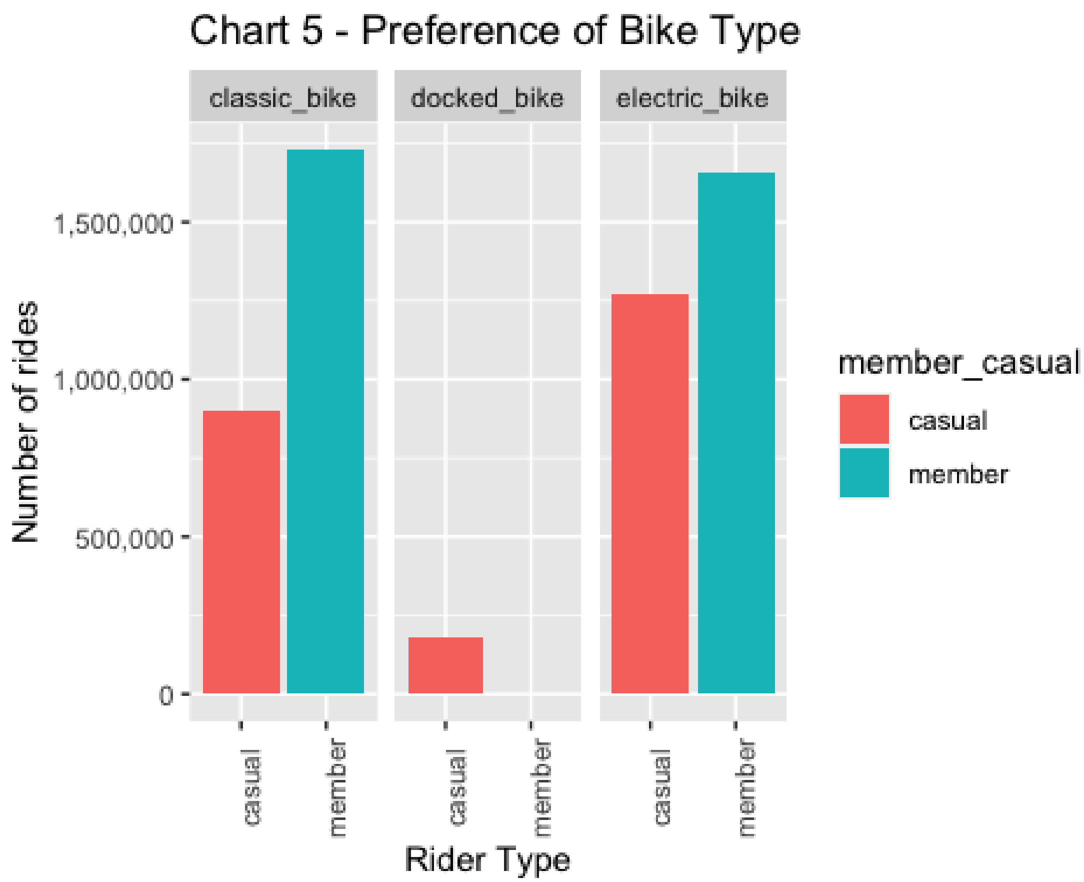
Riders By Bike Type

```
bicycle_data %>% group_by(member_casual, rideable_type) %>%
select(member_casual, rideable_type) %>% count()
```

```
## # A tibble: 5 × 3
## # Groups:   member_casual, rideable_type [5]
##   member_casual rideable_type       n
##   <chr>         <chr>           <int>
```

```
## 1 casual        classic_bike   898613
## 2 casual        docked_bike    180477
## 3 casual        electric_bike 1267786
## 4 member        classic_bike  1729886
## 5 member        electric_bike 1656689
```

```
bicycle_data %>% group_by(member_casual, rideable_type) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  ggplot(aes(x= member_casual,y= number_of_rides, fill = member_casual )) +
  geom_col(position = "dodge") +
  facet_wrap(~ rideable_type) +
  scale_y_continuous(labels= scales::comma) + theme(axis.text.x =
element_text(angle = 90)) +
  labs(title="Chart 5 - Preference of Bike Type", y= "Number of rides", x=
"Rider Type")
```



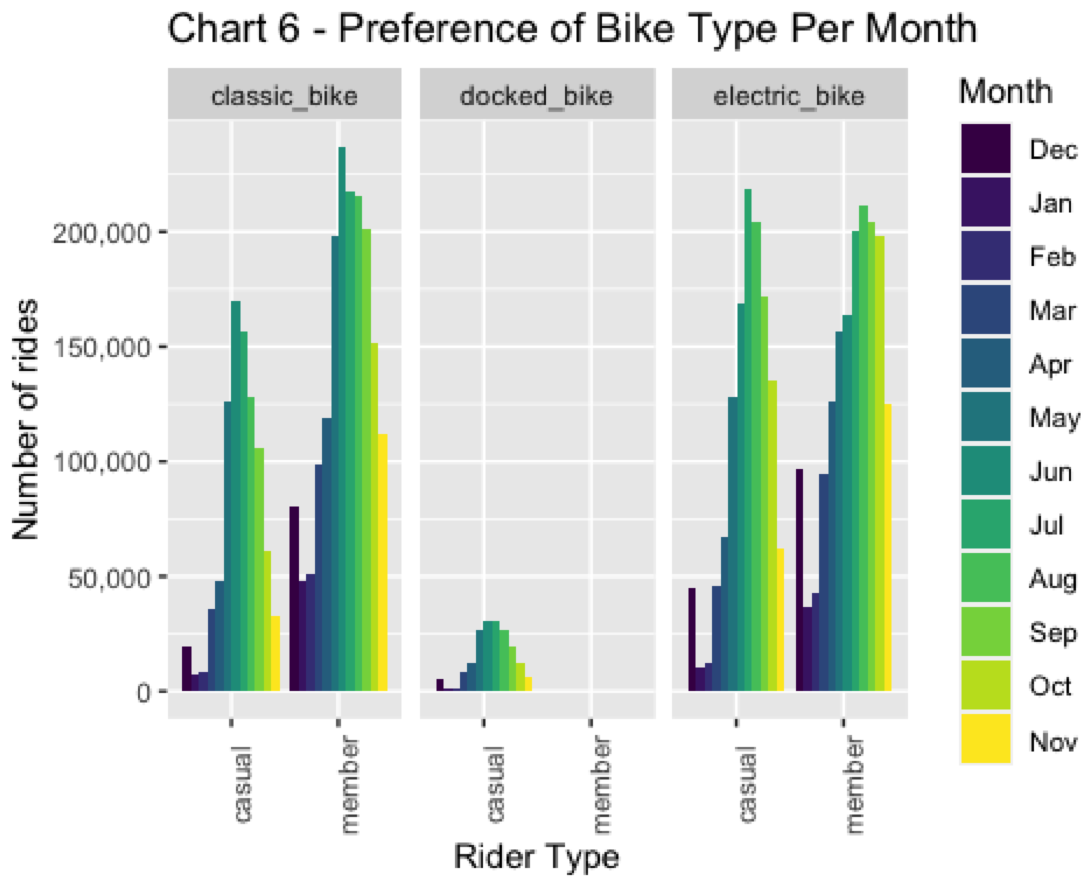
- It is clear that the electric bike is the most used, followed by the classic bike, and then the docked bike.

Bike Type Per Month

```
bicycle_data %>% group_by(member_casual, rideable_type, Month) %>%
select(member_casual, rideable_type, Month) %>% count()
```

```
## # A tibble: 60 × 4
## # Groups:   member_casual, rideable_type, Month [60]
##    member_casual rideable_type Month      n
##    <chr>         <chr>         <ord>  <int>
##  1 casual        classic_bike  Dec    19806
##  2 casual        classic_bike  Jan     6974
##  3 casual        classic_bike  Feb     8107
##  4 casual        classic_bike  Mar    35387
##  5 casual        classic_bike  Apr    47543
##  6 casual        classic_bike  May   126075
##  7 casual        classic_bike  Jun   169996
##  8 casual        classic_bike  Jul   156095
##  9 casual        classic_bike  Aug   128635
## 10 casual        classic_bike  Sep   105375
## # … with 50 more rows
```

```
bicycle_data %>% group_by(member_casual, rideable_type, Month) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  ggplot(aes(x= member_casual,y= number_of_rides, fill = Month )) +
  geom_col(position = "dodge") +
  facet_wrap(~ rideable_type) +
  scale_y_continuous(labels= scales::comma) + theme(axis.text.x =
element_text(angle = 90)) +
  labs(title="Chart 6 - Preference of Bike Type Per Month", y= "Number of
rides", x= "Rider Type")
```



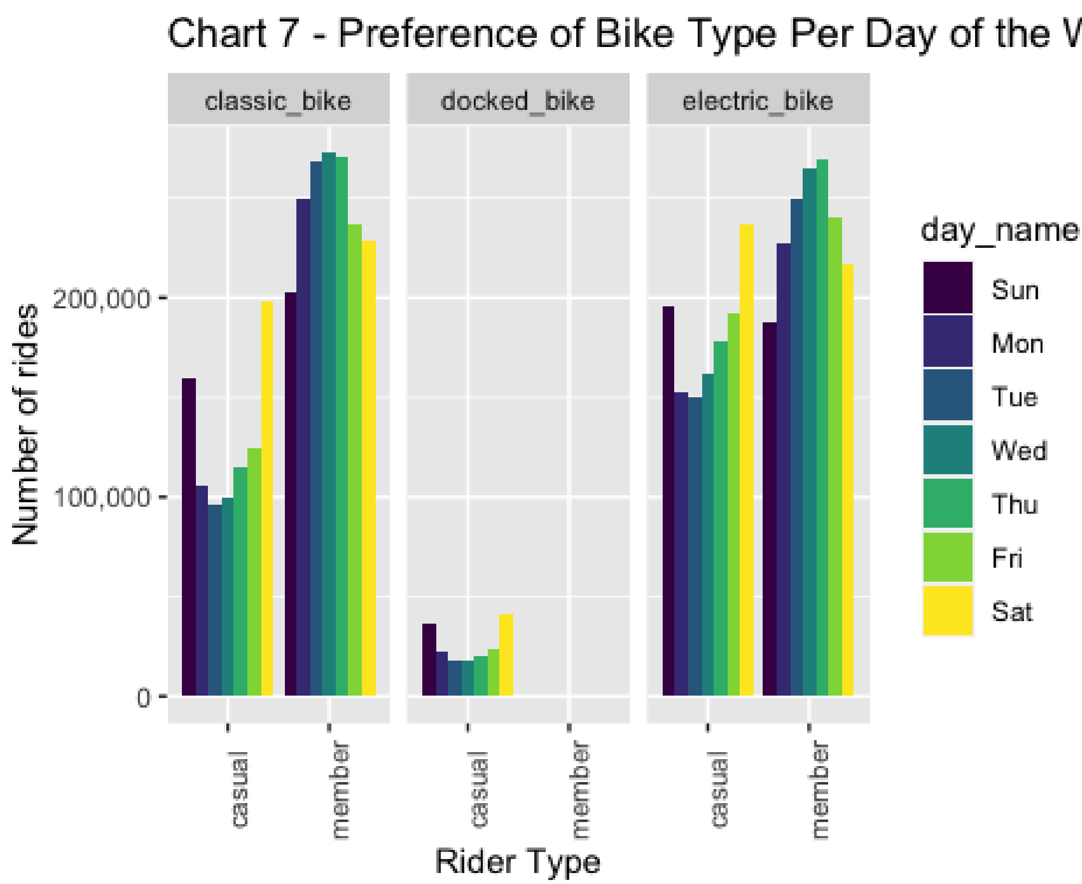
- This chart gives a good indication of what bikes to be advertising based on the month that it is.

Bike Type Per Day

```
bicycle_data %>% group_by(member_casual, rideable_type, day_name) %>%
select(member_casual, rideable_type, day_name) %>% count()

## # A tibble: 35 × 4
## # Groups:   member_casual, rideable_type, day_name [35]
##    member_casual rideable_type day_name      n
##    <chr>         <chr>         <ord>     <int>
##  1 casual        classic_bike  Sun      159587
##  2 casual        classic_bike  Mon      105067
##  3 casual        classic_bike  Tue       96240
##  4 casual        classic_bike  Wed       99737
##  5 casual        classic_bike  Thu      114936
##  6 casual        classic_bike  Fri      124709
##  7 casual        classic_bike  Sat      198337
##  8 casual        docked_bike   Sun       36197
##  9 casual        docked_bike   Mon       22909
## 10 casual        docked_bike   Tue       17866
## # … with 25 more rows
```

```
bicycle_data %>% group_by(member_casual, rideable_type, day_name) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  ggplot(aes(x= member_casual,y= number_of_rides, fill = day_name )) +
  geom_col(position = "dodge") +
  facet_wrap(~ rideable_type) +
  scale_y_continuous(labels= scales::comma) + theme(axis.text.x =
element_text(angle = 90)) +
  labs(title="Chart 7 - Preference of Bike Type Per Day of the Week", y=
"Number of rides", x= "Rider Type")
```



- This chart gives a good indication of what bike types are the most popular for each day of the week.
- For instance, the company could heavily advertise electric bikes for casual members on the weekends.

Ride Length Summary

```
bicycle_data %>%
  summarise(averageridelength = mean(ride_length, na.rm = TRUE),
            minridelength = min(ride_length, na.rm = TRUE),
            medianridelength = median(ride_length, na.rm = TRUE),
            maxridelength = max(ride_length, na.rm = TRUE) )
```

```
## # A tibble: 1 × 4
##   averageridelength minridelength medianridelength maxridelength
##               <dbl>         <dbl>            <dbl>         <dbl>
## 1              16.5             0             10.3         1440.
```

```
bicycle_data %>% group_by(member_casual) %>%
  dplyr::summarise(averageridelength = mean(ride_length, na.rm = TRUE),
            minridelength = min(ride_length, na.rm = TRUE ),
            medianridelength = median(ride_length, na.rm = TRUE),
            maxridelength = max(ride_length, na.rm = TRUE))
```

```
## # A tibble: 2 × 5
##   member_casual averageridelength minridelength medianridelength
maxridelength
##   <chr>                     <dbl>         <dbl>            <dbl>
<dbl>
## 1 casual                     22.4             0             13.0
1440.
## 2 member                     12.4             0              8.83
1440.
```

Ride Length Per Month & Day

```
aggregate(bicycle_data$ride_length ~ bicycle_data$member_casual +
bicycle_data$day_name, FUN = mean, na.rm = TRUE)
```

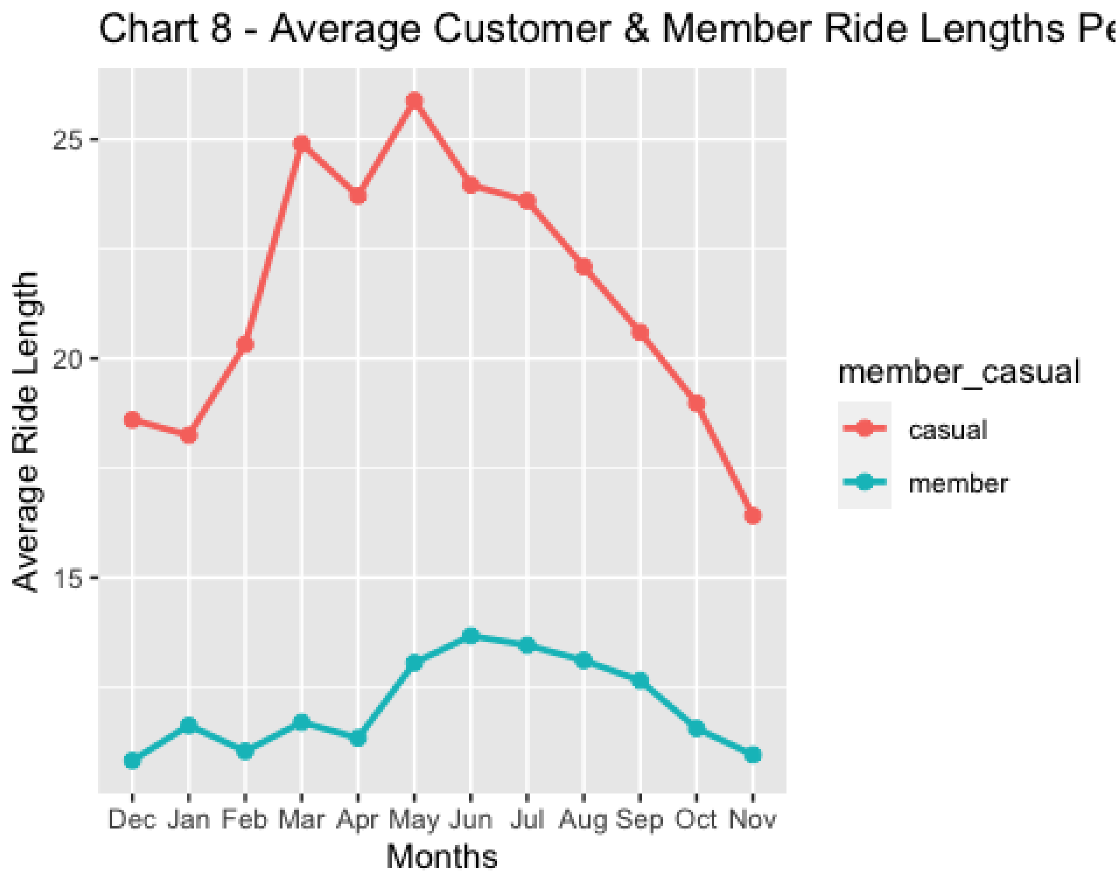
```
##    bicycle_data$member_casual bicycle_data$day_name
bicycle_data$ride_length
## 1                      casual                   Sun
25.78416
## 2                      member                   Sun
13.75114
## 3                      casual                   Mon
22.94757
## 4                      member                   Mon
12.00174
## 5                      casual                   Tue
20.17575
## 6                      member                   Tue
11.85273
## 7                      casual                   Wed
19.26373
## 8                      member                   Wed
11.80619
## 9                      casual                   Thu
19.99613
## 10                     member                   Thu
12.02746
## 11                     casual                   Fri
21.00481
## 12                     member                   Fri
12.21613
```

```
## 13                      casual                    Sat
25.11199
## 14                      member                    Sat
13.83043
```

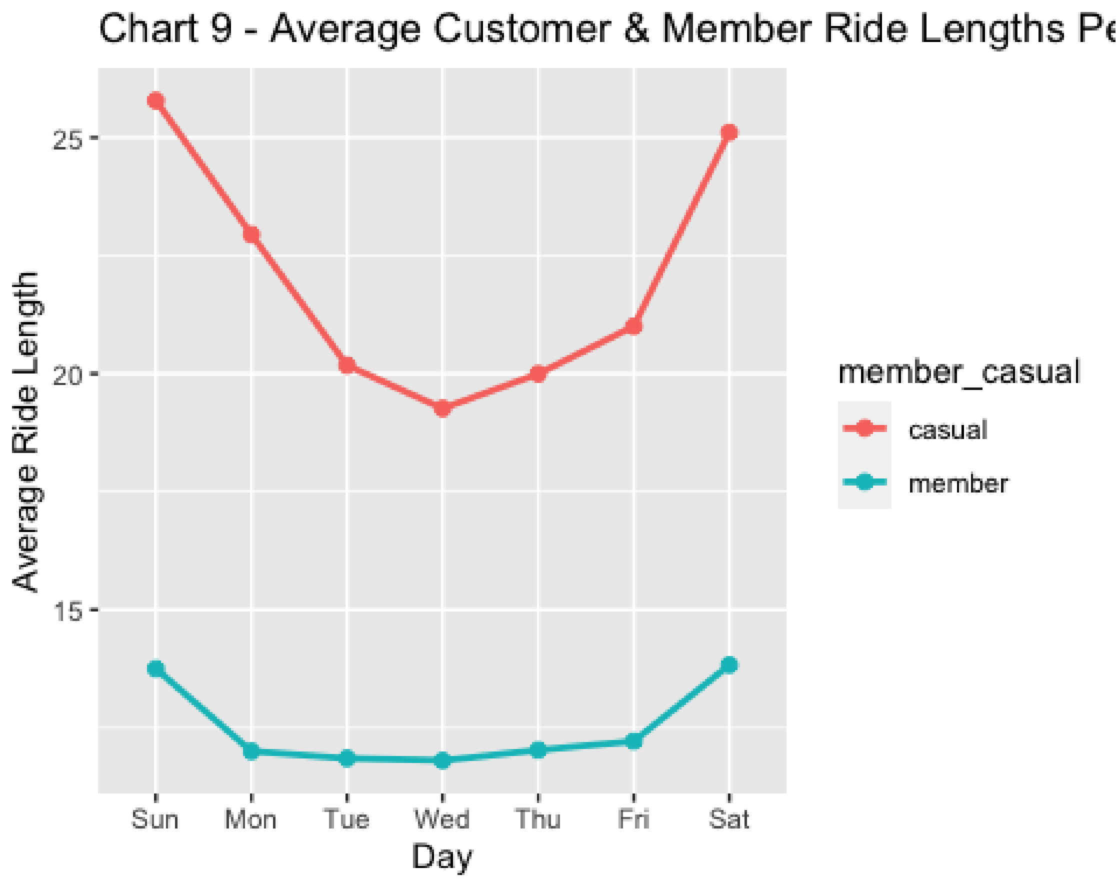
```
aggregate(bicycle_data$ride_length ~ bicycle_data$member_casual +
bicycle_data$Month, FUN = mean, na.rm = TRUE)
```

```
##    bicycle_data$member_casual bicycle_data$Month bicycle_data$ride_length
## 1                      casual                Dec                 18.60055
## 2                      member                Dec                 10.82676
## 3                      casual                Jan                 18.24788
## 4                      member                Jan                 11.62680
## 5                      casual                Feb                 20.32071
## 6                      member                Feb                 11.03861
## 7                      casual                Mar                 24.90026
## 8                      member                Mar                 11.69875
## 9                      casual                Apr                 23.71169
## 10                     member                Apr                 11.33948
## 11                     casual                May                 25.88328
## 12                     member                May                 13.05791
## 13                     casual                Jun                 23.95761
## 14                     member                Jun                 13.67096
## 15                     casual                Jul                 23.59689
## 16                     member                Jul                 13.45617
## 17                     casual                Aug                 22.09648
## 18                     member                Aug                 13.11101
## 19                     casual                Sep                 20.59825
## 20                     member                Sep                 12.65192
## 21                     casual                Oct                 18.98034
## 22                     member                Oct                 11.55873
## 23                     casual                Nov                 16.41345
## 24                     member                Nov                 10.95847
```

```
bicycle_data %>% group_by(member_casual, Month) %>%
  drop_na(ride_length) %>%
  dplyr::summarise(number_of_rides = n(),
            averageride_length = mean(ride_length, na.rm = TRUE)) %>%
  ggplot(aes(x = Month, y = averageride_length, colour = member_casual, group
= member_casual)) +
  geom_line(linewidth =1) +
  geom_point(size = 2) +
  labs(title = "Chart 8 - Average Customer & Member Ride Lengths Per Month",
y = "Average Ride Length", x = "Months" )
```




```
bicycle_data %>% group_by(member_casual, day_name) %>%
  drop_na(ride_length) %>%
  dplyr::summarise(number_of_rides = n(),
                   averageride_length = mean(ride_length, na.rm = TRUE)) %>%
  ggplot(aes(x = day_name, y = averageride_length, colour = member_casual,
group = member_casual)) +
  geom_line(linewidth =1) +
  geom_point(size = 2) +
  labs(title = "Chart 9 - Average Customer & Member Ride Lengths Per Day", y
= "Average Ride Length", x = "Day" )
```



- The charts and graphs show that casual riders have a much longer ride length than members do. This also supports the statement that casual riders are intending to use the bikes in a more activity and leisure way where as the members use the bikes mostly for commuting.

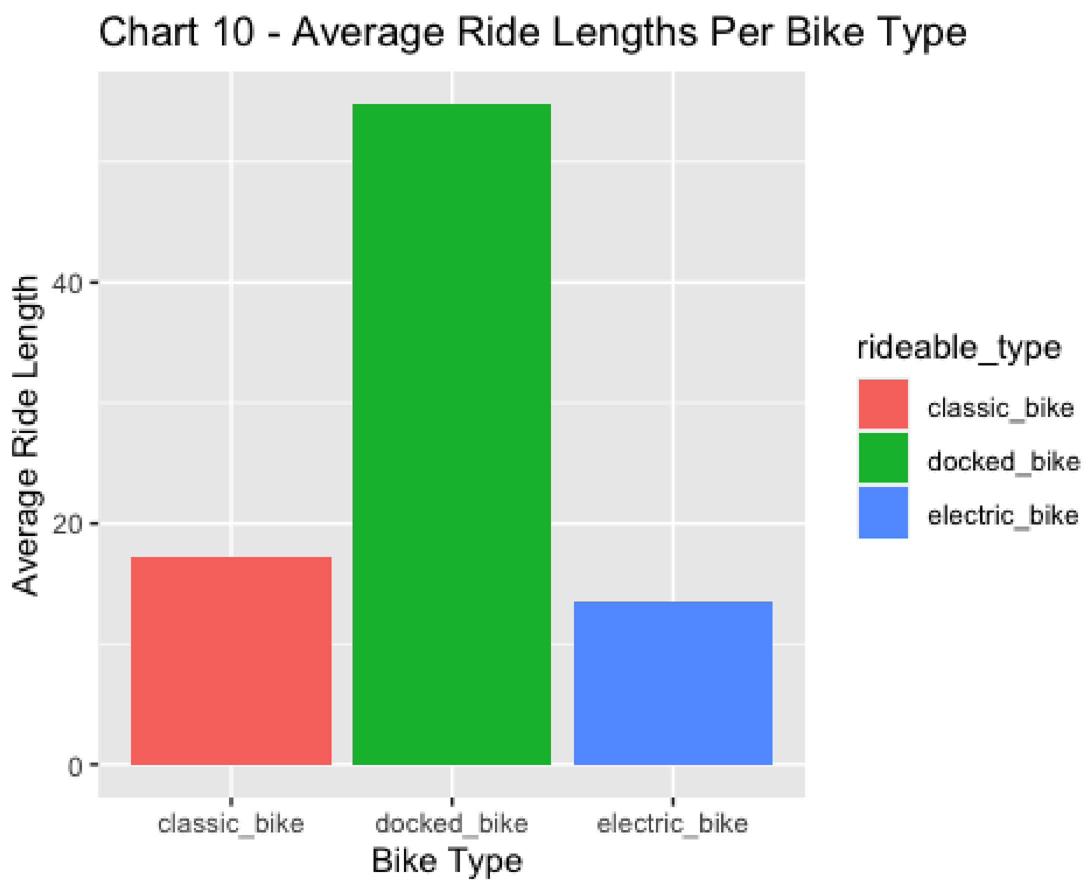
Ride Length Per Bike Type

```
bicycle_data %>% group_by(member_casual, rideable_type) %>%
  drop_na(ride_length) %>%
  dplyr::summarise(number_of_rides = n(),
            averageride_length = mean(ride_length) )

## # A tibble: 5 × 4
## # Groups:   member_casual [2]
##   member_casual rideable_type number_of_rides averageride_length
##   <chr>         <chr>                   <int>              <dbl>
## 1 casual        classic_bike           898613               24.6
## 2 casual        docked_bike            180477               54.8
## 3 casual        electric_bike         1267786               16.3
## 4 member        classic_bike          1729886               13.3
## 5 member        electric_bike         1656688               11.5

bicycle_data %>% group_by(rideable_type) %>%
  drop_na(ride_length) %>%
```

```
  dplyr::summarise(number_of_rides = n(),
            averageride_length = mean(ride_length)) %>%
  ggplot(aes(x = rideable_type, y = averageride_length, fill =
rideable_type)) +
  geom_col()+
  labs(title = "Chart 10 - Average Ride Lengths Per Bike Type", y = "Average
Ride Length", x = "Bike Type " )
```




```
bicycle_data %>% group_by(member_casual, rideable_type) %>%
  drop_na(ride_length) %>%
  dplyr::summarise(number_of_rides = n(),
            averageride_length = mean(ride_length)) %>%
  ggplot(aes(x = member_casual, y = averageride_length, fill =
member_casual)) +
  geom_col()+
  facet_wrap(~ rideable_type) +
  labs(title = "Chart 11 - Average Customer & Member Ride Lengths Per Bike
Type", y = "Average Ride Length", x = "Rider Type " )
```

- While the docked bike has significantly less total rides, their ride length is much higher than the other options.
- It is also seen that there is zero rides with the docked bike with members. This is an indication that it is not included with their membership.

Popular Bike Stations For Members

```
bicycle_data %>% subset(member_casual == "member") %>%
  group_by(start_station_name) %>%
  drop_na(start_station_name) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  arrange(-number_of_rides) %>%
  top_n(6)

## # A tibble: 6 × 2
##   start_station_name           number_of_rides
##   <chr>                                  <int>
## 1 Kingsbury St & Kinzie St               25389
## 2 Clark St & Elm St                      22350
## 3 Wells St & Concord Ln                  21595
## 4 University Ave & 57th St               20201
## 5 Clinton St & Washington Blvd           19950
## 6 Ellis Ave & 60th St                    19673
```

```
bicycle_data %>% subset(member_casual == "member") %>%
  group_by(end_station_name) %>%
  drop_na(end_station_name) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  arrange(-number_of_rides) %>%
  top_n(6)
```

```
## # A tibble: 6 × 2
##   end_station_name             number_of_rides
##   <chr>                                  <int>
## 1 Kingsbury St & Kinzie St               24972
## 2 Clark St & Elm St                      22718
## 3 Wells St & Concord Ln                  22189
## 4 University Ave & 57th St               20847
## 5 Clinton St & Washington Blvd           20665
## 6 Clinton St & Madison St                20026
```

Popular Bike Stations For Casual Riders

```
bicycle_data %>% subset(member_casual == "casual") %>%
  group_by(start_station_name) %>%
  drop_na(start_station_name) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  arrange(-number_of_rides) %>%
  top_n(6)
```

```
## # A tibble: 6 × 2
##   start_station_name                 number_of_rides
##   <chr>                                        <int>
## 1 Streeter Dr & Grand Ave                      58599
## 2 DuSable Lake Shore Dr & Monroe St            32535
## 3 Millennium Park                              25913
## 4 Michigan Ave & Oak St                        25428
## 5 DuSable Lake Shore Dr & North Blvd           23778
## 6 Shedd Aquarium                               20568
```

```
bicycle_data %>% subset(member_casual == "casual") %>%
  group_by(end_station_name) %>%
  drop_na(end_station_name) %>%
  dplyr::summarise(number_of_rides = n()) %>%
  arrange(-number_of_rides) %>%
  top_n(6)
```

```
## # A tibble: 6 × 2
##   end_station_name                   number_of_rides
##   <chr>                                        <int>
## 1 Streeter Dr & Grand Ave                      60477
## 2 DuSable Lake Shore Dr & Monroe St            30001
## 3 Millennium Park                              27152
## 4 Michigan Ave & Oak St                        26702
## 5 DuSable Lake Shore Dr & North Blvd           26269
## 6 Theater on the Lake                          19524
```

- These charts show where the company should focus their marketing for both members and casual riders.

Summary & Findings

- Both member and casual riders used the cyclistic bikes more frequently during Q2 and Q3.
- While casual riders used the bikes more frequently on weekends, member riders used the bikes more throughout the week.
- The classic and electric bike were popular among both cyclistic groups where as the docked bike was only used by casual bikers.
- Casual riders have a longer average ride length (22 minutes) compared to members (12 minutes).
- Docked bikes are the least used bikes but have the longest average ride length (55 minutes).
- The stations that are most popular to member riders differ to the most popular stations of the casual riders.

Recommendations

1). The marketing strategy should be targeted at the 2nd & 3rd quarter as these are the most popular time periods among casual riders.

2). Prices could be increased for weekends and docked bikes as these are the most common among casual riders and it could persuade them to buy a membership. Also, offering a different membership package that is more tailored for the casual riders could be another option (Weekend-only membership).

3). Marketing strategies should be focused on the most popular bike stations among casual riders. Marketing campaigns can also be sent either by email or in the docking stations explaining the benefits of the annual membership since docked bikes are more common among casual riders.