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

Background

Cyclistic launched a successful bike-share offering with 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.

Cyclistic's pricing plans include: 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, the director of marketing 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, there is a very good chance to convert casual riders into members. Casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Business task

> What are the best marketing strategies to convert casual riders into annual members?

Break down of the business task with three questions:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

This analysis will focus on the first question:

> How do annual members and casual riders use Cyclistic bikes differently?

Data sources

The dataset can be found here. User data is from a 12 month period from January 2022 - December 2022. (Cyclistic is a fictional company and the dataset has been made publicly available by Motivate International Inc. via license). Riders' personal data is unavailable and prohibited.

Data cleaning, data analysis and data visualization done with R.

Data cleaning

```r
# load libraries used for analysis
library(tidyverse)
library(lubridate)
library(ggplot2)
```

```r
# load datasets
df_202201 <- read_csv('202201-divvy-tripdata.csv')
df_202202 <- read_csv('202202-divvy-tripdata.csv')
df_202203 <- read_csv('202203-divvy-tripdata.csv')
df_202204 <- read_csv('202204-divvy-tripdata.csv')
df_202205 <- read_csv('202205-divvy-tripdata.csv')
df_202206 <- read_csv('202206-divvy-tripdata.csv')
df_202207 <- read_csv('202207-divvy-tripdata.csv')
df_202208 <- read_csv('202208-divvy-tripdata.csv')
df_202209 <- read_csv('202209-divvy-tripdata.csv')
df_202210 <- read_csv('202210-divvy-tripdata.csv')
df_202211 <- read_csv('202211-divvy-tripdata.csv')
df_202212 <- read_csv('202212-divvy-tripdata.csv')
```

```r
# combine all dataframes into one dataframe
df_2022 <- bind_rows(df_202201,df_202202,df_202203,df_202204,df_202205,df_202206,
                     df_202207,df_202208,df_202209,df_202210,df_202211,df_202212)
head(df_2022)

# check for missing data
colSums(is.na(df_2022))

# remove columns that are not needed
df_2022 <- df_2022 %>%
  select(-c(start_lat,start_lng,end_lat,end_lng,
            start_station_id,end_station_id,start_station_name,end_station_name))
```

```r
# column names of the dataframe
colnames(df_2022)

# dimensions of the dataframe
dim(df_2022)

# structure of the dataframe
str(df_2022)

# summary analysis of the dataframe
summary(df_2022)
```

```r
# check for number of member/casual and rideable type
table(df_2022$member_casual)
table(df_2022$rideable_type)
```

```r
# add separate columns for date and time
df_2022$date <- as.Date(df_2022$started_at)
df_2022$month <- format(as.Date(df_2022$date),'%m')
df_2022$day <- format(as.Date(df_2022$date),'%d')
df_2022$year <- format(as.Date(df_2022$date),'%Y')
df_2022$day_of_week <- format(as.Date(df_2022$date),'%A')
df_2022$time <- format(as.POSIXct(df_2022$started_at),format='%H:%M')
```

```r
# create ride_length column to show the time length of each unique ride
df_2022$ride_length <- (difftime(df_2022$ended_at,df_2022$started_at,units='mins'))
df_2022$ride_length <- as.numeric(as.character(df_2022$ride_length))

# check for negative ride lengths and delete them
table(df_2022$ride_length<0)

# delete negative ride lengths
df_2022 <- df_2022[!(df_2022$ride_length<0),]

# check summary of new column
summary(df_2022$ride_length)
```

Data analysis

```r
# calculate the mean, median, max, min for each type of customer
aggregate(df_2022$ride_length ~ df_2022$member_casual, FUN = mean)
```

```
##   df_2022$member_casual df_2022$ride_length
## 1                casual            29.14572
## 2                member            12.71401
```

```r
aggregate(df_2022$ride_length ~ df_2022$member_casual, FUN = median)
```

```
##   df_2022$member_casual df_2022$ride_length
## 1                casual           13.000000
## 2                member            8.833333
```

```r
aggregate(df_2022$ride_length ~ df_2022$member_casual, FUN = max)
```

```
##   df_2022$member_casual df_2022$ride_length
## 1                casual            41387.25
## 2                member             1559.90
```

```r
aggregate(df_2022$ride_length ~ df_2022$member_casual, FUN = min)
```

```
##   df_2022$member_casual df_2022$ride_length
## 1                casual                   0
## 2                member                   0
```

```r
# order the days of week
df_2022$day_of_week <- ordered(df_2022$day_of_week,
                               levels=c('Sunday','Monday','Tuesday','Wednesday',
                                        'Thursday','Friday','Saturday'))

# average ride time by day of week
aggregate(df_2022$ride_length ~ df_2022$member_casual + df_2022$day_of_week, FUN = mean)
```

```
##    df_2022$member_casual df_2022$day_of_week df_2022$ride_length
## 1                 casual              Sunday            34.05795
## 2                 member              Sunday            14.03124
## 3                 casual              Monday            29.18736
## 4                 member              Monday            12.27011
## 5                 casual             Tuesday            25.82287
## 6                 member             Tuesday            12.12949
## 7                 casual           Wednesday            24.75085
## 8                 member           Wednesday            12.10489
## 9                 casual            Thursday            25.54792
## 10                member            Thursday            12.29273
## 11                casual              Friday            28.04425
## 12                member              Friday            12.53077
## 13                casual            Saturday            32.61408
## 14                member            Saturday            14.14006
```

```r
# analyze by type of customers and days of the week
df_2022 %>%
  mutate(weekday=wday(started_at,label=TRUE)) %>%
  group_by(member_casual,weekday) %>%
  summarise(number_of_rides=n(),average_duration=mean(ride_length)) %>%
  arrange(member_casual,weekday)
```

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday number_of_rides average_duration
##    <chr>         <ord>             <int>            <dbl>
##  1 casual        Sun              389011             34.1
##  2 casual        Mon              277671             29.2
##  3 casual        Tue              263731             25.8
##  4 casual        Wed              274354             24.8
##  5 casual        Thu              309327             25.5
##  6 casual        Fri              334698             28.0
##  7 casual        Sat              473185             32.6
##  8 member        Sun              387208             14.0
##  9 member        Mon              473335             12.3
## 10 member        Tue              518618             12.1
## 11 member        Wed              523867             12.1
## 12 member        Thu              532255             12.3
## 13 member        Fri              467083             12.5
## 14 member        Sat              443274             14.1
```

```r
# analyze by type of customers and month
print(df_2022 %>%
        group_by(member_casual,month) %>%
        summarise(number_of_rides=n(),average_duration=mean(ride_length)) %>%
        arrange(member_casual,month),n=24)
```

```
## # A tibble: 24 x 4
## # Groups:   member_casual [2]
##    member_casual month number_of_rides average_duration
##    <chr>         <chr>           <int>            <dbl>
##  1 casual        01              18520             30.4
##  2 casual        02              21416             26.7
##  3 casual        03              89880             32.6
##  4 casual        04             126417             29.5
##  5 casual        05             280414             30.9
##  6 casual        06             369044             32.1
##  7 casual        07             406046             29.3
##  8 casual        08             358917             29.3
##  9 casual        09             296694             28.0
## 10 casual        10             208988             26.4
## 11 casual        11             100747             21.3
## 12 casual        12              44894             22.3
## 13 member        01              85250             12.0
## 14 member        02              94193             11.4
## 15 member        03             194160             12.0
## 16 member        04             244832             11.5
## 17 member        05             354443             13.4
## 18 member        06             400148             14.0
## 19 member        07             417426             13.7
## 20 member        08             427000             13.4
## 21 member        09             404636             13.0
## 22 member        10             349693             12.0
## 23 member        11             236947             11.1
## 24 member        12             136912             10.6
```

```r
# analyze by type of bike
df_2022 %>%
  group_by(member_casual,rideable_type) %>%
  summarise(number_of_rides=n(),average_duration=mean(ride_length)) %>%
  arrange(member_casual,rideable_type)
```

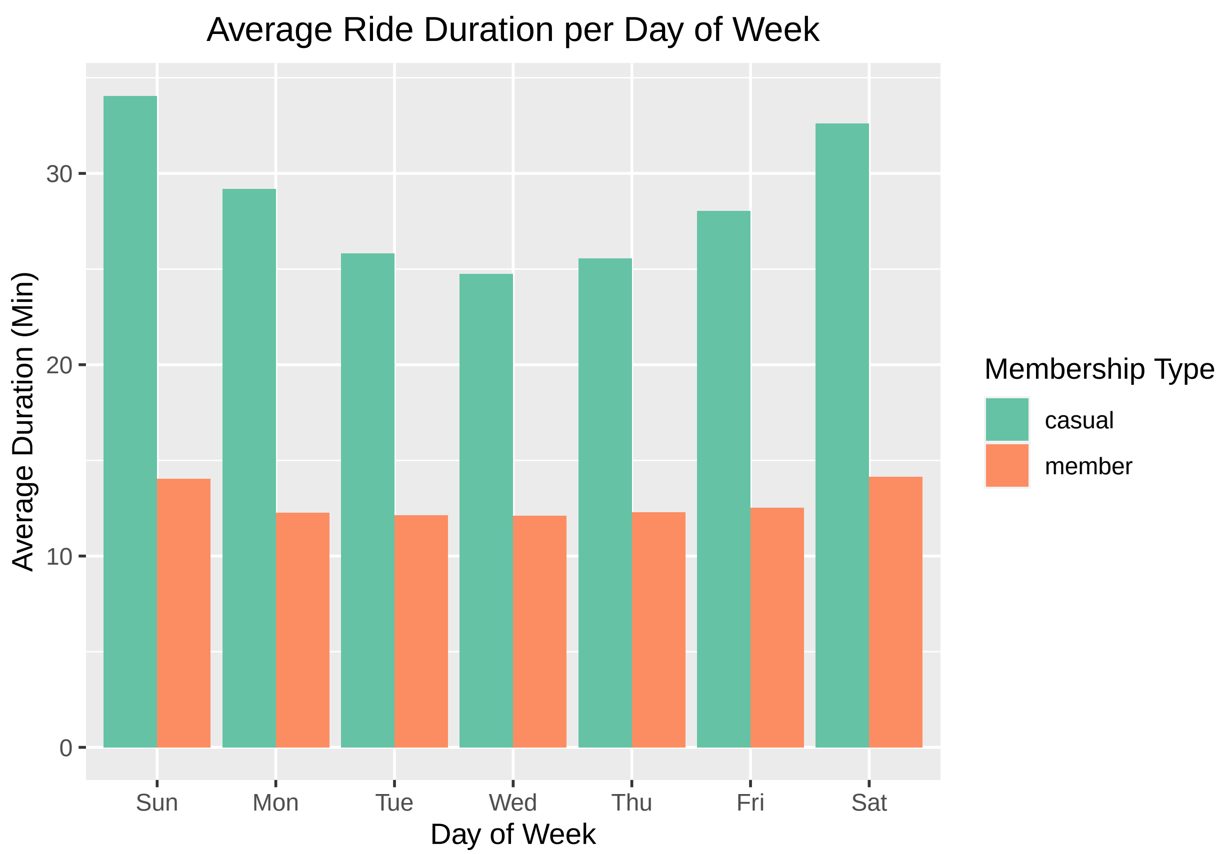
```
## # A tibble: 5 x 4
## # Groups:   member_casual [2]
##   member_casual rideable_type number_of_rides average_duration
##   <chr>         <chr>                   <int>            <dbl>
## 1 casual        classic_bike           891443             28.8
## 2 casual        docked_bike            177474            123.
## 3 casual        electric_bike         1253060             16.2
## 4 member        classic_bike          1709743             13.9
## 5 member        electric_bike         1635897             11.5
```

Data visualizations

```
# bar chart of number of rides by day of week
df_2022 %>%
  mutate(weekday=wday(started_at,label=TRUE)) %>%
  group_by(member_casual,weekday) %>%
  summarise(number_of_rides=n()) %>%
  arrange(member_casual,weekday) %>%
  ggplot(aes(x=weekday,y=number_of_rides,fill=member_casual)) +
  geom_col(position='dodge') +
  labs(x='Day of Week',y='Total Number of Rides',title='Rides per Day of Week',
       fill='Membership Type') +
  theme(plot.title=element_text(hjust=0.5)) + scale_fill_brewer(palette='Set2') +
  scale_y_continuous(breaks=c(100000,200000,300000,400000,500000),
                     labels=c('100k','200K','300K','400K','500K'))
```

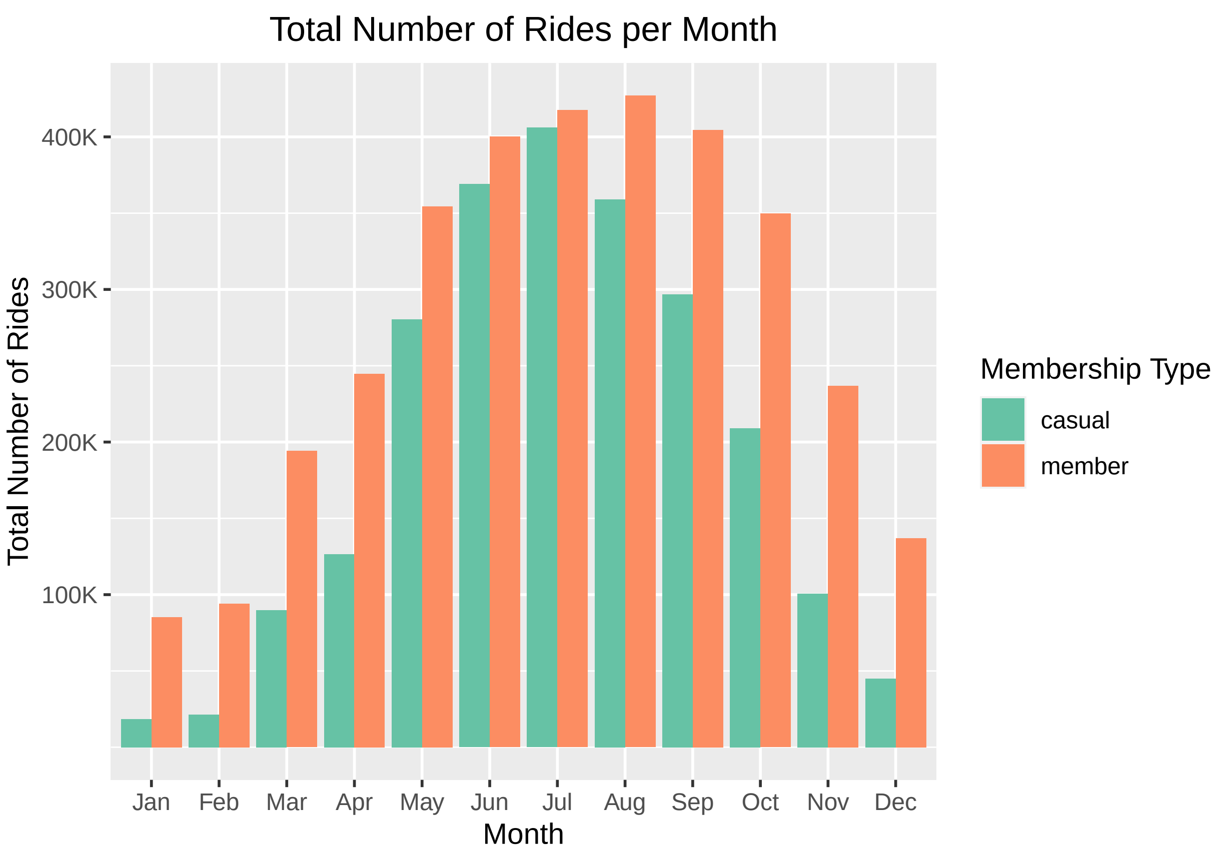
Casual riders bike usage peaks during the weekend while members bike usage peaks during the weekday. This indicates that members use the bike-share service more for commutes while casual riders use the bike-share service more for leisurely purposes on the weekends.

```r
# bar chart of average duration of members/casuals
df_2022 %>%
  mutate(weekday=wday(started_at,label=TRUE)) %>%
  group_by(member_casual,weekday) %>%
  summarise(number_of_rides=n(),average_duration=mean(ride_length)) %>%
  arrange(member_casual,weekday) %>%
  ggplot(aes(x=weekday,y=average_duration,fill=member_casual)) +
  geom_col(position='dodge') +
  labs(x='Day of Week',y='Average Duration (Min)',
       title='Average Ride Duration per Day of Week',fill='Membership Type') +
  theme(plot.title=element_text(hjust=0.5)) + scale_fill_brewer(palette='Set2')
```



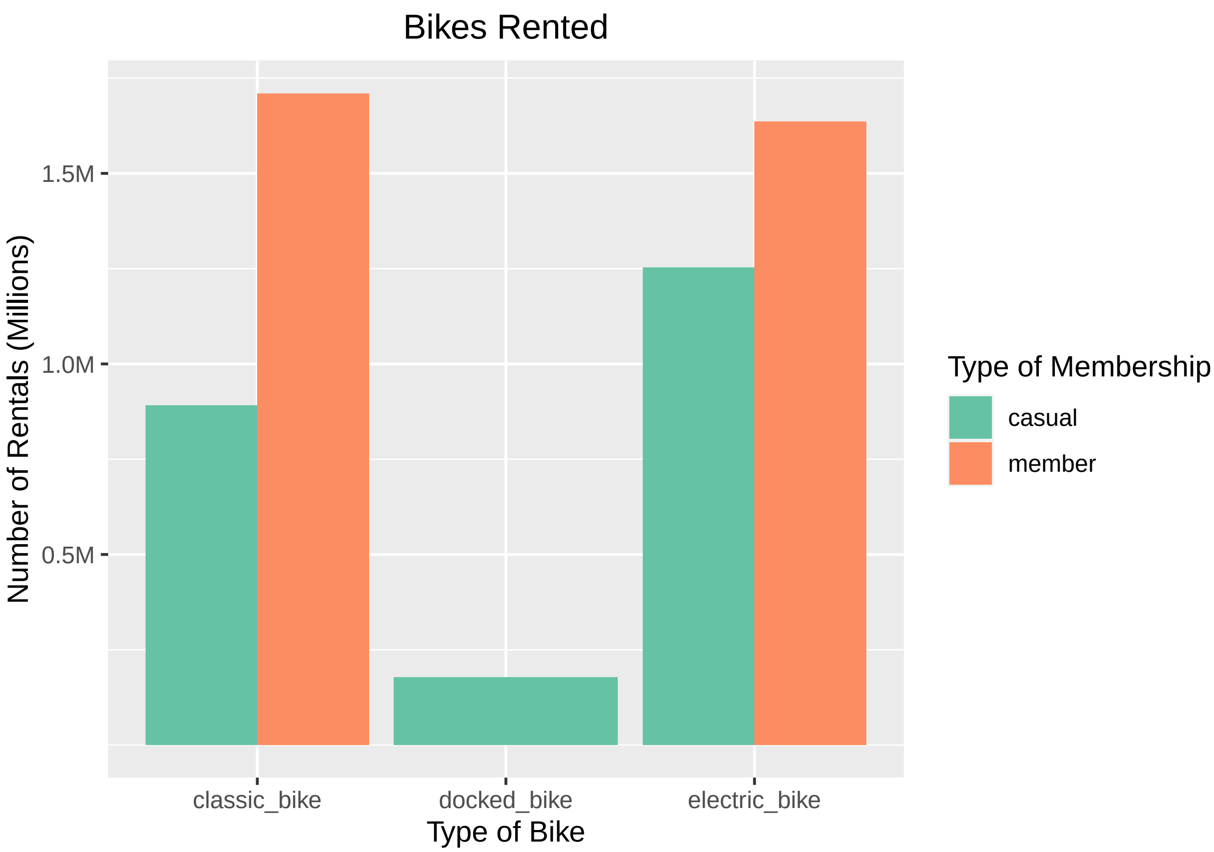
Overall, casual riders use the bike-share service fewer times but for longer rides while members use the bike-share service more often for shorter rides.

```r
# bar chart of number of rides per month
df_2022 %>%
  group_by(member_casual,month) %>%
  summarise(number_of_rides=n()) %>%
  arrange(member_casual) %>%
  ggplot(aes(x=month,y=number_of_rides,fill=member_casual)) +
  geom_col(position='dodge') +
  labs(x='Month',y='Total Number of Rides',title='Total Number of Rides per Month',
       fill='Membership Type') +
  theme(plot.title=element_text(hjust=0.5)) + scale_fill_brewer(palette='Set2') +
  scale_x_discrete(labels=c('Jan','Feb','Mar','Apr','May','Jun',
                            'Jul','Aug','Sep','Oct','Nov','Dec')) +
  scale_y_continuous(breaks=c(100000,200000,300000,400000),
                     labels=c('100K','200K','300K','400K'))
```



For both casual riders and members, the number of rides peak during the warmer summer months of Chicago while the number of rides are fewer in the colder winter months of Chicago, with very few casual riders in the winter.

```r
# bar chart of number of rentals by type of bike
df_2022 %>%
  ggplot(aes(x=rideable_type,fill=member_casual)) +
  geom_bar(position='dodge') +
  labs(x='Type of Bike',y='Number of Rentals (Millions)',title='Bikes Rented',
       fill='Type of Membership') +
  theme(plot.title=element_text(hjust=0.5)) + scale_fill_brewer(palette='Set2') +
  scale_y_continuous(breaks=c(500000,1000000,1500000),labels=c('0.5M','1.0M','1.5M'))
```



Most bikes rented are the classic bikes and the electric bikes by both casual riders and members. Members slightly favor the classic bikes while casual riders favor the electric bikes. Members do not ride docked bikes at all while very few casual riders use docked bikes.

Key Findings

- There are more riders who are members than casual riders.
- Members use the bikes throughout the week but slightly more during the weekdays for their commutes but for shorter rides than casual riders.
- Casual riders use the bikes for longer rides and use it more during the weekends than the weekdays.
- The bikes overall are used more during the warmer months compared to the colder months with very few casual riders using the bikes during the winter months.
- Members use the classic bikes and electric bikes, slightly favoring the classic bike and do not use the docked bikes at all. Casual riders favor the electric bikes over the classic bikes with very few using the docked bikes.

Recommendations

- Introduce monthly or seasonal membership options for casual riders for the warmer months but more costly per month than an annual membership.
- Raise prices on single-ride passes or full-day passes. This will push casual riders who use the bike-share service often to become annual members.