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

Max Dotson

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.

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.

Business Task

Design marketing strategies aimed at converting casual riders into annual members.

Prepare

Cyclistic's historical trip data (September 2021 - September 2022) has been utilized for business findings, and it is organized in 13 .CSV files. The data was made available by Motivate International Inc. under this license.)

Process

- R programming is utilized for the processing and analysis of data due to its capability to handle large data sets, and visualization tools.

```r
# Installation and loading of necessary packages
install.packages("tidyverse")
install.packages("lubridate")
install.packages("ggplot2")

library(tidyverse)
library(lubridate)
library(ggplot2)
```

```r
#=====================
# STEP 1: COLLECT DATA
#=====================

sep_21<-read_csv("202109-divvy-tripdata.csv")
oct_21<-read_csv("202110-divvy-tripdata.csv")
nov_21<-read_csv("202111-divvy-tripdata.csv")
dec_21<-read_csv("202112-divvy-tripdata.csv")
jan_22<-read_csv("202201-divvy-tripdata.csv")
feb_22<-read_csv("202202-divvy-tripdata.csv")
mar_22<-read_csv("202203-divvy-tripdata.csv")
apr_22<-read_csv("202204-divvy-tripdata.csv")
may_22<-read_csv("202205-divvy-tripdata.csv")
jun_22<-read_csv("202206-divvy-tripdata.csv")
jul_22<-read_csv("202207-divvy-tripdata.csv")
aug_22<-read_csv("202208-divvy-tripdata.csv")
sep_22<-read_csv("202209-divvy-publictripdata.csv")
```

```r
#====================================================
# STEP 2: WRANGLE DATA AND COMBINE INTO A SINGLE FILE
#====================================================

sep_21<- mutate(sep_21, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
oct_21<- mutate(oct_21, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
nov_21<- mutate(nov_21, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
dec_21<- mutate(dec_21, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
jan_22<- mutate(jan_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
feb_22<- mutate(feb_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
mar_22<- mutate(mar_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
apr_22<- mutate(apr_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
may_22<- mutate(may_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
jun_22<- mutate(jun_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
jul_22<- mutate(jul_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
aug_22<- mutate(aug_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))
sep_22<- mutate(sep_22, ride_id = as.character(ride_id),rideable_type = as.character(rideable_type))

# All months joined into one data frame
all_trips<- bind_rows(sep_21, oct_21, nov_21, dec_21, jan_22, feb_22, mar_22, apr_22, may_22, jun_22, jul_22, aug_22, sep_22)
```

Step 3: Clean up data and prepare for analysis

- Removal of columns listing latitudes and longitudes

```r
all_trips <- all_trips %>%
  select(-c(start_lat, start_lng, end_lat, end_lng))
```

- Adding columns that list the date, month, day, and year of each ride

```r
all_trips$date <- as.Date(all_trips$started_at)
all_trips$month <- format(as.Date(all_trips$date), "%m")
all_trips$day <- format(as.Date(all_trips$date), "%d")
all_trips$year <- format(as.Date(all_trips$date), "%Y")
all_trips$day_of_week <- format(as.Date(all_trips$date), "%A")
```

- Adding a "ride_length" calculation to all_trips (in seconds)

```r
all_trips$ride_length <- difftime(all_trips$ended_at,all_trips$started_at)
```

- Converting "ride_length" from Factor to numeric so calculations can be run on the data

```r
all_trips$ride_length <- as.numeric(as.character(all_trips$ride_length))
```

- Removing "bad" data
- The dataframe includes a few hundred entries when bikes were taken out of docks and checked for quality by Divvy or ride_length was negative

```r
all_trips_v2 <- all_trips[!(all_trips$start_station_name == "HQ QR" | all_trips$ride_length<0),]
```

Analyze/Share

```r
aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = mean)
```

all_trips_v2$member_casual <chr>	all_trips_v2$ride_length <dbl>
casual	1892.2520
member	779.7345

2 rows

```r
all_trips_v2 %>%
  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)
```

```
`summarise()` has grouped output by 'member_casual'. You can override using the `.groups` argument.
```

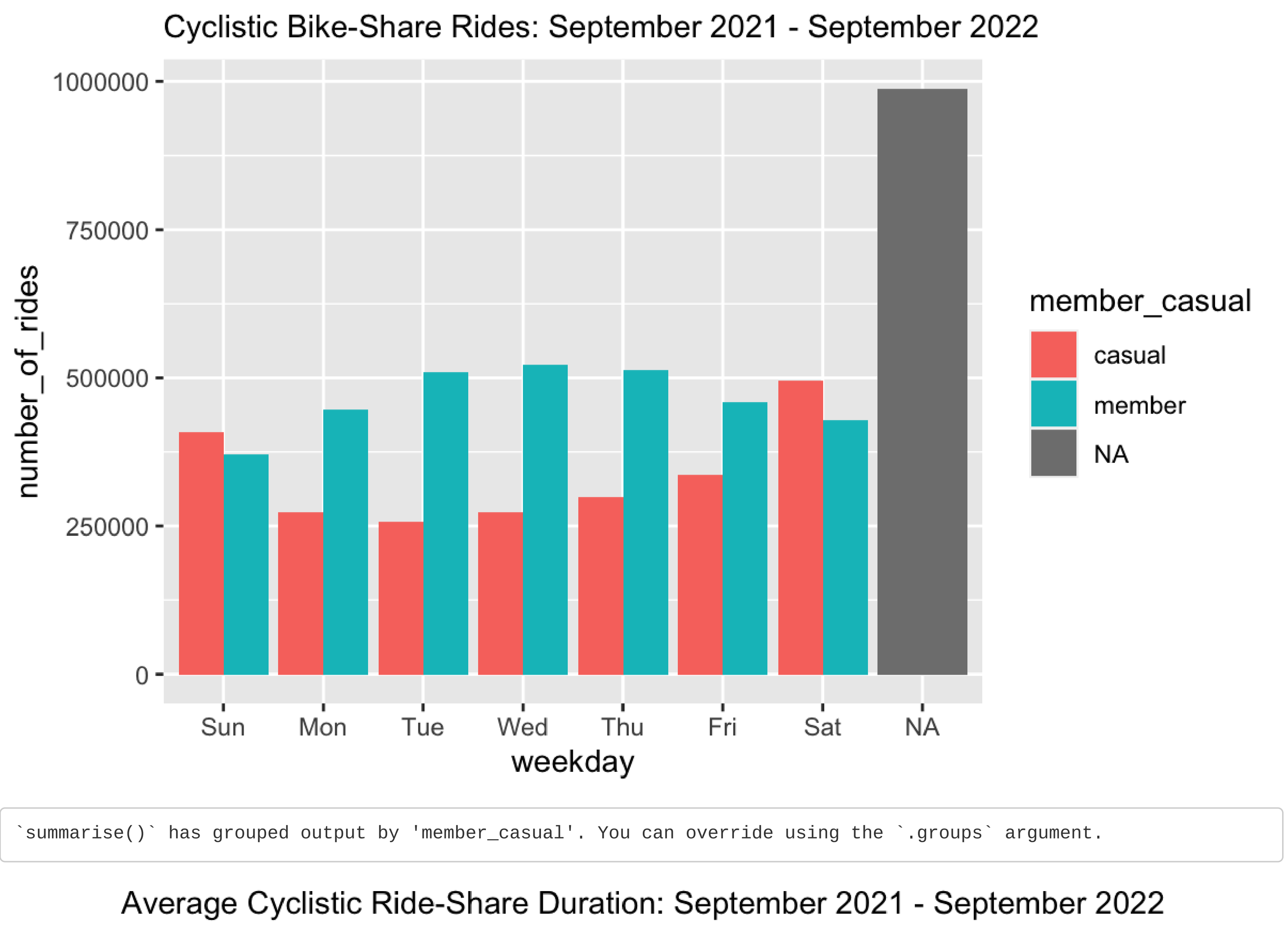
member_casual <chr>	weekday <ord>	number_of_rides <int>	average_duration <dbl>
casual	Sun	408585	2209.3165
casual	Mon	273523	1933.1290
casual	Tue	256927	1655.1013
casual	Wed	273043	1608.9168
casual	Thu	298707	1648.2251
casual	Fri	336911	1811.7737
casual	Sat	495608	2089.1237
member	Sun	371374	876.5319
member	Mon	447086	755.3677
member	Tue	510265	737.0215

1-10 of 15 rows Previous 1 2 Next

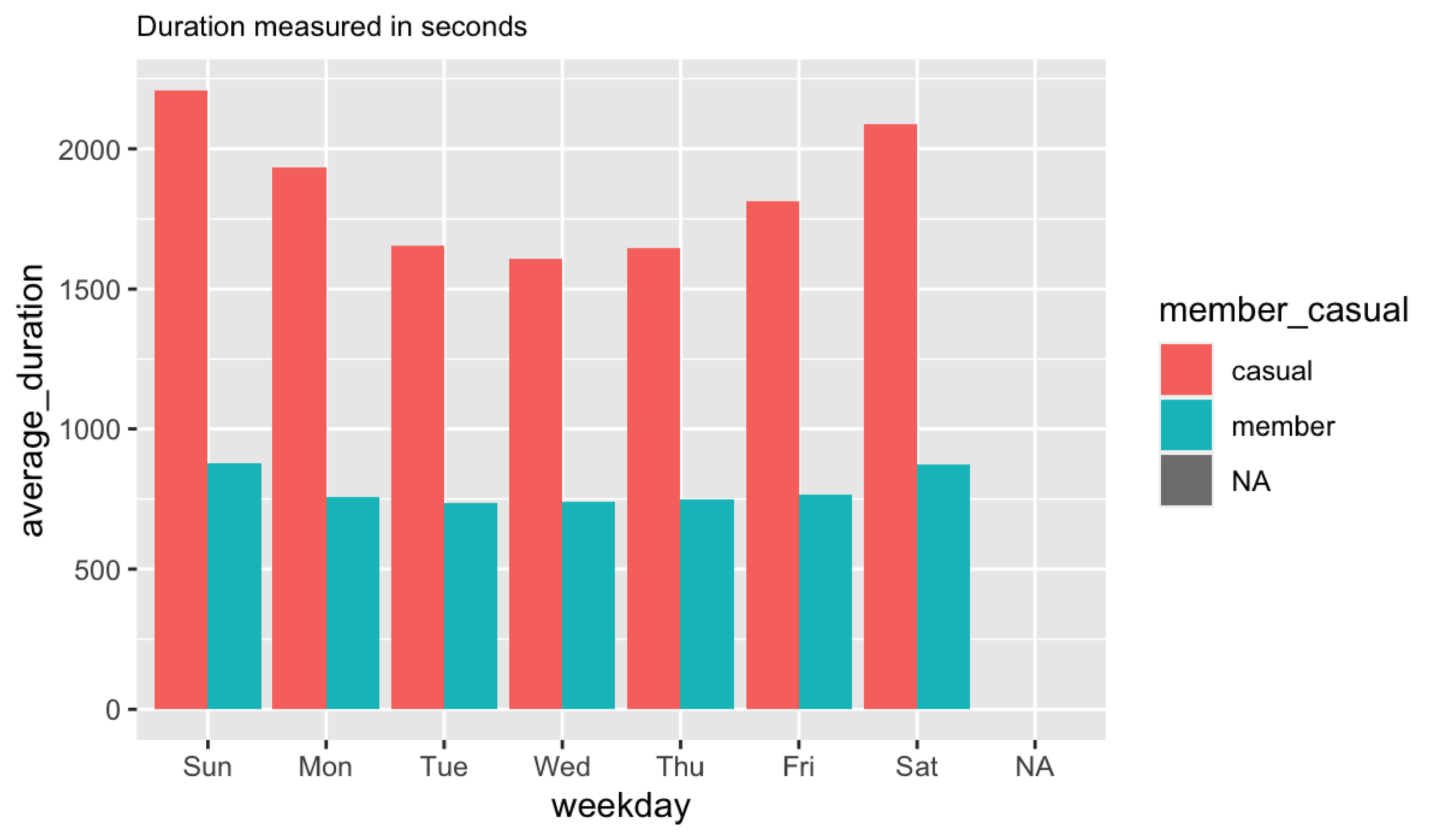
Key Findings:

- Casual riders spend a greater amount of time per ride than those who who are members.
- Saturday and Sunday are the most popular days for casual riders, and also see an increase in average duration per ride.
- Members take a greater amount of trips during the weekdays than they do the weekends, which may indicate that a majority of memberships are for work purposes.

```
`summarise()` has grouped output by 'member_casual'. You can override using the `.groups` argument.
```




```
`summarise()` has grouped output by 'member_casual'. You can override using the `.groups` argument.
```



Act

Top 3 recommendations:

- Implement a weekend focused membership tier that offers discounts to casual riders who switch to members
- Allow money off a membership each month for those who average approx. 1700 (28.3 minutes) seconds of travel per day in a week.
- Offer a chance for casual riders to ride new bikes on weekends, if they switch to a yearly, monthly, or weekend focused membership.