BIKE-SHARE

Ademilola Araoye

2022-08-25

CASE STUDY: How does bike share navigate speedy sucess?

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of bicycles that are geotracked and locked into a network of 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.

This analysis will focus on how member and casual riders use bike differently. And how it can assist the marketing team achieve their goal of turning casual riders into memebers.
in this analysis, i will be using the six analytical phases; **ASK**, **PREPARE**, **PROCESS**, **ANALYZE**, **SHARE**, **ACT**.

ASK Three questions will guide the future marketing program: 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?

PREPARE Prepare: Collect the data, identify how it's organized, determine the credibility of the data.

For this project, I will use the public data of Cyclistic's historical trip data to analyze and identify trends. The data has been made available by Motivate International Inc. under the license.

in this phase, i'll be using R studio to review the several structure of the data. hence, to get a glimpse of the data and also to clean the data to get it ready for analyses . each month contain one csv file and i'll be using 2 month for the analyses.information of the bike rides contain details of the ride_id, rideable_type, start_time and end_time, start and end_ station, latitude and longitude of the start and end stations.

The dataset is reliable, original, and cited.

PROCESS R STUDIO will be used to process the analysis. after getting a glimpse of the data.frame the next stage is to clean the data, since the data structure as been identified. first, we install the necessary packages necessary for cleaning with the install.packages() and load them with library().

```
install.packages("readxl")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```r
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```r
install.packages("writexl")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```r
library(readxl)
library(writexl)
library(janitor)
```

```
##
## Attaching package: 'janitor'

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

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(lubridate)
```

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

```r
library(dplyr)
library(dplyr)
```

```r
df1 <- read_csv("202201-divvy-tripdata.csv",col_names = TRUE)
```

importing data set and assigning it to a variable to make the data structured easy to bind.

```
## Rows: 103770 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df2 <- read_csv("202202-divvy-tripdata.csv",col_names = TRUE)
```

```
## Rows: 115609 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): ride_id, rideable_type, start_station_name, start_station_id, end_...
## dbl  (4): start_lat, start_lng, end_lat, end_lng
## dttm (2): started_at, ended_at
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#since the data has similar structure, then the rbind function can be used to merge the data frames tog
trip_data <-rbind(df1,df2)


##cleaning the data set by removing na field after merging
clean_trip_data <- trip_data %>%
filter(trip_data$started_at<trip_data$ended_at) %>%
drop_na() %>%
clean_names() %>%
remove_empty(c("rows","cols"))
```

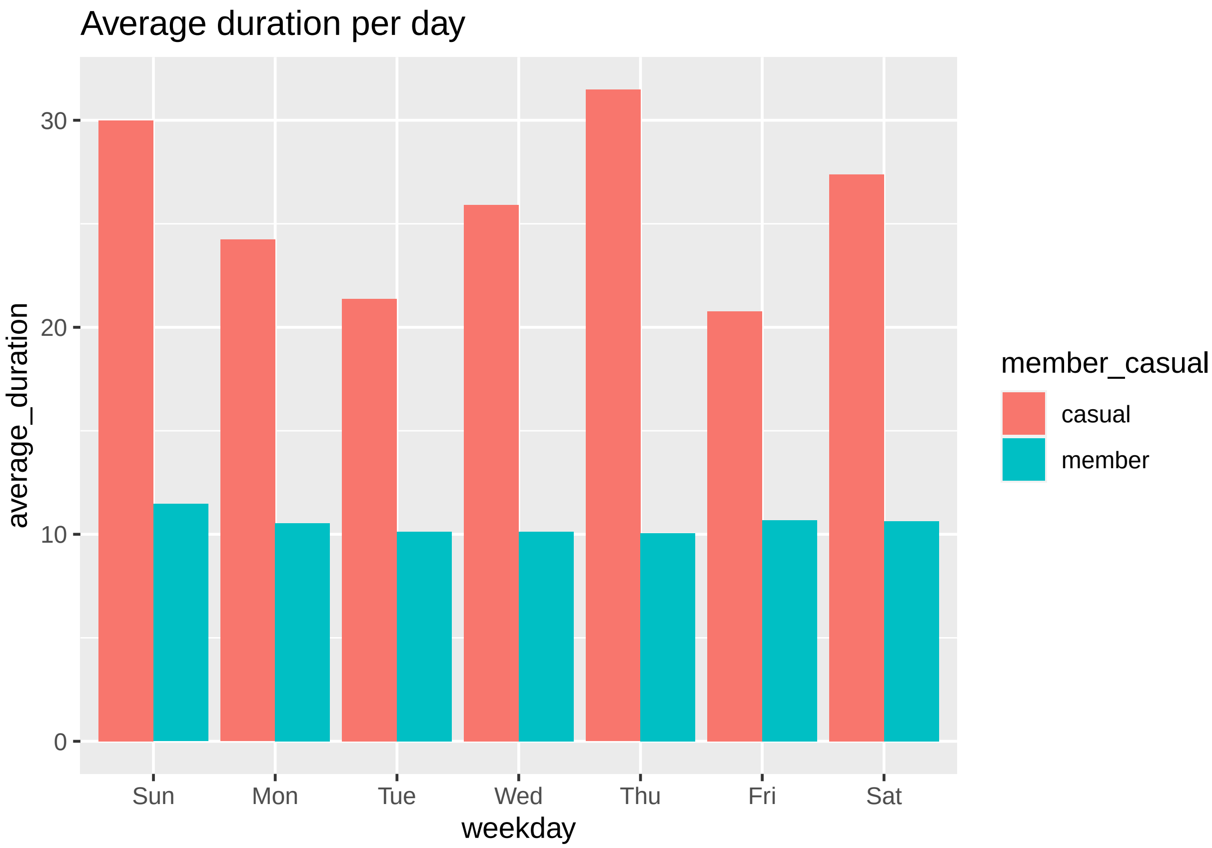
ANALYSIS in this stage, i ran codes to identify trends and display format of the data, i also perform some statistical analysis to answer the question to solve the above problem.

```
##calculating the the differences between the starting_at and ending_at and attributing it t  variable

trip_time <- clean_trip_data %>%  mutate(month = month(started_at, label = FALSE),
weekday = wday(started_at, label = TRUE),
year = year(started_at),
hour = hour(started_at),
ride_length = as.numeric(difftime(ended_at,started_at,units = "mins")))

mean_trip <-trip_time %>%
group_by(rideable_type) %>%
summarize(mean_trip = mean(ride_length))

member_casual_group <- trip_time %>%
group_by(member_casual) %>%
summarize(mean_group = mean(ride_length), mode_group = mode(ride_length), min_group = min(ride_length),
```

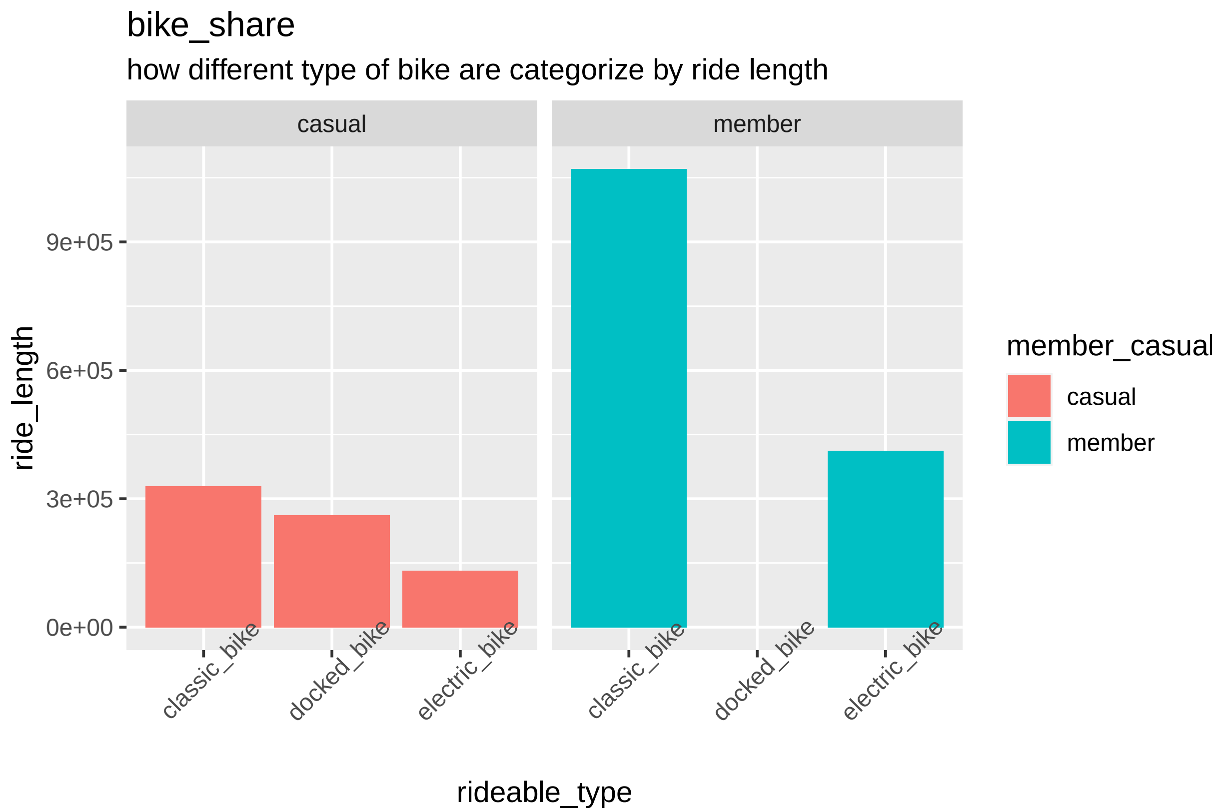
SHARE #####analysing data with visual

```
##analysing by type and week
##table 1
trip_time %>%
  group_by(member_casual, weekday) %>%
  summarise(average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday)  %>%
  ggplot(aes(x = weekday, y = average_duration, fill = member_casual)) +
  geom_col(position = "dodge")+
  labs(title = "Average duration per day")
```

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



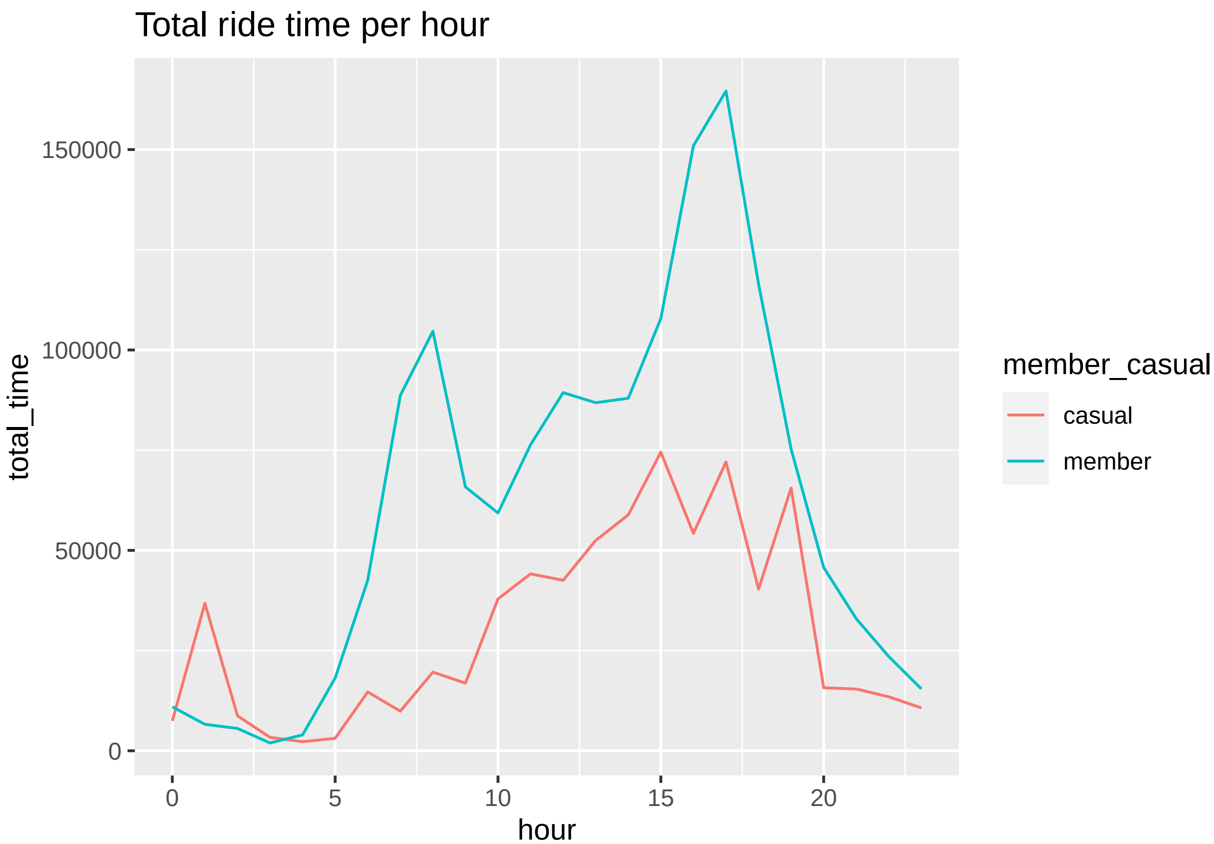

```
##table 2
ggplot(data=trip_time) +
  geom_col(mapping = aes(x=rideable_type, y=ride_length, fill = member_casual))+ facet_wrap(~member_casu
  theme(axis.text.x = element_text(angle=45))+
  labs(title = "bike_share", subtitle ="how different type of bike are categorize by ride length", capt
```



how member and casual cyclist ride bikes differently

```
##numbers of riders and the total ride time between members and casuals on different hours in a day
##table 3
trip_time%>%
  group_by(member_casual, hour) %>%
  summarise(total_time = sum(ride_length)) %>%
  ggplot(aes(hour, total_time, color = member_casual)) +
  geom_line() +
  scale_y_continuous() +
  labs(title = "Total ride time per hour")
```

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



ACT

Recommendation below are some of the recommentaion i suggest from the analyses above.

1.it can be seen from table 2 that classic bike commands demand traffic among members. it also shows a higher demand than others types. i recommend an improved classic bike restricted to only members.

2. limited bike on thursdays and weekends would push casual to become member