Cycling Data Analysis

Elena

2021/8/30

Introduction

This is a analysis on the cycling data of two types of riders of the company Cyclistic in the past twelve months, from August 2020 to July 2021, in order to provide advice on digital marketing campaign aiming at converting casual riders to annual members.

set up the environment

```
# install.packages("RODBC")
# install.packages("ggplot2")
# install.packages("tidyverse")
# install.packages("rmarkdown")
#
# library(dplyr)
# library(tidyr)
# library(RODBC)
# library(ggplot2)
# library(scales)
# library(rmarkdown)
```

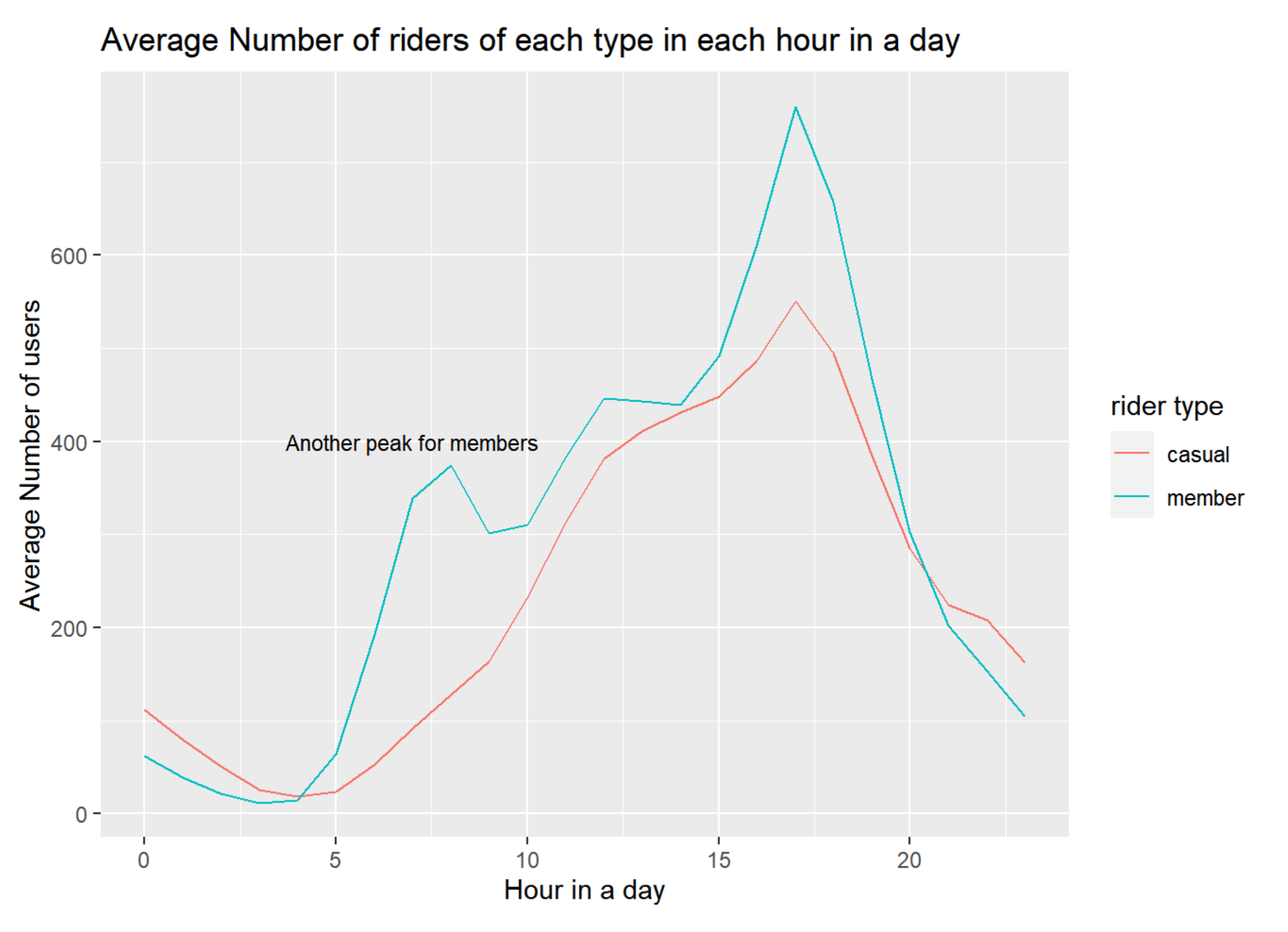
Connect R to MSSQL

```
dbcon <- odbcConnect("EC_SQL", rows_at_time = 1 )

if(dbcon == -1){
  quit("no", 1)
}
```

Average number of riders in each hour of a day

Analyze the number of riders in a day to explore the time difference of how each type of riders use the bikes.

```
sql <- "
SELECT *
FROM CyclingData..[RidePerHour]
"
TotalTrip <- sqlQuery(dbcon, sql)

TotalTrip <- TotalTrip %>% arrange(START_HOUR)

plot01 <- ggplot(TotalTrip, aes(x = START_HOUR, y = AVG_DAY, group = member_casual, color = member_casual)) +
  geom_line() +
  labs(x = "Hour in a day", y = "Average Number of users",
       title = "Average Number of riders of each type in each hour in a day", color = "rider type")

plot01 + annotate("text", x = 7, y = 400, label = "Another peak for members", size = 3)
```



We have two points to pay attention to from the graph above.

First , there are two peaks of usage in a day for members. One is at around 7-9 AM in the morning and the other one is at around 5 - 6 PM in the evening. While there is only one peak for casual riders, which is at 4 - 6 PM.

Second, there are more casual riders than members before 4 AM and after 9 PM. Most of the members use the bikes at between 4 AM to 9 PM, which is the day time.

Compare the average riding duration and the mode of duration the each type of riders

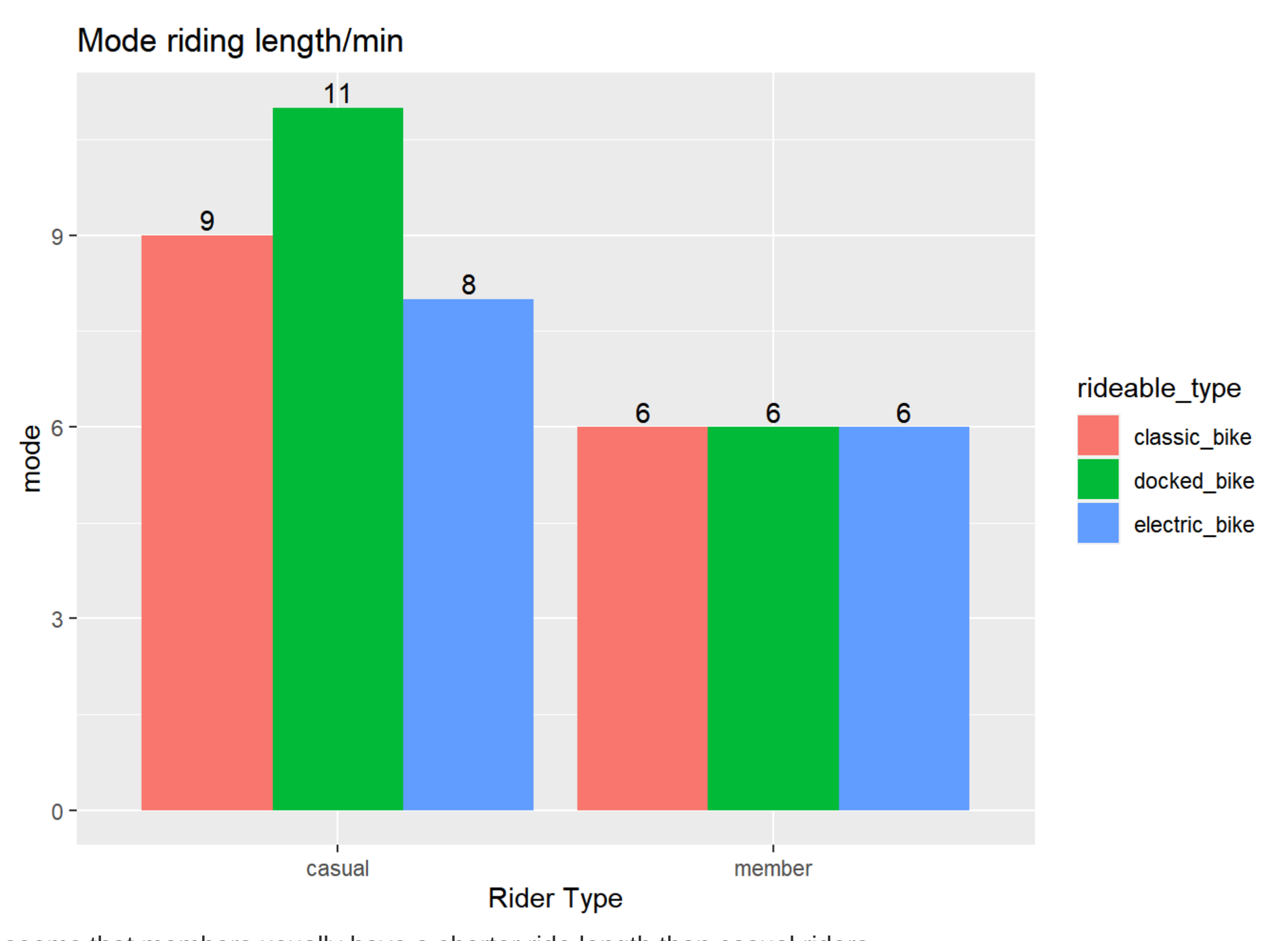
```
sql02 <- "
SELECT *
FROM CyclingData..Summary
"
Ridelength_summary <- sqlQuery(dbcon, sql02)
```

Average

```
mean_plot <- ggplot(Ridelength_summary, aes(x= member_casual, y = mean , fill = rideable_type)) +
  geom_bar(position = "dodge",stat =  'identity') +
  geom_text(aes(label = mean),position=position_dodge(width=0.9),vjust=-0.25) +
  labs(title = "Average riding length/min", x = "Rider Type")

mean_plot
```

Mode

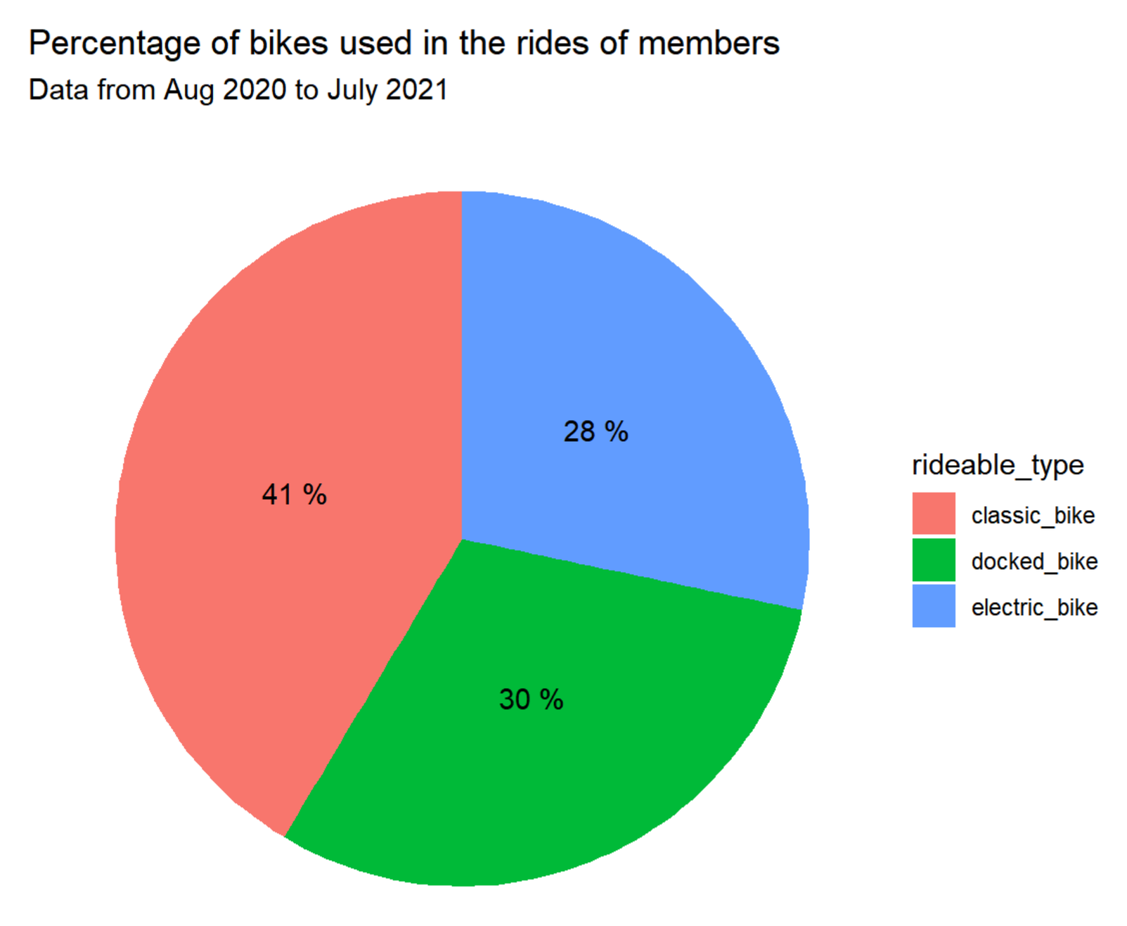
```
mode_plot <- ggplot(Ridelength_summary, aes(x= member_casual, y = mode , fill = rideable_type)) +
  geom_bar(position = "dodge",stat =  'identity') +
  geom_text(aes(label = mode), position = position_dodge(width=0.9), vjust=-0.25) +
  labs(title = "Mode riding length/min", x = "Rider Type")

mode_plot
```



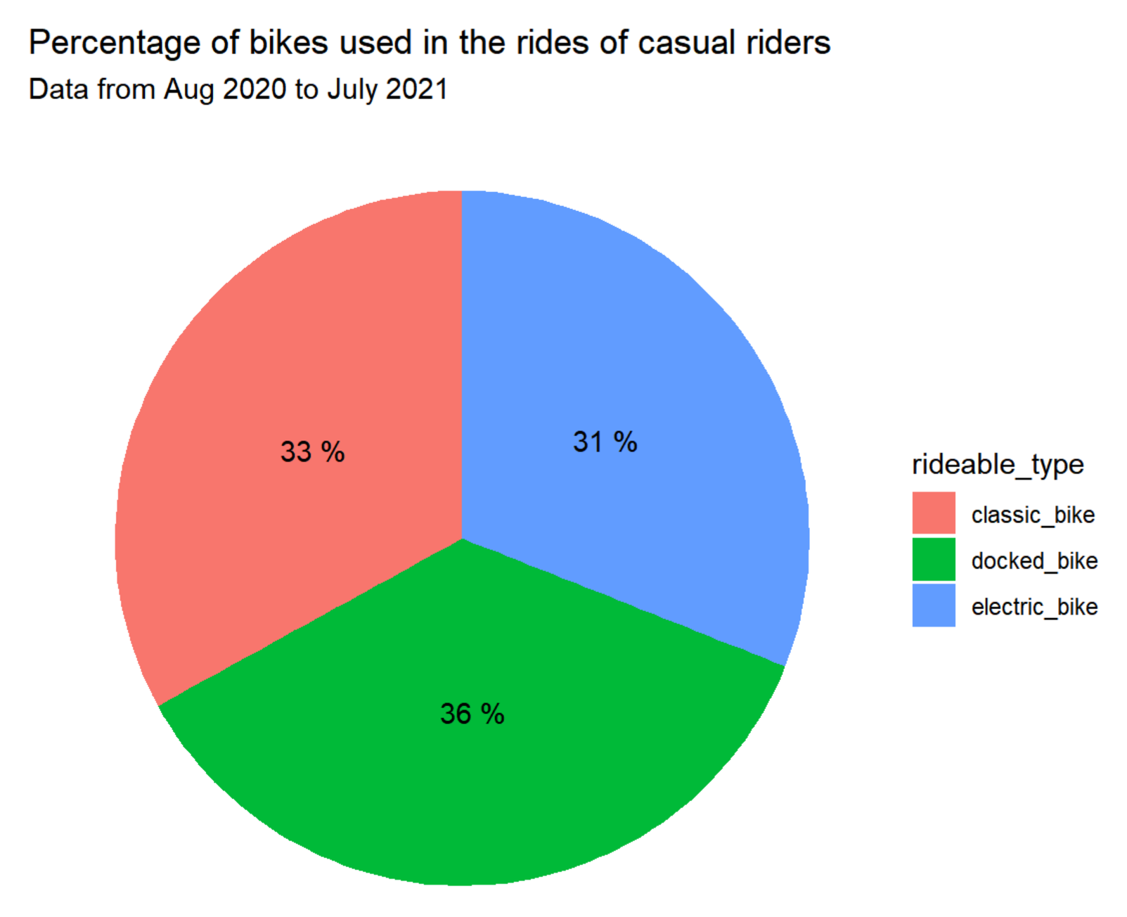
No matter which kind of bicycle, it seems that members usually have a shorter ride length than casual riders.

Compare the percentage of each type of bikes used by two types of riders

```
sql04 <- "
SELECT *
FROM CyclingData..bikeTypeNum"

bikeTypeNum <- sqlQuery(dbcon, sql04)

bikeTypeNum_mem <- bikeTypeNum %>%
  filter(member_casual == "member") %>%
  mutate (percentage = round(bikeTypeNum/sum(bikeTypeNum),digits = 2)*100)


pie_member <-
  ggplot(bikeTypeNum_mem, aes(x = '', y = percentage, fill = rideable_type)) +
  geom_bar(stat = "identity") +
  coord_polar("y", start = 0) + theme_void() +
  geom_text(aes(label = paste(percentage, "%")),
            position = position_stack(vjust = 0.5)) +
  labs(title = "Percentage of bikes used in the rides of members", subtitle = "Data from Aug 2020 to July 2021")


pie_member
```




```
bikeTypeNum_cas <- bikeTypeNum %>%
  filter(member_casual == "casual") %>%
  mutate (percentage = round(bikeTypeNum/sum(bikeTypeNum),digits = 2)*100)

pie_casual <-
  ggplot(bikeTypeNum_cas, aes(x = '', y = percentage, fill = rideable_type)) +
  geom_bar(stat = "identity") +
  coord_polar("y", start = 0) + theme_void() +
  geom_text(aes(label = paste(percentage, "%")),
            position = position_stack(vjust = 0.5)) +
  labs(title = "Percentage of bikes used in the rides of casual riders", subtitle = "Data from Aug 2020 to July 2
021")

pie_casual
```



From the two pie charts above, it's not difficult to notice that there's no preference for a particular type of bicycle among the casual riders. The three types of bikes are equally used in the past twelve months. While for our members, classic bikes are significantly used more than the docked and electric bikes. They account for 40% of the rides alone in the past twelve months.

Comparison of number of each type of riders on each day of a week

```
sql03 <- "
SELECT *
FROM CyclingData..weekday_table
"
weekday_info <- sqlQuery(dbcon, sql03)


weekday_plot <-
  weekday_info %>% arrange(member_casual, desc(count)) %>%
  ggplot(aes(x =weekday_star, y = count, fill = member_casual)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_x_continuous("weekday", breaks = seq(1,7)) + scale_y_continuous(labels= scales :: comma_format(big.marks
 = ",")) +
  labs(title = "The number of riders on the day of a week (Sunday = 1)")


weekday_plot
```

On Sunday and Saturday, there are more casual riders using the bikes than members. While at the weekdays, there are much more members using the bikes than casual riders in total.

Summary

Difference in using the bikes between casual riders and members

```
Member_types <- c("Casual riders",  "Members")
Peak_hours <- c("4-6 PM",   "7-9 AM & 5 – 6 PM")
Ride_Length <- c("From around 20 to 60 minutes on average", "From  around 10 to 15 minutes on average")
Bike_type_preference <- c("None",   "Classic bikes")
Most_active_days_in_a_week <- c("At the weekends","At weekdays")

df <- data.frame(Member_types, Peak_hours, Ride_Length, Bike_type_preference, Most_active_days_in_a_week)

as_tibble(df)
```

```
## # A tibble: 2 x 5
##   Member_types  Peak_hours        Ride_Length                              Bike_type_preference Most_active_da
ys_in_a_w~
##   <chr>         <chr>             <chr>                                    <chr>                <chr>
## 1 Casual riders 4-6 PM            From around 20 to 60 minutes on average  None                 At the weekend
s
## 2 Members       7-9 AM & 5 – 6 PM From  around 10 to 15 minutes on average Classic bikes        At weekdays
```

- Members are mostly people in the working class who use the bikes when they go to work and get off from work from Monday to Friday. They ride the bikes from the office to the bus stop, the metro station, or their home nearby and so on, and the other way around, so the riding length of the members is usually quite short, from 10 to 15 minutes.
- While casual riders are the people who use the bikes for travelling, having some leisure time at the weekend, so their riding length is usually much longer than the members using the bikes for commuting.

Recommendations on the marketing program

- Provide discounts for rides lasting for a longer time for annual members since casual riders usually have a longer ride.
- Perks or discounts for riding at the weekend for annual members since casual riders are most active at the weekends.
- Marketing channel - digital platforms for city sightseeing, leisure, health, etc.

Further actions

- More research about the influence of price in converting casual riders into annual riders.
- Dive deeper into the purposes of casual riders using shared bikes.