Cyclist Project

Ibrahim Sherif Kamal

9/21/2021

The scenario

In this 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.

Business Task

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

Packages

```
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
```

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

```
library("ggplot2")
library("scales")
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor

library("janitor")

##
## Attaching package: 'janitor'

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

library("readr")
```

Importing the Data

the data of last 12 month

```
csv1 <- read.csv("./csv/202004-divvy-tripdata.csv")
csv2 <- read.csv("./csv/202005-divvy-tripdata.csv")
csv3 <- read.csv("./csv/202006-divvy-tripdata.csv")
csv4 <- read.csv("./csv/202007-divvy-tripdata.csv")
csv5 <- read.csv("./csv/202008-divvy-tripdata.csv")
csv6 <- read.csv("./csv/202009-divvy-tripdata.csv")
csv7 <- read.csv("./csv/202010-divvy-tripdata.csv")
csv8 <- read.csv("./csv/202011-divvy-tripdata.csv")
csv9 <- read.csv("./csv/202012-divvy-tripdata.csv")
csv10 <- read.csv("./csv/202101-divvy-tripdata.csv")
csv11 <- read.csv("./csv/202102-divvy-tripdata.csv")
csv12 <- read.csv("./csv/202103-divvy-tripdata.csv")
```

Merging multiple csv files

merging them into a single Data frame

```
biker_ride <- rbind(csv1, csv2, csv3, csv4, csv5, csv6, csv7, csv8, csv9,
csv10, csv11, csv12)
```

Rechecking for empty columns or rows

after cleaning the data over Excel

```
dim(biker_ride)

## [1] 3447609      15
```

```
biker_ride <- janitor::remove_empty(biker_ride,which =  c("cols"))

biker_ride <<- janitor::remove_empty(biker_ride, which = c("rows"))

dim(biker_ride)

## [1] 3447609      15
```

Changing weekdays and ride duration data into a date

by using the lubridate function

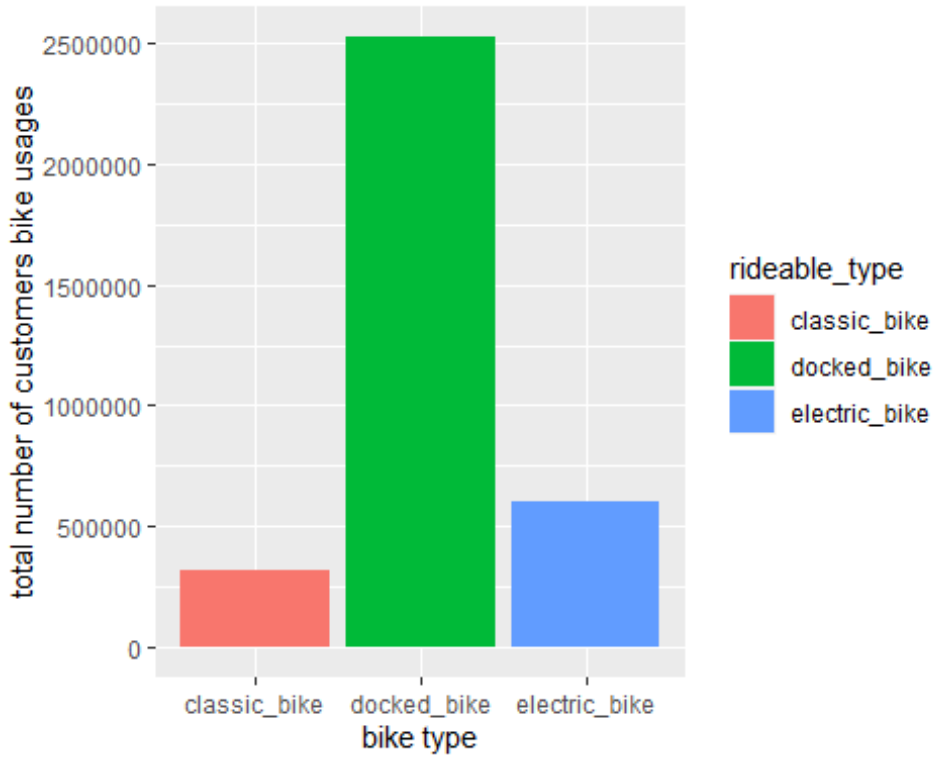
```
biker_ride$day_of_week <- lubridate::wday(biker_ride$day_of_week, label =
TRUE)

biker_ride$ride_length <- lubridate::hms(biker_ride$ride_length)
```

the total number of every bike model

```
total_bike <- biker_ride %>%
  group_by(rideable_type) %>%
  summarize(bike_total_use = n())

print(total_bike)

## # A tibble: 3 x 2
##   rideable_type bike_total_use
##   <chr>                  <int>
## 1 classic_bike          317542
## 2 docked_bike          2528472
## 3 electric_bike         601595
```

*showing the total number of each bike type **visually***

```
ggplot(data = biker_ride)+
  geom_bar(mapping = aes(x = rideable_type, fill =  rideable_type))+
  labs(x = " bike type ", y = "total number of customers bike usages")
```



Total number of bike used by every customer type

```
biker_ride_type <- biker_ride %>%
  group_by(member_casual, rideable_type) %>%
  summarize(member_casual_total_use = n()) %>%
  arrange(rideable_type)

print(biker_ride_type)
```

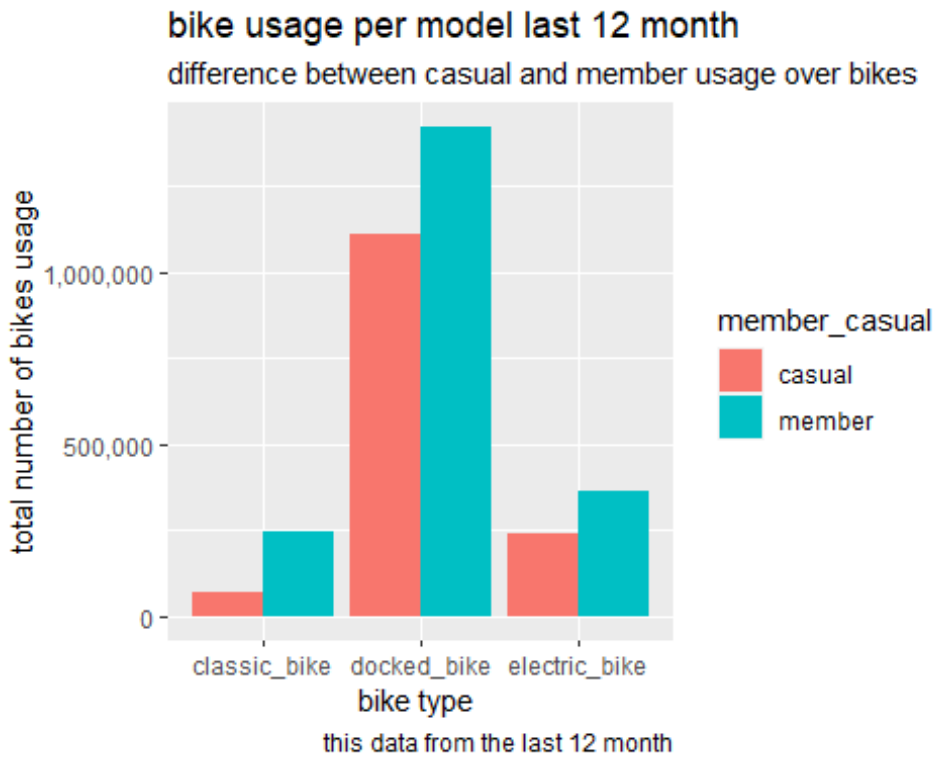
```
## # A tibble: 6 x 3
## # Groups:   member_casual [2]
##   member_casual rideable_type member_casual_total_use
##   <chr>         <chr>                           <int>
## 1 casual        classic_bike                    70474
## 2 member        classic_bike                   247068
## 3 casual        docked_bike                   1108243
## 4 member        docked_bike                   1420229
## 5 casual        electric_bike                  237579
## 6 member        electric_bike                  364016
```

*The difference between every customer type bike use **Visualization***

using ggplot2 library

```
ggplot(data = biker_ride) +
  geom_bar(mapping = aes( x = rideable_type, fill = member_casual), position
= "dodge")+
  labs(title = "bike usage per model last 12 month",subtitle = "difference
```

```
between casual and member usage over bikes", caption = "this data from the
last 12 month")+
  scale_y_continuous(labels = comma)+
  labs(x= "bike type ", y = "total number of bikes usage")
```



The difference between member and casual over days last 12 month

```
bike_ride_week <- biker_ride %>%
  group_by(member_casual, day_of_week) %>%
  summarise( member_vs_casual_per_day = n())
#arrange(day_of_week)

print(bike_ride_week)
```

```
## # A tibble: 14 x 3
## # Groups:   member_casual [2]
##    member_casual day_of_week member_vs_casual_per_day
##    <chr>         <ord>                          <int>
##  1 casual        Sun                           260762
##  2 casual        Mon                           149805
##  3 casual        Tue                           143900
##  4 casual        Wed                           157010
##  5 casual        Thu                           164913
##  6 casual        Fri                           206808
##  7 casual        Sat                           333098
##  8 member        Sun                           262389
##  9 member        Mon                           264703
## 10 member        Tue                           281548
```

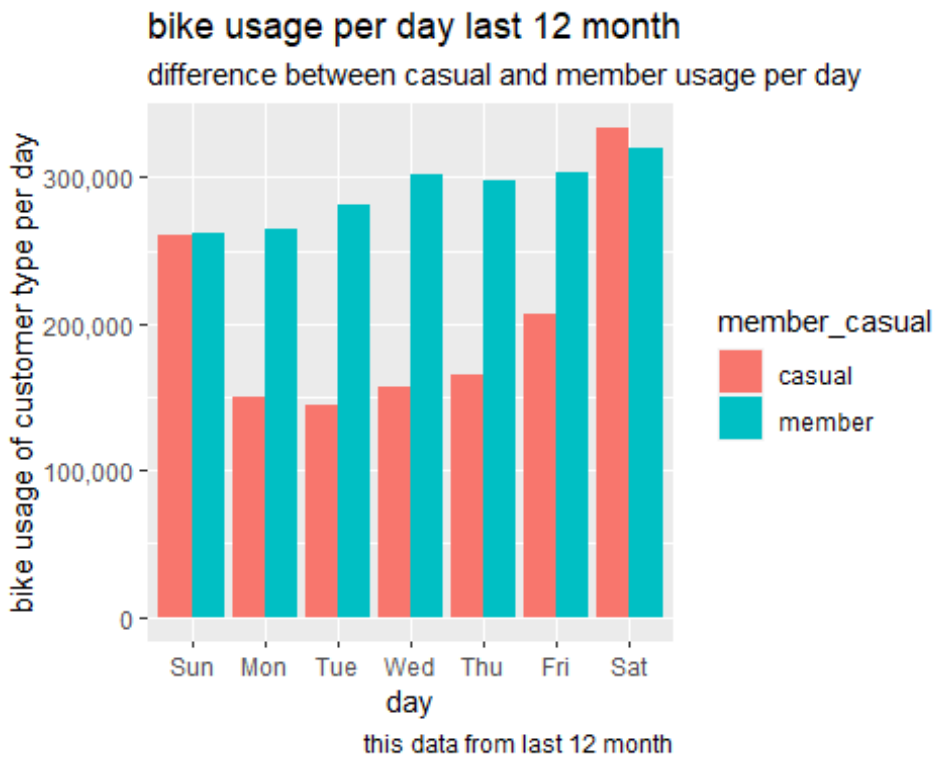
```
## 11 member         Wed                              302126
## 12 member         Thu                              297633
## 13 member         Fri                              303365
## 14 member         Sat                              319549
```

Customer usage over days last 12 month ***Visualization***

customer type:

- Casual
- Member

```
ggplot(data = biker_ride)+
  geom_bar(mapping = aes(x = day_of_week, fill = member_casual), position = "dodge")+
  scale_y_continuous(labels = comma)+
  labs(title = "bike usage per day last 12 month", subtitle = "difference between casual and member usage per day", caption = "this data from last 12 month")+
  labs(x = "day", y = "bike usage of customer type per day")
```



The bike rides duration difference

between casual and member customers

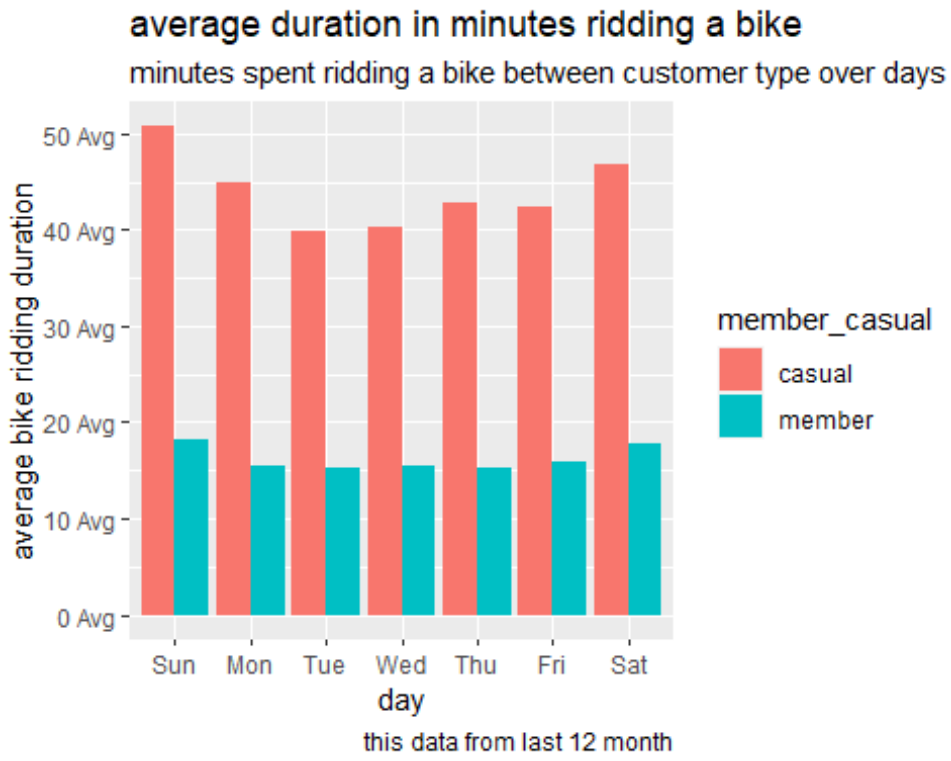
```
biker_ride_length <- biker_ride %>%
  add_column(ride_length_min = as.numeric(biker_ride$ride_length)/ 60)
```

```
biker_ride_duration <- biker_ride_length %>%
group_by(member_casual, day_of_week)  %>%
summarize( max_duration_minutes = max(ride_length_min),
average_duration_minutes = mean(ride_length_min), total_duration_minutes =
sum(ride_length_min)) %>%
arrange(day_of_week)

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

print(biker_ride_duration)

## # A tibble: 14 x 5
## # Groups:   member_casual [2]
##    member_casual day_of_week max_duration_min~ average_duratio~
total_duration_~
##    <chr>         <ord>                  <dbl>            <dbl>
<dbl>
##  1 casual        Sun                    51146             50.7
13217850
##  2 member        Sun                    31169             18.2
4769513
##  3 casual        Mon                    37127             44.9
6729800
##  4 member        Mon                    33421             15.5
4096879
##  5 casual        Tue                    47797             39.9
5734795
##  6 member        Tue                    15105             15.2
4271125
##  7 casual        Wed                    54283             40.4
6339555
##  8 member        Wed                    41271             15.5
4671200
##  9 casual        Thu                    47411             42.8
7055221
## 10 member        Thu                    13920             15.2
4532301
## 11 casual        Fri                    42113             42.3
8745076
## 12 member        Fri                    11897             15.9
4835848
## 13 casual        Sat                    50693             46.9
15622134
## 14 member        Sat                    16504             17.8
5695005
```

bike rides duration difference ***Visualiztion***

```
ggplot(data = biker_ride_duration)+
  geom_col( mapping = aes(x = day_of_week , y = average_duration_minutes ,
fill = member_casual), position = "dodge")+
```

```
  labs(x = "day", y = "average bike ridding duration")+
  labs(title = "average duration in minutes ridding a bike ", subtitle =
"minutes spent ridding a bike between customer type over days", caption =
"this data from last 12 month")+
  scale_y_continuous(labels = label_number( suffix = " Avg"))
```



Key Findings

- The docked bike option is far more popular than both classic bikes and electric bikes
- Annual members rent more bike models overall
- Weekend are the most days casual members rent bikes
- Sunday bike rent is almost equal between casual and members
- Regardless of being a member or not, the most popular day to rent a bike is Saturday
- Saturday is the only day of the week last 12 month that casual users bikes rent exceeds annual members
- casual users spend more time riding a bike, every day of the week last 12 month

Conclusion

Three recommendations based on my analysis

- Offer a **weekend membership** at a different price point than the full annual membership to entice casual users into become members.
- increase number of **electric bikes**, to encourage casual users into becoming members because they tend to spend more time riding bikes.
- create **annual-membership** promotion for customers that spent long time riding bikes, to make casual users realize over their subconscious that they spend long time riding bikes so casual users feel tempted into becoming annual members