Membership Capstone Project

Sevket Asil Kisa

2022-08-24

1. Background

The director of marketing at Cyclistic, a bike-share company in Chicago believes the company's future success depends on maximizing the number of annual memberships. Therefore, the team wants to understand how casual riders and annual members use Cyclistic bikes differently. The team needs to design a new marketing strategy to convert casual riders into annual members.

2. Project Purpose

Understanding how the difference in usage of the two types of users, which are casual riders and annual members, can be leveraged to make conversion from casual to member possible. Support the Cyclistic executives and decision making process by creating a market strategy to convert casual users into members through comprehending why casual riders would buy a membership, and how digital media could affect their marketing tactics.

3. 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. 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 (marketing manager at Cyclistic) 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.

4. Dataset

The dataset used in this case study was obtained from the website (https://rpubs.com/hello-jiarong/874299) and will be covering the period between November 2021 and April 2022 (Note that some of the trips may end in May 2022. (Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this license (https://ride.divvybikes.com/data-license-agreement)). This is public data that you can use to explore how different customer types are using Cyclistic bikes.

5. Steps

- Downloaded the dataset for the period between November 2021 to April 2022 in .csv format.
- Upload all six months' data to BigQuery SQL servers.
- Combined all of the tables and do the necessary cleaning.
- Downloaded as a .csv format and uploaded the combined data into RStudio.
- Analyses done in both RStudio and Tableau is shown below.

6. Analysis (with their steps)

6.1. Packages installed and loaded

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✔ ggplot2 3.3.6     ✔ purrr   0.3.4
## ✔ tibble  3.1.8     ✔ dplyr   1.0.9
## ✔ tidyr   1.2.0     ✔ stringr 1.4.0
## ✔ readr   2.1.2     ✔ forcats 0.5.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(readr)
library(dplyr)
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(lubridate)
```

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

```
library(geosphere)
library(knitr)
```

6.2. Uploaded the data and assigned it to a name

```
tripdata <- read_csv("~/Downloads/bq-results-20220812-074619-1660290494422.csv")
```

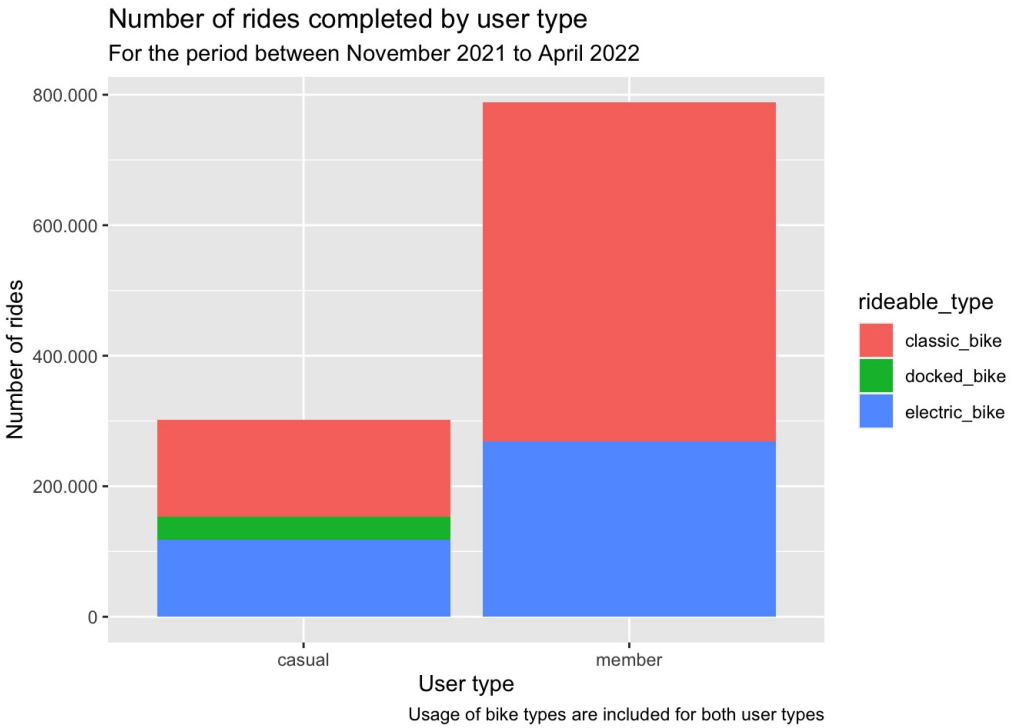
```
## Rows: 1090124 Columns: 13
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (9): ride_id, rideable_type, started_at, ended_at, start_station_name, s...
## dbl (4): start_lat, start_lng, end_lat, end_lng
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
tripdata %>% summarize(first_trip=min(started_at), last_trip=max(ended_at))
```

```
## # A tibble: 1 × 2
##   first_trip              last_trip
##   <chr>                   <chr>
## 1 2021-11-01 00:00:14 UTC 2022-05-01 16:16:59 UTC
```

6.3. Created the first graph for the comparison of usages between members and casuals

```
ggplot(tripdata) + geom_bar(mapping = aes(x=member_casual, fill=rideable_type)) +
  scale_y_continuous(labels = comma_format(big.mark = ".", decimal.mark = ",")) +
  labs(title = "Number of rides completed by user type",
       subtitle = "For the period between November 2021 to April 2022",
       caption = "Usage of bike types are included for both user types",
       x = "User type",
       y = "Number of rides")
```



Members made more than double the rides than casual users, as they have free usage of up to 45 minutes rides covered in yearly payment.

6.4. Added four new columns for usage times, distances traveled in each trip, day of the week, and time that trip started at

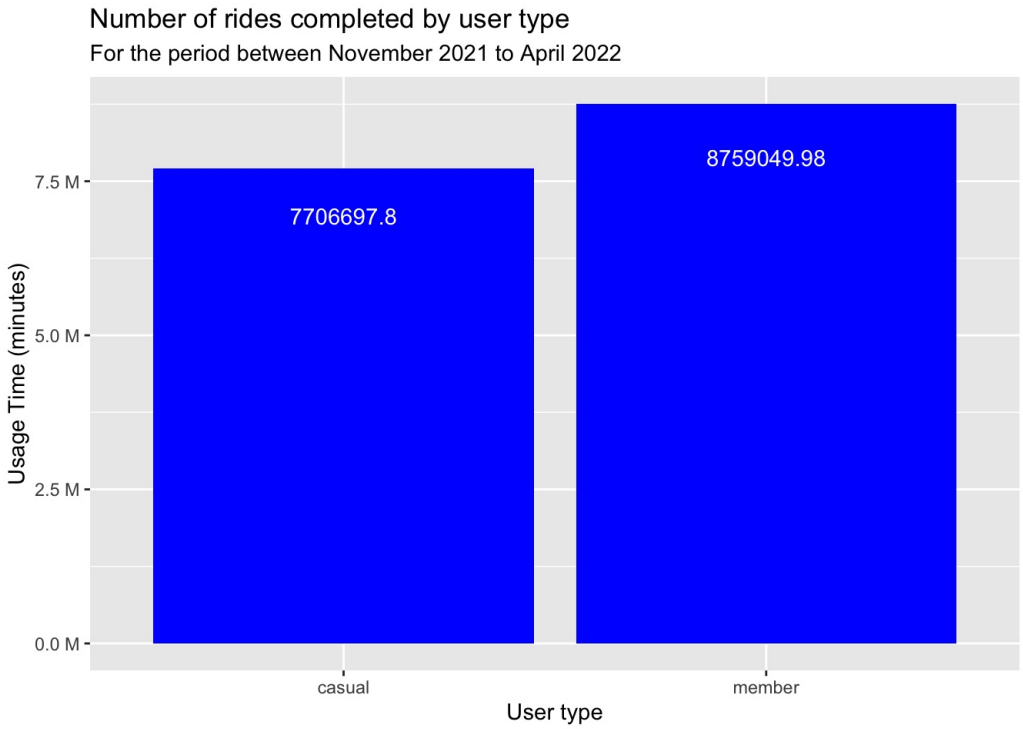
```
tripdata_new <- mutate(tripdata, usage_time = difftime(ended_at, started_at, units = "mins")) %>%
  mutate(tripdata, distance_km = distHaversine(cbind(start_lng, start_lat), cbind(end_lng, end_lat))*0.001) %>%
  mutate(tripdata, day = wday(tripdata$started_at, label = TRUE, abbr = FALSE)) %>%
  mutate(tripdata, time = format(as.POSIXct(started_at), format = "%H"))
head(tripdata_new)
```

```
## # A tibble: 6 × 17
##   ride_id        ridea…¹ start…² ended…³ start…⁴ start…⁵ end_s…⁶ end_s…⁷ start…⁸
##   <chr>          <chr>   <chr>   <chr>   <chr>   <chr>   <chr>   <chr>     <dbl>
## 1 465D19A186F3E… electr… 2022-0… 2022-0… Montic… 301     Conser… 518        41.9
## 2 1B0C2AA538C5D… electr… 2022-0… 2022-0… Narrag… 309     Narrag… 309        41.9
## 3 EDA510CE23B73… electr… 2022-0… 2022-0… North … 310     Austin… 16921      41.9
## 4 800C44F1CA934… electr… 2022-0… 2022-0… Lockwo… 312     Kilpat… 358        41.9
## 5 8DB2013FE94BA… electr… 2022-0… 2022-0… Lockwo… 312     Kilpat… 358        41.9
## 6 970E0D0AFEA11… electr… 2022-0… 2022-0… Lockwo… 312     Kilpat… 358        41.9
## # … with 8 more variables: start_lng <dbl>, end_lat <dbl>, end_lng <dbl>,
## #   member_casual <chr>, usage_time <drtn>, distance_km <dbl>, day <ord>,
## #   time <chr>, and abbreviated variable names ¹rideable_type, ²started_at,
## #   ³ended_at, ⁴start_station_name, ⁵start_station_id, ⁶end_station_name,
## #   ⁷end_station_id, ⁸start_lat
## # ℹ Use `colnames()` to see all variable names
```

6.5. Created the second graph for the comparison of usage times between members and casuals

```
data_bar2 <- tripdata_new %>% group_by(member_casual) %>% summarise(time=sum(usage_time))
ggplot(data_bar2, aes(x=member_casual, y= time)) + geom_bar(stat = "identity", fill="blue") +
  scale_y_continuous(labels = label_number(suffix = " M", scale = 1e-6)) +
    labs(title = "Number of rides completed by user type",
       subtitle = "For the period between November 2021 to April 2022",
       x = "User type",
       y = "Usage Time (minutes)") +
  geom_text(aes(label=round(time,2)), position = position_stack(vjust = .9), color="white")
```

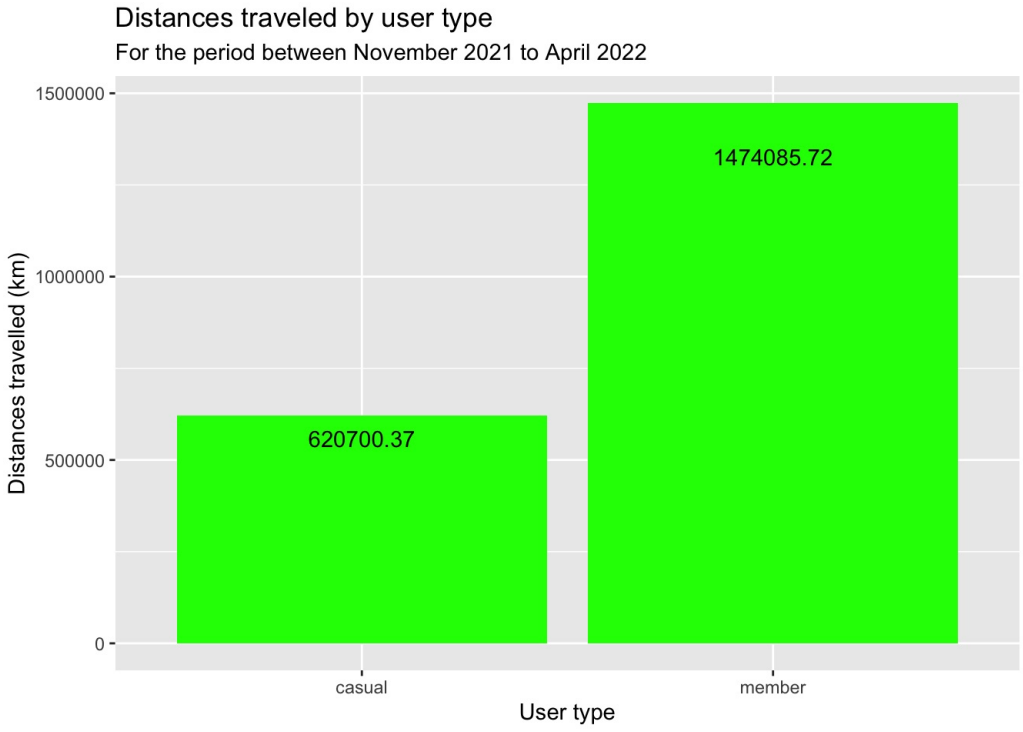
```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```



Even the usage number is much higher for members than casuals, usage time differs only 17.539 hours (1.052.352 minutes). This shows us that casual riders prefer longer rides.

6.6. Ceated the third graph for the comparison of distances traveled between members and casuals

```
data_bar3 <- tripdata_new %>% group_by(member_casual) %>% summarise(length=sum(distance_km))
ggplot(data_bar3, aes(x=member_casual, y=length)) + geom_bar(stat = "identity", fill="green") +
  labs(title = "Distances traveled by user type",
       subtitle = "For the period between November 2021 to April 2022",
       x= "User type",
       y= "Distances travelled (km)") +
  geom_text(aes(label=round(length,2)), position = position_stack(vjust = .9), color="black")
```



The distances traveled between members and casuals are again very high which is 14.120.156 km.

6.7. Calculated average time used and average distance traveled for each user type

```
data_bar2.1 <- tripdata_new %>% group_by(member_casual) %>% summarise(average_time=mean(usage_time))
head(data_bar2.1)
```

```
## # A tibble: 2 × 2
##   member_casual average_time
##   <chr>         <drtn>
## 1 casual        25.53104 mins
## 2 member        11.11177 mins
```

```
data_bar3.1 <- tripdata_new %>% group_by(member_casual) %>% summarise(average_distance_km=mean(distance_km))
head(data_bar3.1)
```

```
## # A tibble: 2 × 2
##   member_casual average_distance_km
##   <chr>                       <dbl>
## 1 casual                       2.06
## 2 member                       1.87
```

6.8. Created the fourth graph for the number of rides completed by month for each user type

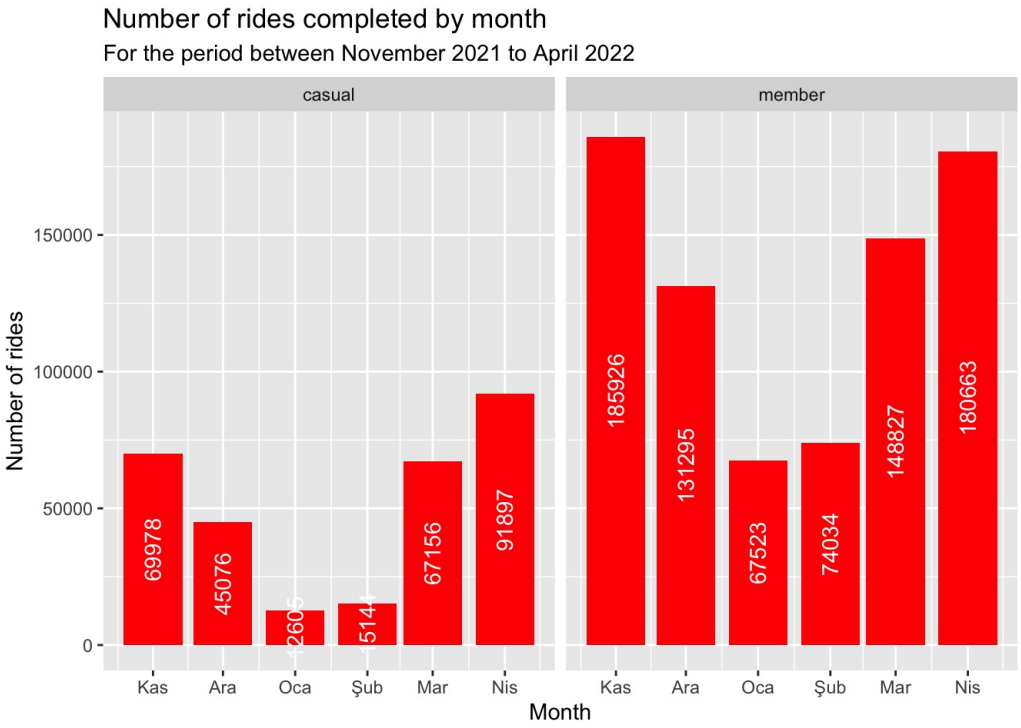
```
data_bar4 <- mutate(tripdata_new, month = floor_date(as_date(started_at), "month")) %>%
  group_by(month, member_casual) %>%
  summarise(count_of = n())
```

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

```
head(data_bar4)
```

```
## # A tibble: 6 × 3
## # Groups:   month [3]
##   month      member_casual count_of
##   <date>     <chr>            <int>
## 1 2021-11-01 casual           69978
## 2 2021-11-01 member          185926
## 3 2021-12-01 casual           45076
## 4 2021-12-01 member          131295
## 5 2022-01-01 casual           12605
## 6 2022-01-01 member           67523
```

```
ggplot(data_bar4, aes(x=month, y=count_of)) + geom_bar(stat = "identity", fill="red") +
  facet_wrap(~member_casual) +
  labs(title = "Number of rides completed by month",
       subtitle = "For the period between November 2021 to April 2022",
       x= "Month",
       y= "Number of rides") +
  geom_text(aes(label=count_of), position = position_stack(vjust = .5), color="white", angle = 90)
```



For both users, similarly, spring and summer rides (not included dataset because of the limitations) are more popular than winter rides because of the weather conditions.

6.9. Created the fifth graph for the number of rides completed by day for each user type

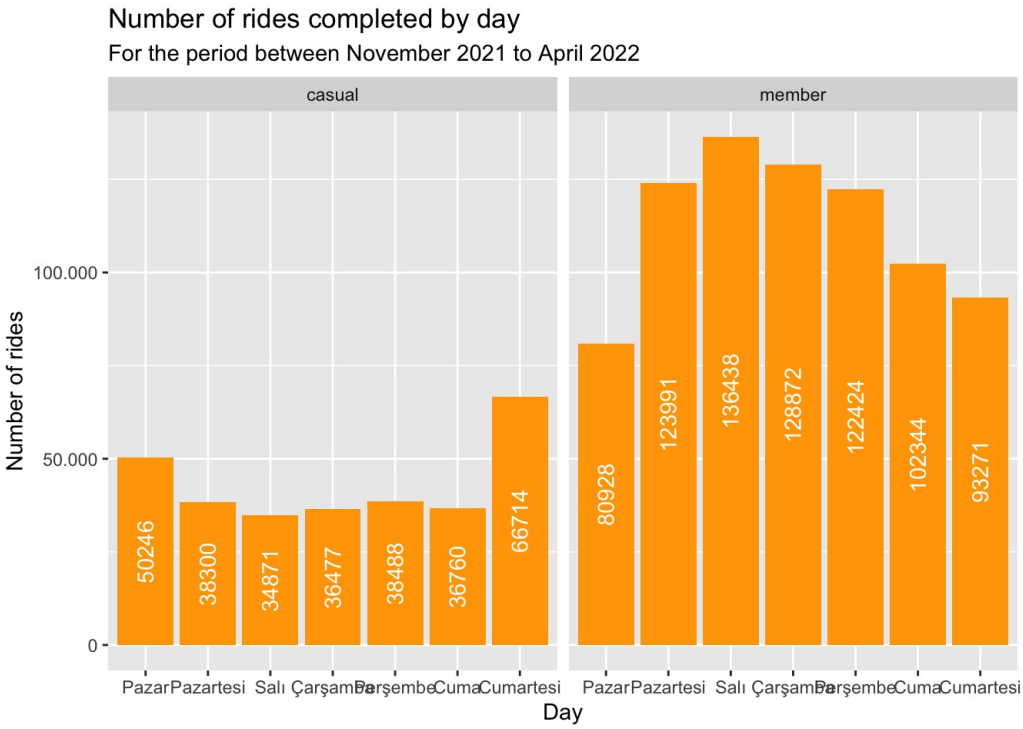
```
data_bar5 <- mutate(tripdata_new) %>% group_by(day, member_casual) %>% summarise(count_of = n())
```

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

```
head(data_bar5)
```

```
## # A tibble: 6 × 3
## # Groups:   day [3]
##   day       member_casual count_of
##   <ord>     <chr>            <int>
## 1 Pazar     casual           50246
## 2 Pazar     member           80928
## 3 Pazartesi casual           38300
## 4 Pazartesi member          123991
## 5 Salı      casual           34871
## 6 Salı      member          136438
```

```
ggplot(data_bar5, aes(x=day, y=count_of)) + geom_bar(stat = "identity", fill="orange") +
  facet_wrap(~member_casual) +
  labs(title = "Number of rides completed by day",
       subtitle = "For the period between November 2021 to April 2022",
       x= "Day",
       y= "Number of rides") +
  geom_text(aes(label=count_of), position = position_stack(vjust = .5), color="white", angle = 90) +
  scale_y_continuous(labels = comma_format(big.mark = ".", decimal.mark = ","))
```



Bike usage in weekdays is more if we look at the members, on the other hand, for the casual riders, usage on weekends is higher. This can be observed because the members mostly use it as a transportation method to their work or final destination rather than casual users who are mostly using it for entertainment purposes.

6.10. Created the sixth graph for the number of rides completed by time for each user type

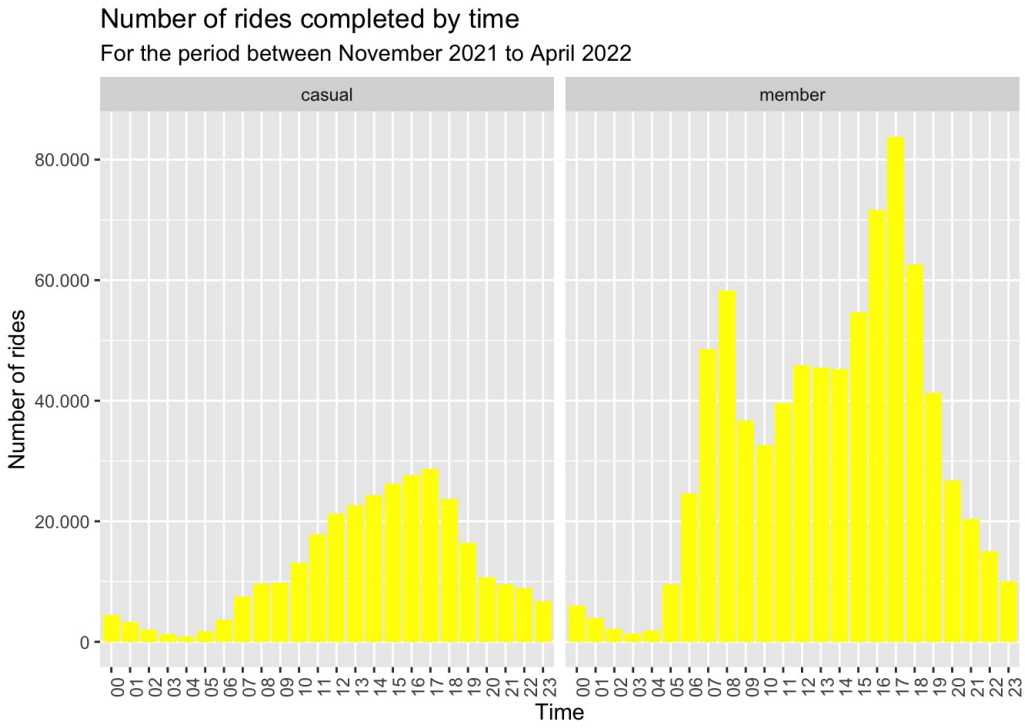
```
data_bar6 <- mutate(tripdata_new) %>% group_by(time, member_casual) %>% summarise(count_of = n())
```

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

```
head(data_bar6)
```

```
## # A tibble: 6 × 3
## # Groups:   time [3]
##   time  member_casual count_of
##   <chr> <chr>            <int>
## 1 00    casual            4427
## 2 00    member            5933
## 3 01    casual            3272
## 4 01    member            3904
## 5 02    casual            2072
## 6 02    member            2094
```

```
ggplot(data_bar6, aes(x=time, y=count_of)) + geom_bar(stat = "identity", fill="yellow") +
  facet_wrap(~member_casual) +
  labs(title = "Number of rides completed by time",
       subtitle = "For the period between November 2021 to April 2022",
       x= "Time",
       y= "Number of rides") +
  scale_y_continuous(labels = comma_format(big.mark = ".", decimal.mark = ",")) +
    theme(axis.text.x = element_text(angle = 90))
```



Again, members tend to use more at starting and ending of their work hours, besides casuals use more between 12 pm to 6 pm.

6.11. Top 5 start stations for both user types

```
table1 <- tripdata_new %>% group_by(member_casual, start_station_name) %>%
  summarise(count_of=n()) %>% arrange(desc(count_of))
```

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

```
table1.1 <- table1 %>% filter(member_casual == "casual") %>% slice(1:5)
print(table1.1)
```

```
## # A tibble: 5 × 3
## # Groups:   member_casual [1]
##   member_casual start_station_name                count_of
##   <chr>         <chr>                                <int>
## 1 casual        Streeter Dr & Grand Ave               7867
## 2 casual        DuSable Lake Shore Dr & Monroe St     5480
## 3 casual        Millennium Park                       4238
## 4 casual        Shedd Aquarium                        3693
## 5 casual        Michigan Ave & Oak St                 2510
```

```
table1.2 <- table1 %>% filter(member_casual == "member") %>% slice(1:5)
print(table1.2)
```

```
## # A tibble: 5 × 3
## # Groups:   member_casual [1]
##   member_casual start_station_name           count_of
##   <chr>         <chr>                           <int>
## 1 member        Kingsbury St & Kinzie St         9174
## 2 member        Ellis Ave & 60th St              7432
## 3 member        Clinton St & Washington Blvd     7189
## 4 member        University Ave & 57th St         7146
## 5 member        Clark St & Elm St                7135
```

As we will also see on the final image which is the map of the start destinations of both users, the trips happened in the coast side and touristic places are mostly by casual riders rather than members. Although members also ride in those places, the trips happen near the businesses, and residential areas are mostly by members.

6.12. Top 5 end stations for both user types

```
table2 <- tripdata_new %>% group_by(member_casual, end_station_name) %>%
  summarise(count_of=n()) %>% arrange(desc(count_of))
```

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

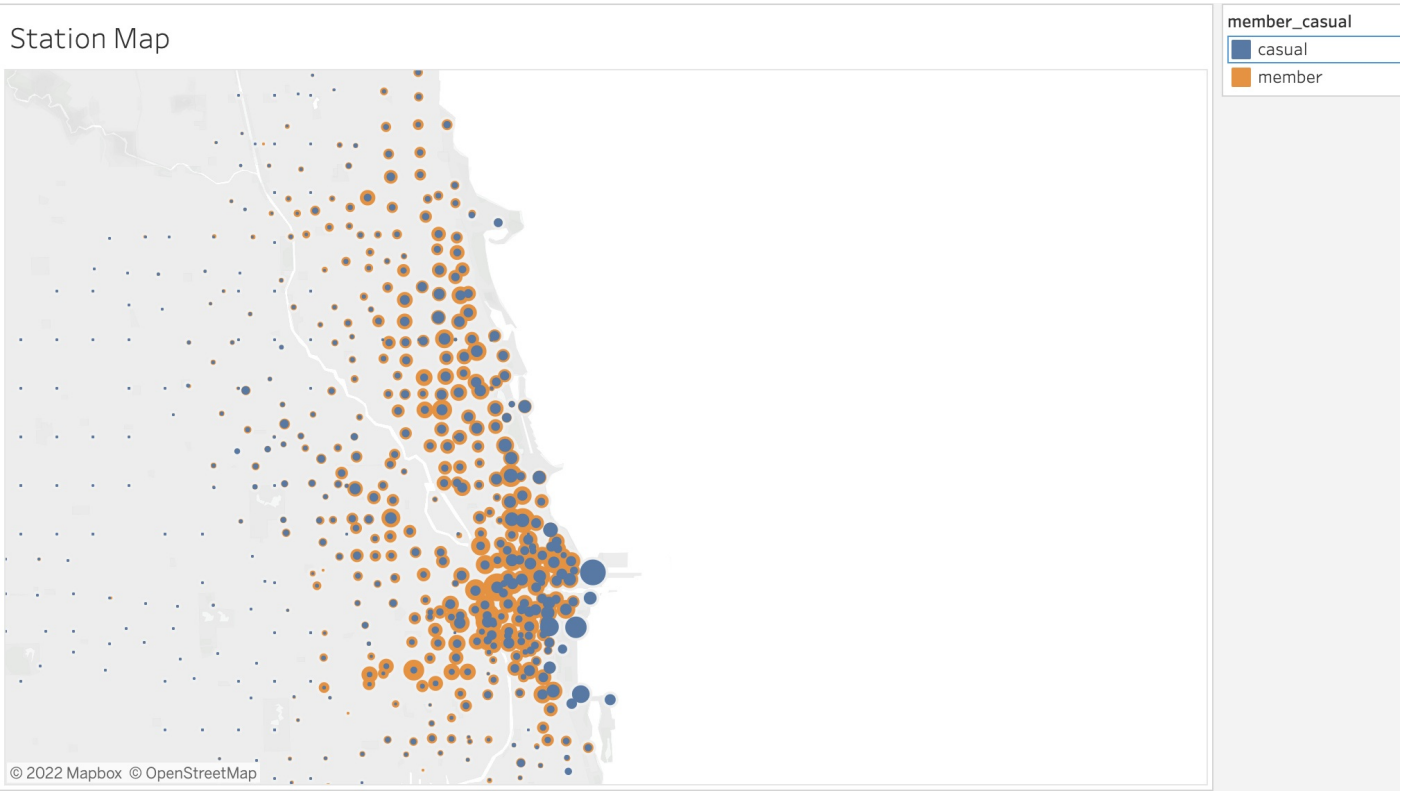
```
table2.1 <- table2 %>% filter(member_casual == "casual") %>% slice(1:5)
print(table2.1)
```

```
## # A tibble: 5 × 3
## # Groups:   member_casual [1]
##   member_casual end_station_name                count_of
##   <chr>         <chr>                              <int>
## 1 casual        Streeter Dr & Grand Ave             8564
## 2 casual        Millennium Park                     4811
## 3 casual        DuSable Lake Shore Dr & Monroe St   4740
## 4 casual        Shedd Aquarium                      3304
## 5 casual        Michigan Ave & Oak St               2901
```

```
table2.2 <- table2 %>% filter(member_casual == "member") %>% slice(1:5)
print(table2.2)
```

```
## # A tibble: 5 × 3
## # Groups:   member_casual [1]
##   member_casual end_station_name            count_of
##   <chr>         <chr>                          <int>
## 1 member        Kingsbury St & Kinzie St        8853
## 2 member        University Ave & 57th St        7772
## 3 member        Clinton St & Washington Blvd    7691
## 4 member        Clinton St & Madison St         7278
## 5 member        Ellis Ave & 60th St             7196
```

6.13. Start station map prepared in Tableau



The start destinations of both users, the trips happened in the coast side and touristic places are mostly by casual riders rather than members. Although members also ride in those places, the trips happen near the businesses, and residential areas are mostly by members. Orange dots represent members, and blue dots represent casual riders. The bigger the circles, more rides happened in that specific area or station.

7. Recommandations

- The casual riders whose starting and using habits are similar to members can be obtained by discounts which can convert them to the members.
- Marketing campaigns can be increased in spring and summer where most of the casuals are using in summer.
- Marketing campaigns can be increased in both top 5 stating and ending stations for casuals and give information and statistics including prices about how it will be more profitable if they subscribe and buy a membership.
- To increase the sales of casual rides, Cyclistic company can partner up with tourist guides and tourism agencies.
- Alternative subscription method can be introduced which is a weekend pass where this subscription members can ride bikes unlimited on the

weekends and it might be more suitable for casual riders who prefer to ride their bikes on weekends.

- Benefits of memberships can be increased like adding a reservations option which will be handy in peak months, also giving monthly gifts to members etc.