Cyclistic Case Study

Analyzing bike riding data

Shivank Sadwal

04/07/2022

About Cyclistic

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.

What are my objectives?

Discover how annual and casual members use Cyclistic differently during the year 2019 to help the future marketing team make informed decisions.

How will I answer this question?

I will analyse publicly available bike riding data from Cyclistic’s database, using Spreadsheets, R Studio and Tableau.

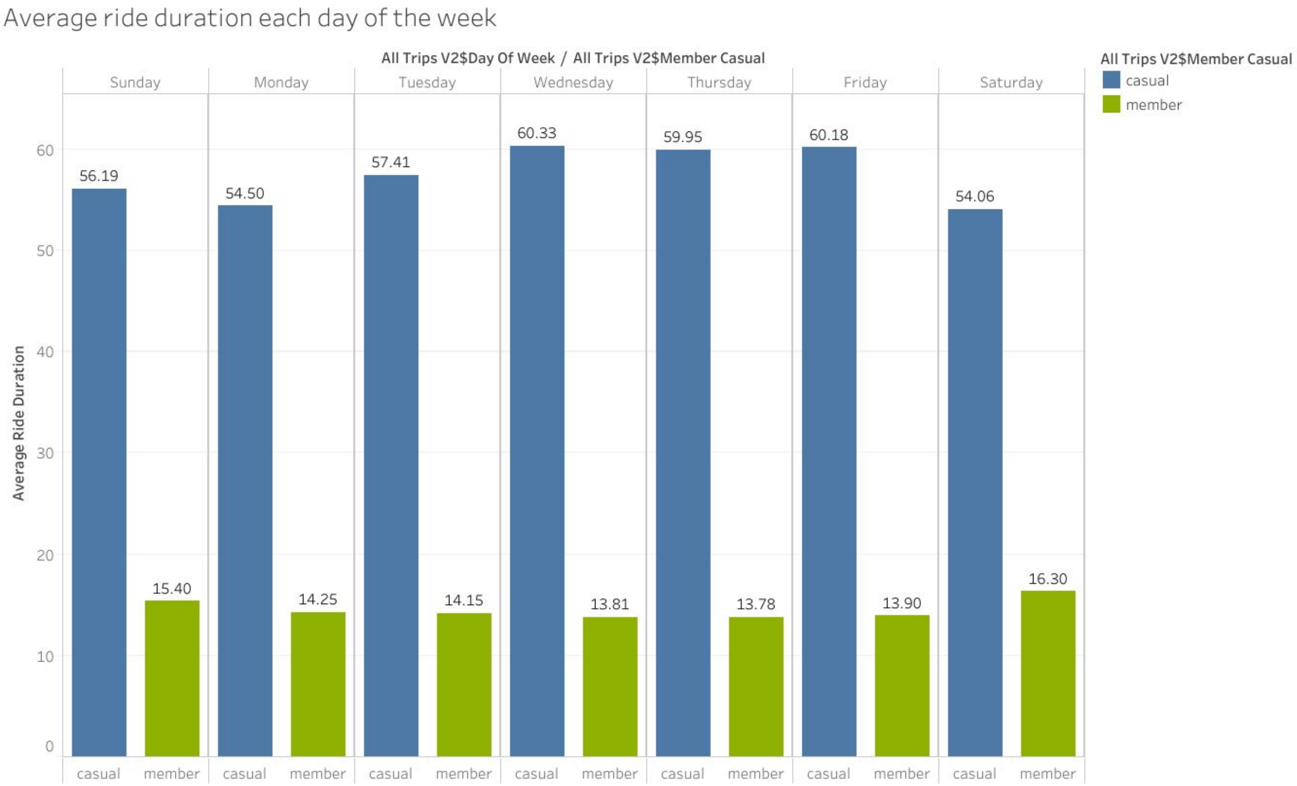
Recommendations

Subsidised cost for extended duration rides

Subsidised cost for extended duration rides

Trends

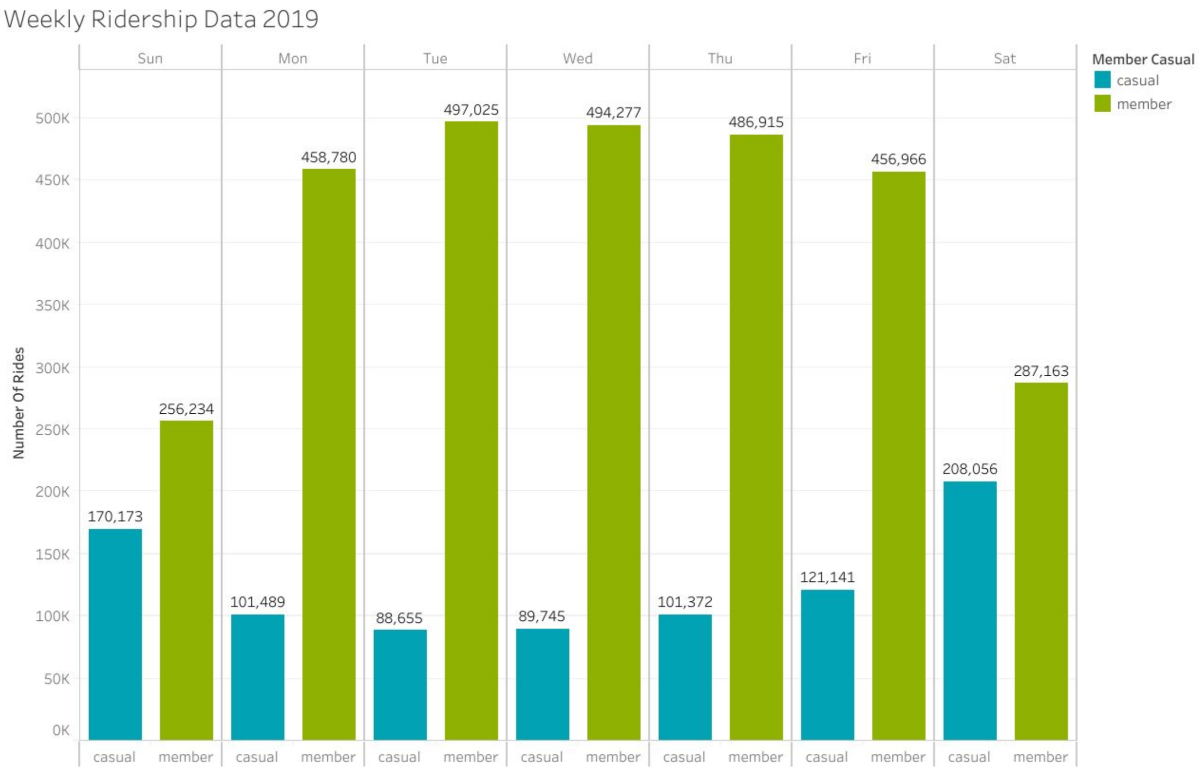
- Casual riders have longer 7 day average duration than members.(57.51 for casual riders and 12.49 for riders with annual subscription)



Subsidised cost for extended duration rides

Trends

- Members bike usage peaks midweeks and declines during the weekends.
- Casual users bike usage does the opposite as it peaks during the weekends and declines during weekdays.(implying they use the bikes for leisure)



Subsidised cost for extended duration rides

Conclusion

Casual members use cyclistic more during the weekends than during the midweeks additionally their average ride durations are significantly higher than the annual members.

I propose rewarding those with an annual membership in the following manner

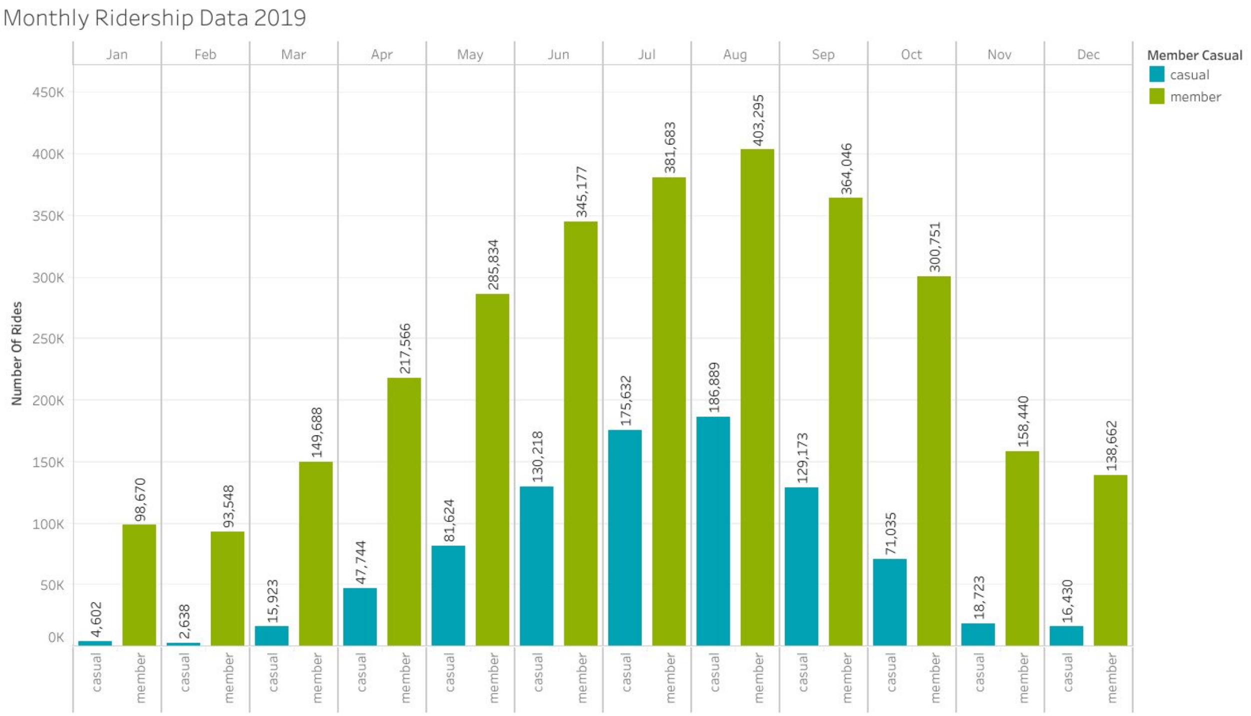
- Decreasing hourly rates for annual members for the weekends
- Decreasing hourly rates for longer ride durations
- Rewarding annual members with ‘X’ free minutes each weekend.

Dynamic Seasonal Promotion

Dynamic Seasonal Promotion

Trends

- Total ridership plummets towards the end of the year and into february.(during the colder months)
- Casual riders decrease at a much steeper rate.



Dynamic Seasonal Promotion

Conclusion

Overall ridership decreases during the colder months of the year, especially for the casual riders.

I propose a seasonal benefit for those with an annual membership using Cyclistic during the winters

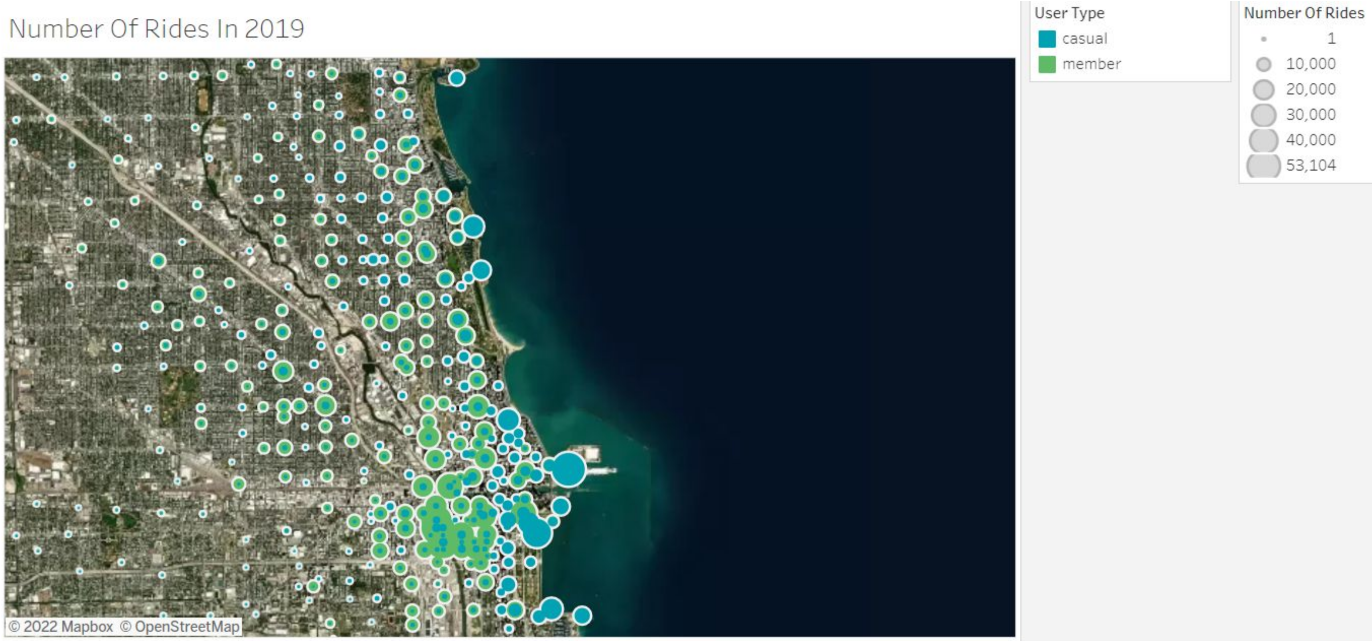
- Decrease overall rental charges during the winter seasons to promote more usage during the cold.

Target Areas For Advertisements

Target Areas For Advertisements

Trends

- This geographic data signifies that the maximum number of casual rides are concentrated near the coastline of chicago.(at beaches , parks and amusement parks)
- This further strengthens my previous suspicion that the casual members use the service more for leisure.



To View Entire Visualization:
https://public.tableau.com/shared/SXJGTB6S4?:display_count=n&:origin=viz_share_link

Target Areas For Advertisement

Conclusion

The geographic data gives us a clearer picture for the areas of Chicago where maximum number of casual riders start their trips ie. near the coastline and parks.

I propose the following

- If Cyclistic wishes to approach a more traditional advertising strategy of distributing fliers or renting billboards it is suggested to start with near cycle stations where there is maximum footfall and usage from casual users
- Few stations with high casual user number of rides are streeter dr & grand ave (53,104 casual rides), S Lake Shore Dr & E Monroe St (39,238 casual rides) etc.

Limitations & Further Analysis

Limitations & Further Analysis

- **Specific Recommendations :** Making specific recommendations is difficult only with the ride data. To further provide a more customer oriented recommendation will require pricing and customer data
- **Data Cleaning :** The criteria used for data cleaning were created based on inferences made from the data itself. For further analysis , liaising directly with the company to clear up confusion will avoid assumptions.

Cleaning & Preparations of Data

Combining Quarterly Tables into a single twelve month table

1. Getting the quarterly datasets

```
q1_2019 <- read.csv("Divvy_Trips_2019_Q1.csv")
q2_2019 <- read.csv("Divvy_Trips_2019_Q2.csv")
q3_2019 <- read.csv("Divvy_Trips_2019_Q3.csv")
q4_2019 <- read.csv("Divvy_Trips_2019_Q4.csv")
```

The data frame names were q1_2019 to q4_2019 respectively

Combining Quarterly Tables into a single twelve month table

2. Rename column names to match the new naming convention from Cyclistic

```
#Renaming all columns to latest naming scheme
q1_2019 <- rename(q1_2019,ride_id = trip_id
                  ,rideable_type = bikeid
                  ,started_at = start_time
                  ,ended_at = end_time
                  ,start_station_name = from_station_name
                  ,start_station_id = from_station_id
                  ,end_station_name = to_station_name
                  ,end_station_id = to_station_id
                  ,member_casual = usertype)
```

This was done for all four datasets.

Combining Quarterly Tables into a single twelve month table

3. Convert ride_id and rideable_type to character so that they can stack correctly

```
# Convert ride_id and rideable_type to character so that they can stack correctly
q4_2019 <-  mutate(q4_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
q3_2019 <-  mutate(q3_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
q2_2019 <-  mutate(q2_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
q1_2019 <-  mutate(q1_2019, ride_id = as.character(ride_id)
                   ,rideable_type = as.character(rideable_type))
```

This was done for all four datasets after observing their structure.

Combining Quarterly Tables into a single twelve month table

4. Stack individual quarter's data frames into one big data frame using bind_rows() function

```
all_trips <- bind_rows(q1_2019,q2_2019, q3_2019, q4_2019)
```

all_trips was the name chosen for the data frame to house the year round data

Cleaning the data frame

1. Removing all columns that are not part of the new Cyclistic schema

```
all_trips <- all_trips %>%
  select(-c("birthyear","gender","X01...Rental.Details.Duration.In.Seconds.Uncapped",
            "X05...Member.Details.Member.Birthday.Year", "Member.Gender", "tripduration"))
```

Cleaning the data frame

2. Creating DATE, MONTH, DAY and YEAR as columns to make aggregation more easier. We use as.Date() functions

```
all_trips$date <- as.Date(all_trips$started_at)  #default format is yyyy-mm-dd

#for custom formats we use format function

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")
```

started_at column was used because it is already in a format that the as.Date() function can use.

started_at
2019-01-01 00:04:37

Cleaning the data frame

3. Creating ride_length column to make calculations easier for this we use the difftime() which subtracts started_at column value from ended_at column value to give us effective ride length in minutes.

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

#checking if new columns were added
str(all_trips)

# Convert "ride_length" from Factor to numeric so we can run calculations on the data
is.factor(all_trips$ride_length)
all_trips$ride_length <- as.numeric(as.character(all_trips$ride_length))
is.numeric(all_trips$ride_length)
```

started_at	ended_at
2019-01-01 00:04:37	2019-01-01 00:11:07

Cleaning the data frame

4. Removing “Bad” Data

The data frame includes some entries when bikes were taken out of docks and checked for quality by Divvy or ride_length was negative.

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

We will create a new version of the dataframe (v2) since data is being removed.

Analyzing the data using aggregation

1. Analyze ridership data by type and weekday and make a dataframe “weekly”

```
weekly <- 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)
```

n() : Used to count number of rides   mean() : Used to find average ride length

Analyzing the data using aggregation

2. Analyze ridership data by type and month and make a dataframe “monthly”

```
monthly <- all_trips_v2 %>%
  group_by(member_casual,month) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual,month)
```

n() : Used to count number of rides   mean() : Used to find average ride length

Analyzing the data using aggregation

3. Analyze ridership data by member type and start station name and making a dataframe named “geo” for later extracting geolocation data.

```
geo <- all_trips_v2 %>%
  group_by(member_casual,start_station_name) %>%
  summarise(number_of_rides = n() , average_duration = mean(ride_length)) %>%
  arrange(member_casual,start_station_name)
```

n() : Used to count number of rides   mean() : Used to find average ride length

Exporting the analyzed data

We will used the write.csv() function to export the data frames (all_tripsv2, geo, weekly and monthly) into .csv files for further analysis and tableau visualisation.

```
write.csv(all_trips_v2, file = 'C:/Users/ahlco/Desktop/CASE STUDY 1/Divvy_trips_2019/csv_divvy_trips_2019/all_trips_v2.csv')

#exporting geo data for the year 2019
write.csv(geo, file = 'C:/Users/ahlco/Desktop/CASE STUDY 1/cyclistic_geographic_data_2019.csv')

#Exporting weekly data
write.csv(weekly, file = 'C:/Users/ahlco/Desktop/CASE STUDY 1/cyclistic_weekly_data_2019.csv')

#Exporting monthly data
write.csv(monthly, file = 'C:/Users/ahlco/Desktop/CASE STUDY 1/cyclistic_monthly_data_2019.csv')
```

Adding Longitude and Latitude Data

Imported the cyclistic_geographic_data_2019.csv file to google sheet and used the "GeoCode" extension to add longitude and latitude values in order to plot geographic data on tableau.

Before

A	B	C	D	E	F	G
	member_casual	city	start_station_name	full address	number_of_rides	average_duration
1	casual	Chicago	2112 W Peterson Ave	2112 W Peterson Ave Chicago	121	134.785124
2	casual	Chicago	63rd St Beach	63rd St Beach Chicago	687	80.68624454
3	casual	Chicago	900 W Harrison St	900 W Harrison St Chicago	883	29.14514911

After

A	B	C	D	E	F	G	H	I
	member_casual	city	start_station_name	full address	Latitude	Longitude	number_of_rides	average_duration
1	casual	Chicago	2112 W Peterson Ave	2112 W Peterson Ave Chicago	41.991291	-87.6823108	121	134.785124
2	casual	Chicago	63rd St Beach	63rd St Beach Chicago	41.78203	-87.5733146	687	80.68624454
3	casual	Chicago	900 W Harrison St	900 W Harrison St Chicago	41.8748099	-87.6497943	883	29.14514911

Thank You