cyclistic case study

Aidan Blassage

5/28/2021

Business Task

- The purpose of this analysis is to figure out the difference between casual riders and members and find options to entice casual riders to become members.
- The key stakeholders for this analysis are Lily Moreno and the marketing team

Data Preparation

- The data being used is the last 6 months of data available from Cyclistic
- The data is organized in 13 columns, though some columns are extraneous and may be removed.
- The data does not look like it has any problems with credibility besides maintenance being logged as a ride. I will remove these maintenance rides in the data processing stage.

#Data Processing

- My cleaning process started with loading the datasets into Excel and checking on blank fields within the database. For the fields that had a blank I utilized a Vlookup function to fill those blanks using the start_lat/start_lng/end_lat/end_lng wherever useful. If blanks could not be filled I removed them so they would not impact the data. Total fields removed less than 1% so it should not have a significant impact on the data.
- In excel I added a column to find each individual ride length and day of the week rented columns. These will help in the analysis to find any trends in the data.
- I did extra data cleaning within Rstudio, some of the columns were irrelevant to observation so I removed them. These columns included start_lat, start_lng, end_lat, end_lng, birthyear, and gender.

all_trips <- all_trips %>%

select(-c(start_lat,start_lng,end_lat,end_lng,birthyear,gender))

- The member_casual column has different inputs across the data so I consolidated the names for analysis.

all_trips <- all_trips %>%

mutate(member_casual = recode(member_casual

```
                    ,"Subscriber" = "member"

                    ,"Customer" = "casual"))
```

- Added more granular dates for advanced operations and analysis

all_trips$date <- as.Date(all_trips$started_at)

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")

- The data does not appear to have any significant issues except maintenance on bikes recorded as renting the bikes. This data will be removed from the dataset for that reason.

all_trips_v2 <- all_trips[!(all_trips$start_station_name == "HQQR"|all_trips$ride_length<0),]

Data Analysis

- I began my analysis by calculating the average ride length of members and casual riders, the max ride length, and the most common day of the week.
- Interestingly there are already clear distinctions between members and casual riders. Members ride more often but have a lower average ride time. Inversely casual members have a longer ride time but ride less often. Casual riders increase the most on the weekends and their average ride time is 45:00 compared to members having a more constant ridership with an average ride being 17:30.
- Some initial hypotheses are that members ride with a purpose to get to a specific place and may use it as a primary form of transportation, while casual riders use it as a form of relaxation or exercise. This makes sense to me as the casual riders ride the most on the weekends and for long periods of time.

mean(all_trips_v2$ride_length)

median(all_trips_v2$ride_length)

max(all_trips_v2$ride_length)

min(all_trips_v2$ride_length)

summary(all_trips_v2$ride_length)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = mean)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = median)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = max)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual, FUN = min)

aggregate(all_trips_v2$ride_length ~ all_trips_v2$member_casual + all_trips_v2$day_of_week, FUN = mean)

all_trips_v2$day_of_week <- ordered(all_trips_v2$day_of_week, levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

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)

- After running this code my hypothesis is confirmed. Members ride most often from Monday-Friday and for a relative short period of time. This could mean that their membership is utilized by going to and from work weekdays. Members ride almost 10x more every day of the week over casual riders. However, this drops to 2-3x as much on the weekends. Casual riders however ride for 4x as long as members do, which we can assume means they don't have as direct a goal as the members do when they utilize cyclistic services.

Act

- My recommendations for Cyclistic's marketing department are to try to convert casual members to ride Cyclistics bikes to and from work like members. Members rides peak from Monday to Friday and their ride time is 33% of casual riders. From this we can assume that members are riding to and from work on weekdays as their rides are shorter and more frequent.

- The digital marketing team should highlight the benefits of riding daily, for the people riding, the cost saved from commute, and the saved impact on the environment.
- A next step that would make the data clearer could be a survey sent to members and casual riders about why they choose Cyclistic when they do, this would fully clarify the benefits the company provides making it easier to market.