Case Study: Bike-Sharing-Company Cyclistic

The following report shows the results of a case study, where a fictitious bicycle rental company (Cyclistic) requests an analysis of their rental data, based on the type of users (annual or occasional). The company's final purpose is to develop marketing strategies, to turn occasional riders into annual members.

Key Questions

- How does Cyclistic usage differ between annual members and casual users?
- Why should casual users purchase an annual Cyclistic membership?
- How can Cyclistic use digital media to motivate casual users to purchase a membership?

Dataset

I have analyzed a dataset from the public data source Index of bucket "divvy-tripdata, with the agreement of data use Data License Agreement | Divvy Bikes.

The dataset has more than 4 millions of observations related to the bike rental in the city of Chicago, USA, and is composed of 13 variables:

- ride_id: id of each ride;
- rideable_type: type of bike that was rented;
- Started_at: start date and time of the ride;
- ended_at: end date and time of the ride;
- start_station_name: name of the station where the ride started;
- start_station_id: id of the station where the ride started;
- end_station_name: name of the station where the ride ended;
- end_station_id: id of the station where the ride ended;
- start_lat: geographical latitud of the station where the ride started;
- start_lng: geographical longitude of the station where the ride started;
- end_lat: geographical latitud of the station where the ride ended;
- end_lng: geographical longitude of the station where the ride ended;
- member_casual: type of user.

For this report I used six months of data, from May-2023 to October-2023 (one table per month).

Process data

To prepare data for the analysis I worked with the tools MS Excel and R.

Before starting the cleaning process, I glanced through the database to get more familiar with the information.

As a first step I looked for duplicated values in the **ride_id** variable with the MS Excel options in "Conditional Formatting", to make sure that each trip was unique.

Then, I looked for missing values in the columns **started_at** and **ended_at**, using MS Excel filter options. Since there were not missing data, I proceeded with the analysis. I reviewed the type of data of these columns and made sure that they were type "date and time".

After that, I created a new column assigning a numeric value to the weekdays, where 1 corresponds to Sundays, 2 to Mondays, 3 to Tuesdays, 4 to Wednesdays, 5 to Thursdays, 6 to Fridays and 7 to Saturdays. I used the information in the column **started_at** and used of the formula "WEEKDAY()" to determine the corresponding weekday of the ride.

From this point I worked with the tool R in an R Studio environment.

I loaded the MS Excel files to an R Studio environment and I joined the six tables (one per month) to create a unique table with all the data.

I calculated the duration of each trip (**duration_min**) subtracting the end time from the start time (**duration_min = ended_at – started_at**). I saved this information in a new column called **duration_min**. I transformed the data type to numeric and checked the results. I discovered that many rows had a negative ride duration. Taking a closer look at these negative ride duration values, I realized that the start (**started_at**) and end time (**ended_at**) of trips were exchanged. I corrected the column order for these values.

I analyzed the distribution of the trip duration (**duration_min**) and saw that there were extremely low and high values. These values represented a 2.79% of the database. Since the purpose of this report is to analyze the usage of rental bikes in a city and, since the amount of the extreme values was not representative of the most of the registered trips duration, these rows were eliminated from the analysis dataset. Therefore, the dataset was filtered to include **duration_min** values between 1-240 minutes.

Analysis

To understand better the nature of the data I analyzed the ride duration **(duration_min)** of all types of users (Figure 1). The mean ride duration was 15.63 minutes and most of the rides were shorter that 12 minutes long, approximately.

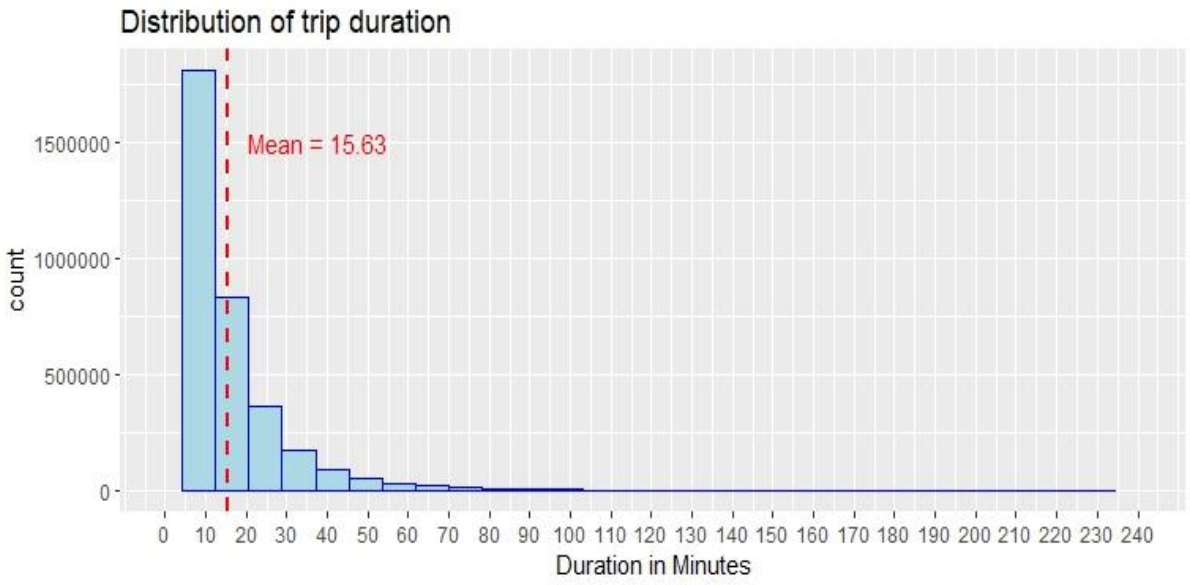

Figure 1: Number of trips registered for each trip duration.

Then, I explored the differences in the trip duration between annual members and casual users (Figure 2). I found that the casual users registered longer trips, with an average of 20 minutes, than annual members, with an average of 12 minutes.

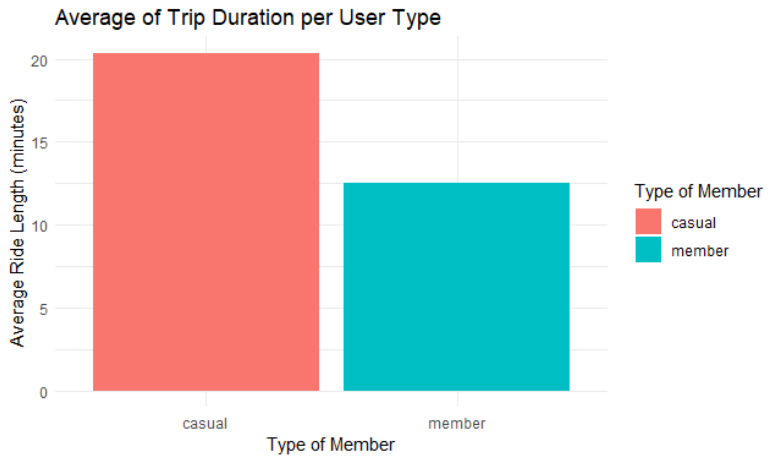

Figure 2: Average ride length for types of members

Furthermore, I analyzed which kind of bicycles are mostly used by each type of users (Figure 3). The results showed that annual members seem to prefer the classical bicycles than the electric bicycles. In contrast, the casual users chose electric bicycles more than classical bicycles. This analysis also showed that docked bicycles are rarely chosen and that only casual users registered trips with this type of bike.

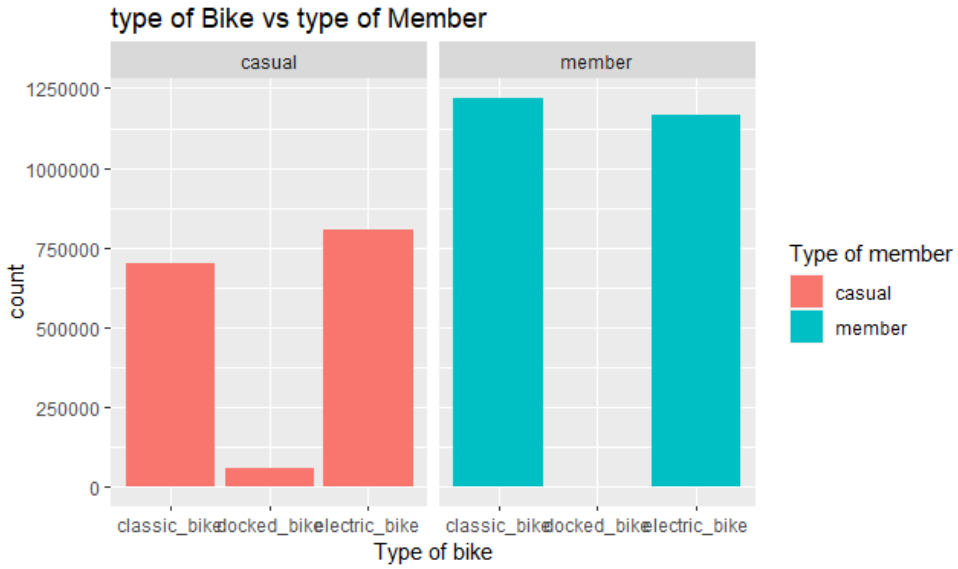

Figure 3: Number of trips registered for different types of bicycles (classic, docked and electric bike) by types of users (casual or member).

In addition, I discovered that annual members registered an overall higher number of trips than casual users as well (Figure 4).

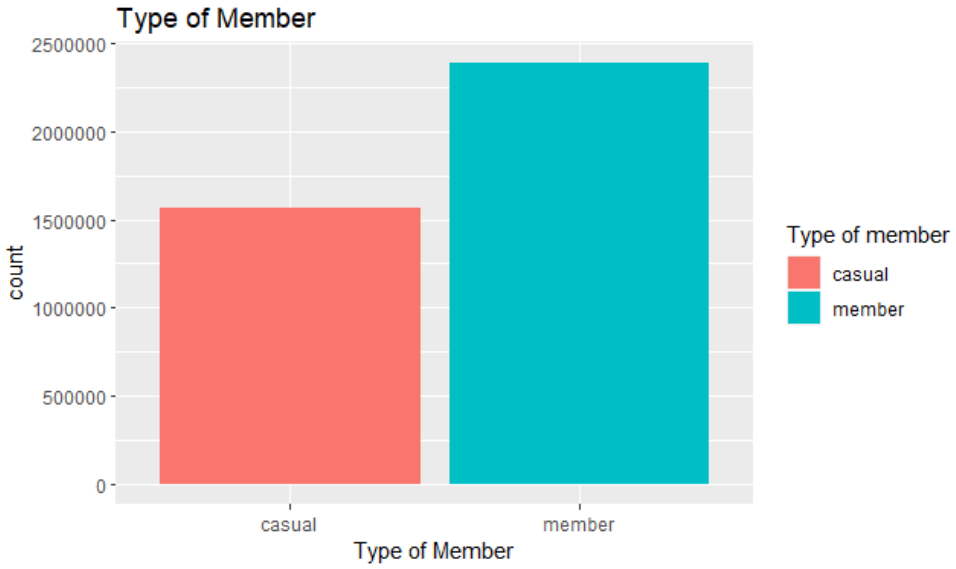

Figure 4: Number of trips registered by types of members (casual or annual member).

The number of trips registered per day revealed that casual users rent more bicycles during the weekends (Saturdays and Sundays) in comparison with the rest of the days (Figure 5). Annual members showed a similar number of registered trips on weekends as casual users. However, the results showed that annual members have a higher use of bikes during the business days (Monday to Friday).

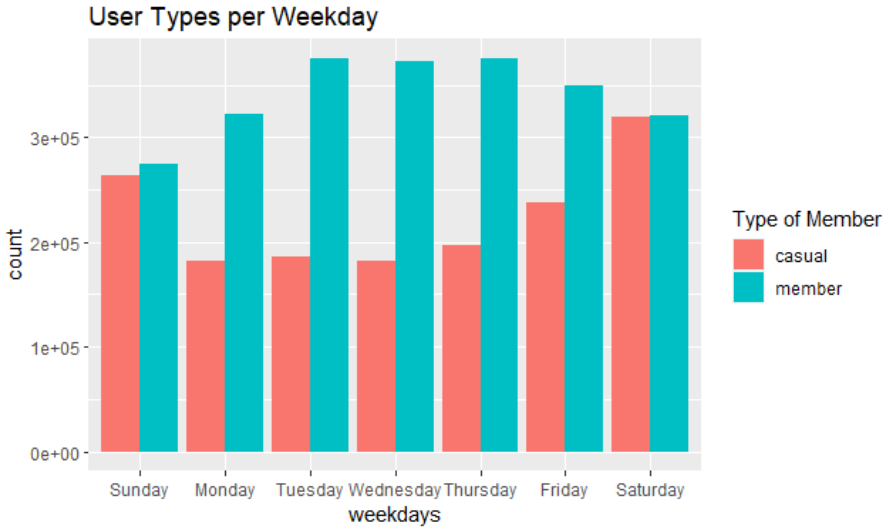

Figure 5: Number of trips registered per Weekday by types of users (casual or annual member).

Exploring the locations of registered trips, I extracted the ten most relevant start trip stations during the week and per type of User (Figure 6). There are five stations that stand out from the rest, especially on weekends and registered by casual members. These are Streeter Dr & Grand Ave, DuSable Lake

Shore Dr & Monroe St., DuSable Lake Shore Dr & North Blvd., Michigan Ave & Oak St. And Theater on the Lake. The annual members did not reveal a clear tendence in the choice of start station.

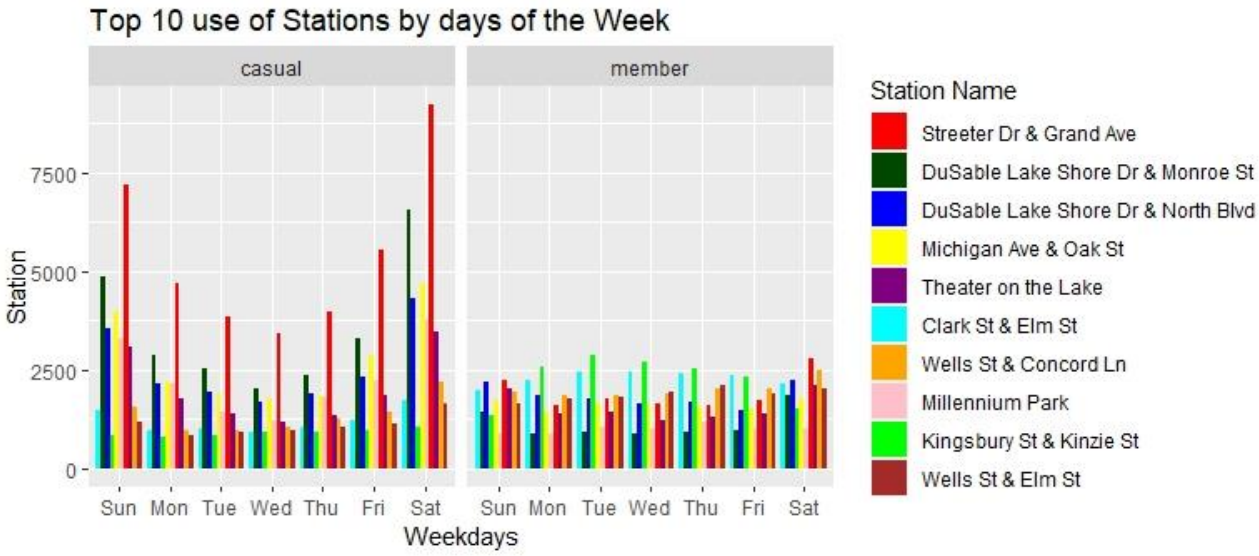

Figure 6: Number of trips started at each station per weekday, in the top 10 stations.

With the use of the geographical information provided for each station, I constructed a map showing the top 10 start stations (Figure 7). These stations are located close to each other and in the central part of the city, where touristic, commercial and cultural offer is found.

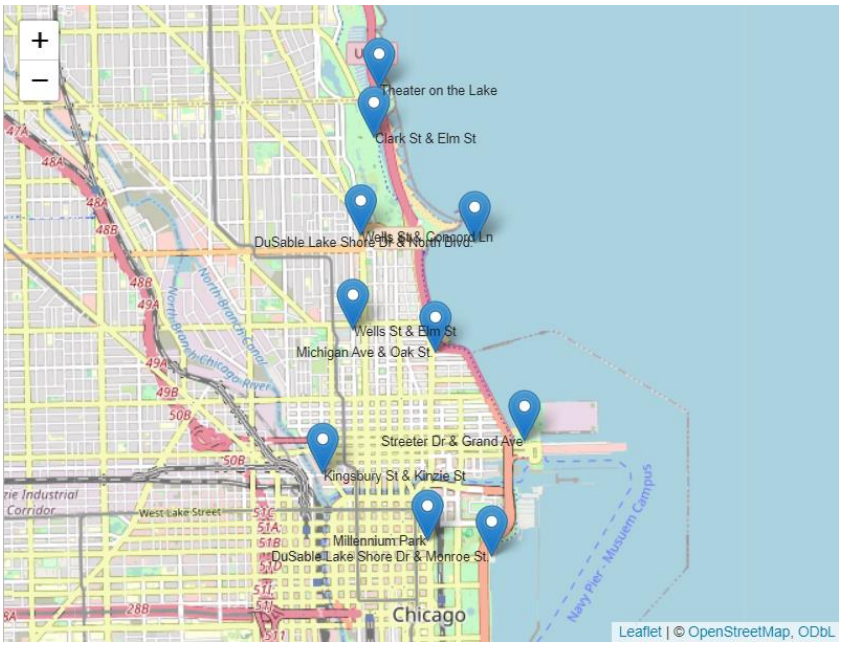

Figure 7: Geographical location of top 10 stations.

Main conclusions

The analysis of the database revealed that, in general, annual members registered a higher number of trips than casual users (Figure 4). The number of trips, however, was not equally distrubuted during the weekdays (Figure 5). Both casual and annual members registered a similar a mount of trips during the weekends. Nevertheless, while the use of rental bikes during business days was lower for casual users in comparisson to weekends, the annual members registered the higher number of trips from Monday to Friday. This could indicate that more annual members may use the rental bikes in business-related activities, such as travelling to work or school, in comparison to casual members. In contrast, the increased number of trips registered by these type of users during the weekends, the duration of the trips (Figure 2) and the location of the top 10 stations (Figure 7) could be linked to a more recreational usage of the bikes. In addition, annual members rented more classical bicycles that electrical ones, while for casual users the most chosen type of bike was the electrical bicycle (Figure 3).

In light with the findings of this analysis, I would recomend exploring marketing alternatives related with the duration of the trips and the weekdays where the bikes are rented, and to focus these strategies nearby the top 10 stations.