01 Final Report

S A Nawash Akhtar

Cyclistic bike-share analysis case study

This case study is a part of Google Data Analytics Certificate. The bike share company "Cyclistic" mentioned in this case study is a fictional company which is based on the "Divvy Bikes" of Chicago.

1. Introduction

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.

Cyclistic has three pricing plan.

- Single-ride pass
- Full-day pass
- Annual membership

Depending on the purchase the company has two user types.

1. **Casual riders** for purchasing Single-ride or Full-day pass
2. **Cyclistic members** for purchasing annual membership

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. The director of the marketing 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, the director believes there is a very good chance to convert casual riders into members.

This report will explore the observation based on the data available from "Cyclistic" and will provide recommendations from the analysis and insights.

Goal: Design marketing strategies aimed at converting casual riders into annual members.

Specific task: Determine how annual members and casual riders use Cyclistic bikes differently.

2. Prepare data

The Cyclistic's historical trip data is stored on the AWS site and can be used by this licence Click Here (https://www.divvybikes.com/data-license-agreement). For reference, Cyclistic is a fictional company so "Divvy Bikes" data will be used for this case study.

The data has separate csv files in the aws server for every year and data from Jan'20 to Oct'22 will be used for this analysis. The analysis was done in the month of Nov'22.

In the dataset each ride data is represented by a unique ride id which contains the necessary information for a single ride. The ride data contains the bike type used by the user, start and end station name, start and end station id, user type, start and end time of the trip. Bike stations also have longitude and latitude to locate them.

This first-party data is collected by Cyclistic itself and can be cross-checked with station name and id, location, start and end time to ensure that individual ride data is reliable, accurate and unbiased. For the purposes of this case study, data was analyzed from 2020 to 2022, but beyond these three years, there are over 7 years of ride data stored on AWS servers up to the year 2013. Cyclistic is updating the database on a monthly basis and analysis was done on current data that is relevant to the task. More than 10 million ride data were analyzed over a three-year period containing critical information that helped the analysis reach its final conclusions.

3. Process data

Due to the large amount of data R is used as the primary tool for analysis and visualization. During the data cleaning process, data type mismatches were found especially for the 2020 data, for which the data could not be merged. So the datatype had to be modified for analysis. Also, a new variable named "ride duration" was created to find the difference between ride start and end times. But after analysis, some ride duration values were found to be negative which was not possible. So those values had to be deleted for future use.

The start and end station names were checked and blank values of these variables were removed. Moreover, since the ride ID is unique for each ride, duplicate values were checked and removed accordingly.
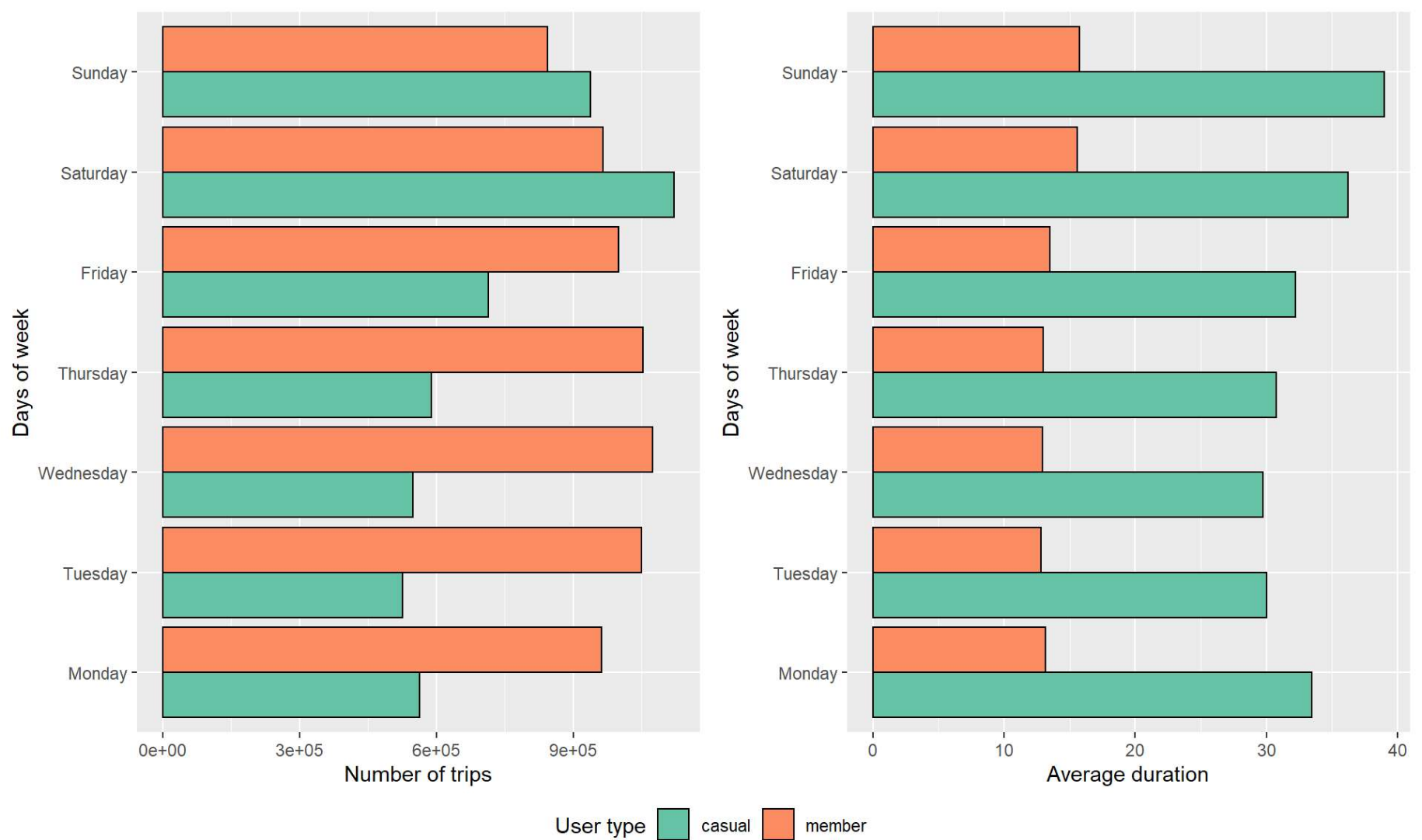
All data cleaning processes are explained in the **"Process Data"** file.

4. Analyze data

The task was to find out the different uses of the bike among casual and member riders. Basically, this analysis will determine how users ride the bike differently.

All the analysis processes are explained in the **"Analyze Data"** file.

4.1 Popular days of week

The figure describes how casual and member riders use the bike differently for the days of the week. Here it can be seen that throughout the week, member users ride about the same amount on weekdays and weekends. And for casual users, ride intensity increases on weekends and is almost constant across weekdays but lower than member riders on weekdays. Because members ride bikes for commuting and casual riders mostly bike for leisure.

As for the average times, the times for members are consistently the same throughout the week, and casual riders also have similar levels throughout the week. Here the average duration of casual riders is higher than that of members but the number of trips is higher for members than casual riders.

4.2 Popular months
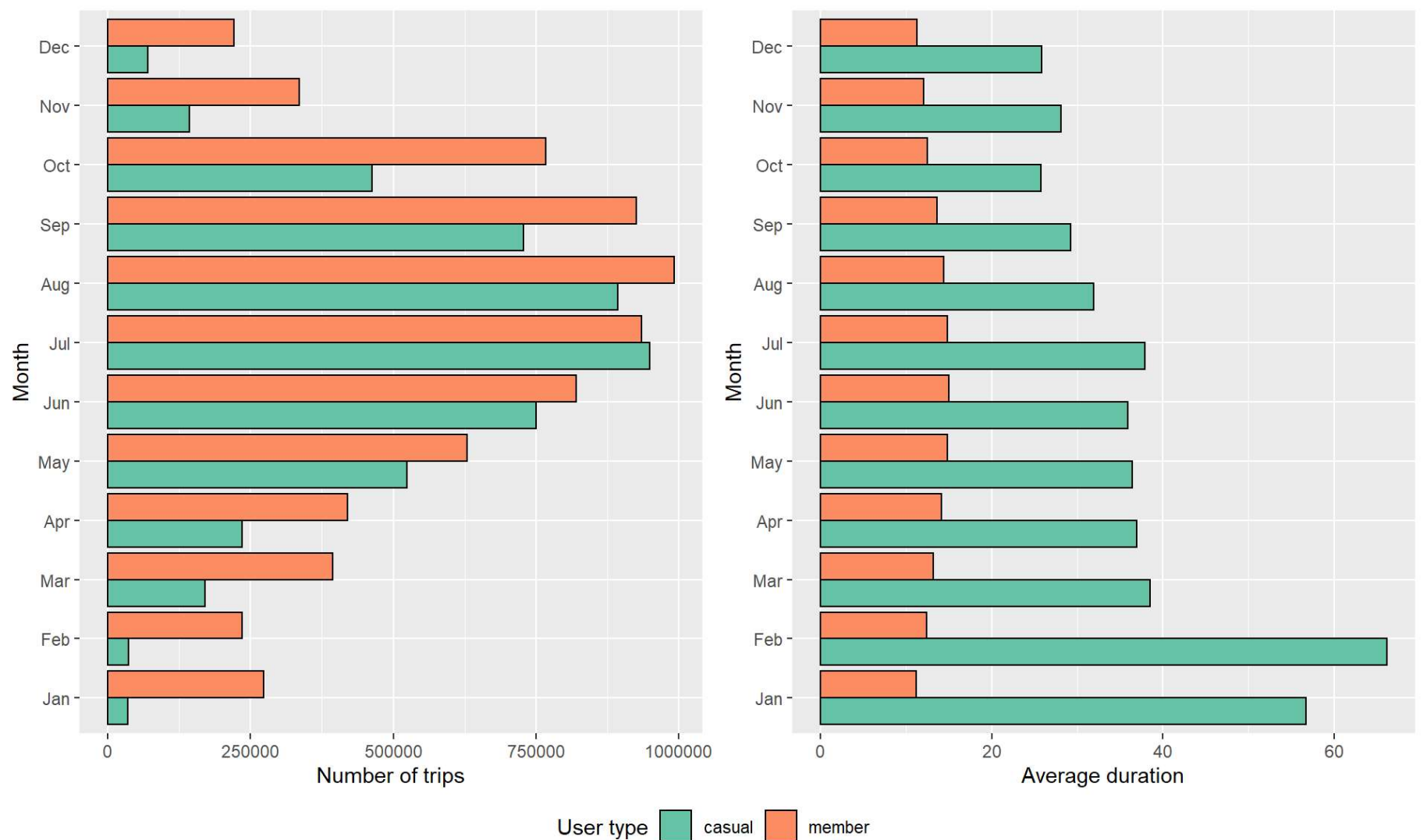
From the figure, we can see that the number of trips for a year increases during the summer months and decreases when there is no summer. The number of trips for members is always higher than for casual riders. Leisure travel increases during the summer months so the number of trips increases. From the average ride duration, we can see that the average ride duration of members throughout the year including summer months is about 10 minutes while the average ride duration of casual riders is from 30-40 minutes throughout the year but increases only in January and February. During these two months, the average ride duration goes up to 50-70 minutes.

4.3 Popular days of month
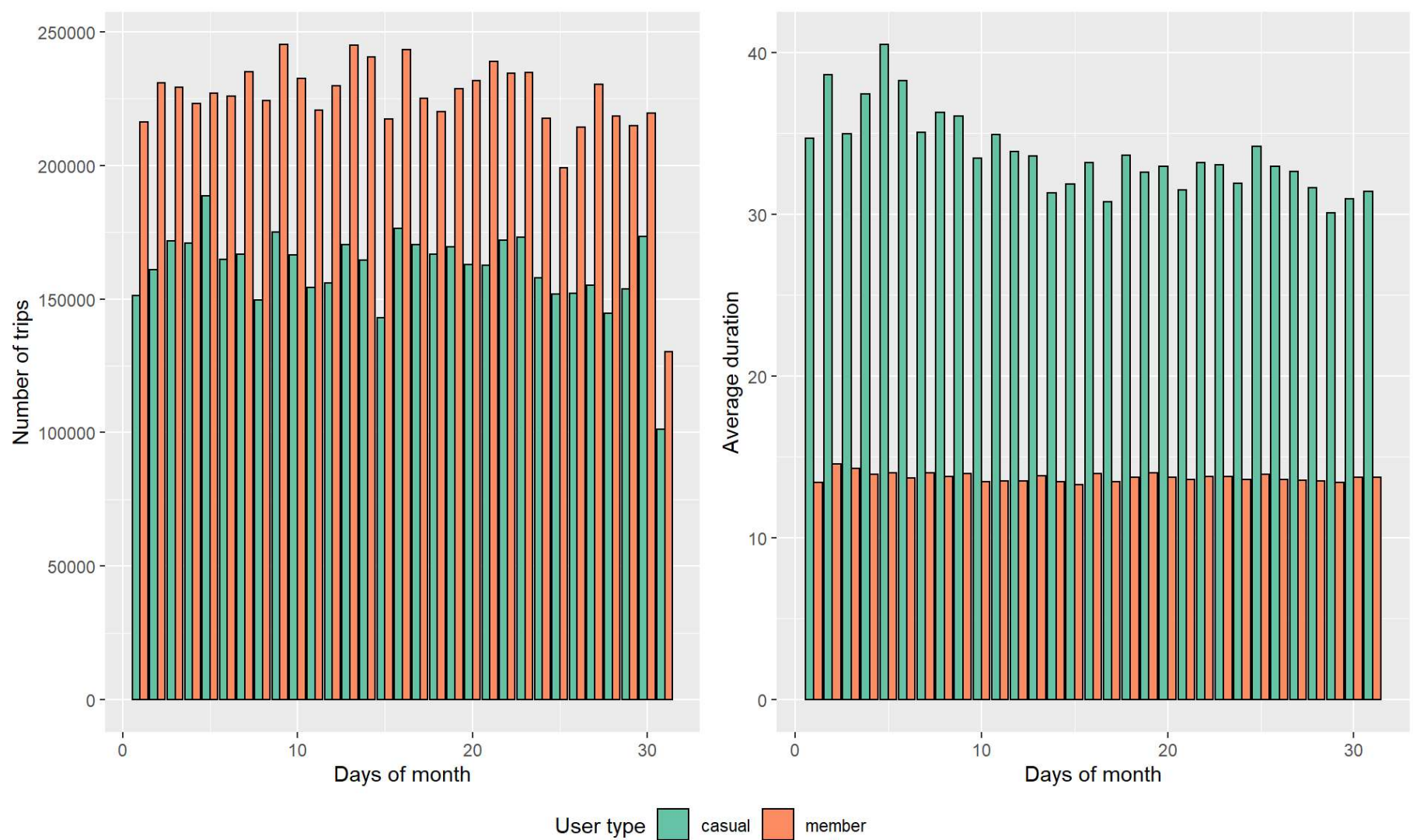
From the figure, we can see that the number of trips and the average duration of monthly usage are almost the same for both riders. It does not drastically change in any circumstances. Although we can say that the number of trips for members is always higher than for casual riders and the average duration is always higher for casual riders than for members.

4.4 Yearly ride data analysis
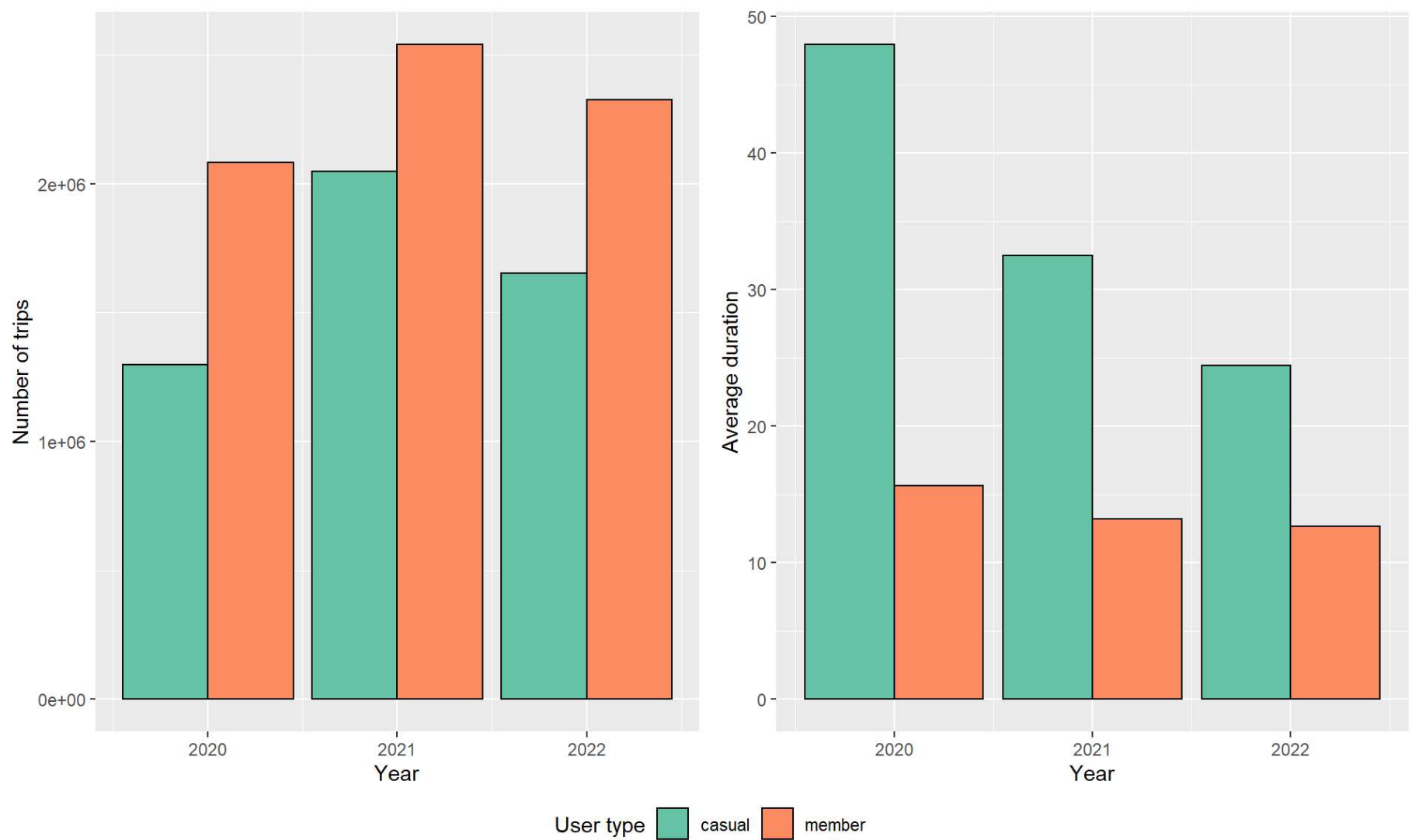
Following the Covid-19 pandemic, the number of trips has increased for the next two years 2021 and 2022. This shows that people are getting more confident to take rides than ever before. But it also appears that although the number of trips is increasing year on year, the average duration is decreasing each year. Since the epidemic subsided, the average duration has been decreasing every year even as restrictions have been eased. The decrease rate is seen more for casual than members. Analysis shows people are taking more rides but for less duration.

4.5 Bike usage heatmap for days of week and months (2020 - 2022)
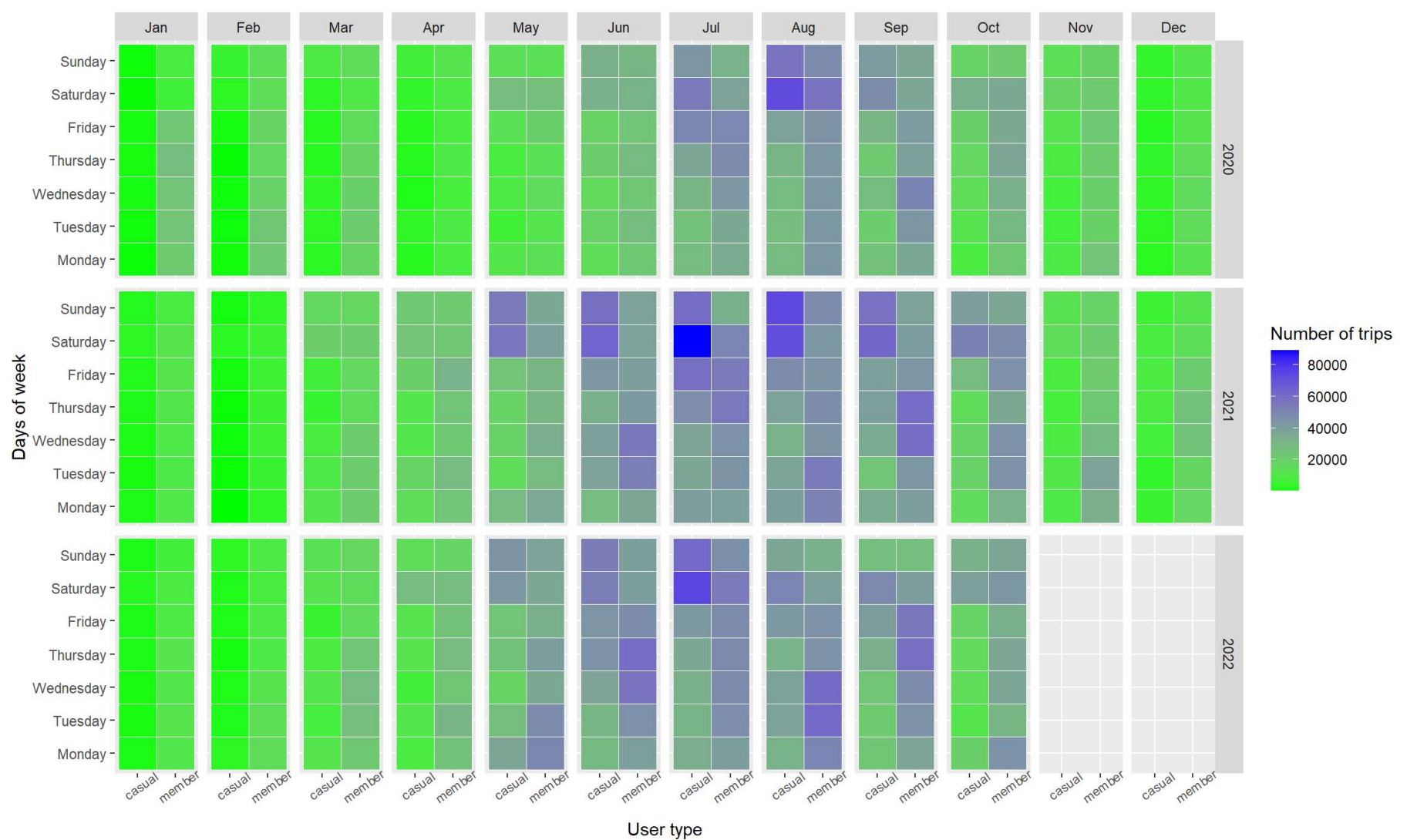
The heat map shows that June, July, August and September are the most popular months for all users. And Saturdays and Sundays are the most popular for casual riders, especially during the summer months. Basically, this means that casual riders mostly use the bike for leisure. Also, members mostly use the bike on weekdays and ramp it up a bit on weekends. During the summer months, members increase their use of a bike for the whole week but interestingly, the days of the week that members use the most during the summer months are Tuesday, Wednesday and Thursday.

4.6 Popular time of the day
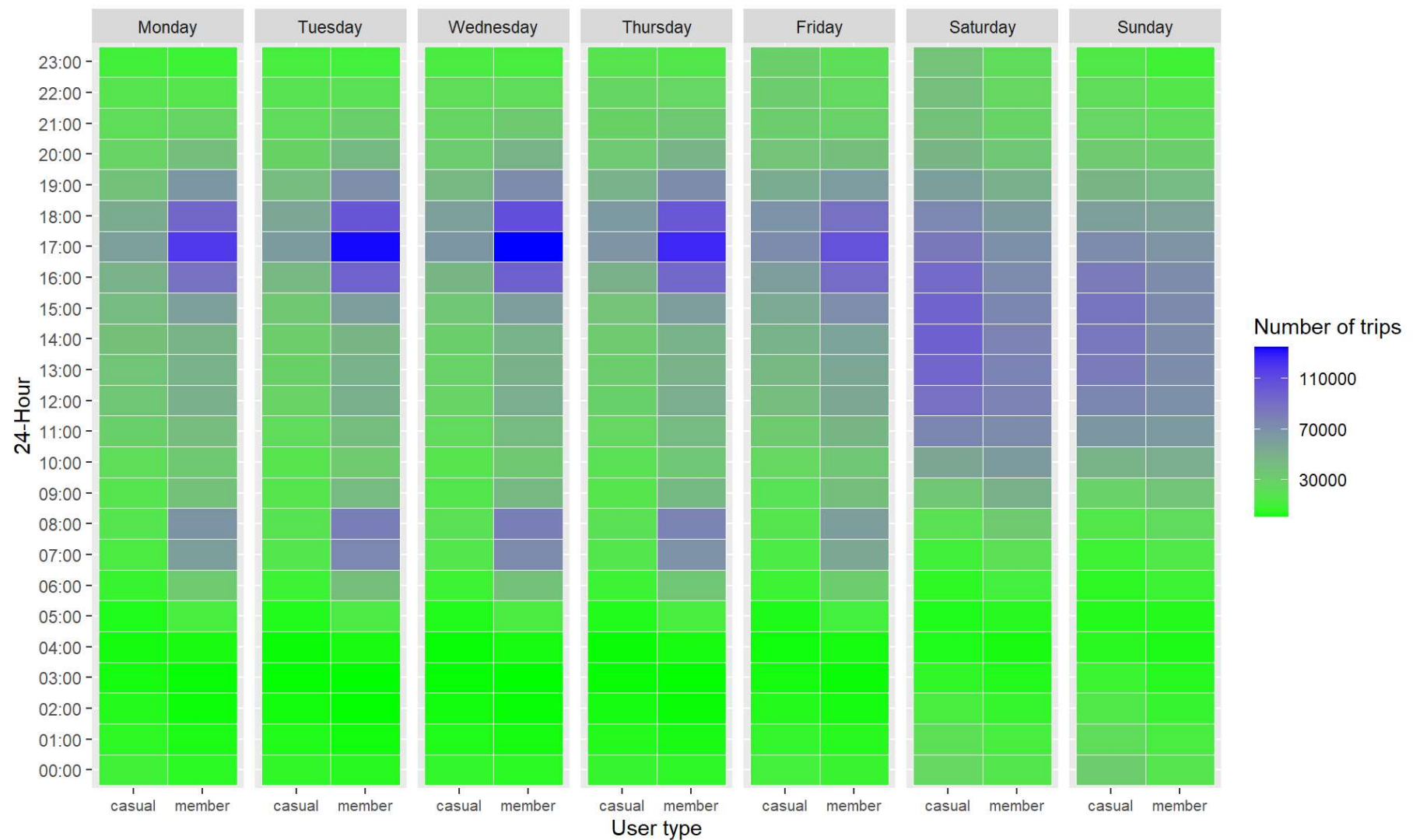
From the figure, we can see that members commute to work using their bikes. It appears that 7am - 8am and 4pm - 7pm are used by most of the members as these are mainly commuting times for the general public. And for this same reason, 5pm is the most popular among members. Also on weekends member users mostly use the bike from 9 am to 7 pm. This shows that they not only use more for commuting but also use bikes for leisure time as well.

Casual riders use the bike from 11 am to 8 pm on weekdays. And rides go up to 1 am on Friday and Saturday nights. This shows that casual riders prefer to take the bike after a night out.

4.7 Most popular station

start_station_name	member_casual	number_of_trips	start_lat	start_lng
Streeter Dr & Grand Ave	casual	143289	41.8923	-87.6120

start_station_name	member_casual	number_of_trips	start_lat	start_lng
Millennium Park	casual	73360	41.8810	-87.6241
Michigan Ave & Oak St	casual	64908	41.9010	-87.6238
Clark St & Elm St	member	62380	41.9030	-87.6313
Kingsbury St & Kinzie St	member	59873	41.8892	-87.6385
Wells St & Concord Ln	member	55886	41.9121	-87.6347
Theater on the Lake	casual	52149	41.9263	-87.6308
Shedd Aquarium	casual	50196	41.8672	-87.6154
Wells St & Elm St	member	49550	41.9032	-87.6343
Dearborn St & Erie St	member	48167	41.8940	-87.6293
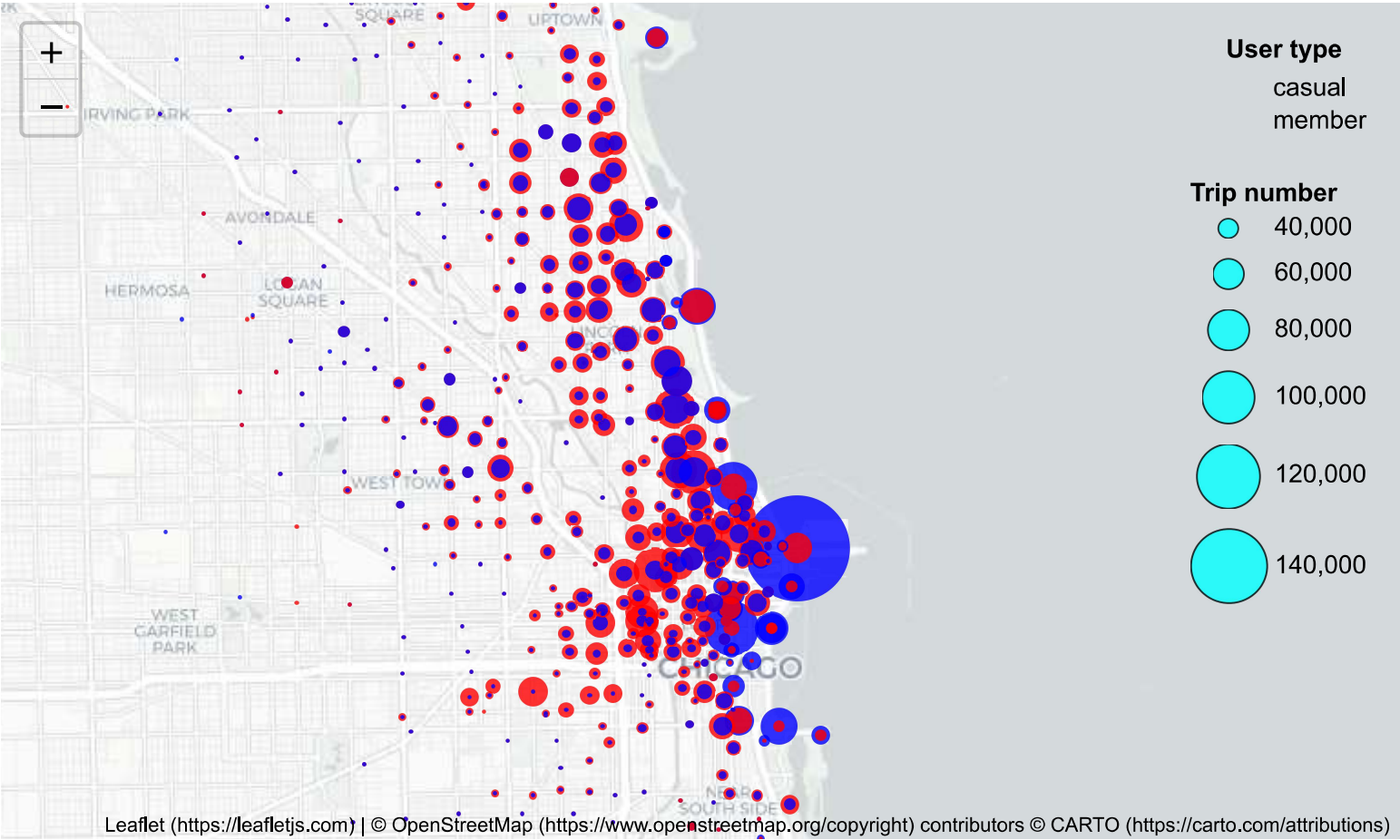
The top three stations for casual users are close to tourist spots. And the number one station for casual users is near Navy Pier. So this confirms the idea that casual riders mainly use the bike for leisure time.

Also, the top three stations for members are around commercial areas and close to any transport stations, especially underground railway stations. So it can be said that members mostly use bike for commuting and they usually come from residential area then use bike to reach their final workplace.

4.8 Most popular bike route

Average duration and maximum duration are shown in minutes.

4.8.1 Popular route for casual riders

member_casual	start_station_name	end_station_name	number_of_trips	average_duration	max_duration
casual	Streeter Dr & Grand Ave	Streeter Dr & Grand Ave	27716	48.20	4035.87
casual	Millennium Park	Millennium Park	14612	49.26	10245.48
casual	Michigan Ave & Oak St	Michigan Ave & Oak St	13878	51.81	1328.63
casual	Lake Shore Dr & Monroe St	Lake Shore Dr & Monroe St	10337	50.50	2874.45
casual	Buckingham Fountain	Buckingham Fountain	9425	67.32	8655.93
casual	DuSable Lake Shore Dr & Monroe St	DuSable Lake Shore Dr & Monroe St	9387	37.09	1517.80
casual	Indiana Ave & Roosevelt Rd	Indiana Ave & Roosevelt Rd	8458	53.91	1607.18
casual	Theater on the Lake	Theater on the Lake	8229	50.00	2828.20
casual	DuSable Lake Shore Dr & Monroe St	Streeter Dr & Grand Ave	7584	28.46	1431.88
casual	Michigan Ave & 8th St	Michigan Ave & 8th St	7386	55.80	2486.65

4.8.2 Popular route for members

member_casual	start_station_name	end_station_name	number_of_trips	average_duration	max_duration
member	Ellis Ave & 60th St	Ellis Ave & 55th St	9314	5.58	1148.15

| member_casual	start_station_name	end_station_name	number_of_trips	average_duration	max_duration
member	Ellis Ave & 60th St	University Ave & 57th St	8716	5.48	1506.53
member	Ellis Ave & 55th St	Ellis Ave & 60th St	8366	5.67	724.32
member	University Ave & 57th St	Ellis Ave & 60th St	8293	5.20	1106.27
member	State St & 33rd St	Calumet Ave & 33rd St	5341	4.70	1010.73
member	Calumet Ave & 33rd St	State St & 33rd St	5295	4.34	1144.65
member	Loomis St & Lexington St	Morgan St & Polk St	4966	5.60	86.57
member	Morgan St & Polk St	Loomis St & Lexington St	4728	6.17	782.73
member	University Ave & 57th St	Kimbark Ave & 53rd St	4025	7.79	455.05
member	MLK Jr Dr & 29th St	State St & 33rd St	3842	7.82	180.40

4.8.3 Bike route visualization
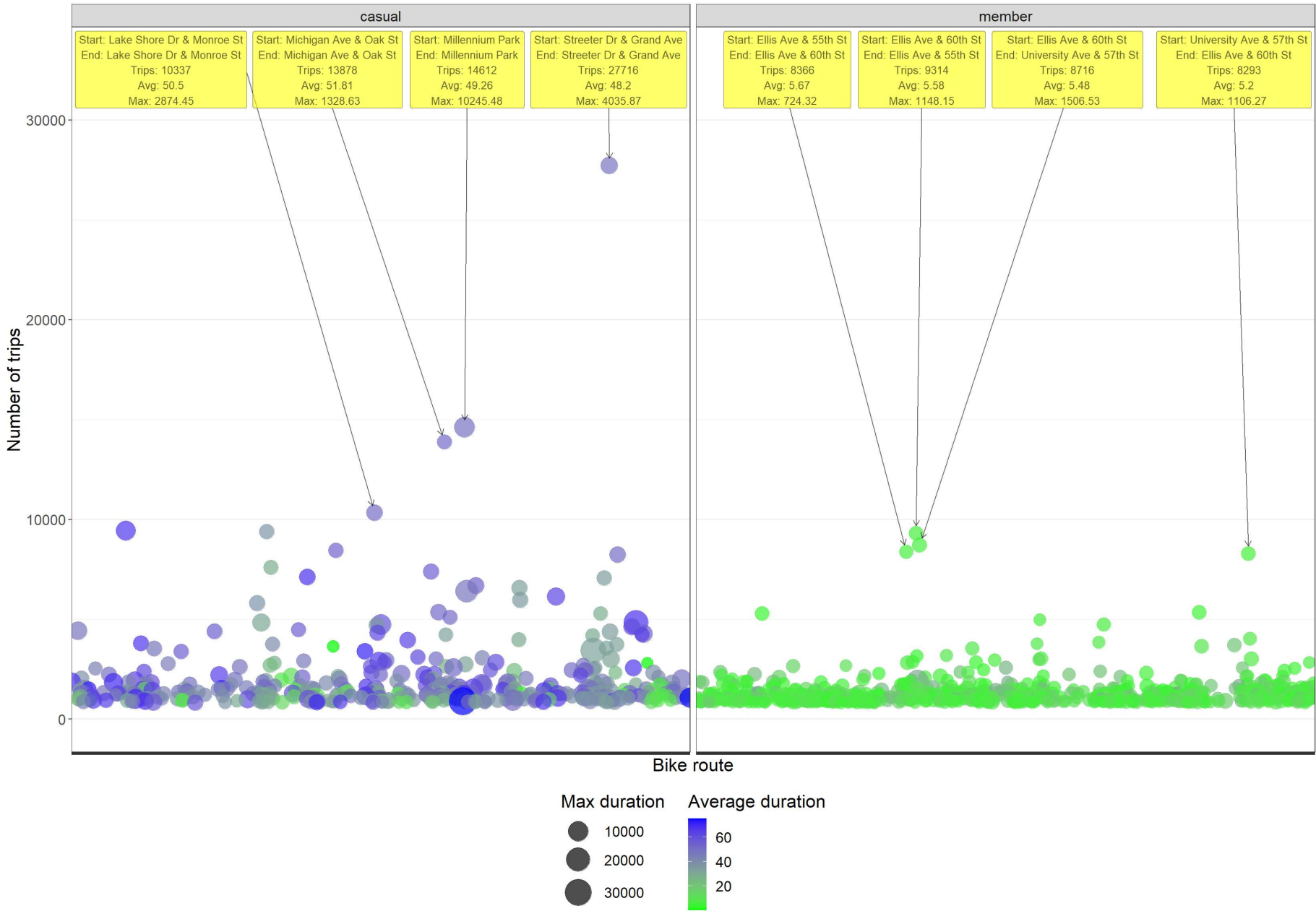
Casual riders usually pick up and drop off from the same station. Also, the average ride duration of the top 5 stations is more than 45 minutes and the maximum duration of a station is around 3 days. Additionally, most stations are located in tourist areas of Chicago. So from the analysis it is clear that most of the time causal riders use bikes in leisure time. And because of that, their average duration is higher than member users.

The top 5 bike routes for members are in the University of Chicago area. Here, university students, teachers and staff mainly go to the destination by short journey by bike. Basically they use bikes to go to different places in the university. For this reason the average duration is less than 10 minutes. Also since they use bikes around the university the maximum duration of this route is not more than 24 hours whereas the maximum duration for casual users is around 72 hours in tourist areas.

4.9 Bike type preference
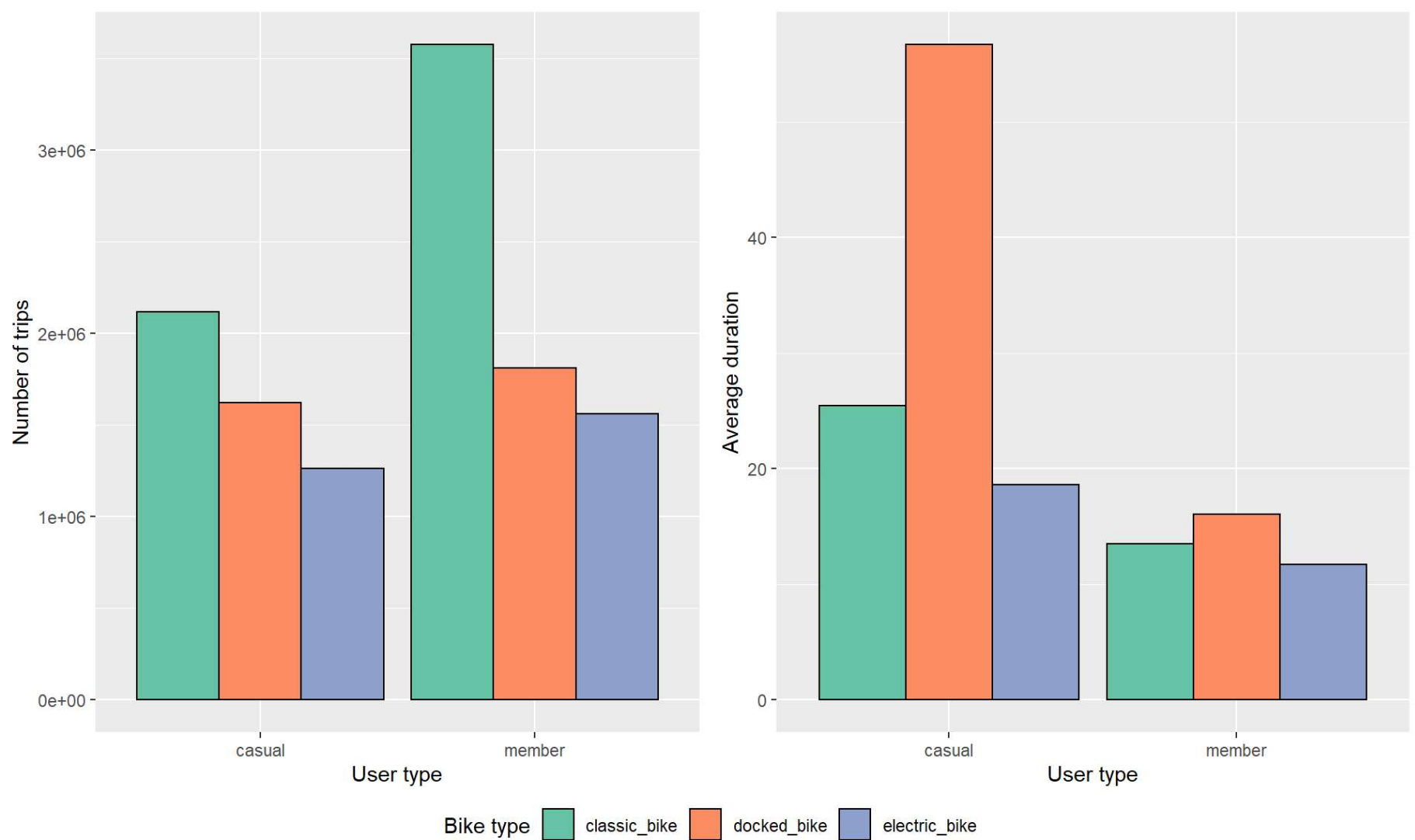
From the figure, it can be seen that there are three different bike types for Cyclistic. Both casual riders and members use classic bikes the most followed by docked bikes and electric bikes. From usage, it can be assumed that the classic bike is the cheapest among them. But interestingly docked bikes are used more on average duration basis and preferred by most users for its price and comfort.

5. Recommendations

The project brief states that the focus should be pointed to the different uses of the bike and the ultimate goal is to encourage casual riders to convert to member users.

Based on the analysis, some recommendations can be given to Cyclistic.

1. From the analysis, it is found that although casual riders mainly use the bike for leisure time, some casual riders also use the bike for weekday commuting. To attract those casual riders, a campaign can be launched that will send app notifications or emails informing them about the benefits of an annual membership that will ultimately save them money in the long run.
2. Casual riders often use the bike for very long periods such as more than a day. So casual riders can be advised that if they convert to a membership, they can use the bike more freely without having to buy a day pass for each day.
3. Some casual riders use the bike more frequently than others so notifications or emails can be sent after completing certain trips to remind them of the money they can save from the membership.
4. Since casual riders use tourist areas more than members, street billboards in tourist areas can be used to attract casual riders to convert into annual members.
5. Finally, casual riders can be offered a promotional code for a limited time to get a discount on their first annual membership price.