Google Data Analytics Capstone Project: How Does a Bike-Share Navigate Speedy Success?

Brainstorming

About the company:

- Location: Chicago

- 5.824 bicycles

- 692 docking stations

- Types of bike: reclining bike, hand tricycle, cargo bike, traditional bike

- Unique point: offering suitable bikes for *people with disabilities* and *riders who can't use a standard two-wheeled bike*

- Pricing plans: single-ride passes, full-day passes, annual memberships

Statistics about the customer habits:

Type of bike:

- Most of users opt for traditional bikes

- ~ 8% use the assistive option

Purpose of use:

- Main use is for leisure

- ~30% use to commute to work everyday

Marketing channel:

- Email

- Social media

Reason for the analyst (new marketing strategy):

- Why: profit from annual members > casual riders → increase the annual mem (maximum)?

- Goal of campaign: converting casual riders into annual members

- How: understand user behaviors, why they should buy a membership…

I. Ask

Identifying business tasks:

These are three questions that need to be answered:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

How to approach?

Finding out the trends and correlation between variables in the data with the help of visualizations and statistical analysis then drawing a conclusion of recommendations and suggestions for Cyclistic company to attract more annual memberships.

II. Prepare

The data was provided by Cyclistic, for the business objective, it should be over the past 12 months from January 2021 to December 2021.

On Cyclistic server, data was separated by months and years and saved as .csv files. I stored all needed data then merged them into one file only called **bikeshare.csv.**

  i. The data included 17 fields (13 original ones, 4 added):

In the analysis, I added 4 fields (***ride_length, week_day, month, start_hour***) to get more insights from data

I	USER	number of users (by mem type, by bike type)
1	**ride_id**	unique ID per ride
2	**rideable_type**	type of bicycle customer used
3	**member_casual**	type of customer taking the ride, member or casual

II	TIME OF RIDE	- frequency, top weekday, top month, what is special in top time? → find common point of 2 types of user - avg length per day (check how long each mem type use a day, maybe concern the start time?), whether they different or not (the basis of determining whether casual turn into annual or not)
4	**started_at**	the time bike was checked out
5	**ended_at**	the time bike was checked in
6	**ride_length**	the total time of the trip calculated as ended_at - started_at
7	**week_day**	day of week the ride started
8	**month**	month of year the ride started
9	**start_hour**	start time of the ride
III	STATION	top departure and end station, their locations, why people start and end there? → their habits, reason of riding
10	**start_station_name**	name of departure station
11	**start_station_id**	unique id of start station
12	**end_station_name**	name of station at the end of the ride
13	**end_station_id**	unique id of end station
14	**start_lat**	latitude of start station
15	**start_lng**	longitude of start station
16	**end_lat**	latitude of end station
17	**end_lng**	longitude of end station

III. Process

I used **Python** for the whole project. The process is detailed in **cyclist.ipynb** file

**Note: I didn't drop null values in this project to get a better overview at this dataset.**

IV. Analysis

I broke the analysis section into 3 part: *Number of rides by time, The length of ride* and *Number of rides by station*

1. Number of rides by time and bike type

According to the number of rides by member type in 2021, member riders commuted more than casual riders, the difference was about 10%.

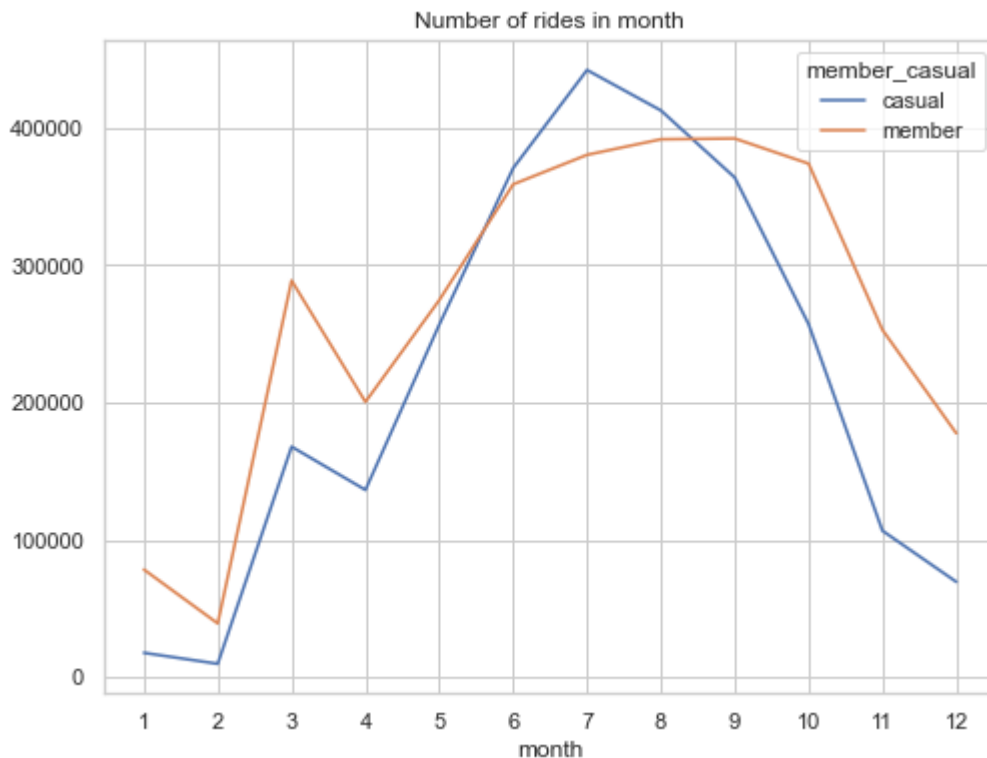| Member type	Number of ride
casual	2613038
member	3210521
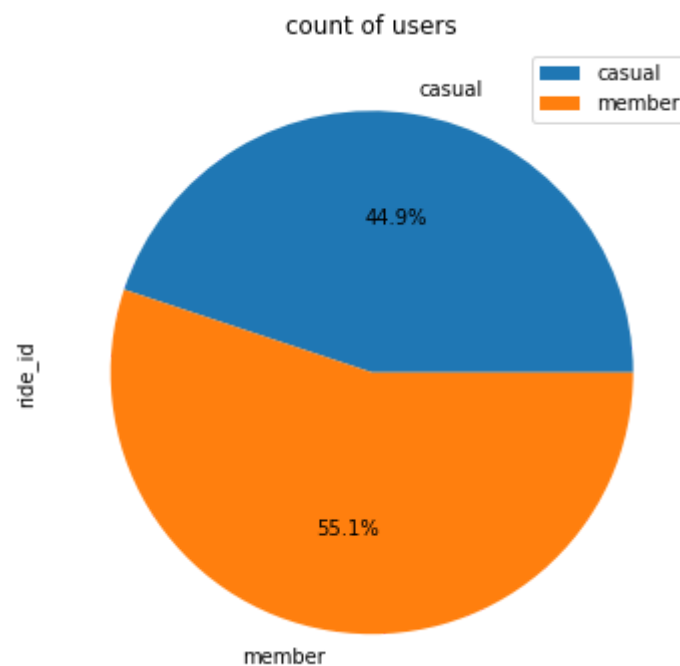
Number of rides by month

- Casual riders rode less in Jan - Feb, rode more from Jun to Sep. The reason for this change was because Jun - Sep was the time for holiday and they wanted to ride more for leisure and casual riders tended to decrease the need of using bike in the winter time.

- Regarding the member riders, their number of ride peaked in Jun - Oct and decreased in Jan - Feb.

→ *The trend of 2 member types was the same, they rode more in Summer - Fall time and rode less in Winter time.*

Number of rides in month

For the number of rides by bike type, classic bike and electric bike was the two most common bike type used by riders.

- Casual riders: they preferred classic bike to electric bike. Electric bike uses outnumbered classic bike from late Sep to Dec.

- Member riders: Used classic bike more most of the time, electric bike outnumbered in Nov - Dec.

*→ Classic bike was most common type, electric bike tended to be used more in the end of the year.*
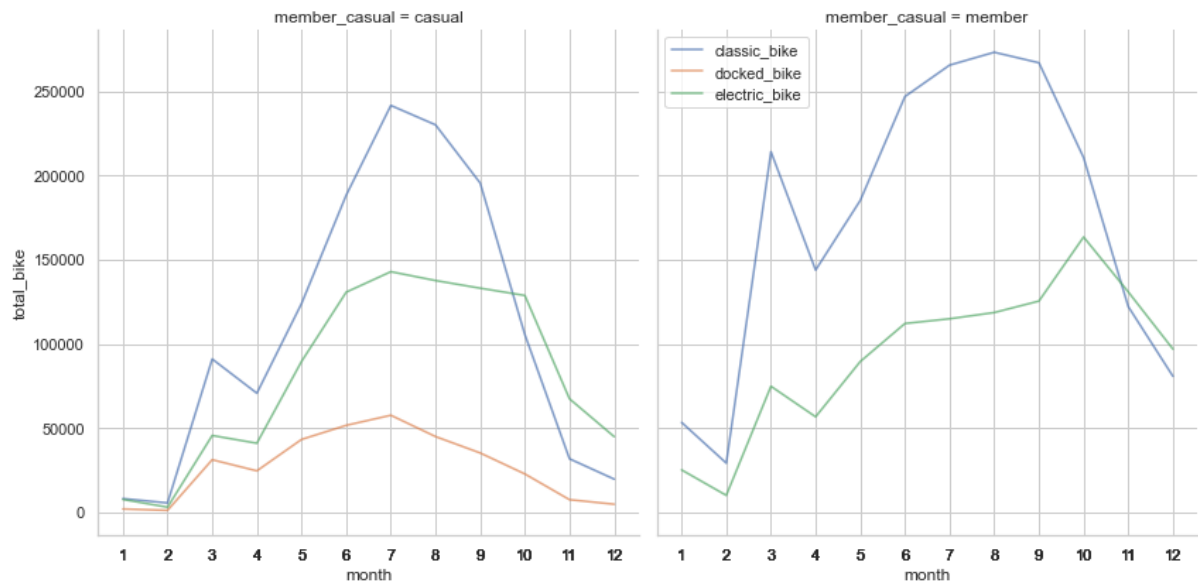
Number of rides by weekday

Weekday code:

Number	Weekday
1	Saturday
2	Sunday
4	Tuesday
5	Wednesday

- Casual riders rode more at the weekend.

- Member riders' number of rides distributed evenly but decreased at the weekend.

→ Casual riders mainly used for leisure at the weekend, member riders mainly used for commuting to work (weekday).
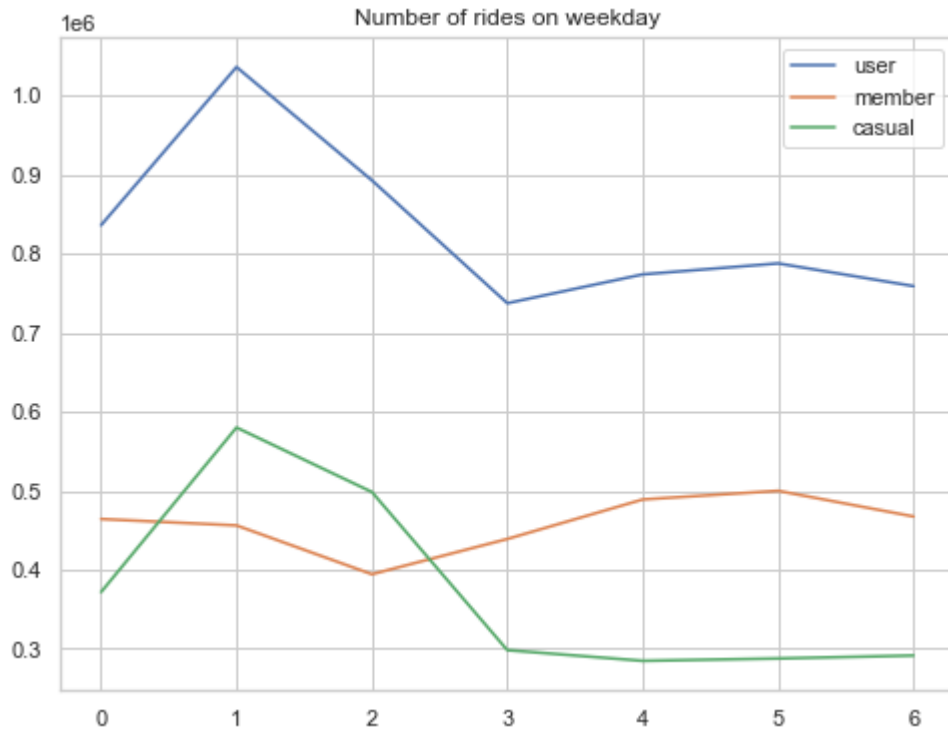
For the number of rides by bike type, classic bike was favorite type for both leisure and work.
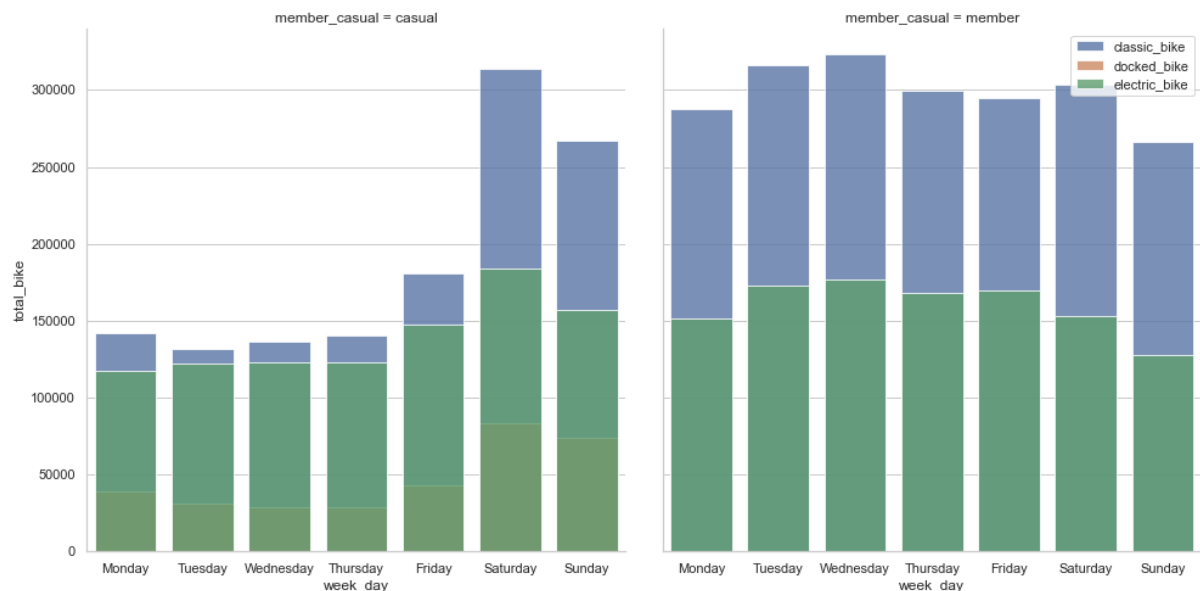
Number of ride by start time

- Member rider: there were 3 peak in the number of rides changes by start time, at 8 am, 12 am and 17 pm. These are all shift time at work.
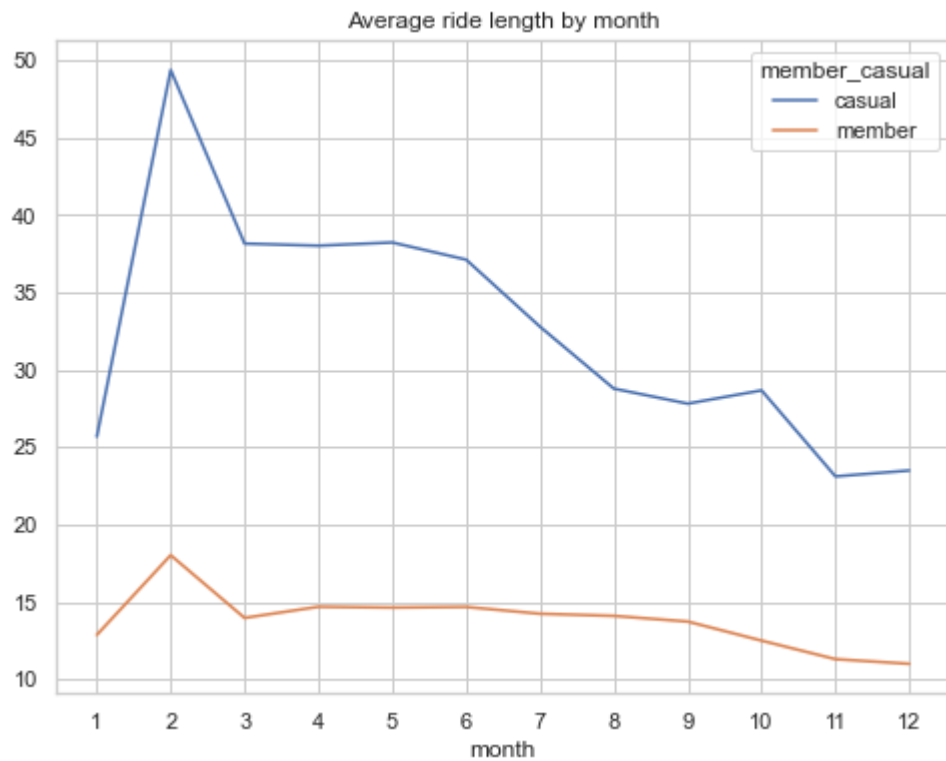
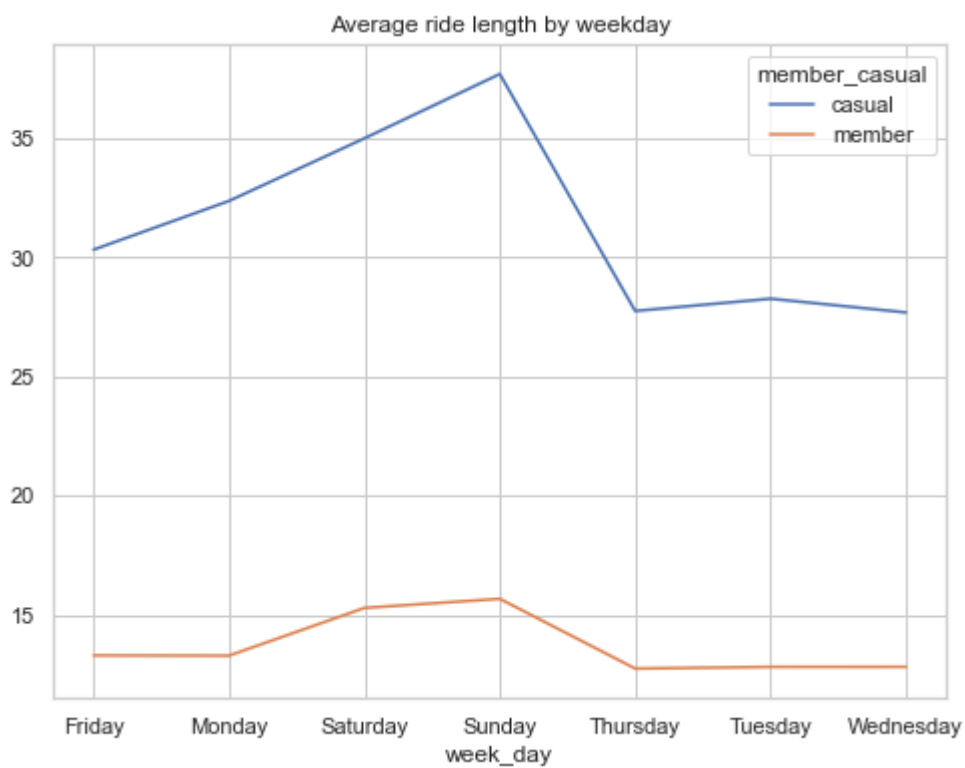- Casual rider: they rode more in the afternoon.

Number of rides in start time

2. The length of ride

The average ride length by month of casual riders was twice to fourth as much as member riders'.

Average ride length by month

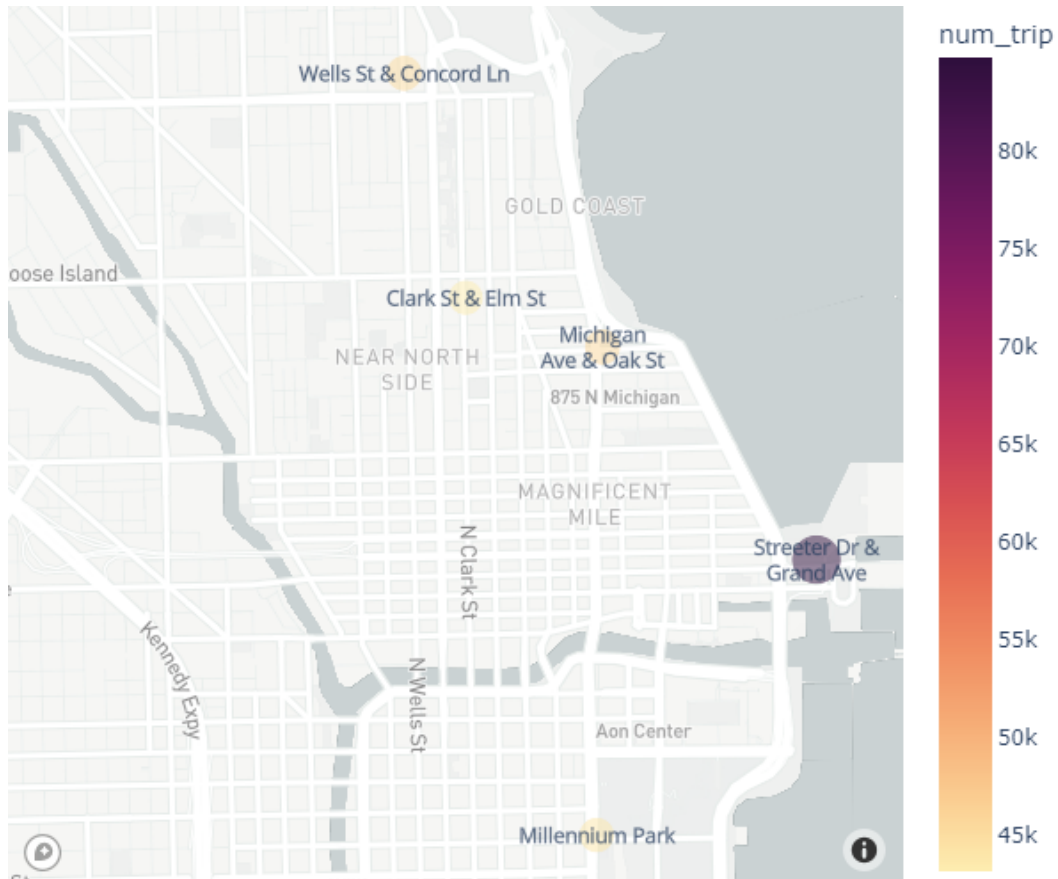This trend was the same by weekday.

Average ride length by weekday

→ *The number of ride of casual riders was fewer than member ones but their average ride length was much longer.*
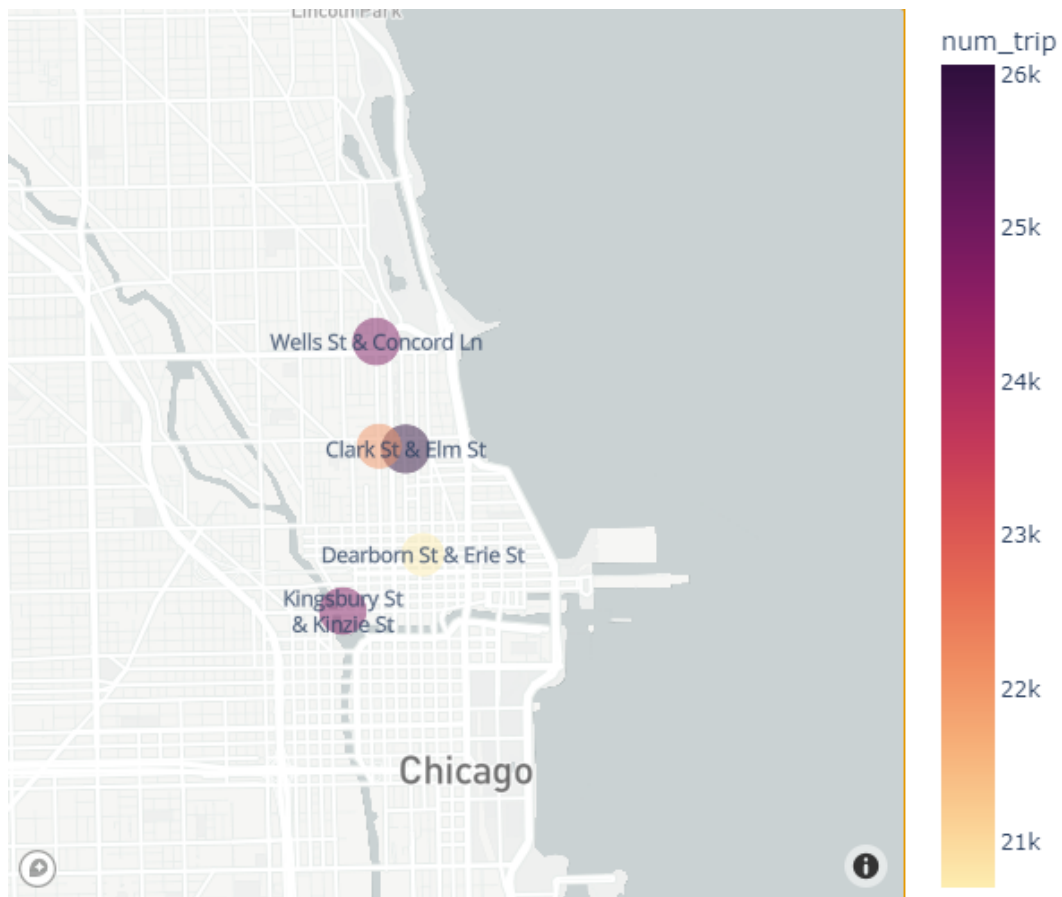
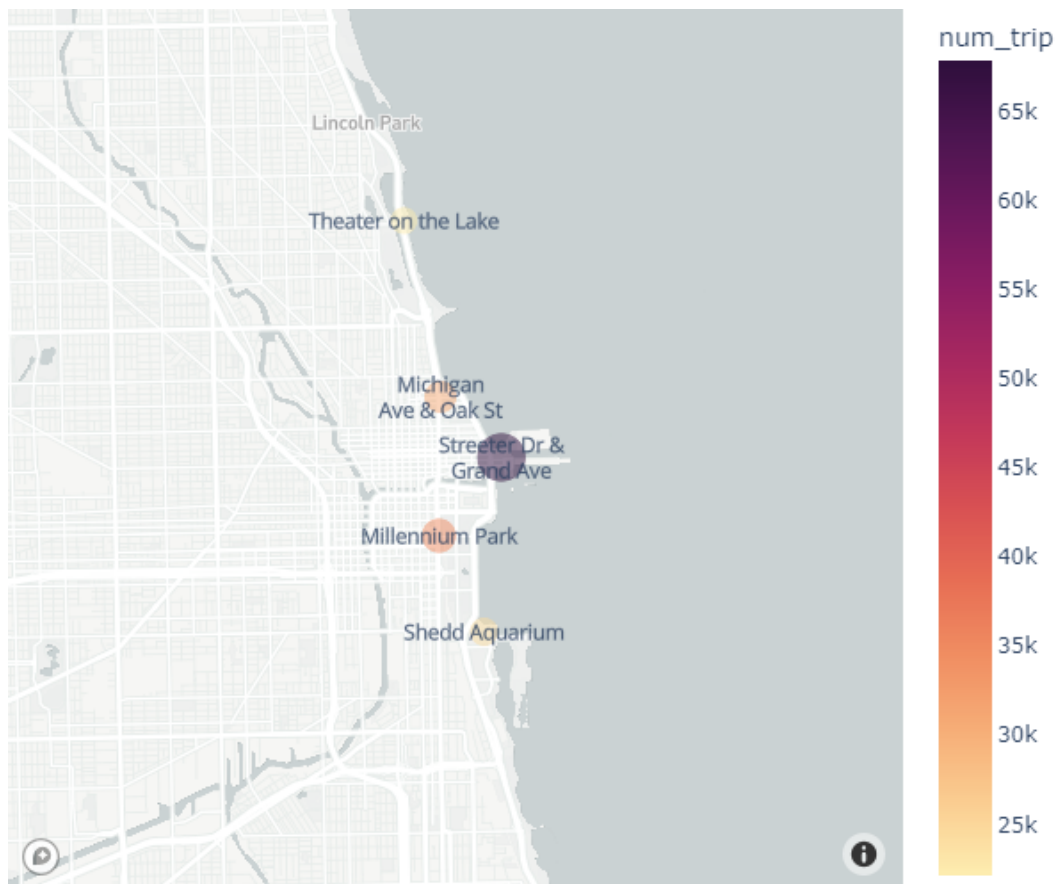# 3. Number of rides by station

Start station

Top 5 start stations used by users were all near the Lake Michigan.

Member rider: they were likely to start at inner street, which made it convenient to go to work.
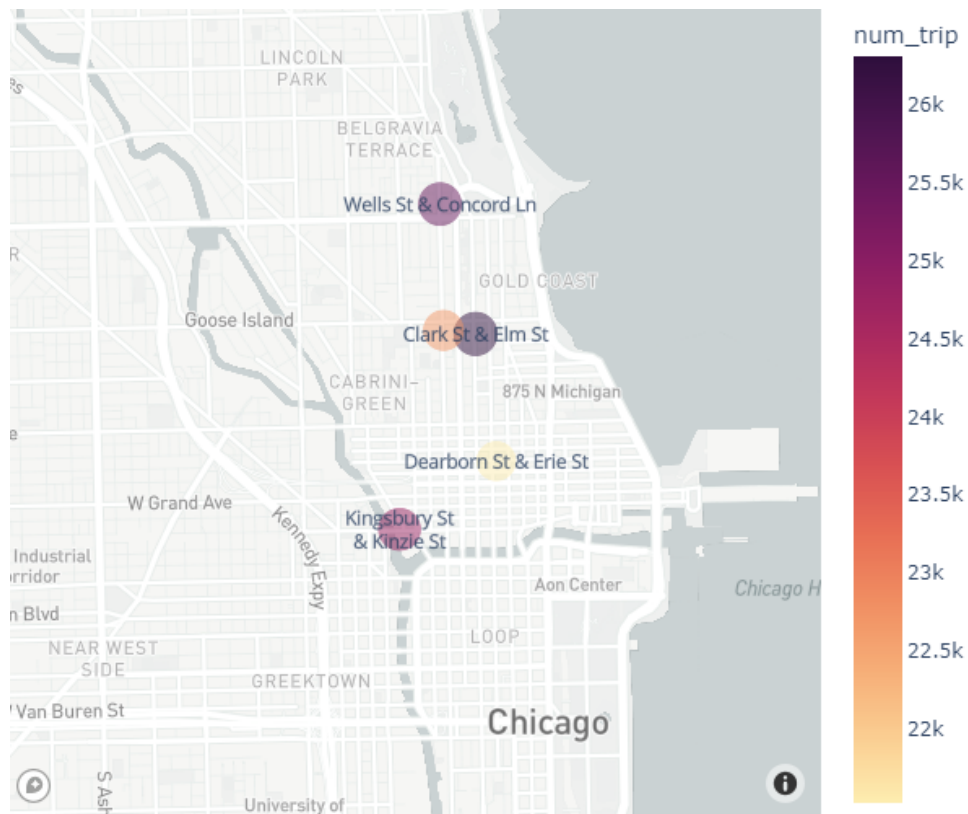
Casual member: was likely to start at the station near the Lake Michigan or Park, rode along the Lake or in the park.
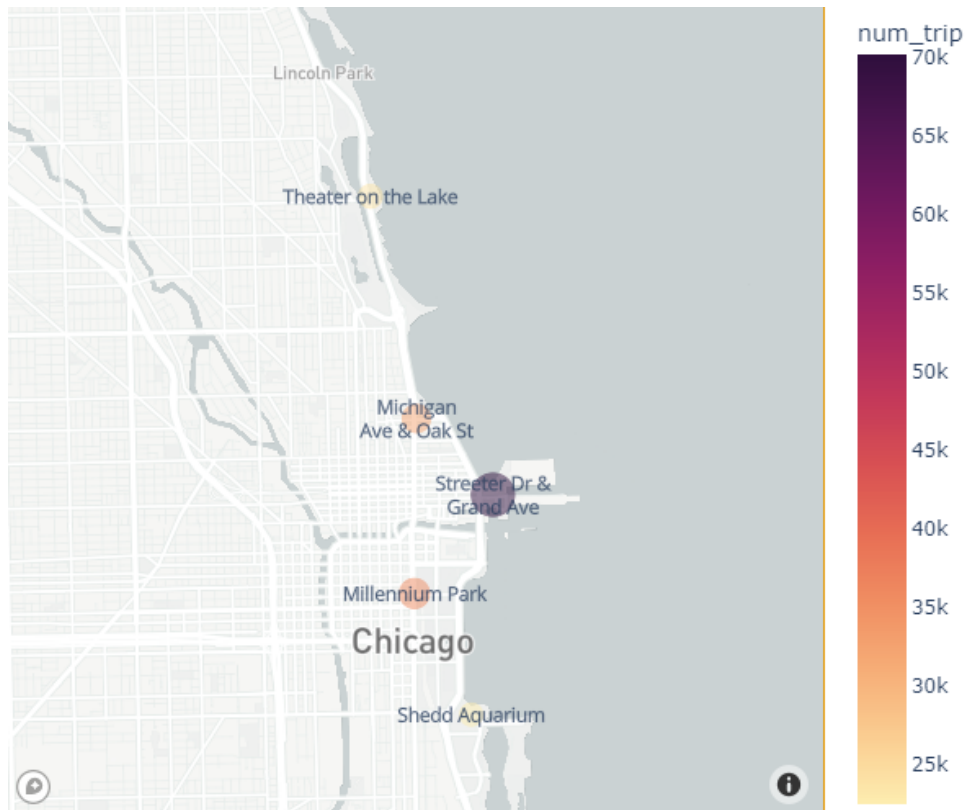
End station

Member rider: top 5 end stations were the same with start station → *many member riders lived near these streets.*

Casual riders also tended to end the ride at the start station.

→ *They only came here for leisure therefore after the ride, they returned bike at the start point.*

V. Recommendation

1. Redeem points for gifts, the point is calculated by the total length, ride more → more point.

2. Casual riders tend to use the service more during the weekends, therefore maybe a promotion through Facebook, Google, Tiktok,... to attract them to use service more on weekdays would be reasonable.

3. Casual riders ride mainly in the afternoon and weekend → can attract them by the annual or monthly packet with first-time special promotion on these time.

4. Their average ride length was more than member riders → they rode more so it would be economical for them to choose the annual packet.

5. Casual riders are likely to ride more in summer and autumn time so the promotion during this time would be more effective.

6. Top 5 start stations used by casual riders are all riverside, advertisements along river would appeal the riders.