

## Kuan Yew Cheng

### Background and Aim

In this case study, I am a junior data analyst in the marketing analyst team working for Cyclistic, a fictional bike-share company based in Chicago. Since its launch in 2016, Cyclistic 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's marketing strategy relied on appealing to broad consumer segments. Cyclistic has 3 flexible pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic's finance analysts have concluded that annual members are much more beneficial than casual riders. In order to provide future growth, the company comes up with new marketing strategies to increase the annual membership from casual riders. Thus, I am been told to analyze Cyclistic historical bike trip data to understand how casual riders and annual members use Cyclistic bikes differently, to identify trends and insight for creating data driven strategies.

### Data Source

Data was collected and stored at <https://divvy-tripdata.s3.amazonaws.com/index.html>. This dataset is an actual public data. It has been made available by Motivate International Inc. under this license at <https://www.divvybikes.com/data-license-agreement>. The data used was made from October 2020 to September 2021, which is stored as monthly data. The monthly data is a CSV file with 13 variables. The data contains the following columns:

No.	Column Name	Description
1	ride_id	Unique identifier of each ride booking
2	rideable_type	Type of bicycle used
3	started_at	Datetime of the start of the trip
4	ended_at	Datetime of the end of the trip
5	start_station_name	Name of the start station
6	start_station_id	Unique identifier of the start station
7	end_station_name	Name of the end station
8	end_station_id	Unique identifier of the end station
9	start_lat	Latitude of the start station
10	start_lng	Longitude of the start station
11	end_lat	Latitude of the end station
12	end_lng	Longitude of the end station
13	member_casual	User type (either annual member or casual)

### Data Preparation

The dataset was imported into a Microsoft SQL Server database. If the "started\_at" and "ended\_at" columns are not correctly formatted as a datetime Conditional split was used to ensure that the rows where these columns are filtered out, as the data in these columns are crucial for analysis. There were no errors in these two columns and all the rows were successfully imported. The dataset was also imported into Google Spreadsheet in order to examine data type and missing data.

Furthermore, the data for 12 months was combined into one table or one dataset and the following new columns was created:

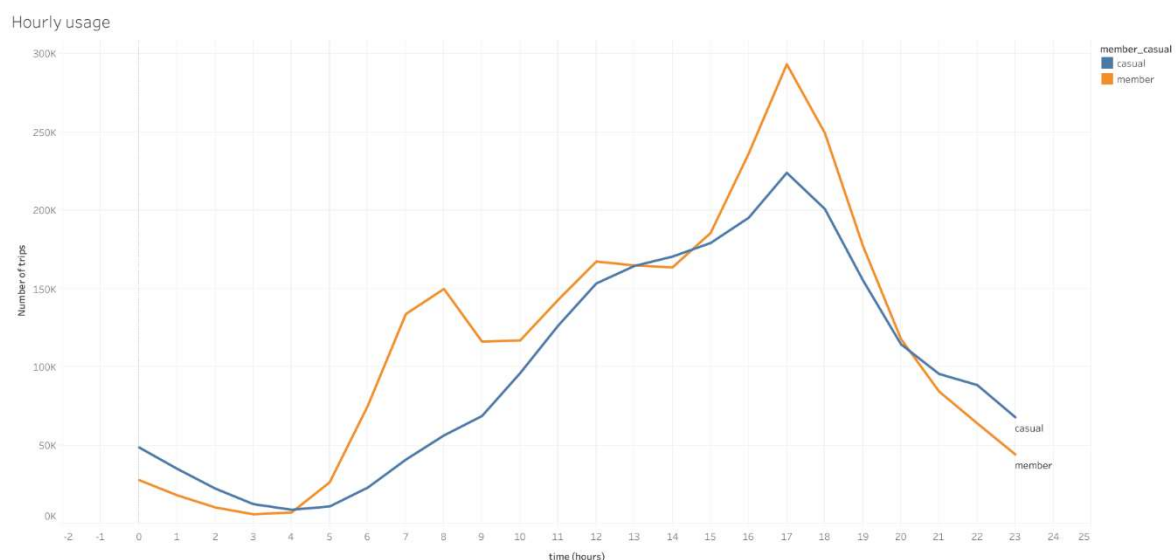
No.	Column Name	Description
1	ride_length	Duration of trip in minutes
2	day_of_week	Day of the trip
3	start_hour_of_day	Hour of the day when the trip starts
4	end_time_of_day	Hour of the day when the trip ends

The important information about the data:

- The dataset contains a total of 5,184,423 rows.
- For simpler calculation, the huge dataset was separated for simple analysis. In this case, a dataset that spanned from October 2020 to February 2021. 59,497 rows of 926,398 rows have a trip duration of less than 3 minutes, which is 6 % of the data. For the analysis, only trips 3 minutes or greater will be taken into account to remove any potential bookings made in error by the user.
- 71,340 rows have null values in the start\_station\_name columns in the small dataset. These rows will still be used in the overall analysis but will be filtered out when analyzing the most popular stations.

## Analysis

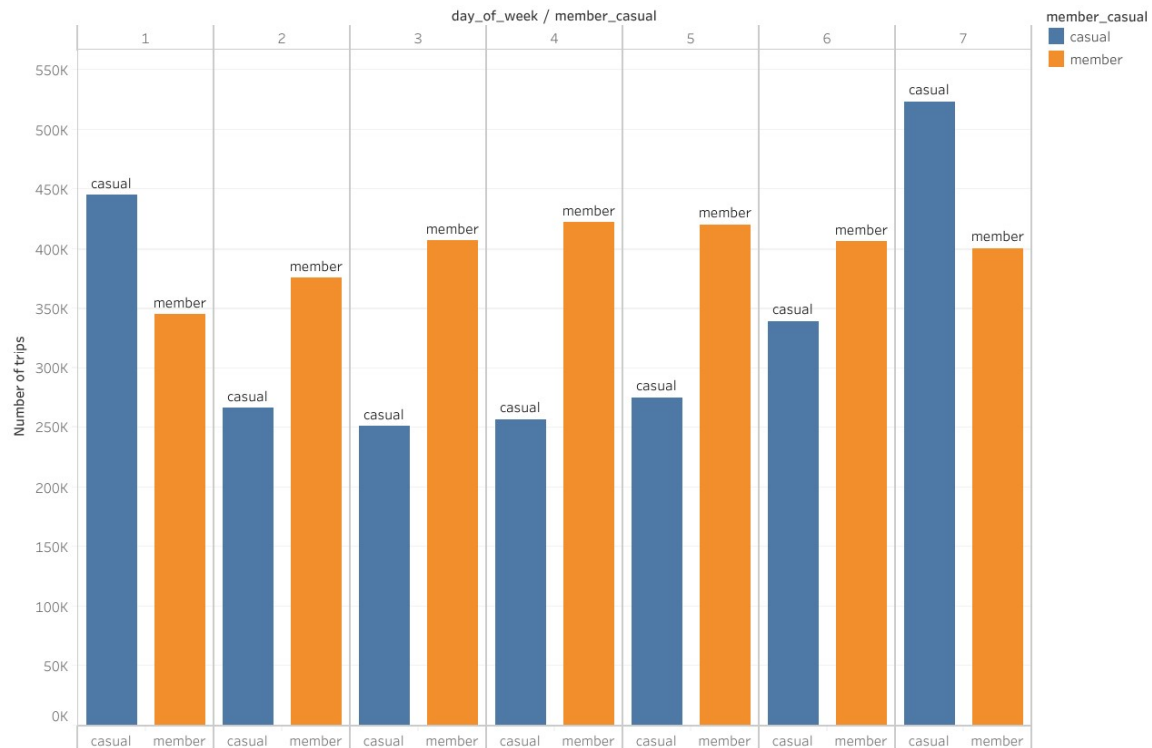
First, let's look at the hourly usage trends. Here we can see that members' usage has two peaks, the first around 8 a.m. and the second around 5 p.m. corresponding with the start and end of the workday. Casual riders on the other hand, start using the bikes more beginning from mid-day until evening.



Then, we can see an interesting trend whereby the number of trips made by members are slightly more than the trips made by casual riders in the 12 months. This is possibly because members use the bikes just to get from point A to point B, while casual riders use them for leisure.

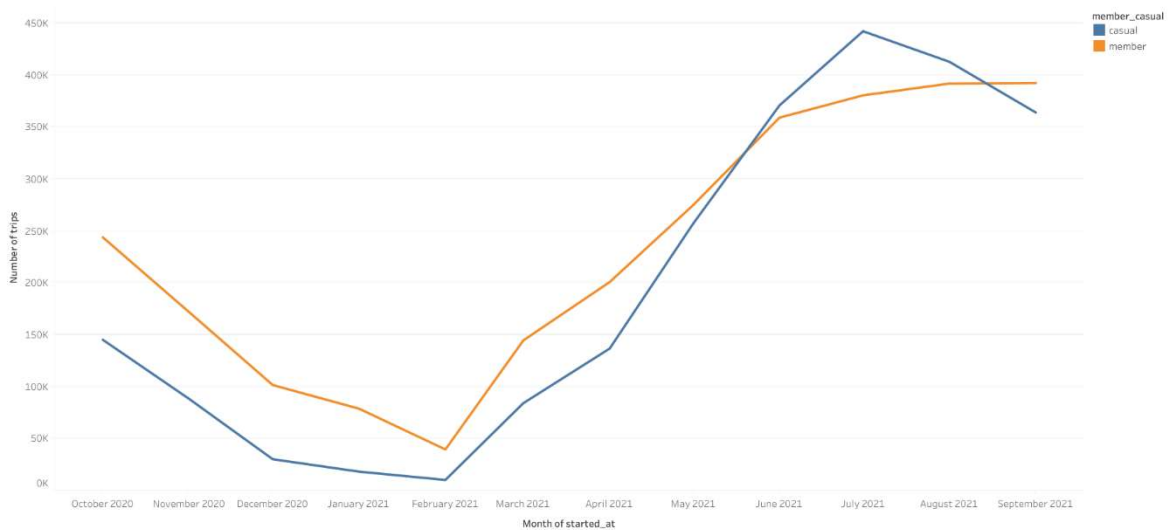
In terms of daily usage, the data shows that members' usage trend remain fairly consistent throughout the week. However casual riders use the bikes more during Monday and the weekend, with the number of trips on Sunday even surpassing that of members.

Daily usage



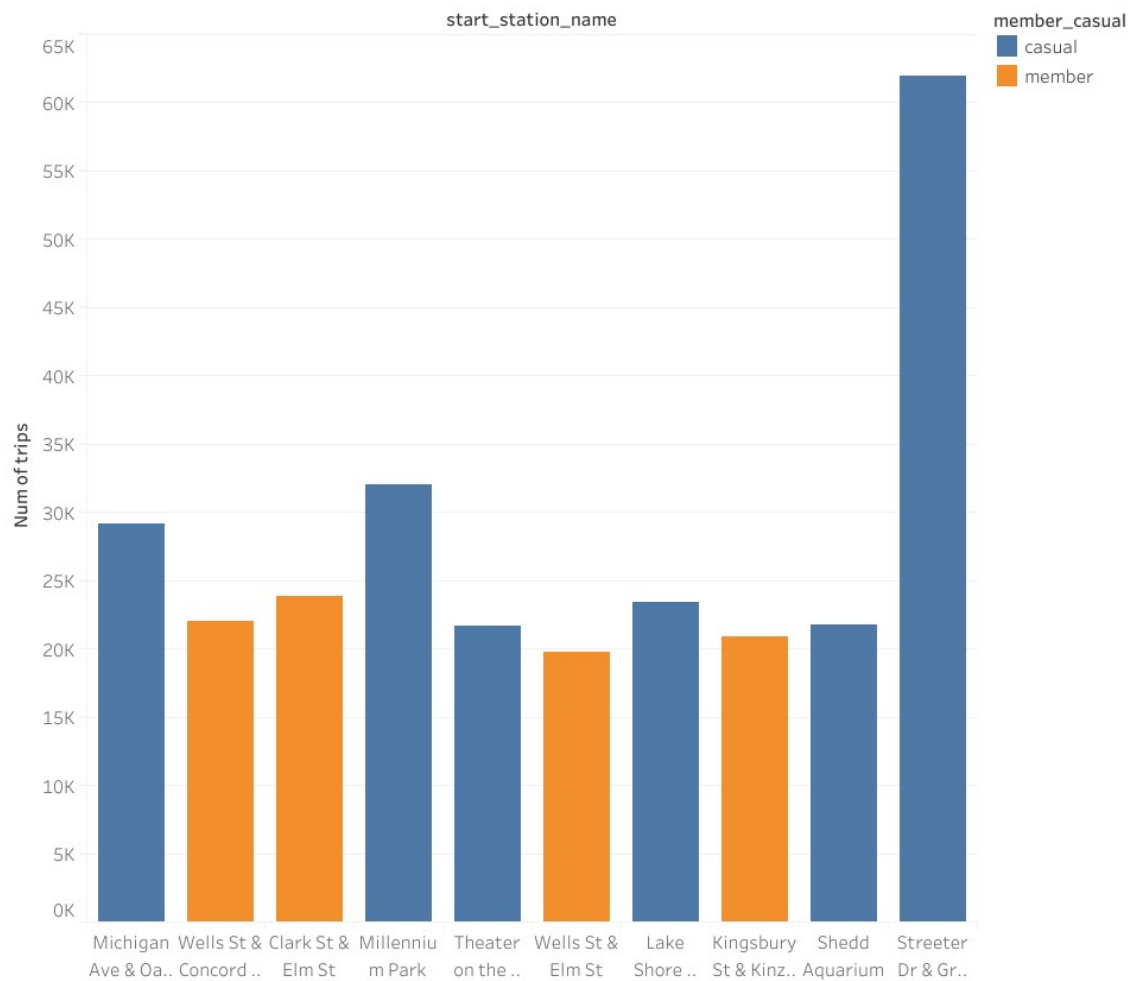
In this monthly usage chart, we can see that both members and casual riders show a similar trend with more trips made in the warmer months, peaking in July, and less trips during winter, with the least trips in February.

Monthly Usage

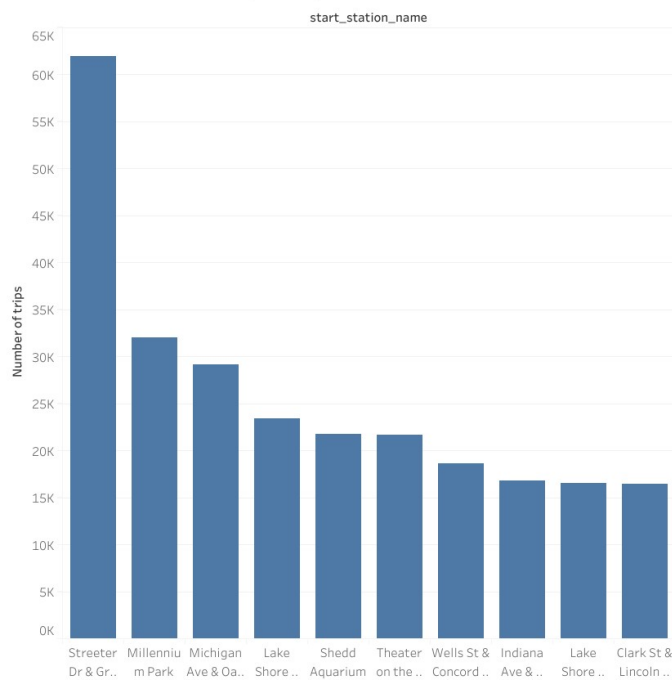


Finally, the following three charts show the most popular starting stations in Cyclistic's network. It is interesting to note that Streeter Dr & Grand Ave, Michigan Ave & Oak St and Millennium Park are significantly more popular among casual riders than they are among members. The fact that these stations are located in a tourist area overlooking Lake Michigan could be a reason behind this observation.

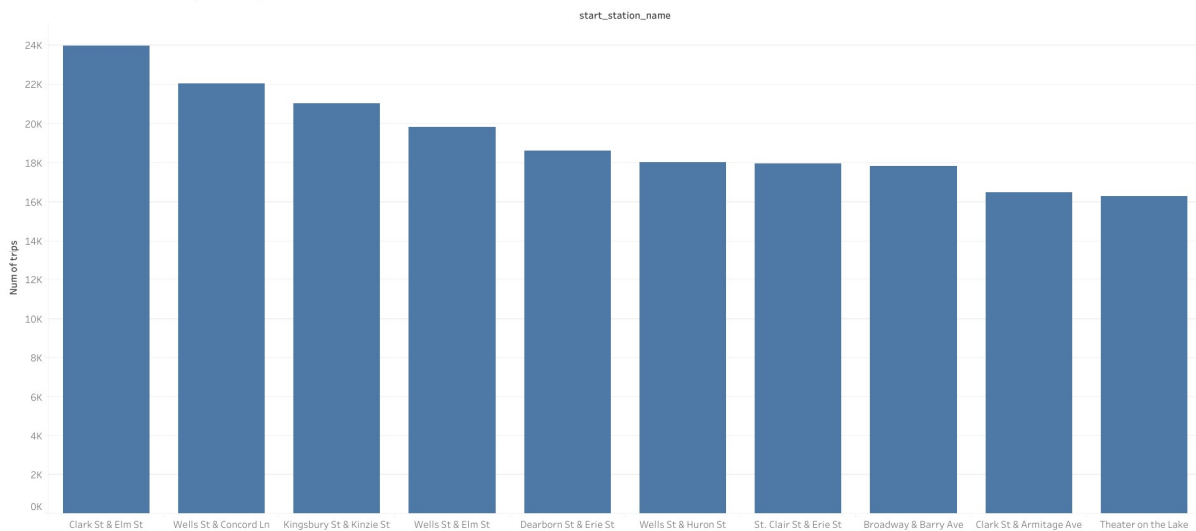
Top 10 Starting Stations (Total)



Top 10 Starting Stations (Casual)



Top 10 Starting Stations (Member)



View the dashboard for this case study at my Tableau Public profile:

<https://public.tableau.com/app/profile/kuan.yew/viz/Cyclisis/Dashboard1#1>

## Conclusion

To guide the marketing campaign to attract more casual riders to convert into annual members, we now have some data driven insights on how casual riders and annual members use Cyclistic bikes differently. The key findings and my recommendations for the marketing campaign are as follows:

1. Casual riders prefer to take longer trips compared to members.

Use this statistic to show casual riders how they could save more money in the long run by becoming a member instead of paying for rides based on trip duration.

Introduce a member only rewards program for example gift redeem or coupon redeem based on trip duration to incentivize casual riders to sign up as members and be eligible for the rewards.

2. Casual riders prefer to use Cyclistic bikes on the weekends where the number of users are almost twice as much as users in the middle of the week.

Develop a weekend membership plan whereby rides on the weekends are included in the base price and cheaper hourly rate while members have the option to book weekday rides at a lower rate.

3. Bicycles in tourist areas are more likely to be used by casual riders than members.

Develop partnerships with the Chicago tourism department or businesses at these locations to offer promotions for Cyclistic members.

Produce advertisements targeting users who frequent stations located in tourist areas. Tourist pass as membership plan can be offered as value plan for casual riders.