

# Case Study : How Does a Bike-Share Navigate Speedy Success?

Tina Huang 2021-05-24

## Scenario

Cyclistic, a bike-share company in Chicago, features more than 5,800 bicycles and 600 docking stations. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, the analysts want to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, the analysts will design a new marketing strategy to convert casual riders into annual members.

## Table of Content

1. Business Task
2. Analysis Summary
3. Visualizations and Key Findings
4. Top Three Recommendations
5. Description of Data
6. Documentation of Cleaning or Manipulation of Data

## Business Task

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

## Analysis Summary

Create pivot tables to quickly calculate the data and find the trends.

1. Calculate the number of trips, average of ride length, and maximum of ride length for members and casual riders:

<i>member_casual</i>	COUNTA of member_casual	AVERAGE of ride_length	MAX of ride_length	STDEV of ride_length
casual	119861	0:38:37	796:16:00	0.2399
member	176196	0:14:24	24:13:00	0.0138
<b>Grand Total</b>	<b>296057</b>	<b>0:24:12</b>	<b>796:16:00</b>	<b>0.1533</b>

- Numbers of trips from members are more than from casual riders.
- Casual riders prefer to have longer duration for the trip on average.
- The maximum of ride length of casual riders lasts more than a month, which is much longer than the member's ride.
- The standard deviation of the member is smaller than the casual riders', which means the duration of each trip of the members is more stable.

2. Calculate the count of trips for each day of week (Monday=1, Sunday=7):

<i>COUNTA of ride_id</i>	<i>day_of_week</i>							
<i>member_casual</i>	1	2	3	4	5	6	7	Grand Total
casual	13998	17836	10300	10371	19702	24899	22755	119861
member	24475	28092	22348	24391	31323	23332	22235	176196
<b>Grand Total</b>	<b>38473</b>	<b>45928</b>	<b>32648</b>	<b>34762</b>	<b>51025</b>	<b>48231</b>	<b>44990</b>	<b>296057</b>

- Casual riders have the most numbers of trips on Saturday, while members have the most numbers of trips on Friday.
- Members have more trips than the casual riders on each day of week.

3. Calculate the average of ride length for each day of week (Monday=1, Sunday=7):

<i>AVERAGE of ride_length</i>	<i>day_of_week</i>							
<i>member_casual</i>	1	2	3	4	5	6	7	Grand Total
casual	0:37:13	0:40:34	0:39:44	0:24:06	0:41:58	0:36:50	0:43:07	0:38:37
member	0:14:09	0:14:25	0:13:07	0:12:59	0:13:47	0:16:09	0:16:30	0:14:24
<b>Grand Total</b>	<b>0:22:32</b>	<b>0:24:34</b>	<b>0:21:31</b>	<b>0:16:18</b>	<b>0:24:40</b>	<b>0:26:50</b>	<b>0:29:58</b>	<b>0:24:12</b>

- On average, both casual riders and members have the longest ride length on Sunday and the shortest ride length on Thursday.
- Casual riders have longer ride length on every day of week.

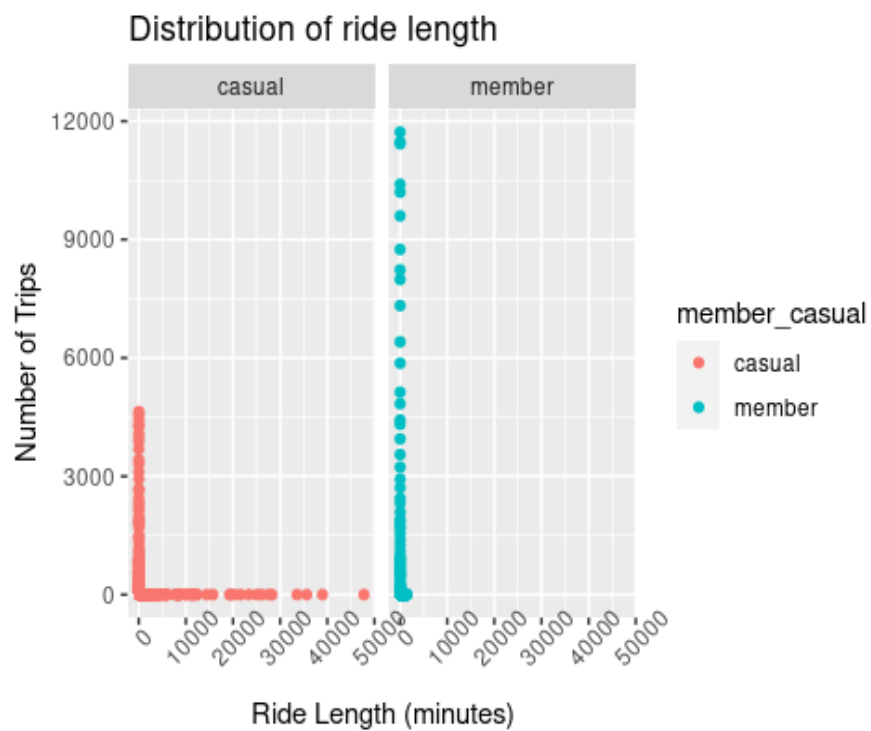
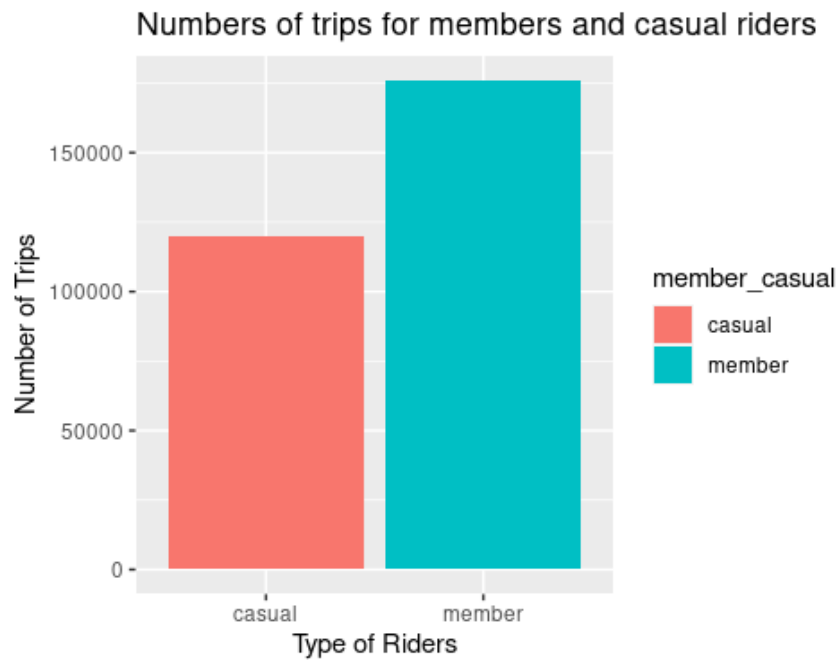
## Visualizations and Key Findings

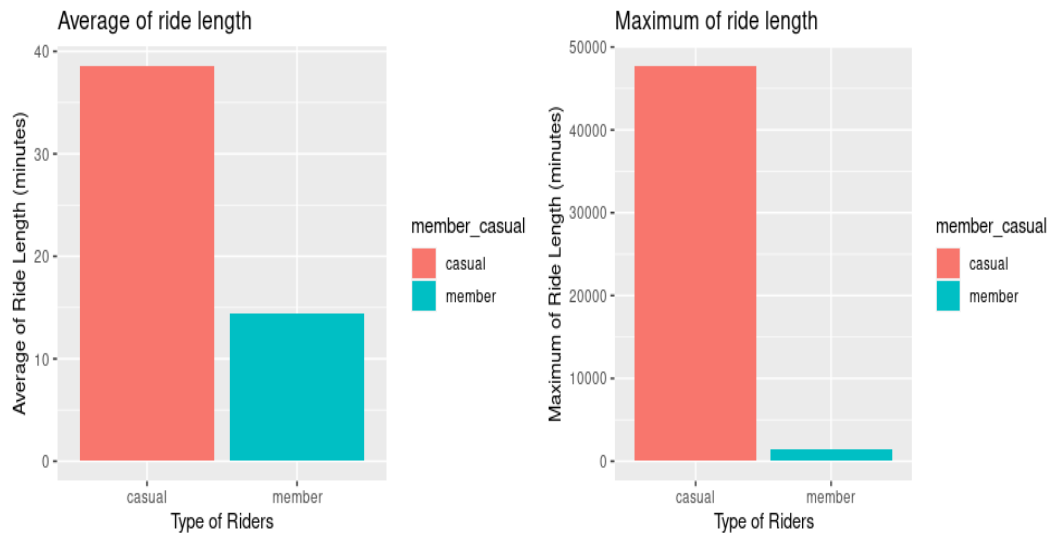
1. Glance of data visualization

There are more annual members than casual riders using bikes in April 2021. The viz of 'Distribution of ride length for members and casual riders,' shows that there is a big difference between the two types of riders.

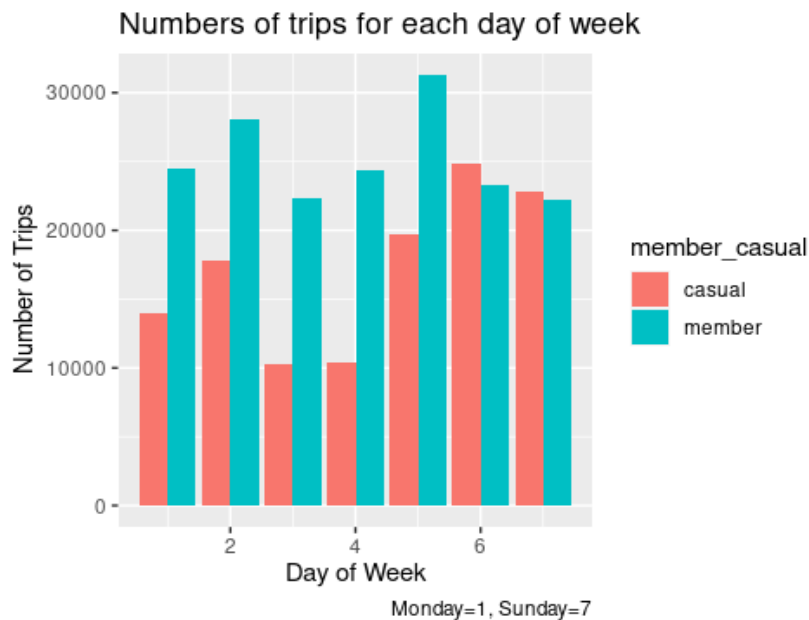
Casual riders have a variety of ride length for each trip. On the other hand, there is a vertical line on the graph for members, having the similar ride length for each trip. The average of ride length for casual riders is two times longer than the members. And the maximum ride length for casual riders is over one month which is much longer than the members.

Casual riders tend to have longer trips but fewer numbers of trips. As the need for using bikes of casual riders is not as frequently as members, casual riders have no incentive to become annual members.





- After grouping by the day of week, I found that casual riders are more active on Saturday and Sunday, tending to use bikes on weekends. That indicates casual riders prefer to use bikes during their leisure time instead of working time. However, members use bikes on weekdays often. It indicates that members tend to commute by Cyclistic's bikes. Since members pay the fees for a whole year, most of the members would like to ride bikes even on weekends. Besides, both of casual riders and members have the most trips on Friday.



- From the graph below, casual riders have longer average of ride length than members for each day of week. It's worth to note that casual riders have a lower average ride length on Thursdays compared to other days. Members are used to ride bikes every day with fixed duration trips.



## Top Three Recommendations

### 1. Price discrimination for different ride length

Since casual riders tend to have longer ride length than members, the marketing team could provide price discounts for members when the ride length exceeds a particular duration to attract casual riders to be annual members. And inform casual riders the discount with emails, pop-up ads on the bike-share apps or other digital marketing tools.

### 2. Price discrimination for different day of week

Since casual riders have fewer trips on weekdays and lower average ride length on Thursdays, the marketing team could provide price discounts for casual riders on weekdays, especially Thursday to increase customer adoption leading to annual membership subscriptions.

### 3. Promote the benefits of riding a bike

Social media is an effective form of advertisement to increase user adoption. The health benefits of bike riding could be promoted with articles to increase bike usage among commuters and as a source of leisure activity.

## Description of Data

### 1. Data source

The dataset is generated from Cyclistic itself, including the Cyclistic trip data in April 2021. (Note: The datasets have a different name because Cyclistic is a fictional company.)

### 2. Data privacy and security

The data has been made available by Motivate International Inc. under this [license](#).

### 3. Data at a glance

```
glimpse(bike_share_202104)
```

```
## Rows: 337,230
## Columns: 13
## $ ride_id      <chr> "6C992BD37A98A63F", "1E0145613A209000", "E498E15508~
## $ rideable_type <chr> "classic_bike", "docked_bike", "docked_bike", "clas~
## $ started_at   <chr> "2021-04-12 18:25:36", "2021-04-27 17:27:11", "2021~
## $ ended_at     <chr> "2021-04-12 18:56:55", "2021-04-27 18:31:29", "2021~
## $ start_station_name <chr> "State St & Pearson St", "Dorchester Ave & 49th St"~
## $ start_station_id <chr> "TA1307000061", "KA1503000069", "20121", "TA1305000~
## $ end_station_name <chr> "Southport Ave & Waveland Ave", "Dorchester Ave & 4~
## $ end_station_id  <chr> "13235", "KA1503000069", "20121", "13235", "20121",~
## $ start_lat      <dbl> 41.89745, 41.80577, 41.74149, 41.90312, 41.74149, 4~
## $ start_lng      <dbl> -87.62872, -87.59246, -87.65841, -87.67394, -87.658~
## $ end_lat        <dbl> 41.94815, 41.80577, 41.74149, 41.94815, 41.74149, 4~
## $ end_lng        <dbl> -87.66394, -87.59246, -87.65841, -87.66394, -87.658~
## $ member_casual  <chr> "member", "casual", "casual", "member", "casual", "~
```

```
colnames(bike_share_202104)
```

```
## [1] "ride_id"      "rideable_type"      "started_at"
## [4] "ended_at"      "start_station_name" "start_station_id"
## [7] "end_station_name" "end_station_id"     "start_lat"
## [10] "start_lng"     "end_lat"            "end_lng"
## [13] "member_casual"
```

## Documentation of Cleaning or Manipulation of Data

1. Identify unique ride\_id and remove duplicates.
2. Filter data by condition (is empty) and look up the data, deleting the rows with empty value which can't be derived.
3. Outliers are explainable, then keep the outliers.
4. Create a column called "ride\_length." Calculate the length of each ride by subtracting the column "started\_at" from the column "ended\_at" (for example, =D2-C2) and format as HH:MM:SS
5. Delete the rows with "ride\_length" less than 0.
6. Delete the rows with "ride\_length" equal to 0, which means the client did not use the bike.
7. Create a column called "day\_of\_week," and calculate the day of the week that each ride started using the "WEEKDAY" command (for example, =WEEKDAY(C2,2)). Format as a number with no decimals, noting that Monday=1, Sunday=7.