

## **Case Study: How does a bike-share navigate speedy success?**

Using the Ask, Prepare, Process, Analyze, Share and Act for the data analysis. I will go over this case study to make a business decision in a bike-share company in Chicago.

### **Keys notes:**

- Maximize the number of annual memberships to increase profits.
- Analyze the difference between a casual rider and an annual member.
- Design a marketing strategy to convert casual riders into annual members.

### **Stakeholders:**

- Marketing Team
- Lily Moreno – Marketing Director
- Executive Team

### **Business Task:**

Analyze the historical trip data of Cyclistic to identify trends, so the marketing team will be able to evaluate the decision to create a target marketing campaign focusing on converting casual riders in annual members.

---

There are three questions to be responded:

- How do annual members and casual riders use Cyclistic bike differently?
- Why would casual riders by Cyclistic annual membership?
- How can Cyclust use digital media to influence casual riders to become members?

In this analysis I will focus on the first question. How do annual member and casual riders use Cyclistic bike differently?

--

### **Prepare:**

The data was made available by Motivate International Inc. And I will be using the data available for one year from 2019.

And for this analysis saved locally on ~/Desktop/Case Studie - Bike Share.

---

### **Process:**

- Before uploading the data tables into R, I checked the CVS file on a text format, and realized that the table for 2019\_Q2 had different columns names than the others. So, I proceed to rename the columns.
- Tables were uploaded to RStudio on May 9<sup>th</sup>,2024.
  - o Bikes\_Trips\_2019\_Q1
  - o Bikes\_Trips\_2019\_Q2
  - o Bikes\_Trips\_2019\_Q3
  - o Bikes\_Trips\_2019\_Q4

<i>Data Entries</i>	
<i>2019_Q1_Bikes</i>	365,069
<i>2019_Q2_Bikes</i>	1,108,163
<i>2019_Q3_Bikes</i>	1,640,718
<i>2019_Q4_Bikes</i>	704,054
<i>Total Entries</i>	3,818,004

- Then proceed to create one data frame 'Total\_Bike\_Trips' to include all tables. And confirmed that the row counts are the same as the separate tables.
- Let's confirm the structure of the data frame.

```
> str(Total_Bike_Rides)
'data.frame': 3818004 obs. of 12 variables:
 $ trip_id      : int  21742443 21742444 21742445 21742446 21742447 21742448 21742449 21742450 21742451 21742452 ...
 $ start_time   : chr   "2019-01-01 00:04:37" "2019-01-01 00:08:13" "2019-01-01 00:13:23" "2019-01-01 00:13:45" ...
 $ end_time     : chr   "2019-01-01 00:11:07" "2019-01-01 00:15:34" "2019-01-01 00:27:12" "2019-01-01 00:43:28" ...
 $ bikeid       : int   2167 4386 1524 252 1170 2437 2708 2796 6205 3939 ...
 $ tripduration : chr   "390.0" "441.0" "829.0" "1,783.0" ...
 $ from_station_id : int  199 44 15 123 173 98 98 211 150 268 ...
 $ from_station_name: chr   "Wabash Ave & Grand Ave" "State St & Randolph St" "Racine Ave & 18th St" "California Ave & Milwaukee Ave" ...
 $ to_station_id   : int   84 624 644 176 35 49 49 142 148 141 ...
 $ to_station_name  : chr   "Milwaukee Ave & Grand Ave" "Dearborn St & Van Buren St (*)" "Western Ave & Fillmore St (*)" "Clark St & Elm St" ...
 $ usertype         : chr   "Subscriber" "Subscriber" "Subscriber" "Subscriber" ...
 $ gender           : chr   "Male" "Female" "Female" "Male" ...
 $ birthyear        : int  1989 1990 1994 1993 1994 1983 1984 1990 1995 1996 ...

> head(Total_Bike_Rides)
  trip_id start_time end_time bikeid tripduration from_station_id from_station_name to_station_id
1 21742443 2019-01-01 00:04:37 2019-01-01 00:11:07 2167 390.0 199 Wabash Ave & Grand Ave 84
2 21742444 2019-01-01 00:08:13 2019-01-01 00:15:34 4386 441.0 44 State St & Randolph St 624
3 21742445 2019-01-01 00:13:23 2019-01-01 00:27:12 1524 829.0 15 Racine Ave & 18th St 644
4 21742446 2019-01-01 00:13:45 2019-01-01 00:43:28 252 1,783.0 123 California Ave & Milwaukee Ave 176
5 21742447 2019-01-01 00:14:52 2019-01-01 00:20:56 1170 364.0 173 Mies van der Rohe Way & Chicago Ave 35
6 21742448 2019-01-01 00:15:33 2019-01-01 00:19:09 2437 216.0 98 LaSalle St & Washington St 49
  to_station_name usertype gender birthyear
1 Milwaukee Ave & Grand Ave Subscriber Male 1989
2 Dearborn St & Van Buren St (*) Subscriber Female 1990
3 Western Ave & Fillmore St (*) Subscriber Female 1994
4 Clark St & Elm St Subscriber Male 1993
5 Streeter Dr & Grand Ave Subscriber Male 1994
6 Dearborn St & Monroe St Subscriber Female 1983
```

- Cleaning process
  - o Removing N/As, the new row total is 3,279,253.
  - o Since the trip\_id must be unique per trip let's make sure we don't have the any duplicates. We confirmed that there were no duplicates.
  - o Confirm all trip\_duration is more than 0.
  - o Ensure all date are formatted the same.

After completing the data cleaning process, the final data frame to use is called: "Total\_Bike\_Trips\_V3".

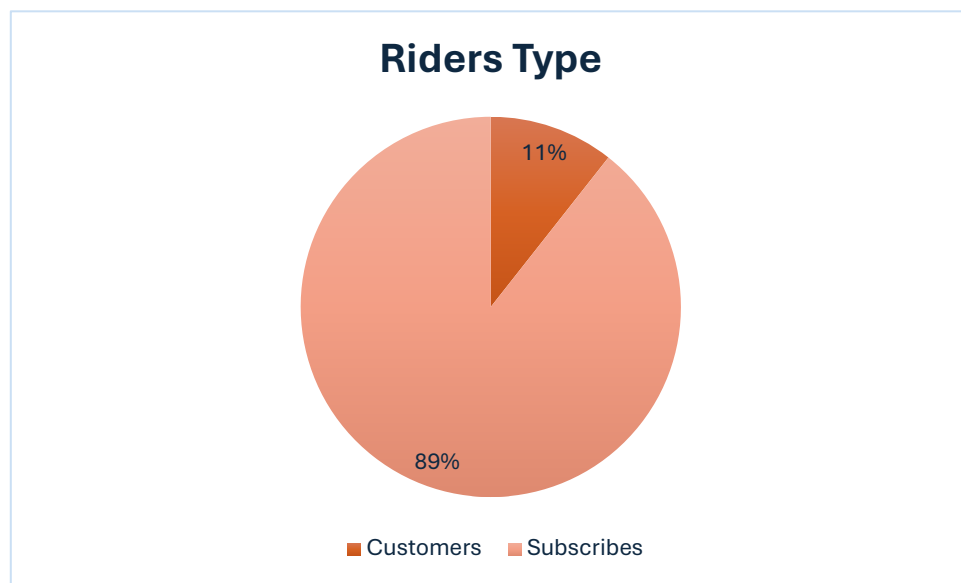
\*\* Refer to markdown document for more details and code.

## Analysis:

We need to verify the differences between the casual riders and the annual members, to do that there are several comparisons that can be made:

- **Riders type**

<i><b>Rider_Type</b></i>	<i><b>Count</b></i>
<i>Customers</i>	348,217
<i>Subscribes</i>	2,931,036



Customer: one-time riders

Subscribers: annual members

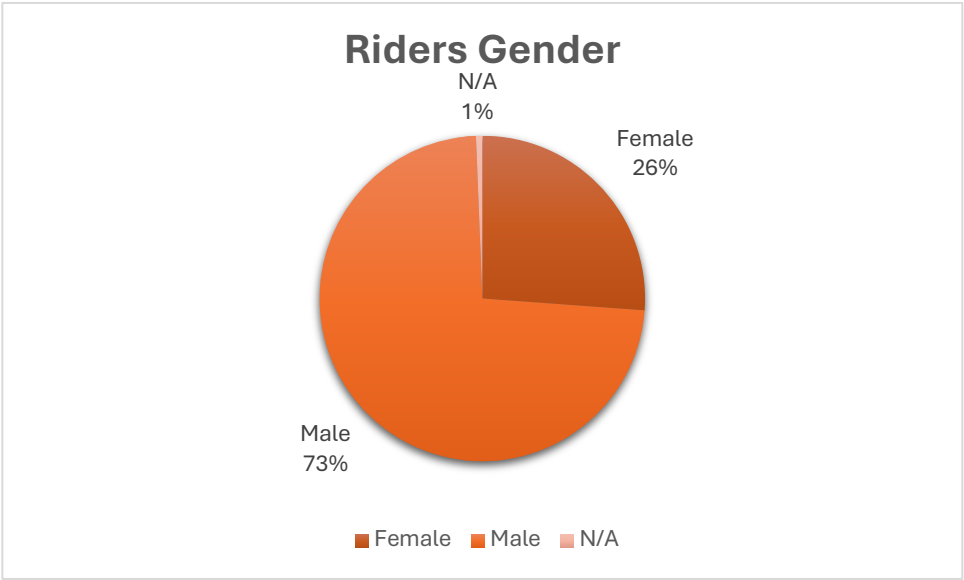
Of all the clients that Cysclist received during 2019, 89% of them were already annual members. And a remain 11% one-time users.

- **Gender per type of rider**

<i><b>Rider_Gender</b></i>	<i><b>Count</b></i>
<i>Female</i>	857,977
<i>Male</i>	2,400,819

N/A

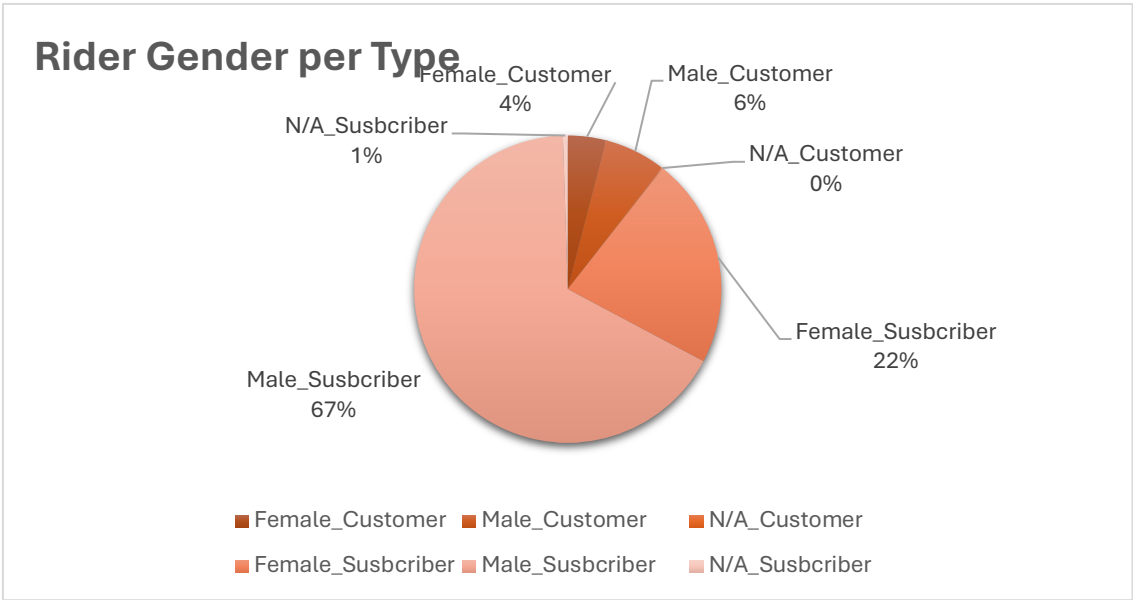
20,457



***Rider\_Gender\_Type***

***Count***

<i>Female_Customer</i>	131,438
<i>Male_Customer</i>	212,742
<i>N/A_Customer</i>	4,037
<i>Female_Suscriber</i>	726,539
<i>Male_Suscriber</i>	2,188,077
<i>N/A_Suscriber</i>	16,420



According to these results the majority of the clients of Clycists are male (73%). And that remains across of the customers and subscribers.

\*\* Sharing the gender information, is the client decision and optional. I am keeping those entries where gender is empty, since the amount is not representative and won't affect the final analysis.

- **Average age per rider type**

Using statistical functions in R to determine the average age per gender, to determine which is the age range of the clients of Cyclists.

<i><b>Riders Age</b></i>	<i><b>Subscriber</b></i>	<i><b>Customer</b></i>
<i>Female</i>	34.1	30.1
<i>Male</i>	35.8	31.5
<i>N/A</i>	31.5	37.5

Based on this, it can be notice that the average rider is around 30 to 37 years all, either if they are subscribers or customers.

- **Rides statistics**

Now let's look to the ride's statics per type of user, so we can identify patterns and trend on the usage.

- **Length of the trips per rider type**

<i><b>User Type</b></i>	<i><b>Trip length (S)</b></i>
<i>Subscriber</i>	859
<i>Customer</i>	2945

The trip length is measure in seconds. At looking at the result of the mean, we can affirm that the costumer usually rides longer than the subscribers.

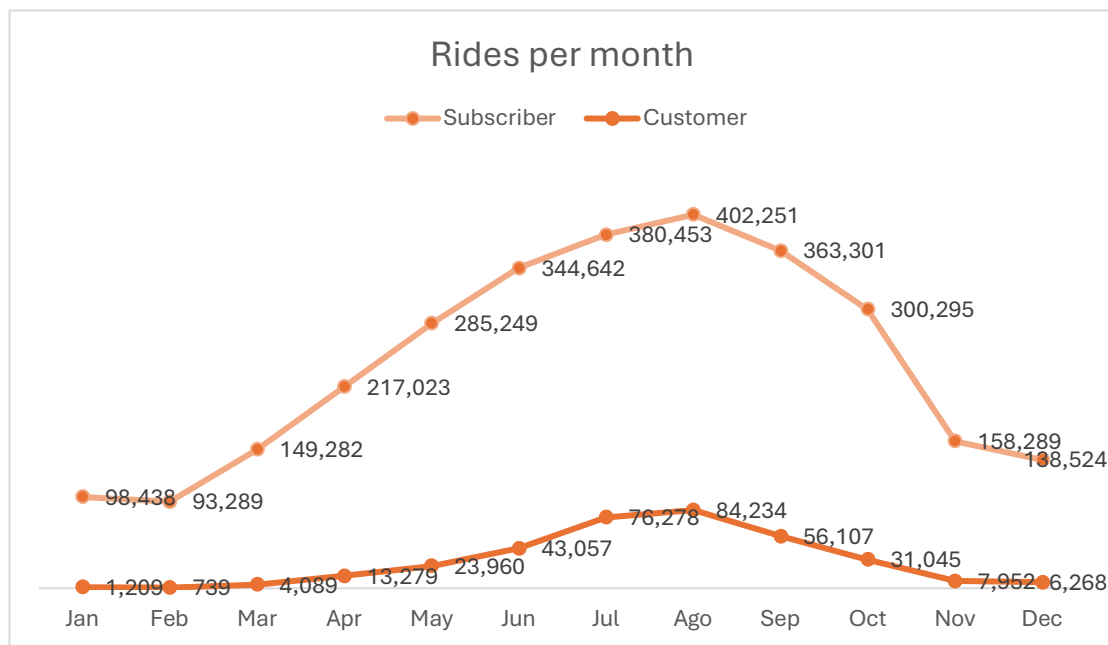
- **Trips per month per rider type**

<i><b>Month</b></i>	<i><b>Subscriber</b></i>	<i><b>Customer</b></i>
<i>Jan</i>	98,438	1,209
<i>Feb</i>	93,289	739
<i>Mar</i>	149,282	4,089
<i>Apr</i>	217,023	13,279
<i>May</i>	285,249	23,960
<i>Jun</i>	344,642	43,057
<i>Jul</i>	380,453	76,278
<i>Ago</i>	402,251	84,234
<i>Sep</i>	363,301	56,107

<i><b>Oct</b></i>	300,295	31,045
<i><b>Nov</b></i>	158,289	7,952
<i><b>Dec</b></i>	138,524	6,268

In this data it can be seen that still the subscribers represent the largest usage of the services, representing an 89% of the total of the rides made in 2019.

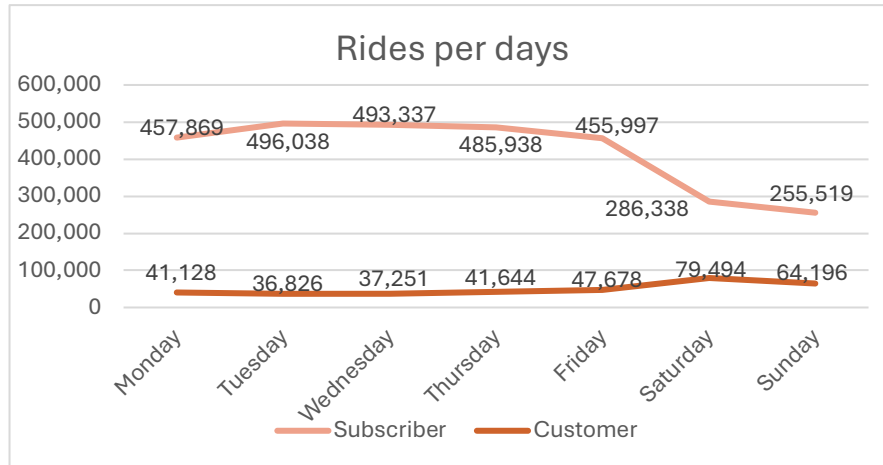
As well, it allows us to see to see that the peak months are from June to October, which indicates that the service is most likely to be used during the summertime and the beginning of fall.



#### ▪ Trips per weekday

<i><b>Day</b></i>	<i><b>Subscriber</b></i>	<i><b>Customer</b></i>
<i><b>Monday</b></i>	457,869	41,128
<i><b>Tuesday</b></i>	496,038	36,826
<i><b>Wednesday</b></i>	493,337	37,251
<i><b>Thursday</b></i>	485,938	41,644
<i><b>Friday</b></i>	455,997	47,678
<i><b>Saturday</b></i>	286,338	79,494
<i><b>Sunday</b></i>	255,519	64,196

It can be shown that for subscriber the peak usage is during weekdays, which can indicate that they use the service to go to and from work. While the customers tend to use the service more over the weekend.



- Which are the most frequent start and end stations?

<i><b>Start_Station</b></i>	<i><b>Frequency</b></i>
<i>Canal St &amp; Adams St</i>	52,283
<i>Clinton St &amp; Madison St</i>	47,706
<i>Clinton St &amp; Washington Blvd</i>	46,604
<i>Columbus Dr &amp; Randolph St</i>	33,961
<i>Franklin St &amp; Monroe St</i>	32,544
<i>Kingsbury St &amp; Kinzie St</i>	32,362
<i>Daley Center Plaza</i>	31,854
<i>Streeter Dr &amp; Grand Ave</i>	31,111
<i>Michigan Ave &amp; Washington St</i>	29,533
<i>Canal St &amp; Madison St</i>	28,310
<i>Canal St &amp; Adams St</i>	52,283
<i><b>End_Station</b></i>	<i><b>Frequency</b></i>
<i>Clinton St &amp; Washington Blvd</i>	49,503
<i>Canal St &amp; Adams St</i>	48,692
<i>Clinton St &amp; Madison St</i>	45,707
<i>Streeter Dr &amp; Grand Ave</i>	33,909
<i>Michigan Ave &amp; Washington St</i>	31,978
<i>Daley Center Plaza</i>	31,669
<i>Kingsbury St &amp; Kinzie St</i>	31,474
<i>Lake Shore Dr &amp; North Blvd</i>	28,100
<i>Franklin St &amp; Monroe St</i>	28,075
<i>Canal St &amp; Madison St</i>	27,240
<i>Clinton St &amp; Washington Blvd</i>	49,503

Taking these stations into consideration, will be a great asset for the marketing team to place promotions in these locations.

## TOP 10 START STATIONS

Canal St & Adams St	Clinton St & Washington Blvd	Franklin St & Monroe St	Daley Center Plaza	Streeter Dr & Grand Ave
Clinton St & Madison St	Columbus Dr & Randolph St	Kingsbury St & Kinzie St	Michigan Ave & Washington St	Canal St & Madison St

## TOP 10 END STATIONS

Clinton St & Washington Blvd	Clinton St & Madison St	Michigan Ave & Washington St	Daley Center Plaza	
Canal St & Adams St	Streeter Dr & Grand Ave	Kingsbury St & Kinzie St	Lake Shore Dr & North Blvd	Canal St & Madison St
			Franklin St & Monroe St	

---

**Act:**



Based on this analysis, there is room to improve the numbers of subscribers doing a target marketing campaign to one-time user. Taking in account the followings:

- Peak month of use, offering a plan with a monthly fee according to the usage during that month.
- Considering that the service is offer in Chicago, where winters can be extremely cold, maybe it can be offering a lower monthly fee during does months.
- Offer a subscription base on use, instead of a standard monthly fee.
- Make subscriptions offer during peak seasons.
- Place marketing material in the most popular stations.