**Ask**

1. Identify the business task.

What is the reason annual members and casual riders use rental bikes differently? By design a marketing strategy to convert casual riders into annual members. The director of marketing and finance analysts inferred that annual members are way more profitable compared to casual riders.

#### 2. Consider key stakeholders.

Lily Moreno: The director of marketing

Cyclistic executive team

Marketing Analytics team

**Prepare**

1. Download data and store it appropriately.

This is public data and has been downloaded from [Motivate International Inc.](https://divvy-tripdata.s3.amazonaws.com/index.html) under this [license](https://ride.divvybikes.com/data-license-agreement). So, I chose to store the original files in my laptop. Also, I made copies of each dataset, in case the originals need to be referenced.

2. Identify how it’s organized.

All downloaded files are saved in (CSV) format. Columns in total 13:

1. ride\_id

2. rideable\_type: types (docked, electric, and classic bikes)

3. started\_at: the date and time the ride started at

4. ended\_at: the date and time the ride ended at

5. start\_station\_name: station’s name the ride started at

6. start\_station\_id: station’s ID the ride started at

7. end\_station\_name: station’s name the ride ended at

8. end\_station\_id: station’s ID the ride ended at

9. start\_lat: start point

10. start\_lng: start point

11. end\_lat: end point

12. end\_lng: end point

13. member\_casual: casual/member riders

3. Sort and filter the data.

Data has been filtered and sorted (properly) in ascending order(A to Z), alphabetically, and numerically.

4. Determine the credibility of the data.

The data is organized by day, month and year. The format of the data is standard. The data is current, cited, original however might not be reliable.

**Process**

1. Check the data for errors.

First, before I start cleaning my data, I make sure all duplicates have been deleted. Data > Remove duplicates(Excel). Then I filtered my data (A to Z) and I found there are some data inconsistencies(### symbols). A question raised, how can a ride start at 4/29/2020 16:54 and end at 4/29/2020 16:51? There is no logic. For these two columns I wanted to make sure that ended\_at is greater than started\_at because of course it does not make any sense if the starting time is greater (after) the finish time.

For that I temporarily created a new column where I could check if the condition ended\_at > started\_at was TRUE or FALSE for each row, so I typed in a cell =E2>C2 and then I filtered by FALSE and removed the data that returned FALSE.

In this case, I could rearrange it (copy and paste data in appropriate columns. But the best option would be to ask someone for clarification and accuracy of the data.

Besides, I found some null values and errors(### appears when I created a new ride\_length column “=D2-C2) in the data when I sorted it from smallest to largest. I deleted errors entry data by filtering it.

2. Choose your tools.

For this analysis, I chose Excel because it enables me to format, organize and calculate data in a spreadsheet. By organizing data using software like Excel, I can make information easier to view as data is added or changed.

3. Transform the data so you can work with it effectively.

I used Excel sheets(12 sheets) to transform the data because excel can’t support more than 1,048,576 rows.

4. Document the cleaning process.

Created a new column “WEEKDAY” =WEEKDAY(C2,1) and chose format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday and did the same for a column called “ride\_length” (subtracting the column “started\_at” from the column “ended\_at” (=D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37:30:55. I did exactly the same manipulation for each (12 files).

Also, I made sure that all data in both started\_at and ended\_at columns had the same timestamp format: ‘dd-mm-yyyy hh:mm:ss’. =TEXT(VALUE(A1),"dd/mm/yyyy hh:mm:ss") using this formula.

#### **Analyze**

1. Aggregate your data so it’s useful and accessible.

Initially, I wanted to use a big query(in excel) to manipulate data(merge 12 files into one worksheet), unfortunately excel can’t support more than 1,048,576 rows. So, I did it manually. With the help of 12 separate sheets, I could manually calculate all rows 3489748(summarizing rows of each month and performing the rest of the calculations).

2. Organize and format your data.

For this step, I used proper formats and filters for my data. For example, for a new column I ‘ve created “WEEKDAY”, I chose format as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday and did the same for new a column called “ride\_length” (subtracting the column “started\_at” from the column “ended\_at” (=D2-C2) and format as HH:MM:SS using Format > Cells > Time > 37:30:55. I did exactly the same manipulation for each (12 files).

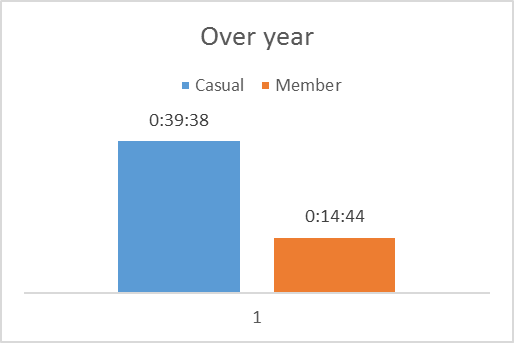
3. Perform calculations.

I calculated the mean of ride\_length(=AVERAGE(Count/Sum), max ride\_length(=SUM), mode of day\_of\_week(=MODE(date)). And on top of that, I created a pivot table to quickly calculate and visualize the data.

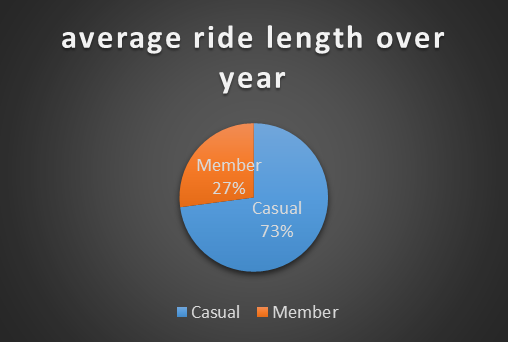
**Share**

[worksheet summary.xlsx](https://docs.google.com/spreadsheets/d/1lpEjuESGmq-Pv-R9wV2vtyKUDI0UXZMJ/edit?usp=drive_web&ouid=101836341365806940625&rtpof=true)

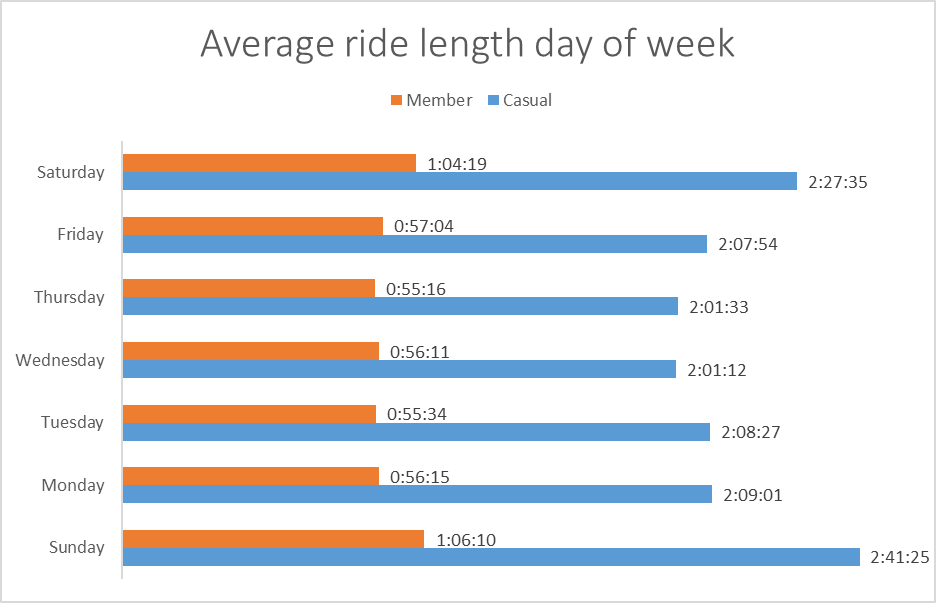
Average ride\_length for members and casual riders().



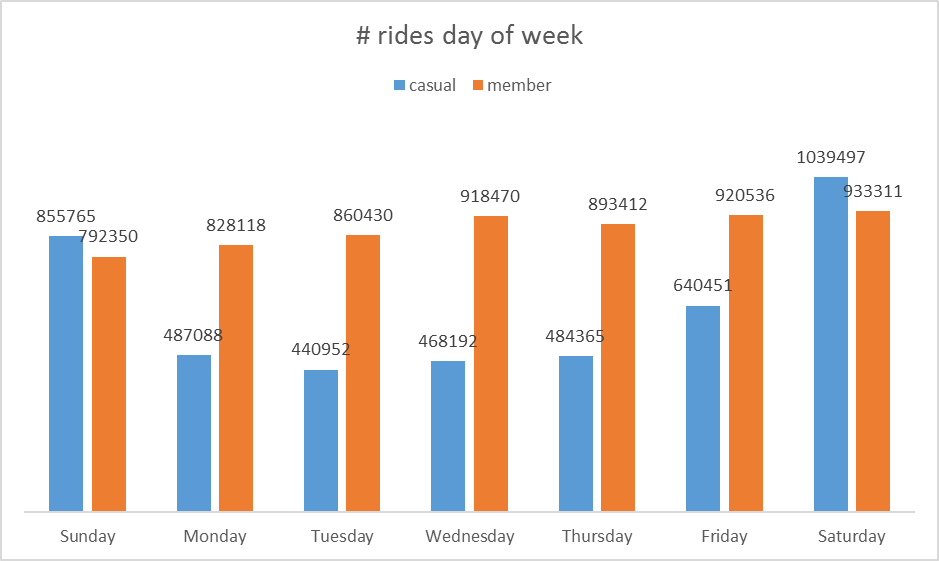
We can see that Casual riders use bikes far more(2.7 to 3x more) despite the fewer rides. Whereas members use bikes less but consistently over a year.



Average ride\_length for users by day\_of\_week().



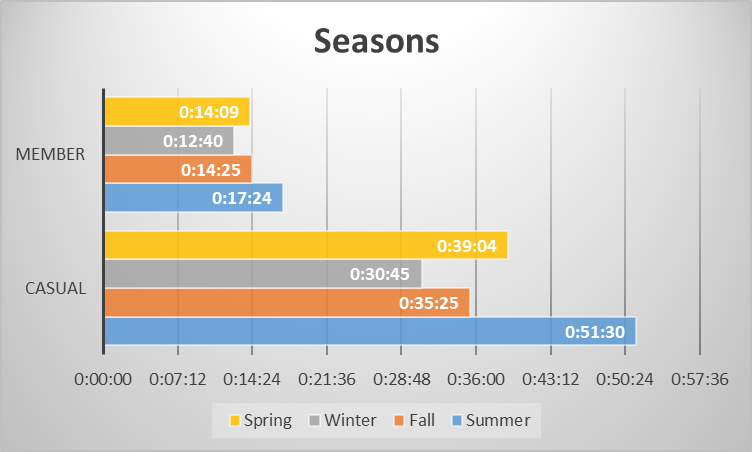
Number of rides for users by day\_of\_week().



We can observe from the graph a considerable increase of casual rides on weekends while the member has regular service usage throughout the week. This indicates members mainly use the bikes for their commutes and not leisure.

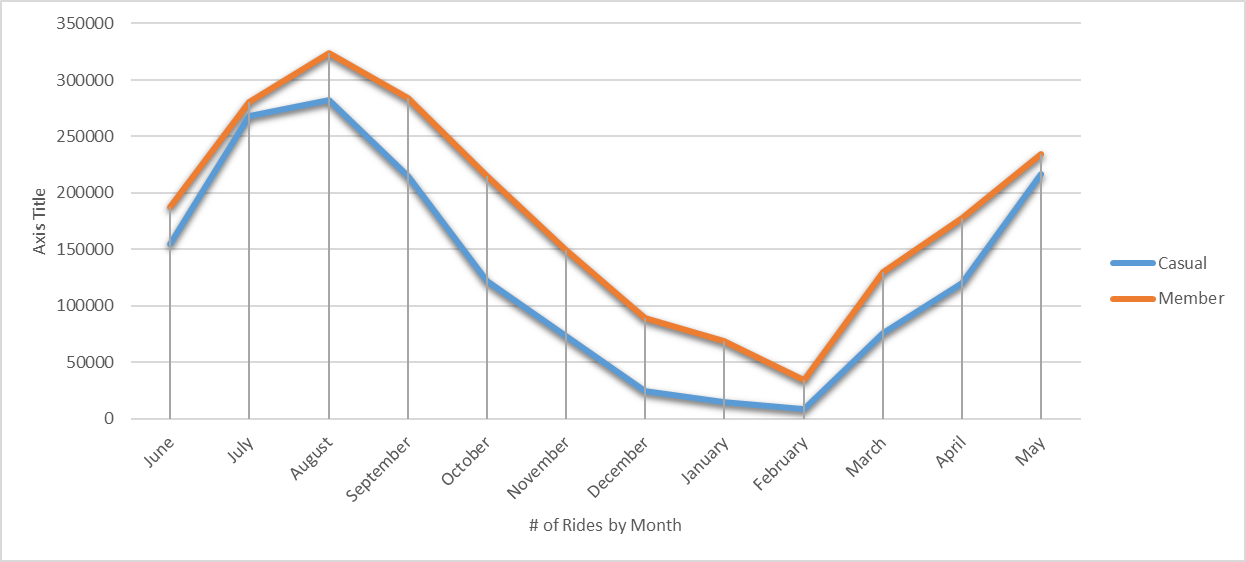
I also created another file and performed the same descriptive analysis steps to explore different seasons to make some initial observations.

Average ride length for four different seasons



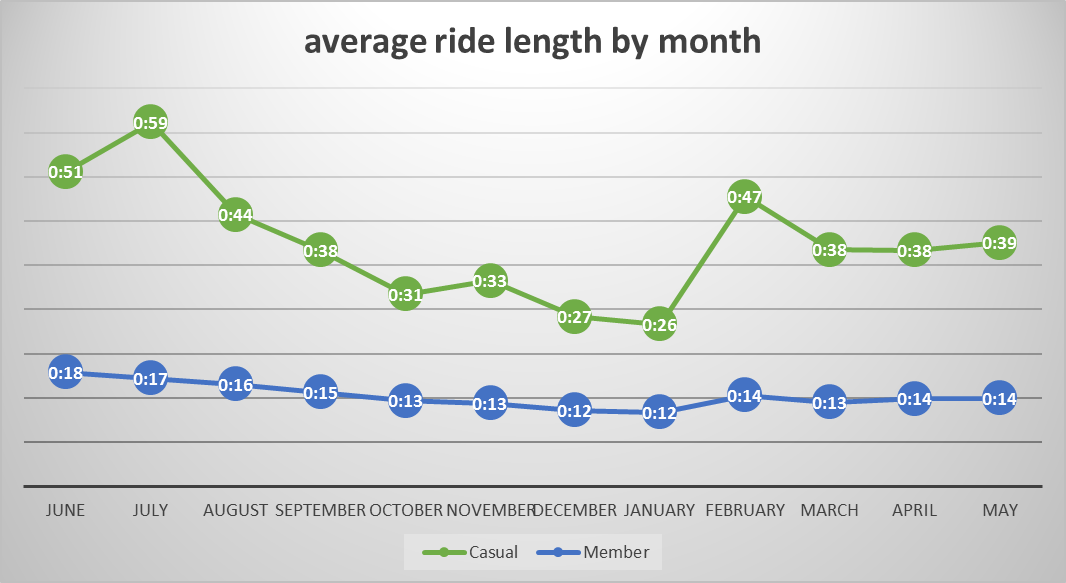
For casual riders, we see that summer is the time when bicycles have been rented most. As for members, it is consistent throughout all seasons except summer(a little bit higher).

Number of rides for users by month().



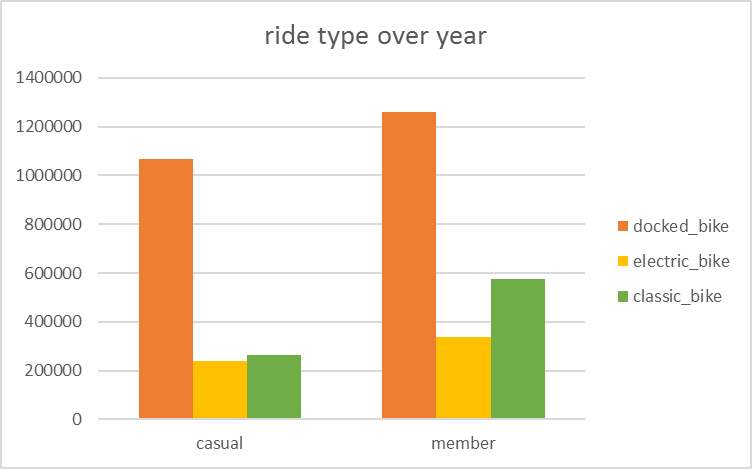
We can see a clear trend “for both casual and member” of how the count of rides starts to go up in February, peaks in August and from there starts to drop down.

Casual/members average ride length by month



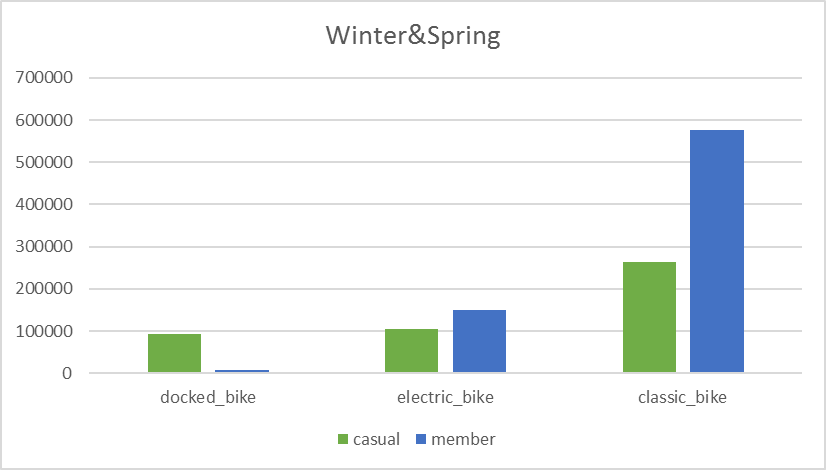
We can see a trend for casual rides of how the average of rides starts tremendously going up in January, peaks in June and from there starts to drop down. As for the members it’s consistent throughout the whole year.

Type of bike used over year



As we can notice, for both casual/members docked types of bikes have been rented most. Next is a classical, and then electric type.

Type of bicycles used in different seasons.



When I analyzed all four seasons, I found out in winter & spring seasons both riders casual/members preferred classical bicycles far more compared to the rest types(especially member users). Next users rented electric types of bikes.

## **Act**

#### *Key takeaways*

* On average, each ride is about 27:11 minutes:
  + Casual users ride for 39:38 minutes on average.
  + Members ride for 14:44 minutes on average.
* Whether being a member or casual rider, Saturday is the day when bikes are rented most.
* Bike rentals start low on Mondays and reach peak on Saturdays with a slight drop off on Sundays.
* Members ride less on the weekend compared to casual riders.
* Members rent bikes consistently throughout the entire week, whereas casual rentals are low Monday through Thursday and peak towards the weekend.
* The casual riders spend more time despite the fewer rides and vice versa for members.
* Casual riders do not use the service during the winter months(a sharp decline noted) compared to annual members.
* Casual users ride 2.7-3 times longer than members.
* The docked bike type is popular among all types of bikes both in terms of number of rentals and average ride duration. However, classical types of bikes were not launched during the winter & spring seasons. When I counted the # of rides during winter & spring seasons I noticed classical bikes are far more popular compared to the rest where the majority were members. Also, members use classic bikes much more than casual members. The electric bike use is nearly identical but casual riders apparently use more docked bikes.
* People who rent bikes follow a seasonal pattern(weather). The lowest usage is in the winter(February) with rentals starting to increase in the (March) spring. Peak usage is in (summer) August before it starts to decline again during the fall season.

#### *Top three recommendations*

In my opinion, members rent bikes mostly for regular commuting, either to go to work or school. The reason is the # of rides being consistent every single day and average ride duration. Whereas, casual rentals are lower Monday to Friday and increase from Friday and reach its peak on the weekend. It might be casual users using the bikes for leisure activities(touring the city, sightseeing).

1. Offer a discount to casual riders or decrease price during weekdays since ride & duration more during weekends. This will evoke interest of casual riders and entice them to convert to annual membership; offer one week of free ridership to feel the full benefits of annual membership.

2. Launch a loyalty program where both riders get bonuses or rewards, but for casuals making limitations - they can’t use it. It can encourage casual riders to subscribe to yearly membership; modify members plan - by giving them coupons to extend membership period of time.

3. Launch ads and promotions from the beginning of March till the end of August where annual members get rewards for achieving certain milestones and for casual riders who change subscription type get bonuses. This will inspire casual riders to become annual members and leverage it.