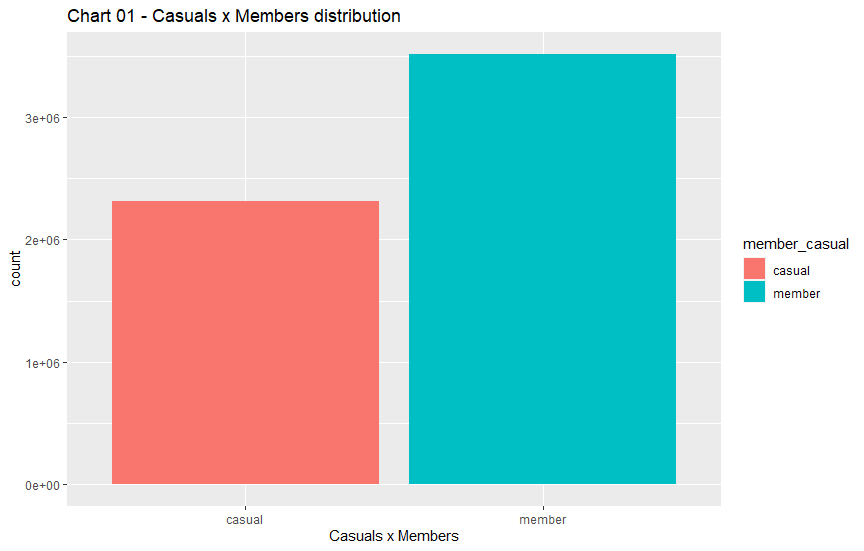
Bicycle Member Analysis

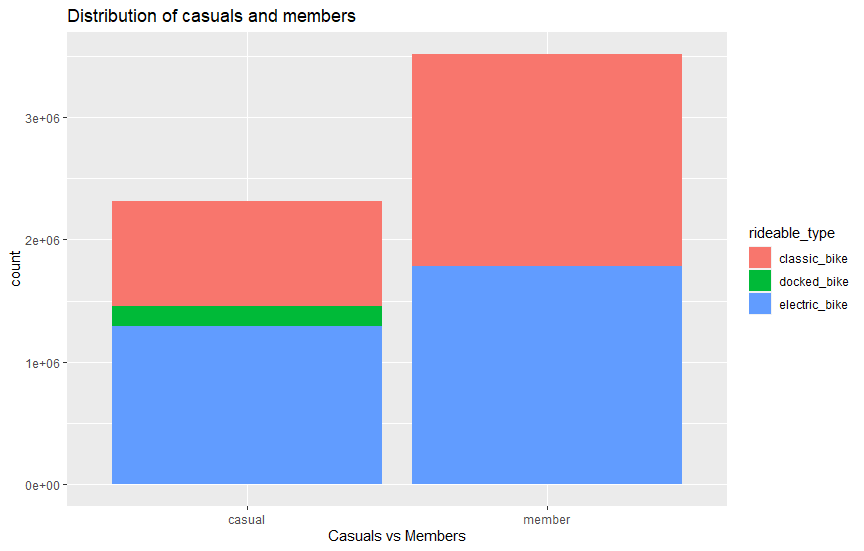
Forest H.

This is a project that analyzes the differences in behavior/patterns between members and nonmembers of a cycling company that offers bike-share in an effort to see what makes members and casual riders different so that a recommendation can be made to increase membership, and ideally made to be members with the annual pass as that is the most profitable. I have taken a more conversational tone rather than a professional one to illustrate my thoughts and process more clearly.

Of course, the first step after clarifying goals and expectations is to clean the data, I started creating a backup, then opening the data in excel and looking for blanks, incorrect data types in columns, converting start and end times to a more usable format, creating a day of the week column from the data.

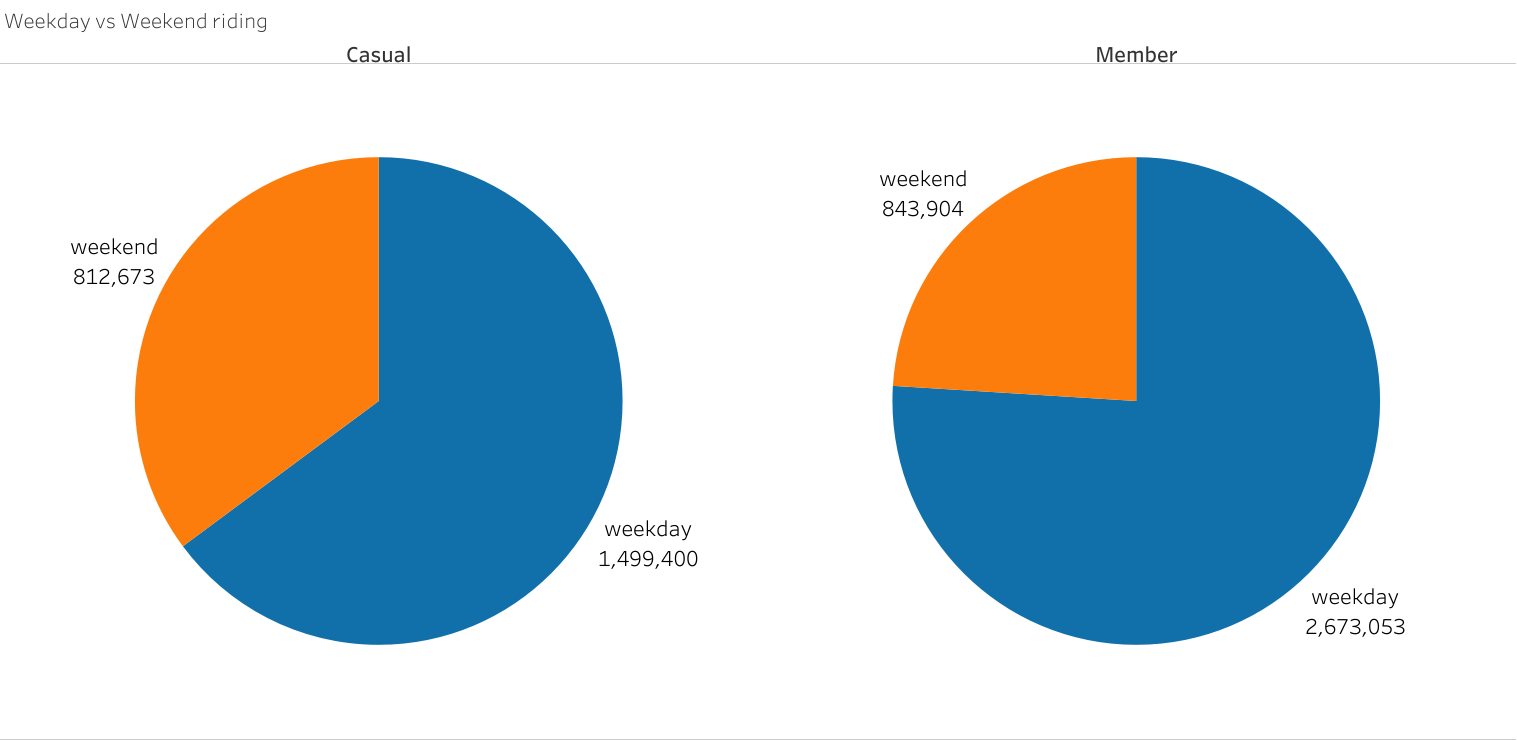
Then I moved into R studio and compiled the various csv files into one data frame and removed duplicates, created a ride time in minutes from start and stop times, created a year\_month column, converted some of the date-time formatting, and created a new cleaned and modified csv. From that I started to compare the groups.

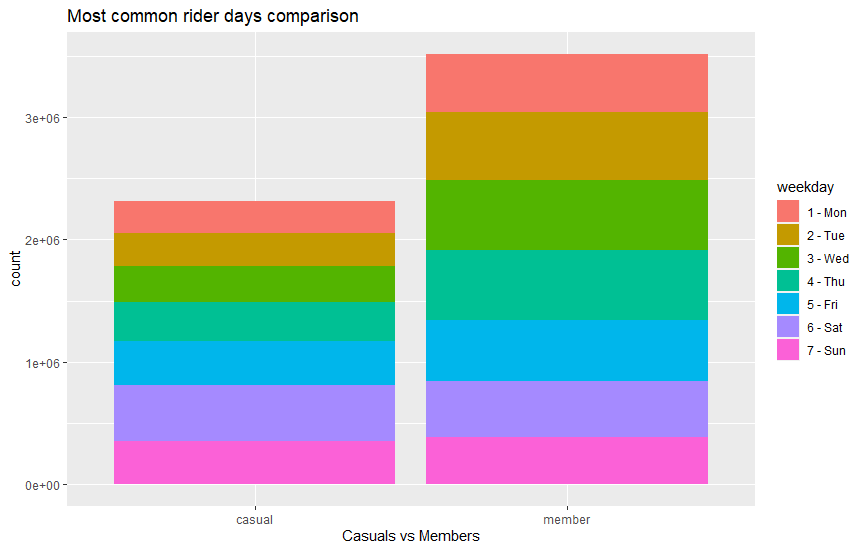




I soon found out that, as you can see the type of bike they rode wasn’t much different between groups, aside from members not using docked bikes.

My initial thoughts were that Members likely ride more often or ride more for commute.





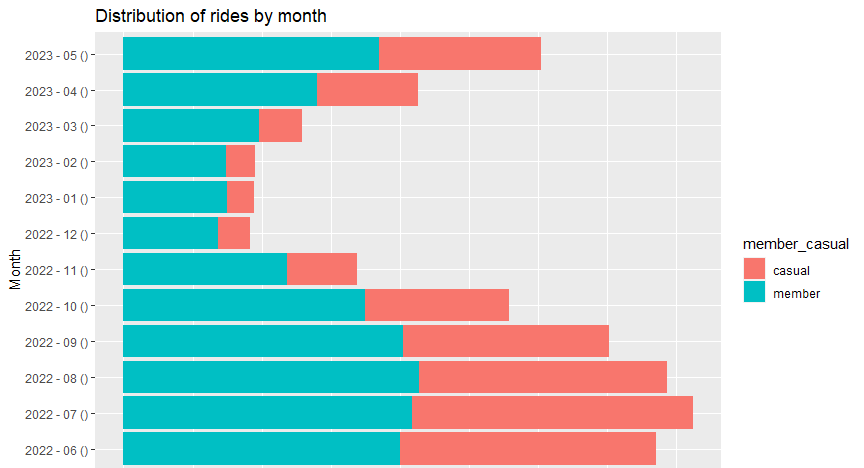
Lo and behold the Members ride more during weekdays than weekends per the above pie chart, and common rider days shows that while the weekend ridership is very similar between members and casuals, the weekday amount is higher Mon-Fri in Members. This isn’t enough to confirm my guess though after all they could just love cycling after work or before, during lunch etc. These charts only look at the number of rides.

I wanted to compare GPS data to find rides consistency between starting and ending locations in conjunction with time of day to further pursue the commute hypothesis but found that the accuracy of GPS I had available to me was a little over a kilometer. So technically if someone lived really close to work this wouldn’t show up.

I would have liked to compare specific riders, but the data only has ride ID not rider ID, so a new ID every ride. I compared average ride time members average 12.46 minutes while casuals average 28.20 minutes, maybe useful later but not very insightful right now. But while already looking at time how about time of day while riding.



The spike is at about 11 am but oddly there is also a spike close to midnight, this spike also doesn’t entirely align with commute times, its close in the morning but you would expect some spike at hours people return from work. Though more members are riding at those times during the weekdays it is not entirely conclusive.

Well continuing with time 

The company is based in Chicago, and it looks like overall ridership decreases significantly during the cold weather, though members are less bothered by this.

To answer the questions asked:

How do Annual members and casual bike differently?

Members are more committed to riding regardless of the day or month. Casuals are more put off by temperature and prefer to ride more on the weekends.

Why would casuals buy membership?

Casuals would be more inclined to ride if the temperature was less of an issue and there was more incentive to ride more regularly, commute while not conclusively what separates members and casual riders would create more riding through the week, which is a pattern members exhibit.

How can Cyclistic use digital media to influence casual riders to become members?

Well to put it all together, I believe a collaboration with a company that provides warm riding gear during the colder months may give a slight discount if you buy our membership and their product. To create more frequent riders during the weekdays some incentive programs say every 5th day you ride you are place on a leaderboard or get a small discount like at some coffee places. Promoting commute riding as sustainable and cheaper alternative in conjunction with ads or more exposure to affordable warm riding gear to push the casual audience towards behaviors more exhibited in members. All in all, encourage more frequent riding during the weekdays and cold months by exposing audiences to products that reduce their apprehension to the cold, and incentives to ride more frequently.

Unfortunately, some data that would have been helpful was in such poor condition it was unusable, for instance station IDs and names could provide insight but is almost completely empty, even if the start id is present often the end isn’t. It may be possible to match what ID/names we know to the coordinates, but it would be a fairly large ordeal, the kind of thing I would ask my manager about before just attempting largely because I don’t even know if all the names of the possible stations in Chicago are present. I thought latitude and longitude was my golden ticket at first but Chicago has a diameter of 25km so an accuracy of roughly 1.1 kilometers isn’t necessarily usable, if working on a team this is something I would ask about.

Here is a snippet describing the data I worked with

ride\_id rideable\_type started\_at ended\_at start\_station\_name

Length:5829030 Length:5829030 Min. :2022-06-01 00:00:04 Length:5829030 Length:5829030

Class :character Class :character 1st Qu.:2022-07-27 14:54:51 Class :character Class :character

Mode :character Mode :character Median :2022-09-23 07:56:12 Mode :character Mode :character

Mean :2022-10-28 01:02:13

3rd Qu.:2023-02-05 15:35:06

Max. :2023-05-31 23:59:58

start\_station\_id end\_station\_name end\_station\_id start\_lat start\_lng end\_lat

Length:5829030 Length:5829030 Length:5829030 Min. :41.64 Min. :-87.87 Min. : 0.00

Class :character Class :character Class :character 1st Qu.:41.88 1st Qu.:-87.66 1st Qu.:41.88

Mode :character Mode :character Mode :character Median :41.90 Median :-87.64 Median :41.90

Mean :41.90 Mean :-87.65 Mean :41.90

3rd Qu.:41.93 3rd Qu.:-87.63 3rd Qu.:41.93

Max. :42.07 Max. :-87.52 Max. :42.37

NA's :5961

end\_lng member\_casual start\_time end\_time

Min. :-88.11 Length:5829030 Min. :2022-06-01 00:00:04 Min. :2022-06-01 00:02:38

1st Qu.:-87.66 Class :character 1st Qu.:2022-07-27 14:54:51 1st Qu.:2022-07-27 15:16:15

Median :-87.64 Mode :character Median :2022-09-23 07:56:12 Median :2022-09-23 08:14:32

Mean :-87.65 Mean :2022-10-28 01:02:13 Mean :2022-10-28 01:20:55

3rd Qu.:-87.63 3rd Qu.:2023-02-05 15:35:06 3rd Qu.:2023-02-05 15:50:20

Max. : 0.00 Max. :2023-05-31 23:59:58 Max. :2023-06-07 23:04:26

NA's :5961

ride\_time\_m year\_month weekday start\_hour casual\_weekend

Min. :-10353.35 Length:5829030 Length:5829030 Length:5829030 Length:5829030

1st Qu.: 5.58 Class :character Class :character Class :character Class :character

Median : 9.85 Mode :character Mode :character Mode :character Mode :character

Mean : 18.71

3rd Qu.: 17.62

Max. : 41387.25