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

By: Abishaike Balachandran

Summary

As part of the Google Data Analytics Certificate Program through Coursera, I had completed the case study entitled, "How Does a Bike-Share Navigate Speedy Success?". Within this case study, a bike sharing company named Cyclistic, located in Chicago is trying to develop a strategy to convert casual members into annual members. As a junior data analyst, I have been tasked to solve this problem by implementing the six phases of the data analysis process: ask, prepare, process, analyze, share, and act.

Ask

A clear statement of the business task is "Analyze the past 12 months of ridership data to identify why casual riders are not opting for an annual membership".

Prepare

The public dataset that was used to complete this analysis has been provided by Motivate International Inc. This public dataset covers Cyclistic's trip data for 2022, which consists of 12 separate CSV files accounting for each month. Within each CSV file, the rider's membership type, bike type, start/end time, and start/end locations were captured.

Process

The tool that I used to process and clean the data was *Google Sheets*. I had prepared the dataset by importing the respective csv files into *Google Sheets*, which was a very helpful tool to immediately capture anomalies within the data. This was the driving factor behind using *Google Sheets* to complete the process step of the data analysis process.

I had quickly noticed that there were blanks present for fields concerning station ids and station names. The first action that I had taken to clean the dataset was to remove any instances of an empty field for *start_station_name*, *start_station_id*, *end_station_name*, and *end_station_id*.

Once the blank station ids/names were removed, I had created an additional field to determine whether the start and end times for each trip were identical. This was accomplished by using a

conditional if statement. For the records, where the fields *started_at* and *ended_at* were identical, I had deleted these records. It doesn't make sense for the times to be identical even if the user had returned to the same station because time elapses. This was the reasoning behind the deletion of these entries.

For the next step of the data cleaning process, I had used a pivot table to evaluate if a station name contained multiple entries for its station id. I believe that this is an anomaly because each station name should contain one distinct station id. If a station name had multiple station ids,I would tally up the number of instances based upon station id. The station id with the lowest count would be removed from the dataset.

I had found records where a station id can have multiple station names. In order to resolve this issue, I had determined the distances between the multiple station names and the total number of entries of each station within the dataset.

If the distances between the stations were small, I had changed the station name with the fewest entries to the station name with the most entries. If the distances were large, then the station names with the fewest entries were deleted. Lastly, if the number of entries amongst the station names were relatively equal, I had removed the records containing that station_id from the dataset. The details about this particular analysis can be found on the *Google Sheet* entitled "Analysis STN ID AND STN NAME-Rev 2".

For the final step of the cleaning process, I had removed the fields *start_lat*, *start_lng*, *end_lat*, *end_lng*, and *Does started_at = ended_at* because they will not be required during the analysis step of the data analysis process. However, I had added the fields *ride_length*, *weekday_started_at*, and *weekday_ended_at* which will be required for the analysis step. These are calculated fields, where its operands are the fields *started_at* and *ended_at*.

Analyze

Through analysis, I was able to find a couple of surprises from the dataset. I had immediately noticed that non-members had utilized the docked bikes whereas members did not utilize the docked bikes at all. The number of instances of docked bikes was not a miniscule amount. For example, during the summer months of June and July, there were 29,862 and 30,275 instances. The average ride length for docked bikes ranged from as low as 38:25 in December to as high as 2:40:59 in January. However, it tended to hover around 45:00, which was evident from April till October.

For each month of the year, there were more members using the service than non-members. However, during the months of May till August, the number of users between members and non- members were not that large. For example, in July, there were 306,185 non-members, whereas there were 324,876 members. This makes sense because the months between May and August represent Spring and Summer. During these seasons, most individuals tend to spend more time outside due to the pleasant weather, hence making cycling a viable option.

The number of riders decrease notably during the winter months of November, December, January, and February. During these months, the surprising trend was that there were far more members than non-members. It seems that although there was a drop off in members compared to the summer months, a notable portion of them still use the service in the winter. This may imply that members during the winter months must be avid cyclists or cycling is the only form of transportation within their budget. This trend does not apply to non-members.

Although there were far more members than non-members, the average ride length for non-members was notably higher than members. It seems that members use the service often to cover short distances, whereas non-members would use the service from time to time to cover a larger distance. However, the average ride length between them was not substantial. For example, the average ride length hovered around 13 minutes for members between May and August, whereas it was around 25 minutes for non-members. This can imply that non-members may use the service

for leisure, whereas members use it as a necessary mode of transportation.

Amongst all of the different bike types offered by Cyclistic, the traditional classic bike is the type that has been heavily used. Although the total number of electric bikes was at its highest during the summer months, the margin between classic bikes and electric bikes was not significant during the winter months. For example, for non-members in January, there were 6,733 instances of classic bikes and 4,675 instances of electric bikes. On the contrary, in July, there were 166,132 instances of classic bikes and 90,598 instances of electric bikes. The same trend had also been observed for members as well. This trend implies that users do not want to spend much time in the cold, hence opting for an electric bike since it will enable them to get to their destination a lot faster.

In terms of usage, non-members tend to use the service on the weekends whereas members use the service on the weekdays. For non-members, the only months of the year where it wasn't used as often during the weekends are in November and December. During these two months, non-members are using this service on Thursday. The fact that non-members are using this service mostly on the weekends can suggest that the bikes are being used for leisure. The frequent usage by non-members on Thursday in November and December can be attributed to the holiday seasons, namely Thanksgiving in November and Christmas in December. On the contrary, it seems that members rely on this service as a primary mode of transportation since it's been heavily used during the weekdays.

Share

As stated previously from the Analysis Section, the first notable difference between members and non-members is that non-members use docked bikes whereas members do not use this type of bike, which can be seen in figures 1 and 2. For non-members, the number of docked bikes gradually increases from winter to summer. From these figures, it is evident that the most used bike type is the classic bike. The highest incidence of classic bikes amongst members and non-members occurred in June. In addition, the highest incidence of electric bikes for

non-members happened in July, whereas it was in August for members. During the summer months, the margin between classic bikes and electric bikes is large, especially in May and June. However, as the summer months wane, the margin between classic bikes and electric bikes start to get smaller.

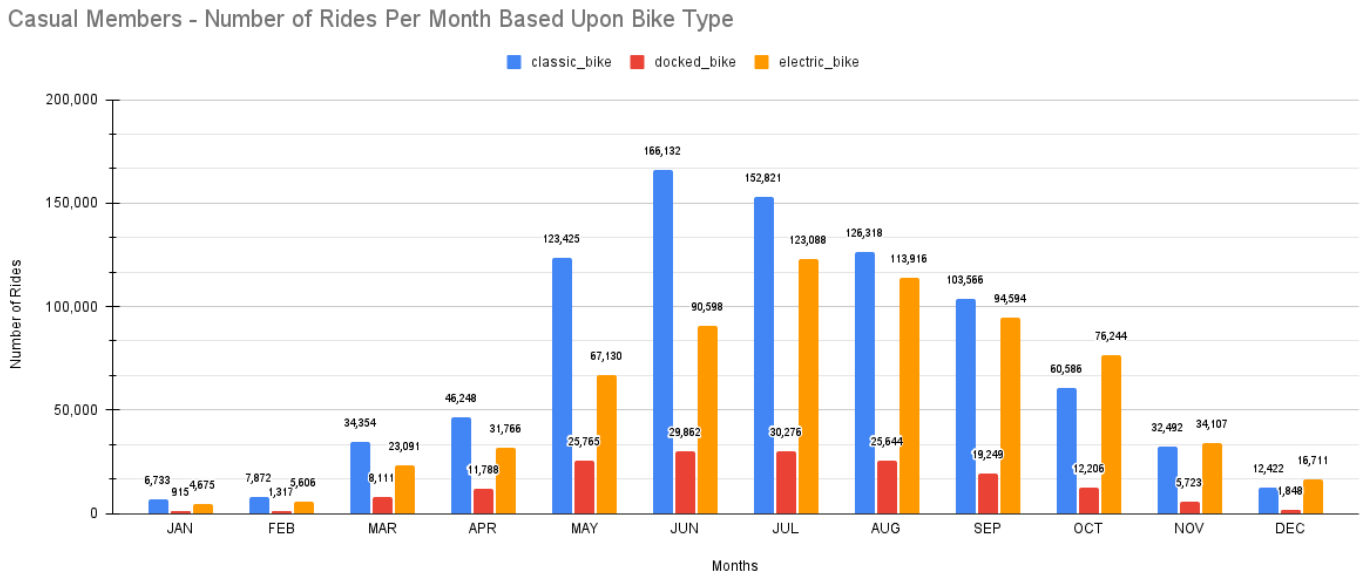

Figure 1: Casual Members - Number of Rides Per Month Based Upon Bike Type

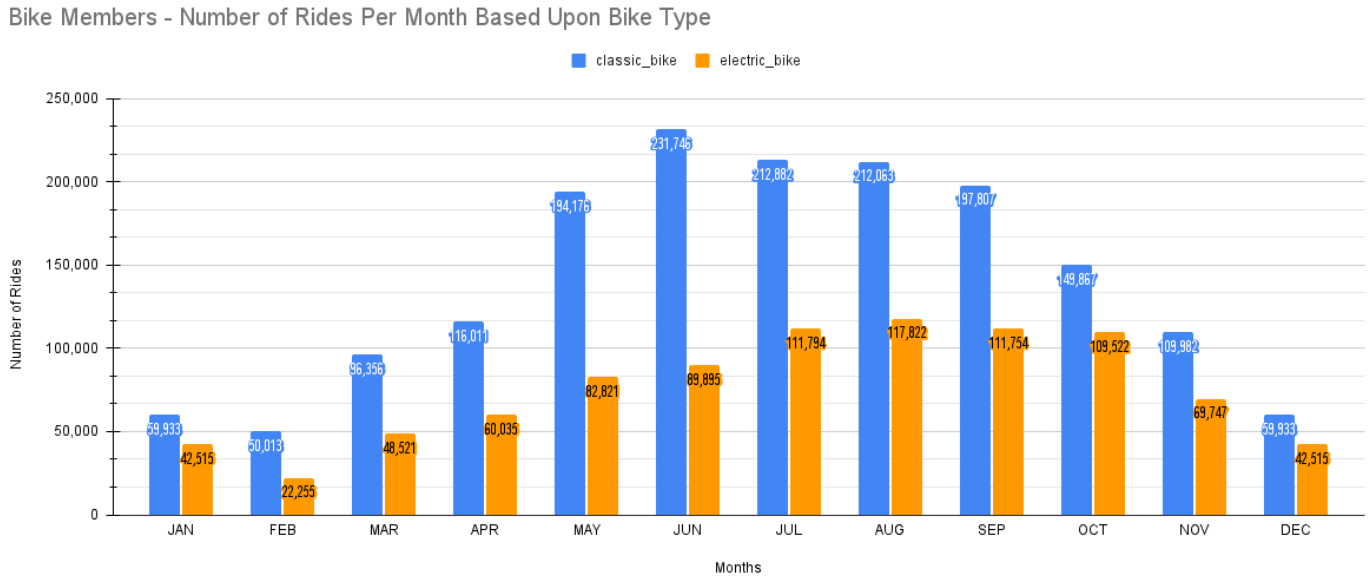

Figure 2: Bike Members - Number of Rides Per Month Based Upon Bike Type

For every month of the year, there were more non-members than members. From January until July, the number of rides for members and non-members had increased. Ridership had then decreased gradually from July until December. This trend aligns with the seasons, where riders tend to bike in Spring and Summer and not during Autumn and Winter. The average ride length for non-members was higher when compared to members. A parabolic trend was observed for non-members where the average ride length gradually increased until June and then it started to gradually decrease from June until December. From January until May, a discernible trend was not apparent. Compared to non-members, there was a sharp decrease in the average ride length from July until December. These trends can be found in figures 3 and 4.

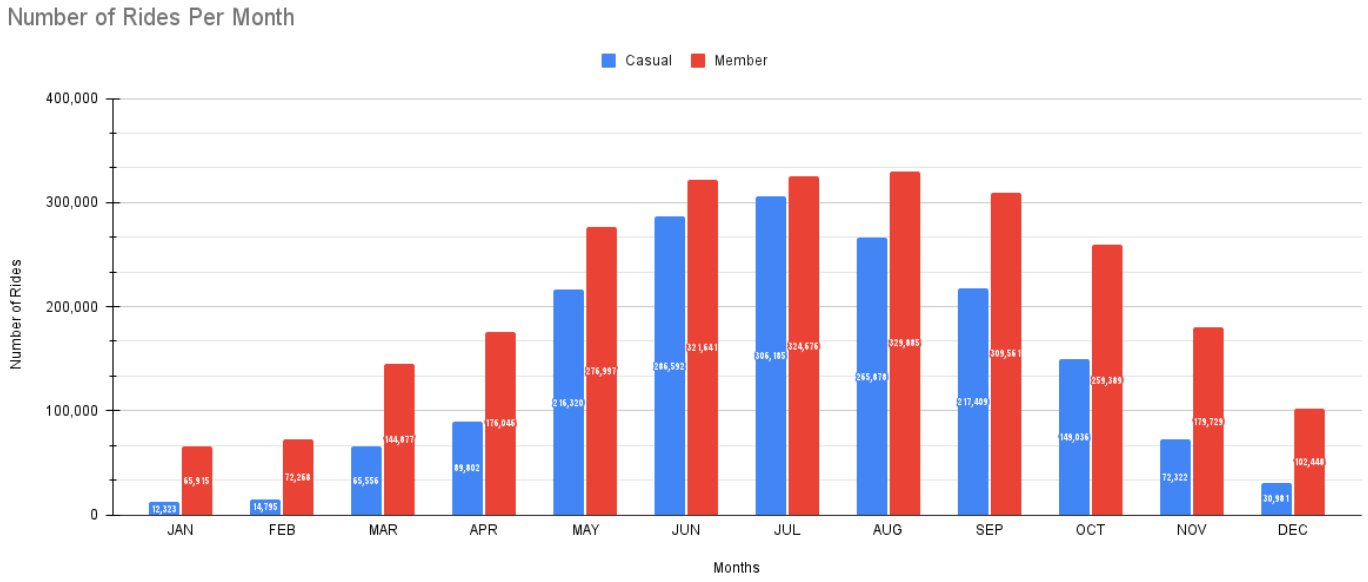

Figure 3:Number of Rides Per Month

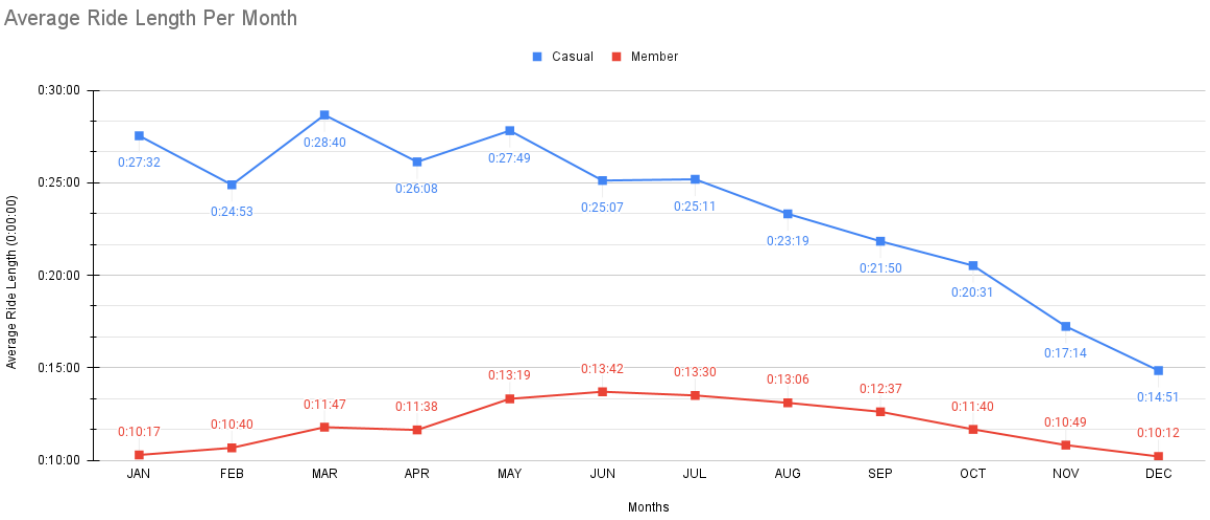

Figure 4:Average Ride Length Per Month

The margin between the average ride length for members between the classic bike and the electric bike was not substantial. However, this was not the case for non-members. This was evident in figures 5 and 6. If one looks at April, the average ride length for members for classic bikes was 0:12:16 whereas it was 0:10:24 for electric bikes. On the contrary, for non-members it was 0:26:17 for classic bikes whereas it was 0:17:30 for electric bikes. The average ride length significantly increased from April to May and drastically decreased from May to June for classic bikes when it came to non-members. However, this trend did not occur for members who used classic bikes from April to June. Regardless of the bike type, the average ride length for non-members is much higher than members.

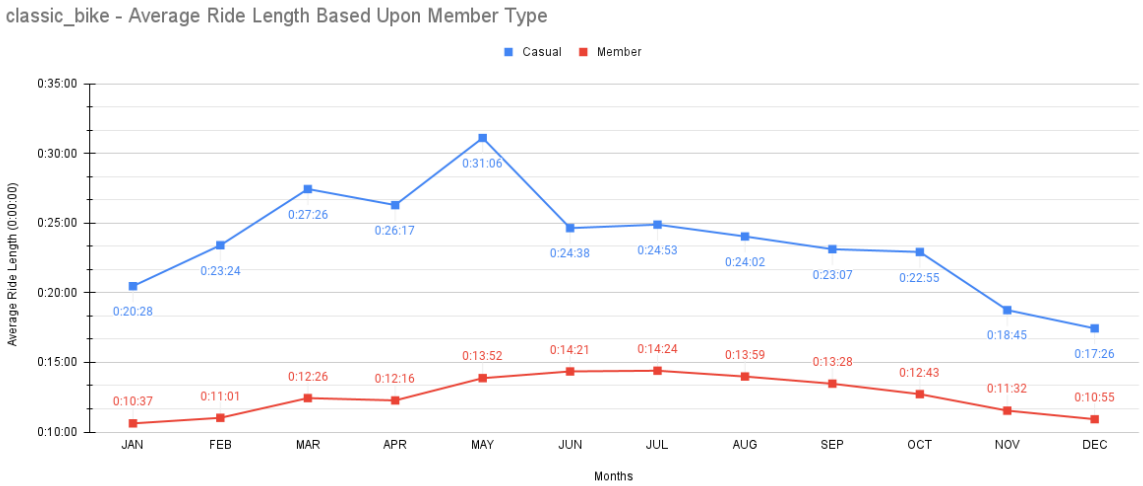

Figure 5:classic_bike - Average Ride Length Based Upon Member Type

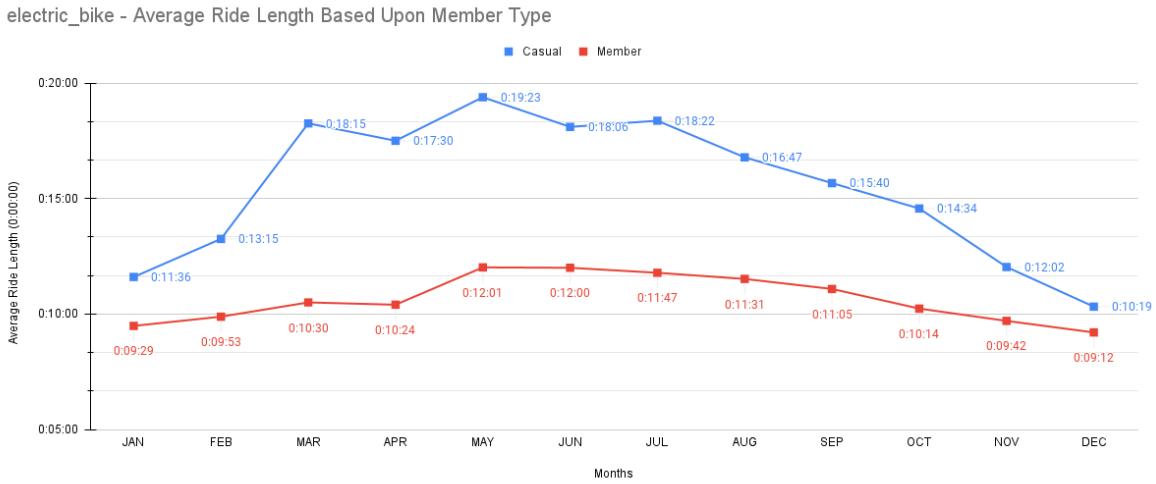

Figure 6:electric_bike - Average Ride Length Based Upon Member Type

Based upon Figure 7, it is evident that non-members use the service primarily on the weekends, where the highest incidence of ridership occurs on Saturday. In fact, in terms of ridership, there are more non-members than members on the weekends. Conversely, in terms of ridership on the weekdays, an inverse relationship occurs where there are more members than non-members. As stated previously, the average ride length is higher for non-members than members. Based upon Figure 8, the trend from Sunday until Saturday is identical for both members and non-members. To elaborate, from Sunday until Tuesday, the average ride length decreased. From Wednesday until Friday, the average ride length gradually increased. Lastly, there is a sharp increase from Friday until Saturday.

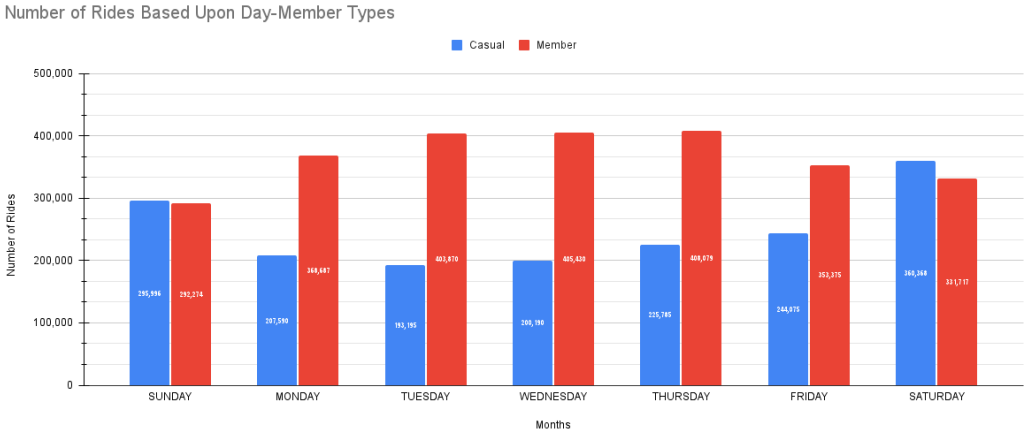

Figure 7:Number of Rides Based Upon Day-Member Types

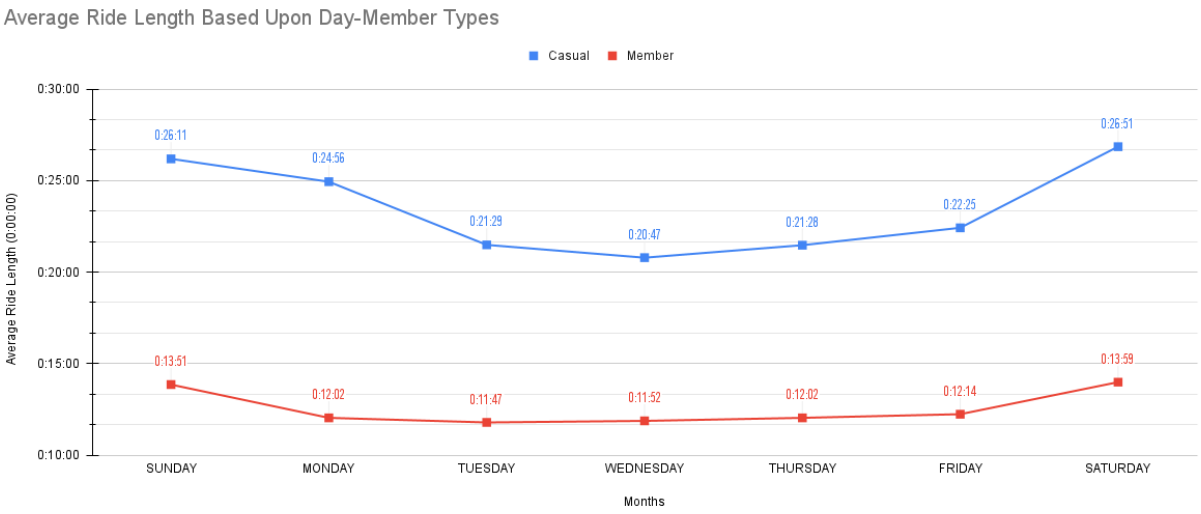

Figure 8:Average Ride Length Based Upon Day-Member Types

Act

The first recommendation that I would suggest to convert casual riders into annual riders is to incentivize the use of docked bikes. Based upon the data, casual members use docked bikes but annual members do not use docked bikes at all. Incentivization of docked bikes may be accomplished by installing docking stations within the Chicago area to facilitate more docked bikes and by offering the use of docked bikes at a discounted rate.

Based upon the dataset, it is apparent that ridership for casual members is low for the periods of January-May and from September-December. During these periods, it would be best to decrease the fare for casual members in order to increase ridership during this period, which will then sway casual members to become annual members. Also, due to the large volume of ridership from casual members during the summer months of June and July, one can slightly increase the fare during this period. I believe that casual members would pay for the slight increase because they enjoy cycling as a leisurely activity in the Summer.

The last recommendation that I would suggest is to implement a weekend rate and a weekday rate. Based upon the data, the majority of casual members tend to use the service on the weekends and the average ride length is greater than annual

members. If a discounted weekend rate gets implemented, ridership amongst casual members will increase, which will persuade a certain cohort of casual members to become annual members. The average ride length for casual members hovers around 20 minutes during the weekdays. A weekday rate can be implemented where if the ride length exceeds 20 minutes, the user will pay a flat rate. This may entice the existing casual members who use the service on the weekdays to eventually opt for an annual membership.