Cyclistic April 2020 Membership Analysis

Elliott Wink

2022-06-15

Contents

Introduction

Cyclistic is a bike-sharing program in the city of Chicago with a fleet of 5,824 bicycles and 692 stations spread across the city. The purpose of this report is to answer the question: "How do annual members and casual riders use Cyclistic bikes differently?" This question will be answered in the "*Analysis*" section. Recommendations of marketing strategies aimed at converting casual riders into annual members will be given in the "*Recommendations*" section. Details of how the data was altered will be covered in the "*Data Preparation and Processing*" section.

Data Preperation and Processing

About the Data

The data set provided ("**202004-divvy-tripdata.csv**") consists of data collected by Cyclistic for the month of April 2020. During the month prior, March 2020, the Covid-19 pandemic drastically changed people's lives in the city of Chicago, and they were ordered to shelter in place with the exception of essential employees and necessary travel, such as purchasing groceries. This shelter in place order was in full effect until the end of July 2020 where these restrictions were only slightly loosened. This stay at home order most likely effected ridership for April 2020 drastically.

The data set originally included 84,776 observations and 13 variables. These variables are: ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, and member_casual. The ***ride_id*** is a unique identifier for each ride taken during the month. The ***rideable_type*** is an identifier for the type of rideable used. This data set only included docked bikes but other bikes such as reclining bikes, hand tricycles, and cargo bikes are also available for riders to use. The ***started_at*** and ***ended_at*** variables include the date and time that each ride started and ended at (in MM/DD/YYYY HH:MM format). The ***start_station_name*** and ***end_station_name*** are the street locations where each ride started and ended at. The *start_station_id* and *end_station_id* are unique identifiers for each docking station. The ***start_lat***, ***start_lng***, ***end_lat***, and ***end_lng*** are the longitude and latitude for the starting and ending docking stations. Finally, the ***member_casual*** is description as to the type of rider that took that ride.

Data Preparation

For this project, the data analysis tools used were the R programming language and Tableau[1]. The R programming language was used for its ability to perform statistical analysis and ability to transform data efficiently. Tableau was used to help further visualize the data where two dashboards were created to display some graphs.

In the data set, there were 99 rides that included missing values. These values did have end times so these missing values may have been due to a software issue with the docking station or a data entry issue. Also there were also 14 rides where the end time was before the start time. Similar to the missing value issue, this may have been a software issue or a data entry issue. There were 40 rides where the start location and end location were different but there was a ride time of zero. This is most likely due to a software or data entry issue. There were 628 rides that had a ride time of zero. This could have been a software issue, data entry issue, or riders may have returned the bike as soon as the rented it.

Data Processing

As stated in the "Data Preparation" subsection, there were 99 rides that included missing values. This is about 0.12% of the original data set, these observations were removed as it will most likely not affect the

[1] https://public.tableau.com/app/profile/elliott.wink/viz/CyclisticApril2020/RidershipStatisticsDashboard

analysis. There was a total of 642 rides that had ride times of zero or less and these observations were also removed. This accounted for 0.76% of the remaining data. The rides that had different start and end locations could have been interpolated so that they closely matched other rides with the same start and end locations but this would not account for rides that take multiple stops before being returned so it was safer to remove them entirely. These changes were saved to another file ("**202004-divvy-tripdata-cleaned.csv**") so that the original data set would not be altered in-case there were mistakes in the data cleaning process. This cleaned data set was the data set used for analysis.

Throughout the analysis section, some of the data is normalized in order to better understand the difference between annual members and casual riders. The normalization function that is used is a min-max normalization[2] which rescales the range of values between [0, 1]. This does not change the original data set but is a technique that is used frequently during this analysis.

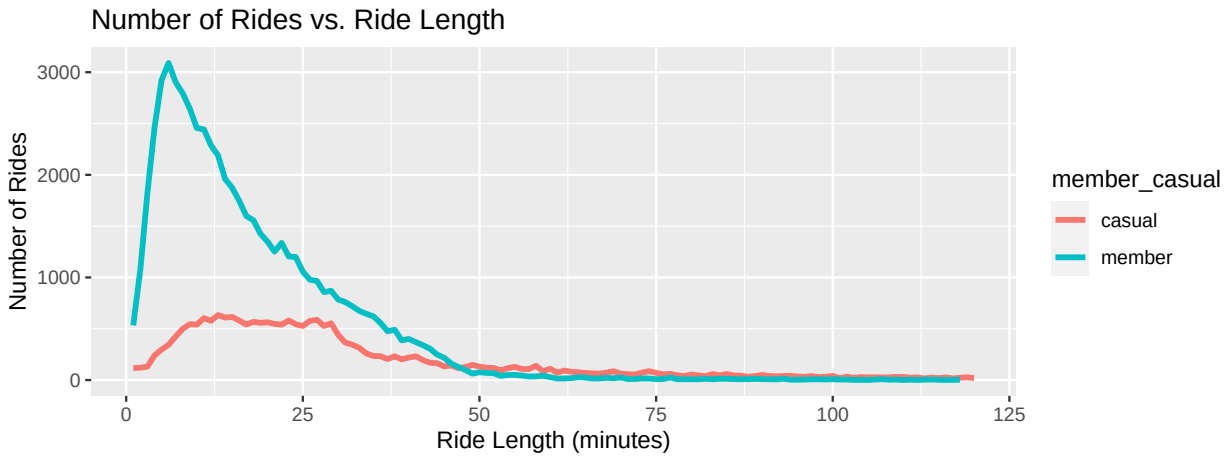
[2]https://en.wikipedia.org/wiki/Feature_scaling#Rescaling_(min-max_normalization)

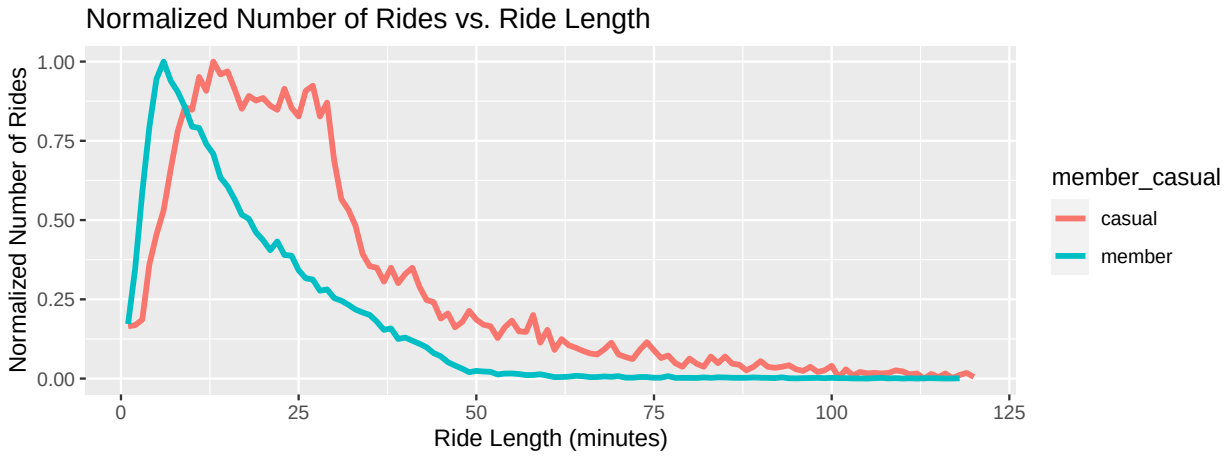
Analysis

Ride Length

The graphs below display the distribution of the length of rides between the annual members and the casual riders. These graphs include ride times under 2 hours as 98.6% of all rides are under 2 hours. The maximum ride length is 978.66 hours and the average ride time over 2 hours is 18.2 hours. The average ride time for annual members is 17 minutes and the average ride time for casual riders is 31 minutes.

The graph labeled "Number of Rides vs. Ride Length" displays the true number of rides lengths at each minute while the graph labeled "Normalized Number of Rides vs. Ride Length" displays the same data but normalized to more easily visualize the difference between annual members and casual riders.



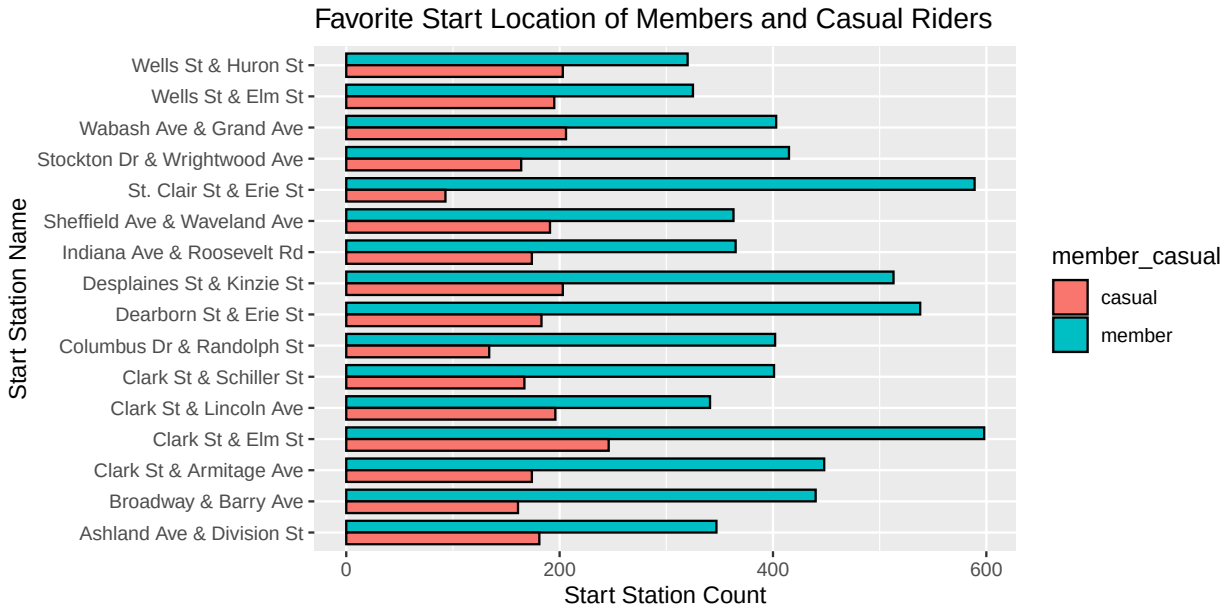


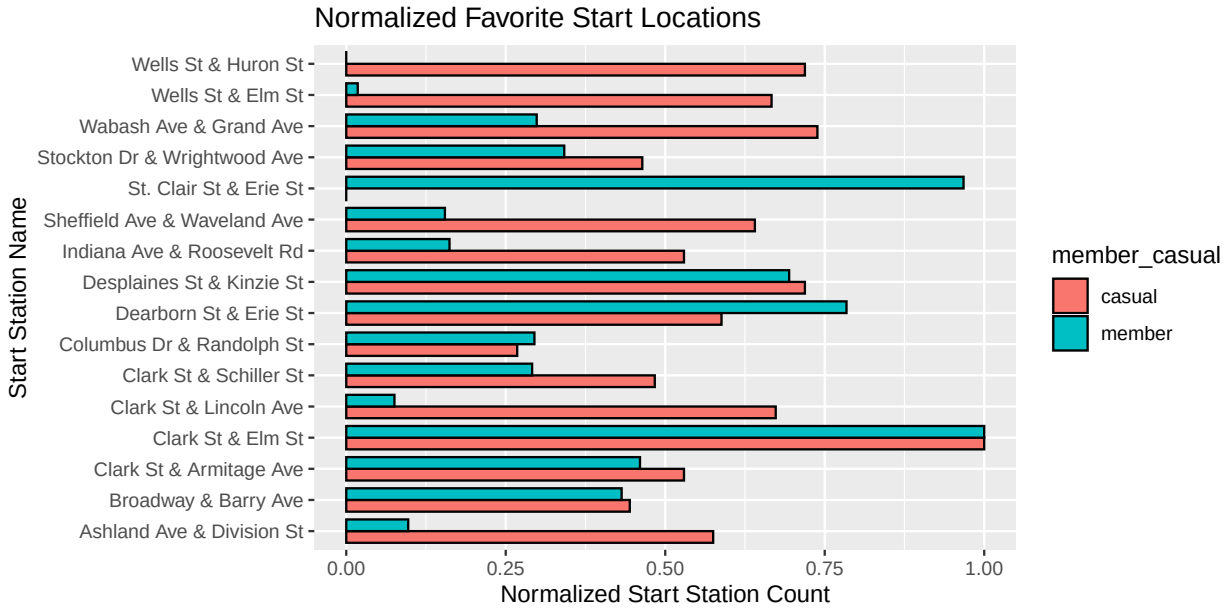
As displayed by these graphs, casual riders tend to ride for a longer duration than annual members. As stated previously, the average ride time for annual members is 17 minutes and the average ride time for casual riders is 31 minutes. This difference may be explained by annual members using bikes to go to work or to travel short distances. Since these riders are not charged per ride, like a sing-ride pass, they have the ability to dock and undock bikes as they please. Full-day pass riders may choose to keep bikes undocked for longer as riders, such as tourists, may use the bikes as a cheap alternative to renting a motor vehicle to explore the city.

Favorite Start Locations

The charts below show the favorite start locations of casual riders and annual members. These charts include the top 10 starting locations for casual riders and annual members with 4 stations being in the top 10 starting locations for both.

The chart labeled "Favorite Start Location of Members and Casual Riders" includes the actual values of the amount of rides starting from these locations. The chart labeled "Normalized Favorite Start Locations" displays the same values but normalized to more easily visualize the difference between annual members and casual riders.





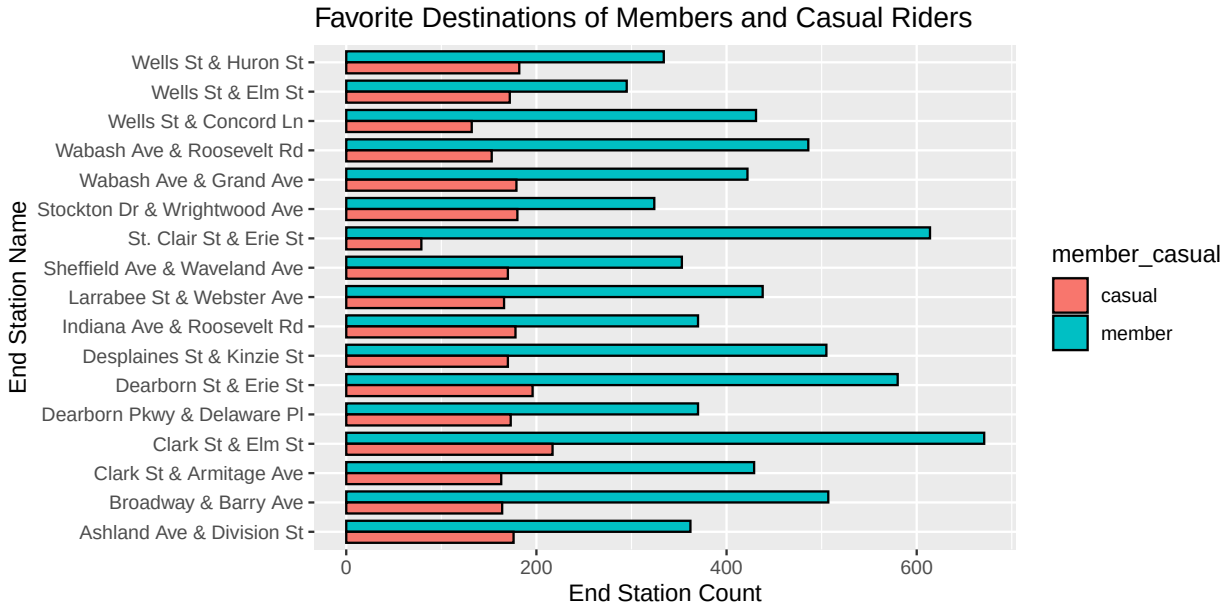
As displayed by these charts, the favorite starting location for annual members and casual riders is Clark St & Elm St. Interestingly, the St. Clair St & Erie St location is very popular among annual members with 589 rides but is not very popular with casual members with only 93 rides. This could be due to a number of factors, such as, this station may be near a high-density residential area and people use this station to go to

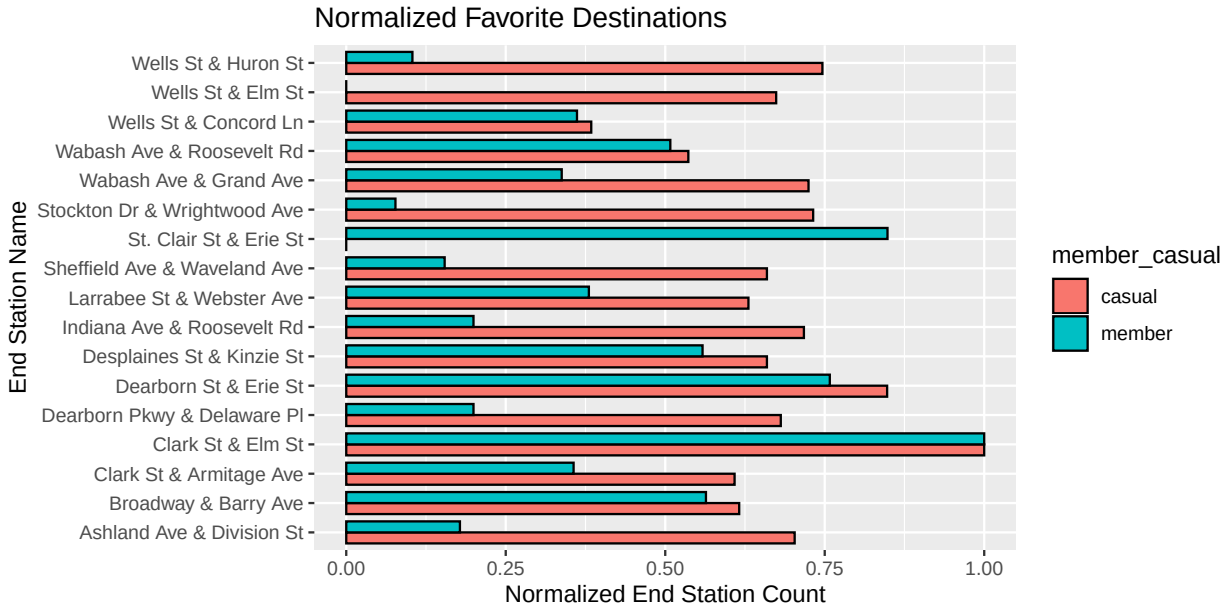
and from work, but would greatly benefit from looking at historic data to identify trends for this particular station.

Favorite End Destinations

The charts below show the favorite end destinations of casual riders and annual members. These charts include the top 10 ending destinations for casual riders and annual members with 3 stations being in the top 10 ending destinations for both.

The chart labeled "Favorite Destinations of Members and Casual Riders" includes the actual values of the amount of rides ending at these locations. The chart labeled "Normalized Favorite Destinations" displays the same values but normalized to more easily visualize the difference between annual members and casual riders.



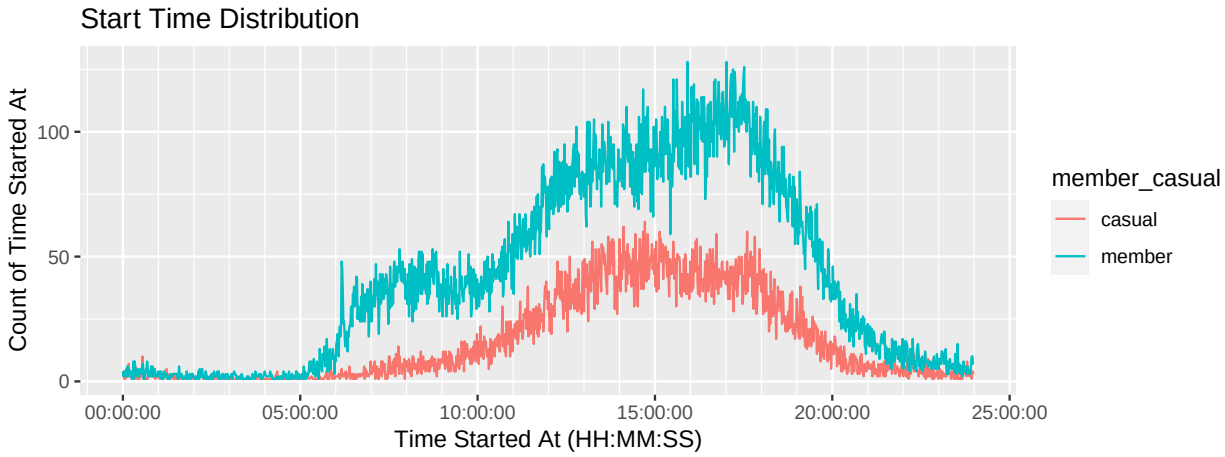


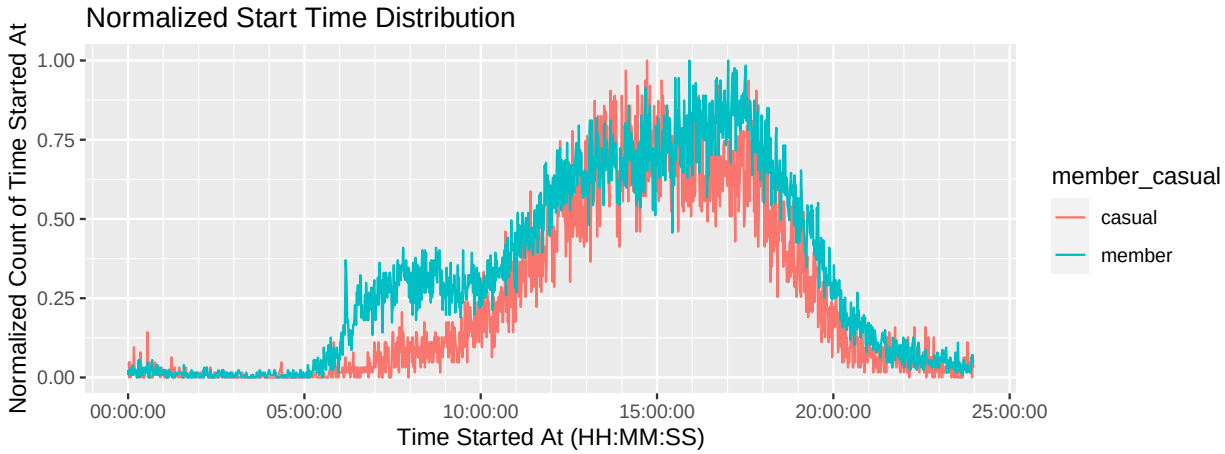
As displayed by these charts, the favorite ending destinations for annual members and casual riders is Clark St & Elm St. The St. Clair St & Erie St location is very popular among annual members with 614 rides but is not very popular with casual members with only 79 rides. As stated in the previous subsection, this could be due to many factors and should be explored more thoroughly in the future. Also, the amount of casual members using these 17 locations is relatively the same besides the St. Clair St & Erie St location having about half of the typical amount of riders.

Start Times

The graphs below display the distribution ride start times between the annual members and the casual riders. The peak times for annual members are at 15:55 (3:55PM) and 17:01 (5:01PM) with 128 rides starting at these times. The peak time for casual riders is at 14:43 (2:43PM) with 64 rides starting at this time.

The graph labeled "Start Time Distribution" displays the actual number of rides at a certain time while the graph labeled "Normalized Start Time Distribution" displays the same data but normalized to more easily visualize the difference between annual members and casual riders.





As displayed by these graphs, peak ridership for both groups is between around 13:00 (1:00PM) and 18:00 (6:00PM). This peak time window should be studied more and it may be beneficial to ask riders of both groups to participate in a survey as to where they are going to better understand why this peak is during such a wide window. A spike in ridership of members happens around 6:00 (6:00AM) and 7:00 (7:00AM) and is most likely due to riders using bikes to travel to work but the number of riders stays consistent until around 10:00 (10:00AM) when the amount of riders increases. As stated previously, this spike in riders in the middle of the day could be better understood with more research and data.

Recommendations

When deciding on strategies to convert casual riders to annual members, the two formats that came to mind were to expand annual member access and/or restrict casual rider access. Since Cyclistic is interested in converting current casual riders to annual membership holders, either strategy could potentially work, but the recommendations provided in this report focus on expanding annual member access.

Electric Bikes

As shown in the "Ride Length" subsection of the analysis, casual riders, on average, use a bike for 14 minutes longer than annual members. One strategy to convert casual riders to annual members would be to offer different bikes such as electric bikes or battery assisted bikes to annual members. These bikes would be appealing to riders that have longer ride times as these bikes reduce the physical strain of long rides. Certain bike docks could be converted to allow the charging of these electric bikes so that riders will always have a fully charged electric bike when they want to use one.

Bike Reservations

During the "Start Times" subsection of the analysis, this report examined the frequency of start times of annual members and casual riders. It is shown that the peak ridership for both annual members and casual riders is between 13:00 (1:00PM) and 18:00 (6:00PM). A strategy to expand member access would be to allow annual members to reserve bikes at locations through the use of a mobile or web application. This can ensure that members have access to bikes when they need them especially at popular locations during busy times of the day and incentivizes casual riders that use these same locations to become annual members.

Bike Infrastructure Improvements

Another strategy to increase annual memberships would be to improve bike infrastructure throughout the city of Chicago. This will most likely have to be a joint effort with the city as well as private donors as increasing bike infrastructure has a large upfront cost but Cyclistic will likely see a return on investment over the following years. The most useful infrastructure changes or improvements would be painted and protected bike lanes, bike path and bike lane snow removal services, and intersection treatments for bikes, such as bike boxes[3] and modifying intersections to give bikes and pedestrians priority over motor vehicles. These changes will likely increase overall ridership throughout the city and, in turn, will increase annual membership in areas of expanded bike infrastructure. Additional research will be needed in order to target specific areas and neighborhoods, such as the areas examined in the "Favorite Start Locations" and "Favorite End Locations" subsections of the analysis, that could benefit from these expansions.

Final Thoughts

For the month of April 2020, casual rides only made up about 28% of the total rides and annual members made up about 72% of the total rides. Without pricing data for these casual rides it is difficult to grasp how much of an impact it would make for Cyclistic if more of these casual riders became annual members. More historical data, such as a year-to-date data set, would help identify if April 2020 displayed typical year-round ridership or if this month differed significantly from other months. More historical data would also improve the ability to discern patterns between casual riders and annual members.

[3] https://nacto.org/publication/urban-bikeway-design-guide/intersection-treatments/bike-boxes/