

| **Case Study Roadmap - Ask** |
| --- |
| **Junior Data Analyst:** Isaiah Gumban  **Date:** September 5, 2022  **Project Summary:**  Since 2016 Cyclistic has had a fleet of 5,824 bicycles which customers can use for single rides, the full day, or the full year. We have discovered over the last 8 years that annual memberships have been the most profitable, and will be key to future growth. If we are able to convince our casual riders to opt-in for the annual membership, we will be able to see not only record revenues year-over-year, but we will also be able to see through future growth of our business.  **Guiding Questions:**   * What are the trends in ridership for casual riders and annual members? * How can we attract new members through understanding of insights from casual and annual members? |
| **Who are our stakeholders in this project?**   * **Primary Stakeholders:** Lily Moreno and the Cyclistic Executive Team * **Secondary Stakeholders:** Cyclistic Marketing Analytics Team |
| **Business Task:**  Identify the number of riders that are using our annual membership, analyze the data of historic ridership, and develop a strategy to increase the number of annual members and decrease casual riders (i.e. single-ride or full-day pass users). We will visualize these findings to identify trends and gain insight as to where the opportunities are to increase annual membership sign ups. In the final stage, three recommendations will be given as a proposal to address how we can increase annual membership sign up. |

| **Case Study Roadmap - Prepare** |
| --- |
| Our data has been obtained from Divvy, which is a Chicago government licensed sharing service which operates under Lyft. The data has been organized by year, month, and day, all of which is from a public data source which can be accessed [here](https://divvy-tripdata.s3.amazonaws.com/index.html). The data has been licensed for public use, and is protected from misuse and security breach of users. Due to these datasets being public data, customer personally identifiable information (such as credit card information, address, name, and purchases), is protected and inaccessible. The terms and conditions of the licensing can be found on the [licensing page](https://ride.divvybikes.com/data-license-agreement) of Divvy.  **Data sources** will be found in the appendix of this case study. |
| **Datasets** can be found in my public Google Drive. Raw data [click here](https://drive.google.com/drive/folders/1hFxAk9aMq_hGMY0osAJMJxfTK6rWeSry?usp=sharing)  The data will be organized by year and month. |
| The data of 12 months will be used for our analysis. To prepare the data, all I have done in this step is ensure there are no duplicates, and null values on Google Spreadsheets. I will be unable to do any further analysis on Spreadsheets going further given the size of these datasets. I plan to use R to clean, manipulate, and visualize the data. |

| **Case Study Roadmap - Process** |
| --- |
| **Tools Used**   * Google Spreadsheets: Used to calculate total time of bike rides, shortest ride length, longest ride length, mode ride length, average ride length for members and casual riders, and average ride length by day of the week * R: Develop data frames to combine tables to discover trends of common ride length times and durations from different years |
| **Data cleaning and manipulation steps**   1. Use Spreadsheets to check for blanks by using conditional formatting to highlight cells red if there are null values. 2. From this point forward, I will use R 3. Step 1 is install packages required to perform analysis (tidyverse, diplyr, data.table, and ggplot2), as well as call their library data for use (library() function). 4. Step 2 is to combine all of the datasets into one large data frame    1. First, I assigned variables to each of the 12 datasets used, with the csv file serving as the dataset    2. Second, I checked the structure of each dataset to check for any column types that were not the same    3. Because April 2020 and May 2020 had their start and end station IDs set to integers instead of char character strings, I changed this using the mutate() function    4. Next, I checked for the number of rows in my new data frame, and checked to verify the column names were the same as the previous 12 datasets    5. Lastly, I created new columns for date, month, day, year, and day of the week 5. Step 3 involved calculating ride lengths for each row    1. The difftime() function was used to get the difference between the start and end time, have it written in seconds, and stored in a new row for ride lengths    2. The ride length was then converted to numeric value 6. Step 4, and the final step of our Process phase, involved cleaning the data by checking for times less than 0/negative time, and duplicates    1. We already filtered out nulls on Spreadsheets    2. The filter() function was used to filter out values for ride length less than 0 or negative    3. distinct() function was used to make sure that there were no duplicate ride instances recorded |

| **Case Study Roadmap - Analyze** |
| --- |
| **Tools Used**   * R: Find the average, shortest, middle, and longest ride length (min, mean, median, and max) Compare average ride lengths of casual riders and members, as well as number of rides of casual riders and members. |
| **Data aggregation steps including analysis of data**   1. Use min(), median(), mean(), and max() functions to find the shortest, middle, average, and max ride lengths 2. The average, longest, and shortest ride lengths of casual riders and members are then compared using aggregate() function 3. We then again use the aggregate() function to compare average ride time of casual riders and members    1. ordered() function is used to sort the ride\_length column by days of the week starting with sunday. This allows comparisons for when analyzing data to be simpler 4. Lastly, we compare number of rides and average ride lengths for casual riders and members for each day of the week |
| **Findings**   * Casual riders use bikes for longer time frames than members * Members use bikes more often, almost double than that of casual riders |

| **Case Study Roadmap - Share** |
| --- |
| **Answers to original guiding questions:**   1. **What are the trends in ridership for casual riders and annual members?**    1. This answer is much more apparent after performing our analysis. In the visualizations below, you will see that there are longer durations of use by casual riders compared to members. This is because our data categorizes casual riders for one time use and day pass users as “casual”, and therefore could contribute to longer ride durations than our members who use our bikes for commute or exercise purposes    2. Contrary to our first finding, there are about half as many casual riders using our bikes than member riders. 2. **How can we attract new members through understanding of insights from casual and annual members?**   Our target from our findings is to find a way to decrease the number of casual riders, or increase the number of rides casual riders make. We will have to assess the cost of casual rides to determine if it would be more practical to encourage more membership enrollments, or decrease the number of casual rider usage. |
| **Visualizations**  **Average ride length - Casual rider vs Member**    **Number of rider per day - Casual rider vs Member** |
| **Key Findings:**   * Longer duration rides by casual riders * More total rides by members |

| **Case Study Roadmap - Share** |
| --- |
| **Conclusion**  In conclusion, if we want to see the future growth of our business through membership enrollments, our target will be to market to casual riders. This should involve a revamped membership program, fees for single use or day passes, and notifications for casual rides. |
| **Three recommendations**   1. **Increase fees for casual and day pass riders**   Companies in interior design utilize memberships to sign up clients for membership, and will provide fees per hour for their service. As of now, single ride passes cost $3.30 US, and day passes cost $15 US per day. If we were to increase single ride passes to $4 US and day passes to $18 US, we would at least be able to cover 2 months worth of membership ride costs per user. This would hopefully offset the number of casual riders, and attract them to opt in for our membership.   1. **Revamp the membership program to lower the cost for longer rides**   Incentivizing our membership program would be a great way to promote growth. As mentioned above, our data tells us that members use their bikes for much shorter periods of time. We could potentially offer them a monthly fee decrease of say 10-15% if they are able to use our bikes for longer periods of time. This not only incentivizes them to go on longer rides, but also creates membership retention for a small fraction of a cost.   1. **Email notifications to casual riders advertising the membership**   One final recommendation I’d make is advertising what casual riders can save by opting in for a membership. This can be included in the email notification confirming the ride details, and also be another way of attracting additional membership enrollment. |