**Bike trip Case Study**

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

(By: Sourabh Kumar)

This case study is the final capstone project for the Google [Data Analytics Professional Certificate] (https://www.coursera.org/professional-certificates/google-data-analytics). We are helping a company located in Chicago called "Cyclistic", and we are working as apart of the data analytics team.

Cyclistic, a bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

This report will walk through the 6 phases of data analysis in detail. I will also include a summary section (Ask Phase) at the end of the recommendations I think Cyclistic should take on board, in order to successfully grow their business further.

Ask Phase: -

Objectives: -

Our primary questions to explore and gather information on are:

\* How do annual members and casual riders use Cyclistic bikes differently?

* Cyclistic has 2 types of members "casual" and "member". Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

\* How can we maximize annual members?

* The more annual members the business can get the better for long-term performance.

Key Stakeholders to Consider

Lily Moreno: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program.

Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

(Other stakeholders include investors, employees, customers, and suppliers)

Prepare Phase: -

The data source:-

The data source is a set of [12 different data sets](https://divvy-tripdata.s3.amazonaws.com/index.html), each one containing the information for every month of the year (April 2020 - March 2021). The data is provided by Motivate International Inc. under this [license](https://www.divvybikes.com/data-license-agreement). Each data set was prior cleaned to remove any rider's personal information, to prevent breaching any data confidentiality regulations.

The data source was used as it meets all the "ROCCC" (Reliable, Original, Comprehensive, Current, Cited) features.

- The data is reliable as it is provided by a reputable source (Google).

- Data is original as it is coming from the primary source, and not from a second/third-party source.

- Data is comprehensive because it contains sufficient data that we can use to gather useful insights.

- The data is current as it provides us with data sets within the past year (at the time of making this report).

- Data sets used have been cited/vetted.

Process Phase: -

Tools I used and why:-

For this case study, I decided the best tool to use was R as it provides easy manipulation and processing of large data sets. I decided to use Tableau for visualizations.

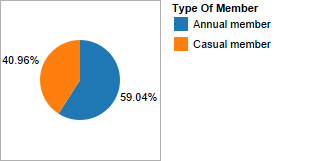
How I ensured data integrity & data is clean to use for analysis?

I first loaded all 12 data sets into R and examined the structure, data types and naming for inconsistency across the 12 data sets.

I noticed ‘start\_station\_id’ and ‘end\_station\_id’ columns are char datatype in dec20 - mar21 (should be double), So I converted then to double to keep consistency when we joining all 12 data sets into 1. After joining all 12 data sets into 1, we can notice that some of the naming of the columns could be more readable, e.g. ‘member casual’ can be changed to ‘type\_of\_member’.

A new column called ‘ride length’ was added to find total trip duration in minutes by finding diff between ended and started dates, this can be used for further analysis later. After adding this new column, we need to check for any negative’s values, also We have been informed that the company had "test" stations used for quality checks So we can see how many there are. Have checking we have gathered there are 3367 tests, and 10552 negative values, so we can remove these from the data-set as they can cause issues in the future. We should create a new data frame for this as we are removing data.

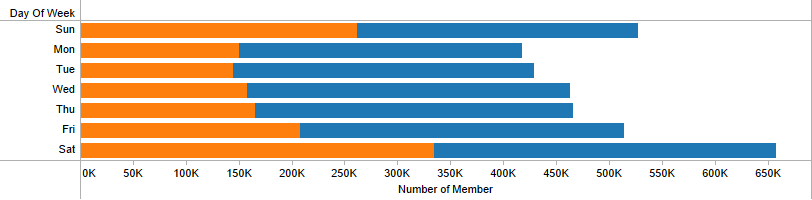
Analyze and Share Phase: -



*Graph represent Percentage of the 2 different types of customers, both member and casual.*

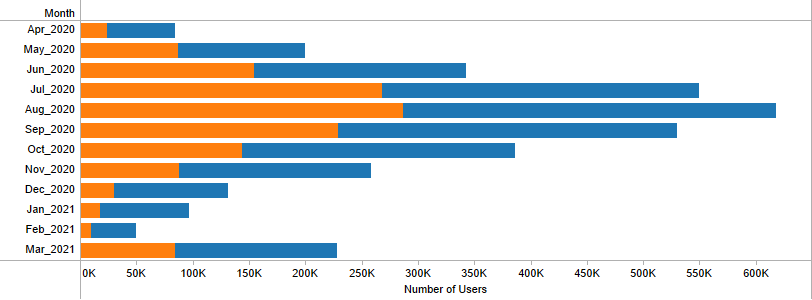
As we can see 59% of riders are members and 40% are casual, which means more than half of the customer base is a member, this is a positive metric, but can be improved. Although getting 100% of the customers as members is very unlikely, aiming to get to 70 - 80 % can be an optimistic goal.

*Days of week Vs Number of Member.*



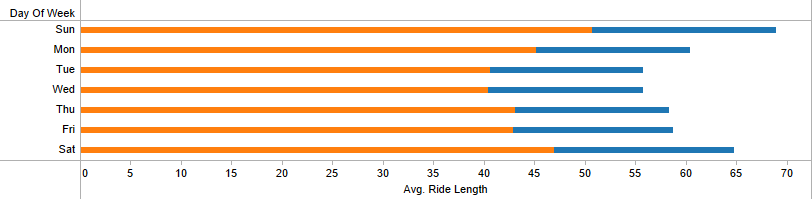
From the graph, the blue color represents Annual member whereas orange represent Casual member. From the visualization we examine that annual member tend to have a consistent number of rides throughout the week, while on the other hand casual riders have more popularity during the weekends (Saturday - Sunday).

*Month (Apr\_2020 to Mar\_2021) Vs Number of Users.*



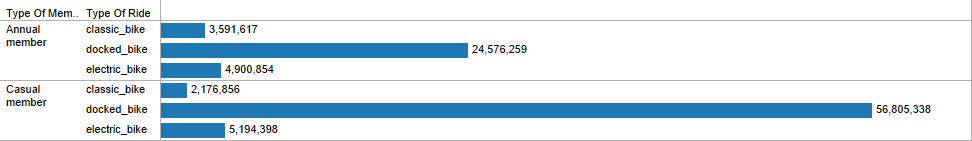
From the graph, the blue color represents Annual member whereas orange represent Casual member. We can see from the second quarter (July - September) was the most popular time of year, August being were number of rides peaked for both customer types. Between July - September accounted for almost 50% of all rides.

*Days of Weeks Vs Average Ride Length.*



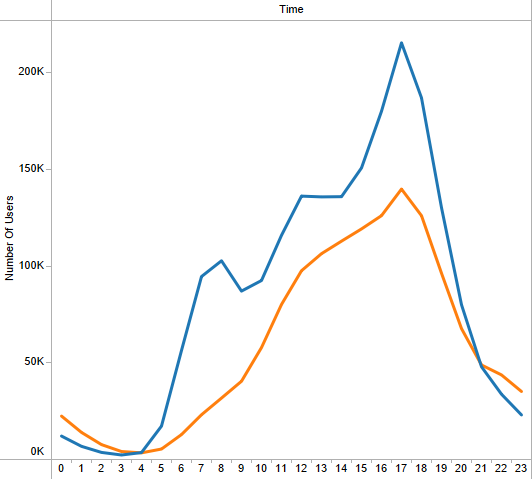
From the graph, the blue color represents Annual member whereas orange represent Casual member. On average trip duration for casual riders is more than twice the amount for annual member. Also, we can see that the trip duration for annual member consistent throughout the days of the week, unlike casual member.

*Bike ride Vs Number of Rides Vs Types of Members*



With a first glance at our analysis, we can see that docked bikes have the most demand, by both annual member and casual member riders. Electric bikes leaning more popularity towards members, come second but by a significant gap, and finally classic bikes have the least demand and is more favored towards member riders.

*Bike popularity throughout the Day*



We can also see that the is a large spike in demand between 4am - 8am, and demand rapidly decreases after 8am. Furthermore, popularity greatly declines after 10pm and stays well below 1000 riders until 4am.

Act Phase: -

Primary takeaways: -

Annual member and casual riders prefer docked bikes (70%) significantly more than electric or classic.

Second Quarter is the most popular quarter of the year for riders at that time climate is normal, not cold or hot.

Winter is the least popular season of the year for riders.

Casual riders have almost twice the average trip duration compared to members.

August the most popular month of the year, while February is the least popular.

My top 3 recommendations for Cyclistic: -

1. Since Second Quarter is the most popular time of year focusing marketing campaigns and advertisements at this time of year could help draw in more customers. Also providing weekly/monthly memberships can help unsure customers get a better sense of if they want to commit to a full annual membership.

2. Including some form of discount for the first year of the annual membership could tempt casual riders to purchase this offer, and that can cause the riders to actively engage more in trips, and potentially cause them to continue renewing their annual membership each year.

3. Currently there is no benefit to being a member apart from financial savings for customers. Since trips are most popular during the weekend for casual riders, and most of the demand is during the evenings this shows there are a lot of parents/children riders. So offering a family discount option or some form of special perk such as completely going on a trip every weekend results in some cash back/free merchandise at the end of that month.

Further analysis options: -

\* Gather data on customer specific details such as age, gender, occupation etc. (with consent), this can allow us to further analysis the type of riders and their habits.

\* More geographic data to see if this is a factor to consider when buying passes/memberships.

\* The pricing plans and strategies data used for the memberships, that way we can see if there is any need to reconsider the pricing plans to encourage more annual membership purchases.