



Google Data Analytics Professional  
Certificate - Capstone Project  
Cyclistic Bike-Share Company Case Study  
Case Study: How does a bike-share navigate  
speedy success?  
Completed by: Thomas Hobson

## **Introduction:**

I have been given the role of junior data analyst who works in the marketing analyst team at The Cyclistic Company in Chicago. Cyclistic is a bike-share company who is looking at alternative transportation by offering a variety of bikes to commute with around the city. They offer the service to members who can choose to use it as a casual rider or an annual member.

I will be using the valuable skills I have learnt through the Google Data Analytics Professional Certificate course to produce the relevant information, insights and recommendations that are required to achieve the desired goal of having casual members convert to annual members.

## **Ask:**

Being an annual member reaps a better financial benefit and important growth in the future for Cyclistic. The job that I have been assigned is to gather data, work through it, clean it, analyze the data and produce valuable insights in how they can convert casual riders into annual members. By doing this, the marketing team will have recommendations for when we design a new marketing strategy. Once we have a new strategy we can then present it to our director of marketing and manager. Once it is approved it will then move to the final step of approval which would be the Cyclistic executive team.

The main question that my team has assigned me is:

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

## **Prepare:**

The data that I have access to is the past 12 month history trip data of Cyclistic and this ranges from 2022/03 up until 2023/03. This data is located in the [divvy-tripdata](#) index and is organized separately at the end of each month under its month's date

ie. 202203-divvy-tripdata.zip. I have then downloaded each month I will be using and stored them in my local drive in chronological order for inspection.

After going through the datasets, I can say that they are safe, reliable, free from any bias and all the data was collected and stored by the company Cyclistic.

This amount of data will allow me, after analyzing, to bring valuable insights to the stakeholders and provide the recommendations that they need in order to create that new marketing strategy to convert casual riders to annual members. At first glance there aren't any problems with the data. The next step in the process stage will clean the data and remove any unwanted errors that may be in the data.

## **Process:**

In this stage of cleaning up the data I have chosen to use BigQuery SQL as we will be working with a large data source. We need to make sure that it is clean for us to be able to analyze it and get the correct information and insights without there being false returns in the process. After reviewing the data I have noticed and summarized what I will be cleaning:

### **Pre-Cleaning Process Notes**

1. We will need to create a table that combines all the months. Joining them all in SQL with the UNION ALL clause removes all duplicates that may be present.
2. Ride\_ID - We will have to see if there are any NULL entries and if all characteristics are the same length. If there are NULL we will be removing them. If there are ones with different characteristic lengths then we will get rid of any extra spaces.
3. Rideable\_type - We need to determine what are the different types of bikes a user can choose from.
4. Started\_at / Ended\_at - Here we will be looking if there are any trips that were longer than a day as that will affect some of our data running over from one month to another, creating incorrect data for some months. We will also remove trips less than 1 minute as these might indicate a wrong choice of machine or cancel of trip.  
This will make our sample size smaller.

5. Start\_station\_name / End\_station\_name / Start\_station\_id / End\_station\_id - We need to assess if there are any NULL variables and remove them from the study. We also need to look if there are any naming inconsistencies and fix them by removing extra spaces before, in and after the text string.
6. Start\_lat / Start\_lng / End\_lat / End\_lng - We need to see if there are any NULL values and remove them from the study.
7. Member\_casual - We need to see if these are the only categories and that there are no miss spellings.
8. We will be adding four columns:
  - 8.1. Ride\_length which will indicate to us how long the trip was.
  - 8.2. Day, Month, Year and these will extract from our started\_at timestamp the day of the week, month of the year and year that the trips took place.

## **Cleaning Process Notes**

1. Started\_at / Ended\_at
  - 1.1. We removed the rows that had trips longer than a day from our dataset and shorter than 1 minute.
  - 1.2. Of this 153 747 rows were removed from our dataset
2. Start\_station\_name/id & end\_station\_name/id
  - 2.1. We removed all the NULL values from our dataset.
  - 2.2. We removed any extra spaces trailing and leading from all the values in our dataset.
  - 2.3. We removed any repair / testing / maintenance stations.
  - 2.4. We cleaned the substrings from the values that had.
3. Columns created
  - 3.1. Ride\_length calculated and stored as minutes
  - 3.2. Day, month and year extracted and shown in separated columns

## **Analyze:**

Here we are going to analyze our clean data and present it in visuals and share our insights with our stakeholders. To do so I will be using SQL to analyze the data and Tableau.

In this analysis I will be focusing on the following:

1. Which bike is the preferred bike
  - 1.1. Amounts all users
  - 1.2. Amongst each user separately
2. When the trips occurred
  - 2.1. Favored Day
  - 2.2. Favored Month
  - 2.3. Favored option - Weekday or weekend
  - 2.4. Total trips per month
  - 2.5. Total trips per day
  - 2.6. Average trip duration
  - 2.7. Longest trip
  - 2.8. Shortest trip

## **Share:**

Here we will be showing our findings from our analysis process and present them in our powerpoint and create compelling visuals to go with it.

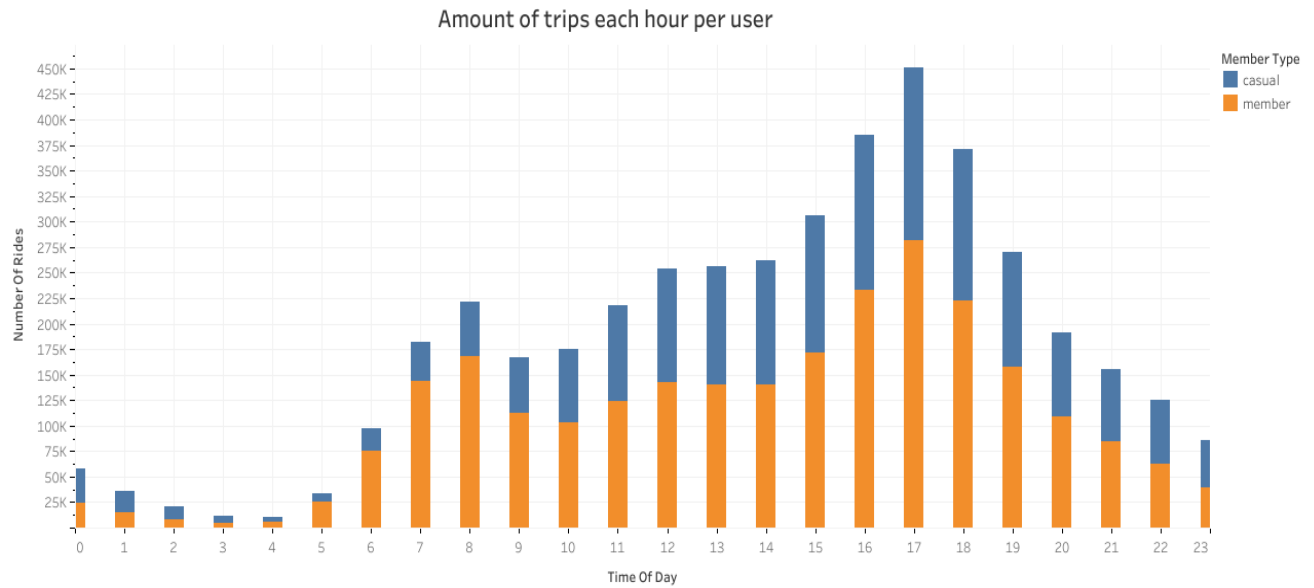
To create our visuals we will be using Tableau and to present our findings we will display it all in a powerpoint presentation.

Here is some of my findings:

When do users travel?

Casual members - travel later in the day suggesting that they are more tourists, students and not from the working class.

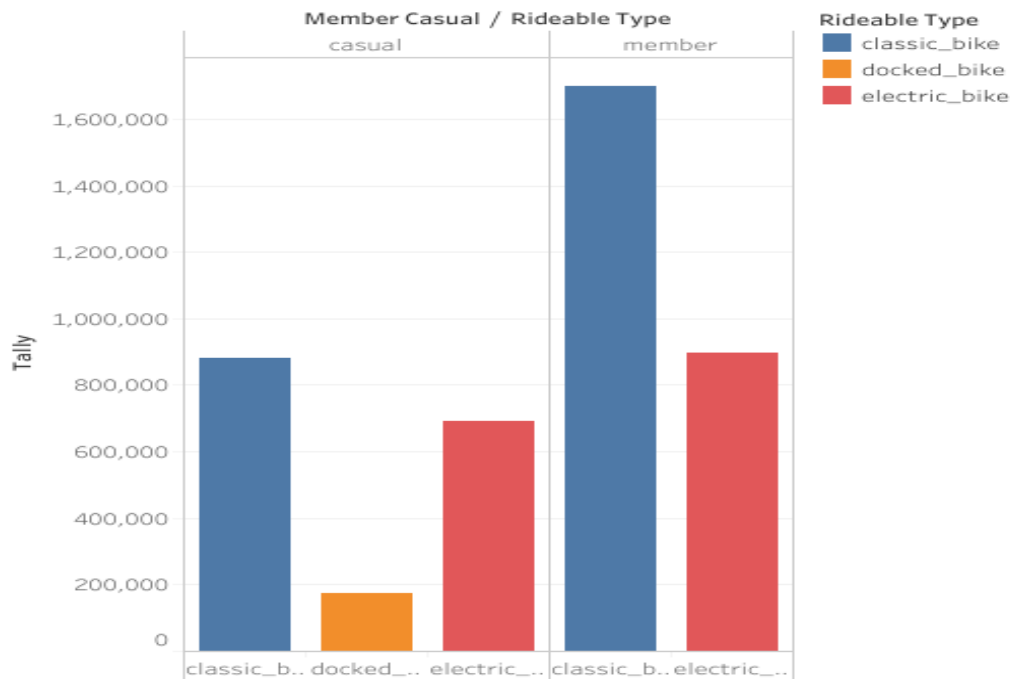
Members - and traveling more around 6 - 8 am and 16 - 18 pm. This will suggest that they are using the bikes for work more than usual



What is the preferred bike type amongst users?:

Both users prefer the classic bike. This could solely come down to the price differences and the value for money. This also gives them the positive of staying healthy and cycling to their destination.

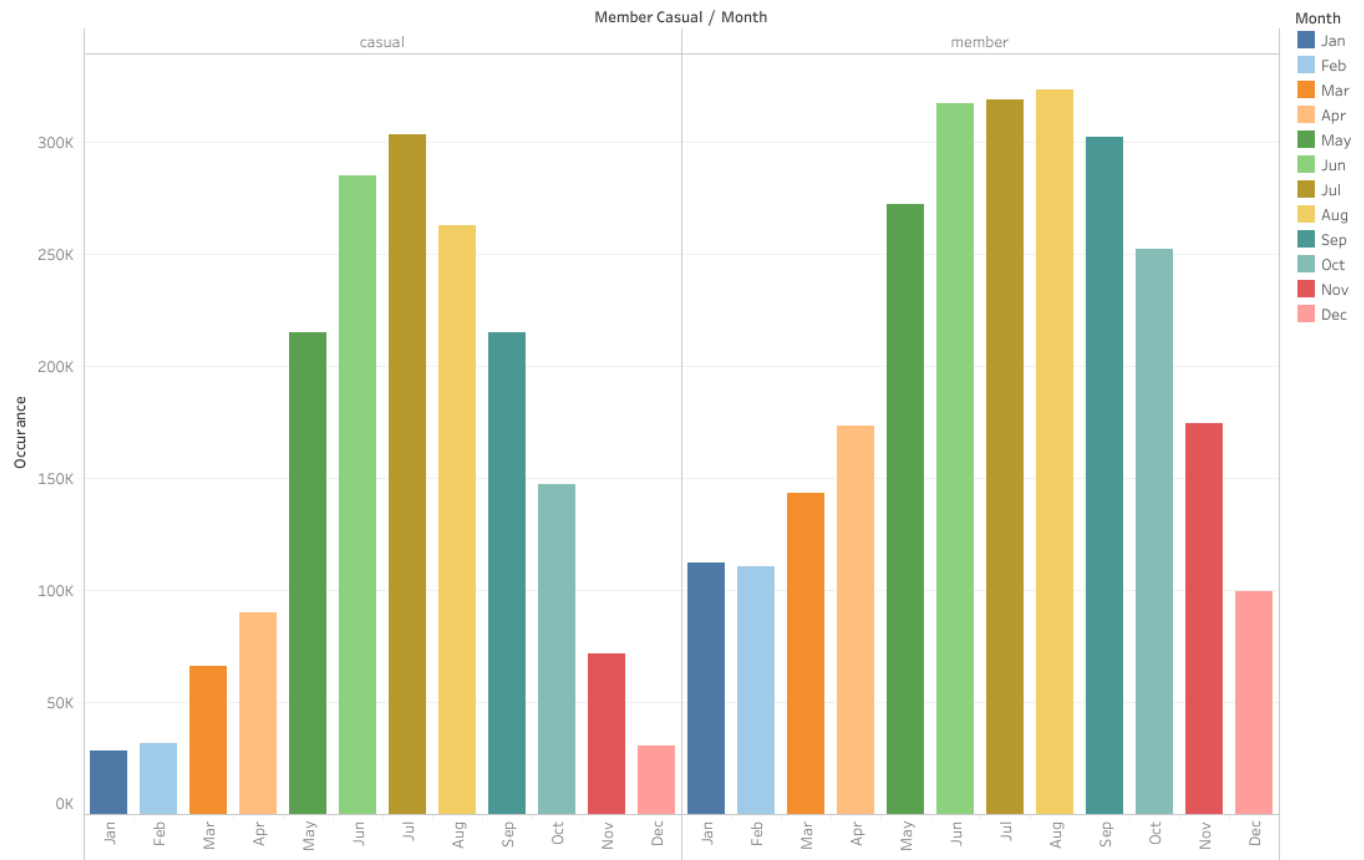
### Preferred Bike



What are the most popular months traveled?

The most traveled months for both users are May - September. This could suggest that it is for the summer holidays as the total amount of trips almost doubles compared to the other months of the year.

Amount of trips per month



What are the most popular days to travel?

When it comes to the members the days in the week are more and come close to being doubled compared to the casuals. This confirms our theory of them using it for work purposes.

Casuals however seem to use it more over the weekend where they use it for their leisure and sightseeing trips.

## Amount of trips per day

Day	Member Casual		Amount	
	casual	member		
Sun	1,414,145	2,097,457	47,480	2,097,457
Mon	59,301	89,573		
Tue	60,824	80,083		
Wed	54,710	80,955		
Thu	61,381	83,347		
Fri	48,346	81,635		
Sat	47,480	84,770		

### **Recommendations:**

The results have shown that the members use the bike share company as a form of transport to work, whereas the casual members use it as leisure.

The members use it on a day to day basis whereas the casual users use it predominantly on weekends and in times that are later in the day.

The marketing strategy is to get the casual members to convert into annual members. To do so I have some recommendations that I think may help achieve this goal:

1. You could create a bundle that is exclusive to users that come for holiday where they can purchase a bundle that has a number of trips per bundle.  
For example you could purchase day trips, week trips, weekend or even month to month trips. This will allow users to book 'passes' that will allow them to travel more and be more beneficial financially.
2. You could create a bundle for casual users that allows them to only purchase for the summer holiday and label it accordingly. This will allow the majority of the trips to be easily accessible to everyone at a price where they find it more affordable to their needs. This will attract more users to the membership options than being casual.
3. My last recommendation will be to market the prices better and make the membership option more beneficial or add perks to it where if you have traveled a

certain amount of days in a row or an allotted time then you get a month free or week free. Entice users to get that option and travel more so that they can get more benefits.