

# **CYCLISTIC BIKE-SHARE ANALYSIS REPORT**

## **(CASE STUDY: HOW DOES A BIKE-SHARE NAVIGATE SPEEDY SUCCESS?)**

By: Ines Elizabeth

for Google Data Analytics Capstone

### **INTRODUCTION**

‘Case Study: How Does a Bike-Share Navigate Speedy Success?’ is a capstone project from the last course of Google Data Analytics Professional Certificate on Coursera. In this case study, I performed tasks of a junior data analyst in a fictional company, Cyclistic, and answered the key business questions by following the steps of the data analysis process: ask, prepare, process, analyze, share, and act.

### **SCENARIO**

Cyclistic is a bike-share company in Chicago that features more than 5,800 bicycles and 600 docking stations. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships by converting casual riders into annual members. Therefore, my team wants to understand how casual riders and annual members use Cyclistic bikes differently.

Cyclistic has flexibility of its pricing plans, there are:

- Single-ride passes
- Full-day passes
- Annual memberships

Based on its pricing plan, Cyclistic customers are divided into two type, there are Casual Riders and Cyclistic Members. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

## DATA ANALYSIS PROCESS

### 1. Ask

Cyclistic's finance analysts have concluded that annual members or Cyclistic Members are much more profitable than Casual Riders. The director of marketing also notes that Casual Riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Business task for this analysis is maximizing the number of annual memberships by converting casual riders into annual members by understanding of how casual riders and annual members use Cyclistic bikes differently.

Then, three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

### 2. Prepare

This case study use Cyclistic's historical trip data to analyze and identify trends. I used the previous 12 months of Cyclistic trip data (January 2022-December 2022) from [divvy tripdata](#). (Note: The datasets have a different name because Cyclistic is a fictional company. The data has been made available by Motivate International Inc. under this [license](#)). Datasets itself are stored and saved by month in csv file.

### 3. Process

After unzipped all data and stored in different folder, I used Excel Workbook to process. I created a column called "ride\_length" and calculated the length of each ride by subtracting the column "started\_at" from the column "ended\_at". It was formatted as 'HH:MM:SS'.

Then, I created a column called "day\_of\_week," and calculated the day of the week that each ride started using the "WEEKDAY" command in each file. It was formatted as General or as a number with no decimals, noting that 1 = Sunday and 7 = Saturday.

I also created a column called "month" using IF command. I used it by inserting the RIGHT function in column "started\_at" as part of IF's logical\_test.

After that, the cleaning process began by inspecting if there were nulls, #####, and blanks. Nulls and ##### were found in “ride\_length”. Blanks were found in “end\_lat”, “end\_lng”, “end\_station\_name”, and “end\_station\_id”. I deleted nulls, #####, and blanks in “ride\_length”, “end\_lat”, and “end\_lng”. Therefore, I did not delete blanks in “end\_station\_name”, and “end\_station\_id” because their “ride\_length” and “day\_of\_week” column were not null or blanks and could be used for further analysis.

At first, I planned to use SQL for my analysis, but several problems happened (some files are over 100MB and need to upgrade the BigQuery sandbox) while importing Excel files into BigQuery. Because of that, I decided to use R language and converted all of my files format from .xlsx (Excel format) to .csv.

I saved Excel file as .csv and reopened using Data tab to get into Text Import Wizard. In step 2 of Text Import Wizard, I unchecked all delimiters to combine all column into one column only. Then, I copied it to Notepad and copied again to new Excel Workbook. Last, I saved the new workbook as .csv file in separated folder.

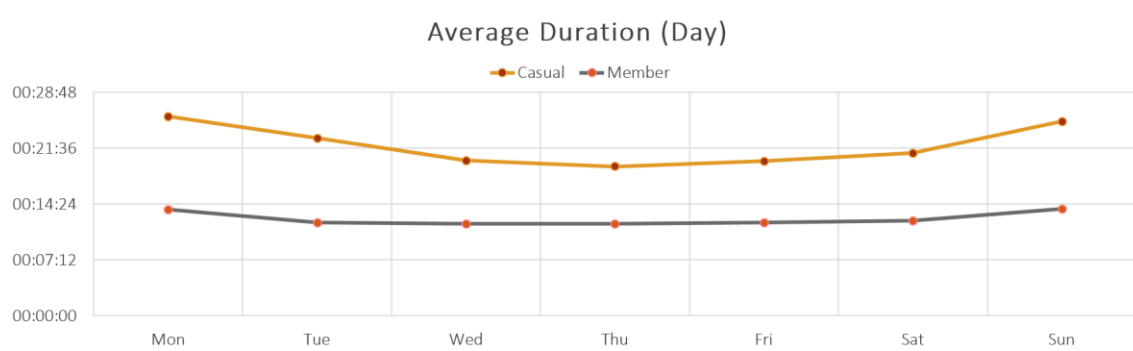
After imported all files into Rstudio Dekstop and ran several inspections using head(), colnames(), str(), glimpse(), and skim\_without\_charts(), there were incorrect formats of “end\_station\_id” on June 2022 and September 2022 data. To avoid error, I converted “end\_station\_id” format to character.

Then, I merged all data into one data set. The merge data set consisted of 5.667.186 rows and 16 columns. I renamed “member\_casual” column name to “rider\_type” to avoid the confusion. I also converted “day\_of\_week” format to numeric for later calculation.

#### **4. Analyze**

Analysis began with conducting a descriptive analysis on merge data set. Descriptive analysis included mean, median, max, and min. These results could also be found by using summary().

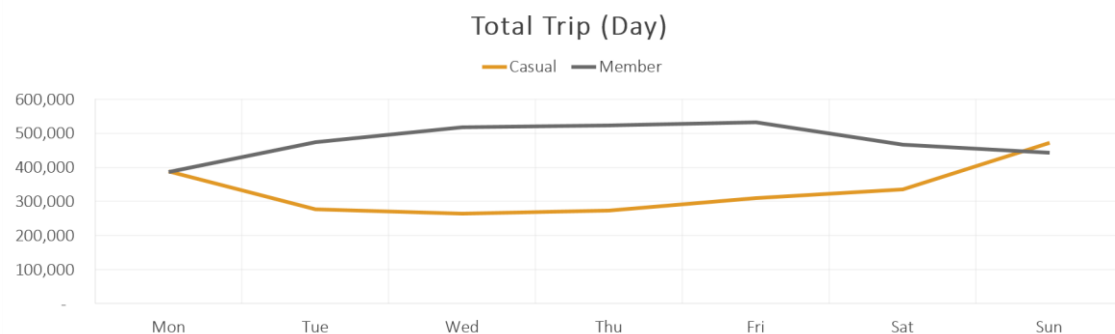
Then, annual members and casual riders could be compared by using aggregate() to look their mean, median, max, and min. Aggregate() was also used to get an average ride time by each day between annual member and casual member (“rider\_type”). The graph below is its visualization using Excel.



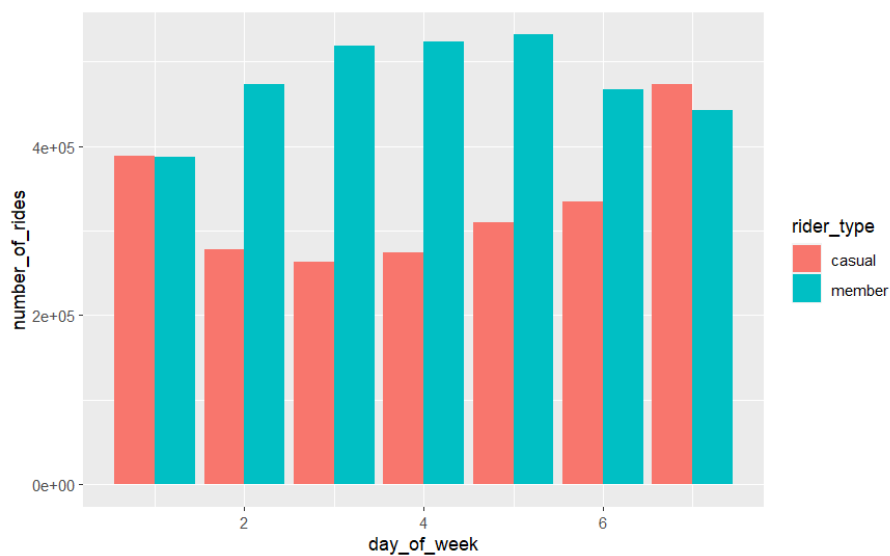
Graph 1. Average Duration (Day) using Excel

The graph shows that both Cyclistic Members and Casual Riders rode longer during weekend, but Casual Riders' average duration in a week was 27.33% higher than Cyclistic Members.

I also analyzed ridership data based on "rider\_type" and "day\_of week". The graph below is visualization for number of ride by "rider\_type" using Excel and ggplot().



Graph 3. Number of Ride or Total Trip (Day) using Excel

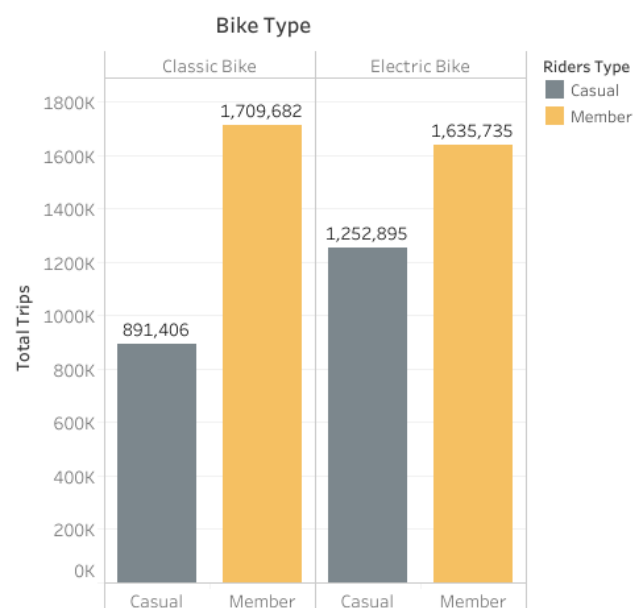


Graph 2. Number of Ride or Total Trip (Day) using ggplot()

Both graphs show Cyclistic Members ride's numbers were more stable across weekdays, but Casual Riders rode more often during weekend.

Graph 1, Graph 2, and Graph 3 indicate most of Cyclistic Members using bike-share for daily use (repetitive activities) such as going to work and school, which the distance and duration were close and short. By choosing annual membership, it helped them to use bike-share easier.

Meanwhile, Casual Riders rode more often and longer during weekend could be indicated that most of Casual Riders usually used bike-share for leisure. Biking around city, hill, or for recreational purpose would be several examples of these. Because of these rare occasions, Casual Riders rather chose single-ride pass or full-day pass than annual membership.



Graph 4. Type of Bike using Excel

Type of bike based on "rider\_type" was exported into csv file for visualization using Tableau and Excel. Graph 4 shows that in 2022, total Casual Riders were 2,144,301 and Cyclistic Members were 3,345,417. Based on analysis, Cyclistic Members rode 21.88% more often than Casual Riders. The reason behind these number could be Cyclistic Members' purpose using bike-share for their daily activities across weekdays.

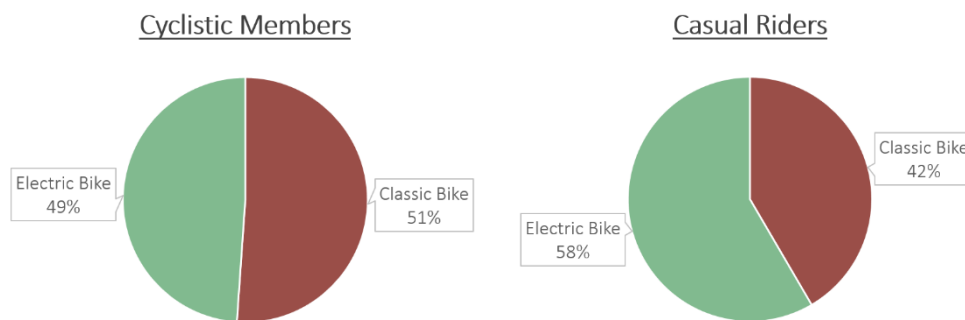
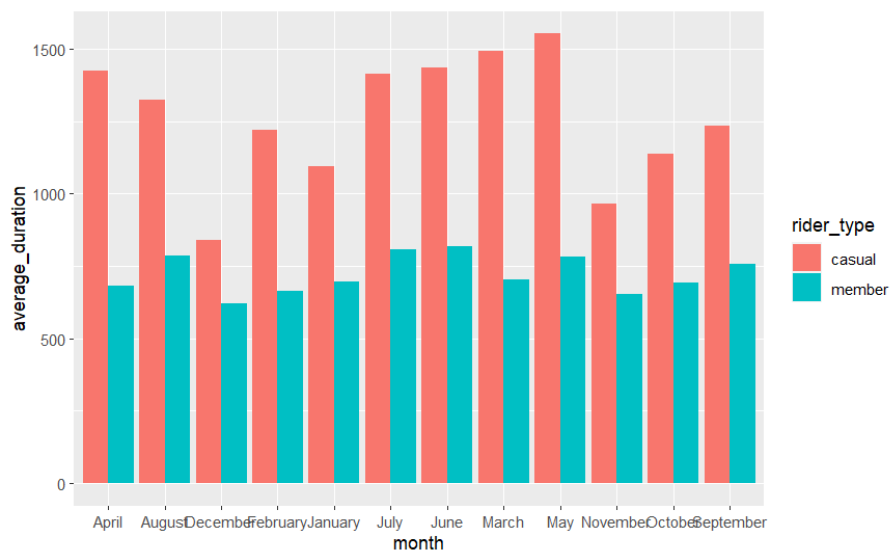


Chart 1. Type of Bike Comparison using Excel

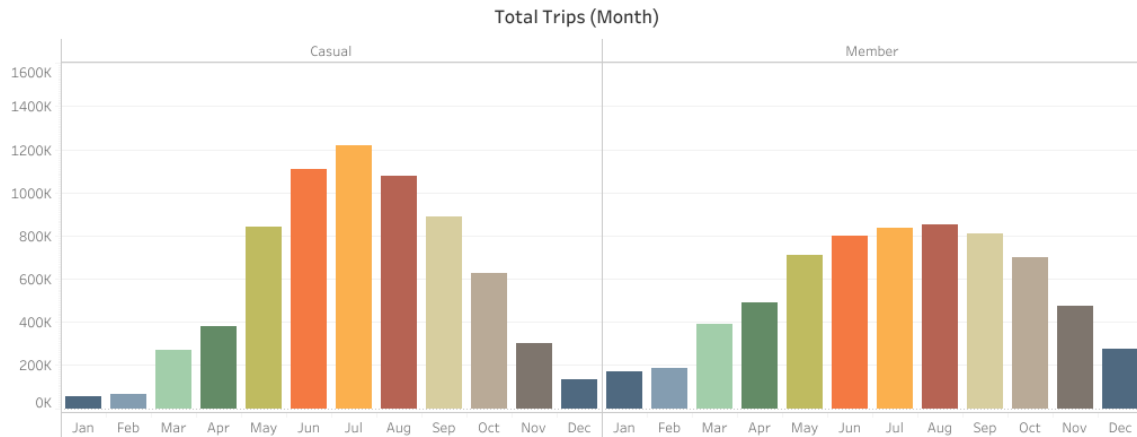
For type of bike, Chart 1 shows that 51% of Cyclistic Members chose to use classic bike and 49% of Cyclistic Members chose to use electric bike. Meanwhile 42% of Casual Riders chose to use classic bike and 58% of Casual Riders chose to use electric. Based on these results, it could be concluded that Cyclistic Members prefer to use classic bike and Casual Riders prefer to use electric bike.

The reason Casual Riders prefer to use electric bike could be the durability of bike to use in several different tracks besides flat roads. Electric bike is also faster than classic bike. Electric bike is the best option for rider who wants to travel longer distances and longer duration.

Two graphs below are visualization for average duration by month by “rider\_type” using ggplot() and total trips by month using Excel.



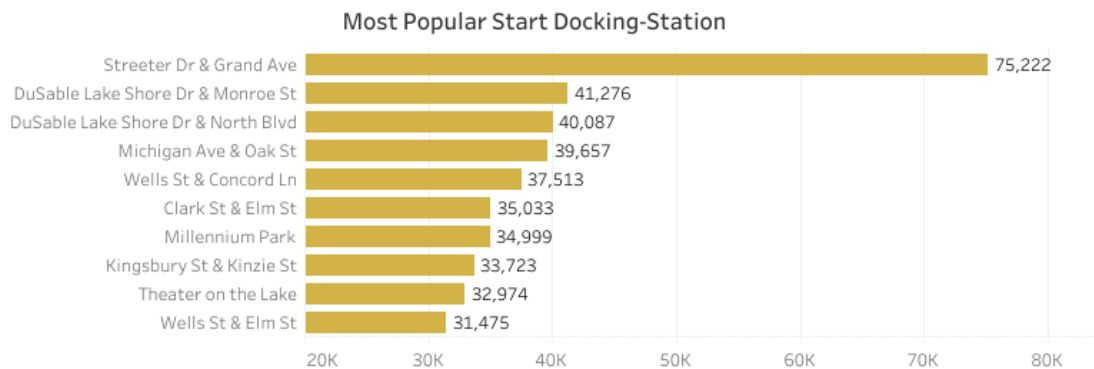
Graph 5. Average Duration (Month) using ggplot()



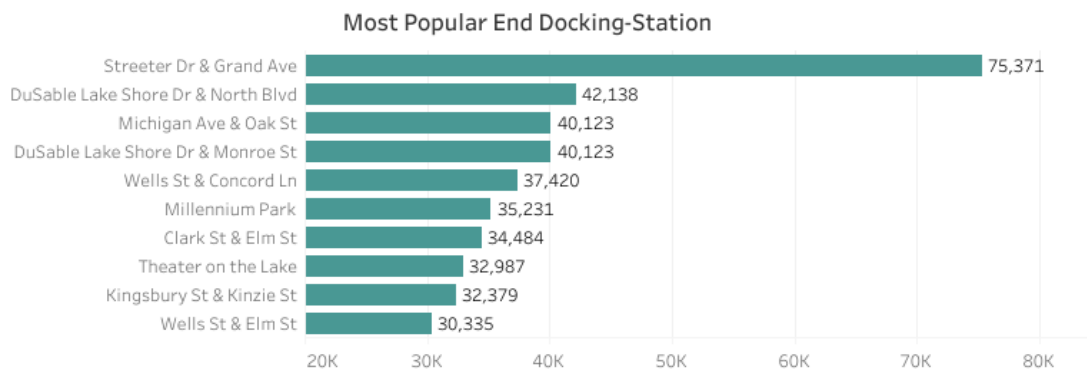
Graph 6. Total Trips (Month) using Excel

Total number of riders per month and total number of riders per day were also exported into csv file for visualization using Tableau and Excel. Graph 5 and Graph 6 show both Cyclistic Members and Casual Riders ride more frequently during warm season (May to October), but Casual Riders has more significant increase than Cyclistic Members.

After that, count() was also used to find most popular start and end docking station. The results were summarized in visualization using Tableau.



Graph 8. Most Popular Start Docking-Station using Tableau



Graph 7. Most Popular End Docking-Station using Tableau

From Graph 7 and Graph 8, Streeter Dr & Grand Ave was the most popular docking-station in 2022. Streeter Dr & Grand Ave is located around popular area in Chicago. There are several recreational areas such as Jane Addams Memorial Park, Ohio Street Beach, Milton Lee Olive Park, Chicago Children's Museum, and Navy Pier. Navy Pier itself is one of most popular destination in Chicago for shopping, dining, or enjoying attractions such as Pepsi Skyline Stage. It also near apartment building named Lake Point Tower and several transit place.

## 5. Share

For visualization, I used Tableau and Excel. Besides this analysis report, the result of these analysis can also be found in PowerPoint format in my GitHub account. Here are my Tableau and GitHub links:

- Tableau : [Tableau Link](#)
- GitHub : [GitHub Link](#)

## 6. Act

Summary and recommendation based on my analysis are included in two sections of the report below.

## SUMMARY

There are several insights found during analysis.

1. Casual Riders preferred to use electric bike 11% higher than Cyclistic Members.
2. Casual Riders' ride number had significant increase during warm season: May to October.
3. Casual Riders rode more often and longer during weekend. It indicated that Casual Riders usually rode for leisure.
4. Casual Riders' average duration in a week was 27.33% higher. It indicated that Casual Riders used Cyclistic to travel longer distances.
5. The most popular docking stations in 2022 was Streeter Dr & Grand Ave, followed by DuSable Lake Shore Dr & Monroe St and the rest were:
  - DuSable Lake Shore Dr & North Blvd



- Michigan Ave & Oak St
- Wells St & Concord Ln
- Clark St & Elm St
- Millennium Park
- Kingsbury St & Kinzie St
- Theater on the Lake
- Wells St & Elm St

Business task in this analysis is maximizing the number of annual memberships by converting casual riders into annual members by understanding of how casual riders and annual members use Cyclistic bikes differently

## **RECOMMENDATION**

Based all of insights above, I recommend four promotion strategies to reach the goal of business task. There are:

1. Provide more model of electric bike and add a special benefit only for annual membership to enjoy the latest model of electric bike.
2. Collaborate with several universities to target Casual Riders who are university students, and offer a special annual memberships price.
3. Choose promotional time based on day and month, when Casual Riders ride more often and longer. Focus the promotion during weekend and warm season month: May to October.
4. Choose promotional location in ten most popular stations, especially Streeter Dr & Grand Ave, and several recreational places. Place bigger advertisement and host a few interactive events about cycling as daily mobility.