



Google Data Analytics

## **CASE I: Cyclistic Bike Share**

Jesús Ponce González

August 2nd, 2021

Certificate Program

## Business Task

Cyclistic is a successful bike-sharing company dedicated in sharing a bike network to all Chicago users. The bikes can be unlocked from one station and returned to any other station in the system anytime. In order to maximize income, Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders.

Our main business task is to maximize income by understanding the key differences between casual and annual members. Using the information obtained by analyzing the data from the past twelve months, we will be able to convert casual members to annual members. The key is not to lose clients, we need to convert them into annual members, so the income increases and make the annual program more appealing.

## Data

The data used for this analysis was obtained from the [company's AWS server](#) and will be stored separately in a physical server with data protection, security and redundancy in case of loss, the data will only be accessed by analysts.

Data is organized by month, and it has the following columns:

1. ride\_id: Rider's ID number
2. rideable\_type: Type of bike used
3. started\_at: Date-time of rider's start trip
4. ended\_at: Date-time of rider's end trip
5. start\_station\_name: Starting station's name
6. start\_station\_id: Starting station's ID number
7. end\_station\_name: Ending station's name
8. end\_station\_id: Ending station's ID number
9. start\_lat: Starting latitude
10. start\_lng: Starting longitude
11. end\_lat: Ending latitude
12. end\_lng: Ending longitude
13. member\_casual: Type of member (casual or member (annual))

Data will be manipulated via SQL and Excel, making only available to analysts who are going to work on the project. Furthermore, we can use this data to analyze trips of the users. We can see how often annual members use bikes compared to casual members. We can also see the trip's final time and what type of trip the rider used it for. This data does not include the rider's personally identifiable information to conserve the rider's identity as private.

## Preparing the data

These are the steps made for preparing the data:

1. The files were downloaded from the AWS server and stored in a separate folder.

2. After decompressing the files, a backup copy was made for each file in order to maintain order.
3. Saved the CSV files to XLS and created two new columns:
  - a. ride\_length: showing the ride's length in minutes and seconds.
    - i. Difference between ride start and ride end
  - b. WEEKDAY: showing the number of days the ride started
    - i. 1 = Sunday & 7 = Saturday

There are some inconsistencies with the data. Some trips start after the end date, or some start trips are ended right away which means that the ride\_length column is not correct. We also noticed that start\_station\_name is not always available. There are no duplicates for the ride trip ID. For all these inconsistencies, the data will be removed as adding zeros will make the calculations not work.

Performing descriptive analysis, the mode for the ride\_length is 13 seconds while the minimum ride was .01 seconds (which might indicate the ride was cancelled right away) and the maximum was 18:38 hours. The day rides are usually more common are on Saturdays. The average ride length is 22 minutes and 46 seconds.

We performed basic statistics for casual and annual members using a pivot table; the average ride length for casual members is 37 minutes while for annual the average length is 14 minutes; the average day for casual members is 4.132 while for annual members is 4.00; with this data we can say that both casual and annual usually ride bikes on Thursday. We also calculated the variance and standard deviation and found that casual members are usually more inclined to ride a bike between Tuesday and Saturday while annual members are more inclined to ride between Tuesday mid-day and Friday mid-day.

## Data visualization

The following visualizations show the importance in the data. (Click the image)

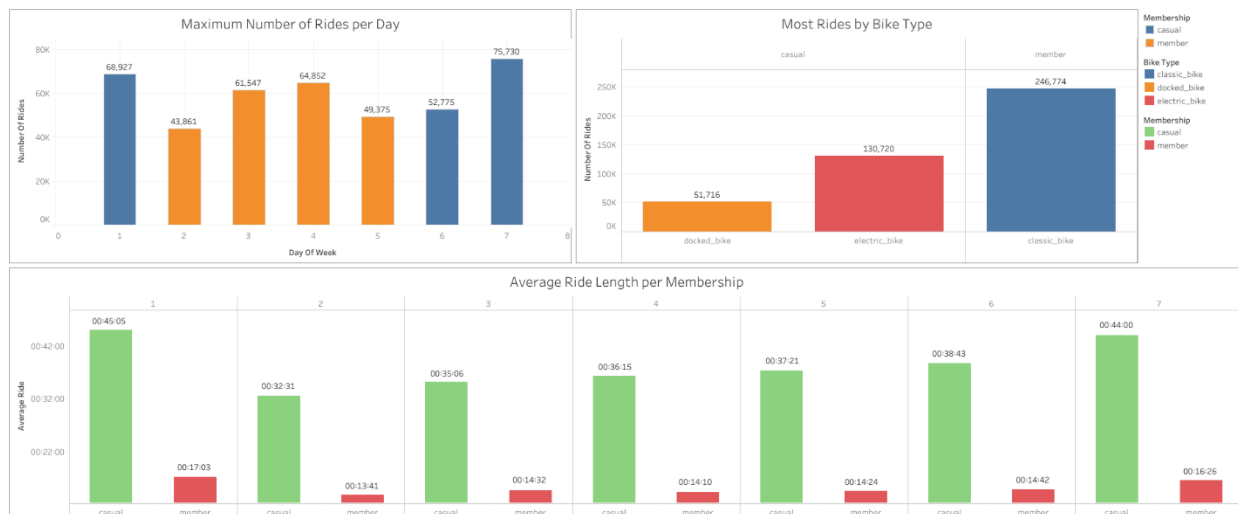


Figure1. Tableau Dashboard

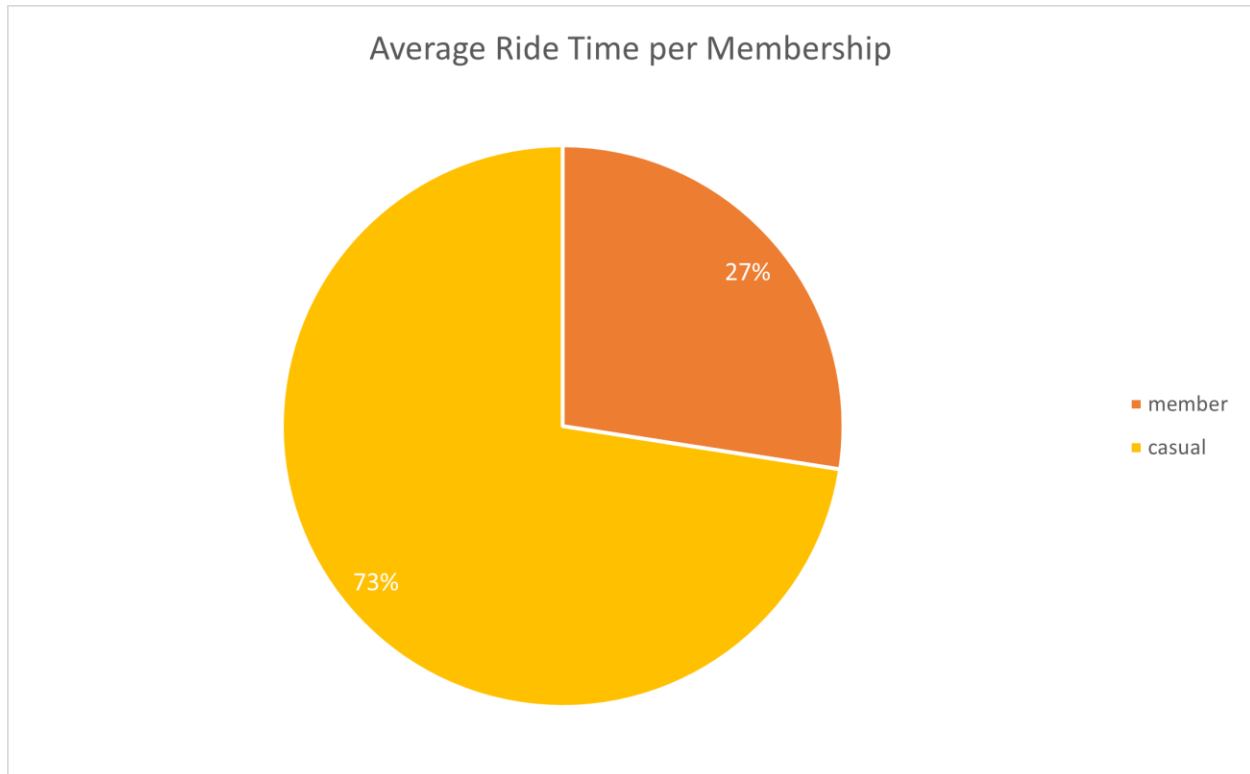


Figure 2. Pie Chart comparing average ride time per membership

The data is showing that casual members use bikes the most while annual use it from time to time. This makes sense since the more common days annual members ride a bike are weekdays while casuals ride it more on weekends. This means that annual riders are usually workers.

We can use this data in order to focus on the casual members who ride bikes on weekdays while make emphasis that when purchasing an annual membership, you save more money since you will not be paying daily or hourly every weekend. The dashboard shows that it doesn't matter which day it is, casual members usually use most of the bikes, which in the long run will make them spend more money.

## Results.

Results show that it the bike scene are mostly for casual members. Casual members ride a bike usually for longer times on weekends while annual drivers drive it less on weekends and more on weekdays. The difference might be that annual drivers are workers who use the bikes for commuting between home and work while casual might be riders who like to ride a bike only, but not tossing the daily casual workers who might use a bike to commute.

We will need to focus on the following:

1. Target the casual riders who ride a bike on weekdays, showing trends and how they can save money.

2. Show benefits of using an annual membership; annual riders use a classic bike more often, compare it between electric and docked, showing that a classic bike is versatile.
3. Target weekend casual riders by giving benefits of using rider through a week.

Overall, casual riders will enjoy riding a bike on weekends, but they can use an annual membership to show that they can save money instead of purchasing a daily or hourly pass. We can also start giving points to annual memberships for riding the bike for long periods of times.