

# Cyclistic Bike Share: Casual Riders vs Annual Riders

By Miguel Antonio Li

## The business task

Cyclistic is a bike-share company located in Chicago. There are two kinds of riders – annual riders and casual riders. According to the finance analysts of the company, riders with annual membership are much more profitable compared to casual riders. Based on their analyzation, the company has decided to create marketing strategies in order to influence their users to become annual riders. The business task I was given, is to analyze the company's data in order to differentiate casual riders and annual riders. The insights I would give from my business task would be beneficial to the creation of marketing strategies as it would help give more information about the annual riders and how they utilize the bike-share application.

One of the stakeholders in this task is Lily Moreno who is the director of marketing and my manager. She is responsible for the development of campaigns and initiative to promote the bike-share program. Another stakeholder I must take into consideration is the marketing analytics team. They are responsible for collecting, analyzing, and reporting data that would help guide the marketing strategy. Last but not the least, the Cyclistic executive team is another stakeholder I must take in consideration as they will be deciding whether to approve the recommended marketing program.

## Data sources used

In order to complete the business task, I have decided to use the previous 12 months of Cyclistic trip data. The data is located online through this link - <https://divvy-tripdata.s3.amazonaws.com/index.html>. In other words, the data I have decided to use is a public data and it has been made available by Motivate International Inc. All 12 months of data have their data stored in an excel sheet. Each data table contains the following columns – the ride id, the type of bike, the date and time of when the riders started and ended their bike session, the station in which the rider started their ride, the station in which the rider ended their ride, the starting latitude, the ending latitude, and the status of the rider that states whether if it's a casual rider or annual rider that rode the bike. Using this data would help me with my business task as it would allow me see the consistent differences between casual riders and annual riders.

Because this data is approved by the Motivate International Inc. which based from my research, is a bicycle sharing company located in New York, I am confident that the data provided is reliable,

original, and cited. As I have mentioned previously that the data contains all sorts of information, I would say that the data is comprehensive. Lastly, the data is current as I mainly focused on the data of last year, 2024. Although the data is in accordance to ROCCC, there are couple issues in the data, one of these issues is that there are some months whose data are incomplete or inaccurate. Some of these missing information would be the start station and end station. Another missing information that I have noticed in some of the data is the dates of when the user The inaccuracy that I have noticed while reviewing the data is that start and end time of the bicycle rides. The end time would be much earlier compared to the start time.

## **Documentation of cleaning or manipulation of data**

The tools that I have used for cleaning and manipulation of data are the following – Excel, Python, and PostgreSQL. To start, I have duplicated the data and in the duplicated files, I have created a new column – ride\_length. The column ride\_length enables me to create a function that calculates of the ride duration with the use of data on what time the rider started and ended their bike session. As I have mentioned earlier, there were some inaccuracies in the start and end time of the ride session in the data. Due to this, there are some results in the calculations that led to an error.

To further clean the data and efficiently remove rows where there are inaccuracies in the data under start and end time, I have used python with the use the pandas library. With the use of my code, I am able to remove those inaccuracies and have the data automatically be sent to PostgreSQL where I would continue with my analysis. After running my python script to clean and upload the data to my database, I have learned that I can categorize the data into which time of day the riders have rode their bike. With the use of python scripts, I have updated the data in my database and updated the table by adding a new column called “time\_of\_day.” The script checks the time of which the rider started to used the bike and sorted it into the following cases - morning, afternoon, evening, and night.

## **A summary of your analysis**

After cleaning and sorting my data, I used PostgreSQL to further analyze the data. To get the most of my analyzation, I have made a list of questions as a guide. These questions are the following – What is the average ride duration for casual riders? What is the average ride duration for annual riders? In what time of day do casual and annual riders ride bikes the most? what kind of bikes do casual riders most often ride? what kind of bikes do annual riders usually ride?

Based on my analyzation, the avg ride duration for the those who are annual is 0:04:44 while for casual riders is 0:02:47. The most common routes for casual riders are the following – around Streeter Dr & Grand Ave, and DuSable Lake Shore Dr & Monroe Street. The most common routes for annual riders are the following - from Ellis Ave & 60th St to University Ave & 57th St, from University Ave & 57th St to Ellis Ave & 60th St. The similarities between casual riders and annual riders is that a great majority of them have been using electric bikes through the 2024. Another similarity is the time of day both casual and annual riders would ride. In the months of January to May, there has been a lot of casual riders and annual riders who would ride in the afternoon. But in the rest of the months, both riders have only been riding during the night.

## Supporting visualizations and key findings

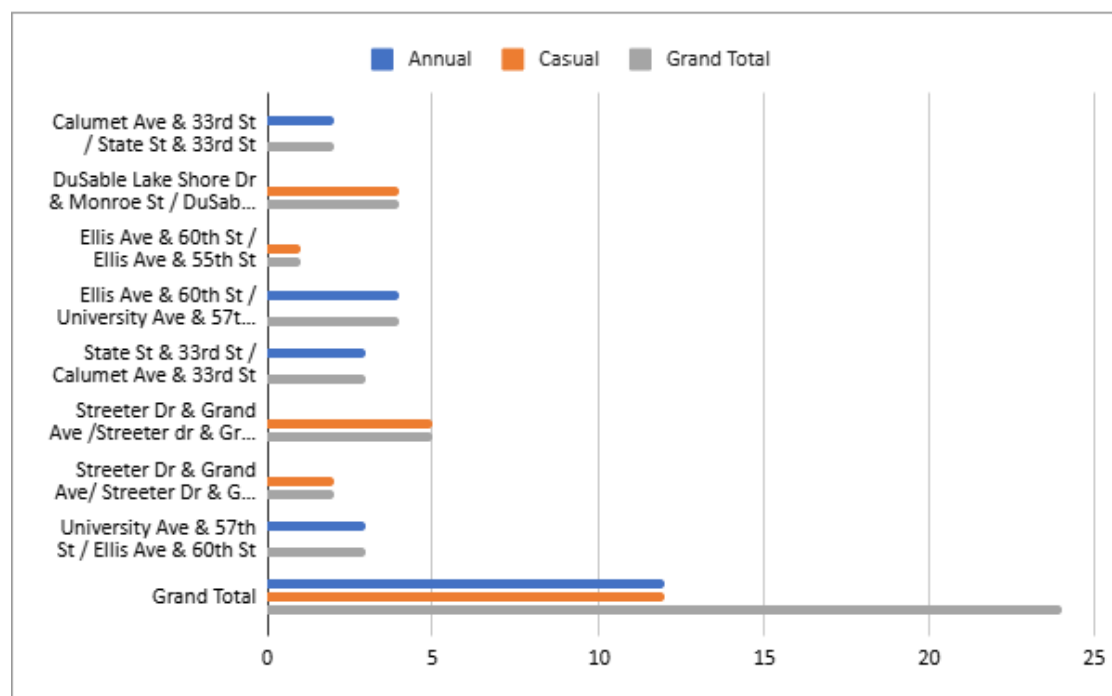
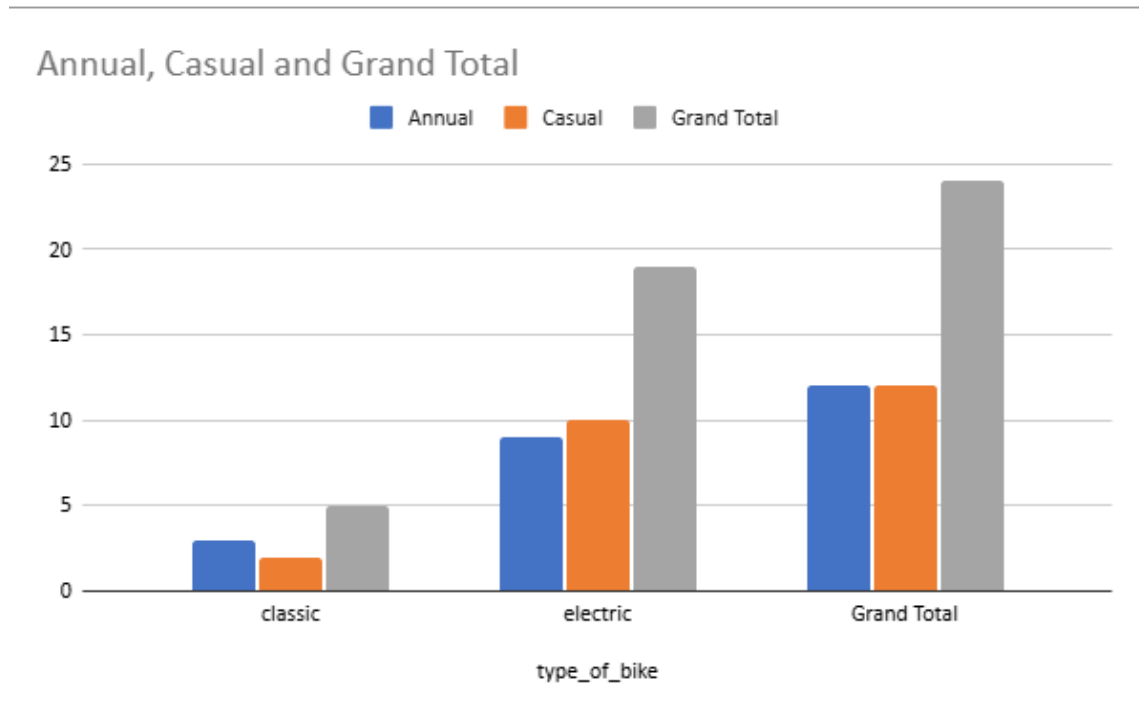


Figure 1 – Data on the Common bike routes of Annual riders and Common riders.

The figure above displays the most common areas that the annual riders and casual riders go around with the Cyclist application. Based on the data, the most common areas the casual riders go around are DuSable Lake Shore Dr & Monroe St / DuSable Lake Shore Dr & Monroe St, Streeter Dr & Grand Ave / Streeter dr & Grand ave, and Ellis Ave & 60th St / Ellis Ave & 55th St. The most common places that annual riders go to are Calumet Ave & 33rd St / State St & 33rd St, Ellis Ave & 60th St / University Ave & 57th St, State St & 33rd St / Calumet Ave & 33rd St, and University Ave & 57th St / Ellis Ave & 60th St.



*Figure 2 – Data on the type of bike the casual and annual riders have chosen for their ride*

The figure above displays the data on what type of bikes the annual and casual riders have been using throughout the year. Based on the data, the most common type of ride that both casual and annual riders ride is electric bikes.

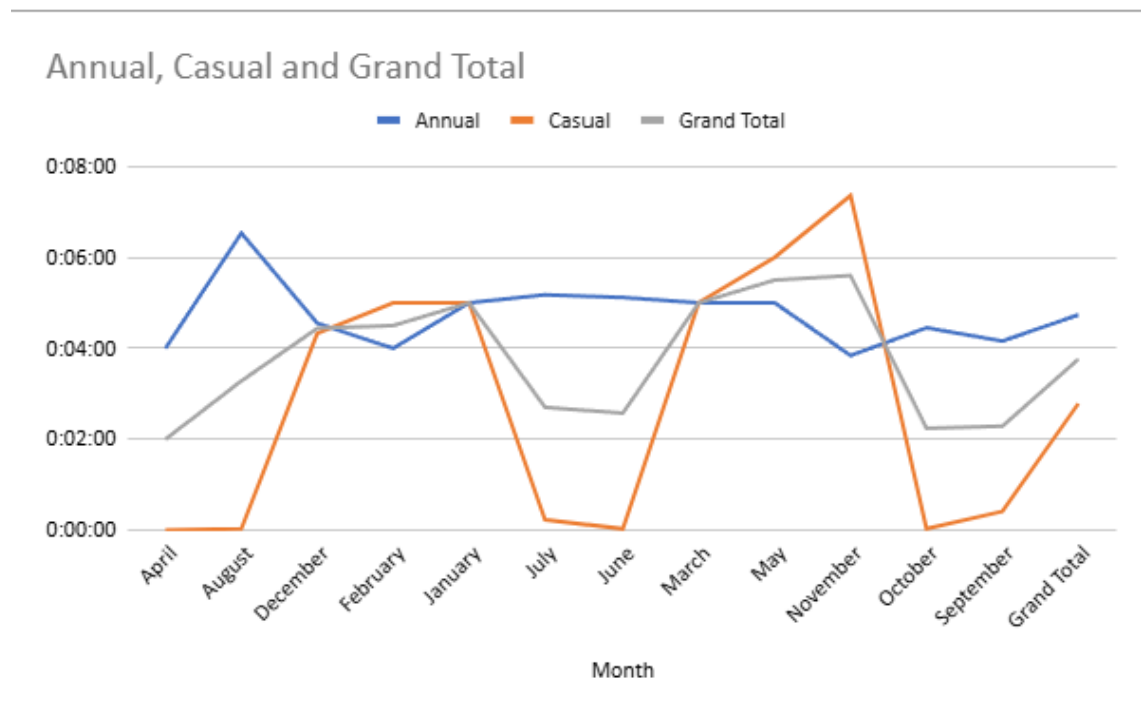


Figure 3 – Data on the average ride duration per month of the annual and casual riders

The figure above displays the monthly ride duration of both the casual and annual riders. Based on the data, the highest ride duration for casual riders occurred in November with the average time of 07:22 minutes. The highest ride duration for annual riders occurred in August with the average of 06:32 minutes. The total average of ride duration for annual riders is 04:44 minutes while for the casual riders, their total average 02:47 minutes.

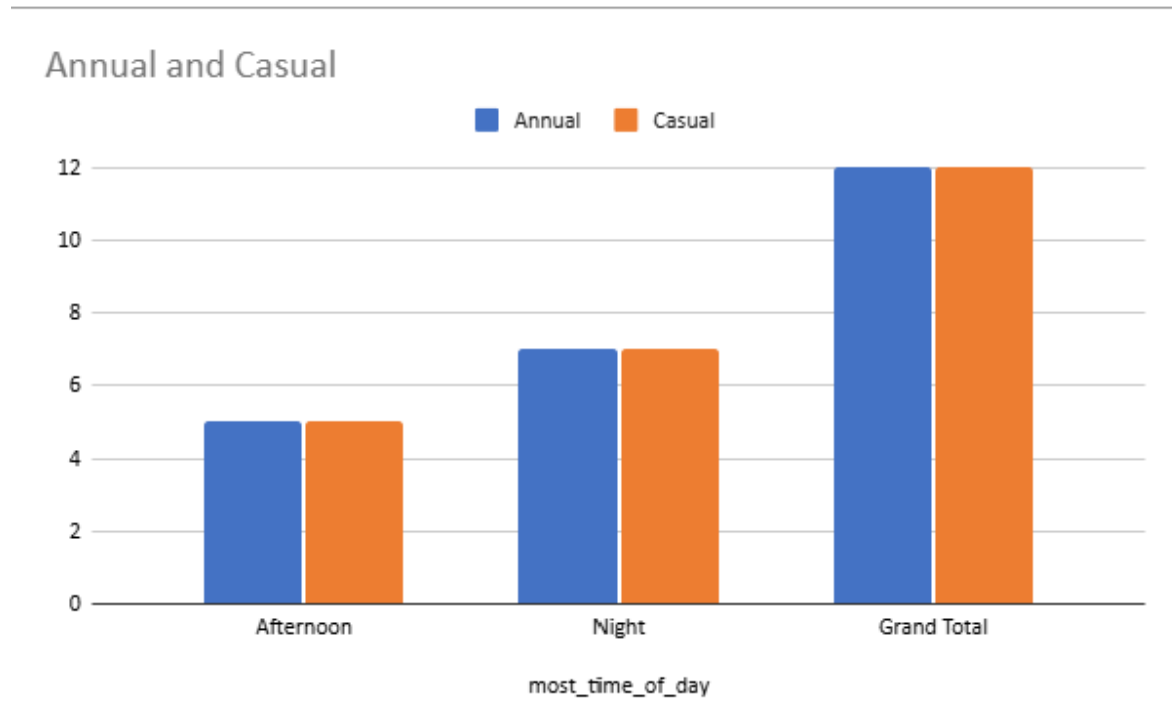


Figure 4 – Time of day which annual and casual riders would choose to ride

Based on the data, there are more casual riders and annual riders who would choose to ride bikes during the night (12am to 5:59am).

## Top three recommendations based on analysis

After my analyzation, I have three recommendations. My first recommendation is to plan events and engagements along around Streeter Dr & Grand Ave, and DuSable Lake Shore Dr & Monroe Street. This is because based from my collected data, majority of casual riders would often use the bike-share application there. Another recommendation is to plan the events at night. This is because in the last few months of last year, majority of bike riders would use the application at night. My last recommendation is to focus on more electric bikes. This is because based on the data I have provided; majority would use electric bikes compared to the classic bike and scooter.