Case Study:

**Differences between casual riders and annual members**

A Case Study about different customer of Cyclistic

by Matteo Barchiesi 01.09.2022

Contents

[**Business Problem 3**](#_Toc113361280)

[**Data Generation, Overview and Storage 3**](#_Toc113361281)

[**Data Cleaning 4**](#_Toc113361282)

[**Analysis 4**](#_Toc113361283)

[**Supporting Visualization and key findings 5**](#_Toc113361284)

[**Key Findings summary and Recommendations 9**](#_Toc113361285)

# Business Problem

The company we are looking at is Cyclistic, a bike-share company in Chicago.

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno (marketing director) believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members.

Three questions will guide the future marketing program:

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

Why would casual riders buy Cyclistic annual memberships?

How can Cyclistic use digital media to influence casual riders to become members?

Moreno has assigned me the first question to answer:

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

# Data Generation, Overview and Storage

For this project I’m going to analyze public data that’s already been gathered and made accessible by Motivate International Inc. under the following [license](https://ride.divvybikes.com/data-license-agreement). The data can be accessed in this page: [Download the previous 12 months of Cyclistic trip data](https://divvy-tripdata.s3.amazonaws.com/index.html).

Note that this is a generic public data set about ride sharing and Cyclist is a fictional company and doesn’t exist.

This is public data that can be used to explore how different customer types are using Cyclistic bikes. But note that data-privacy issues prohibit you from using riders’ personally identifiable information.

This dataset consists of .csv files relative to each month and trimester since 2016, but for this project I’m going to analyze the data relative to the last 12 months period, since 08/2021 to 07/2022.

The data gives information about ride\_id, when the ride started\_at an ended\_at, the starting and arrival station, coordinates and the kind of customer (member or casual) and rideable\_type.

The data has been downloaded in zip format, the decompressed and stored in one folder name CSV files, converted in Xlxs format, and stored in another folder named XLXS file, in order to have both the format and select the most convenient for each task. I also decided to create a backup copy in order to be able to access the initial raw data in every moment.

# Data Cleaning

To prepare the data analysis I used **MS-Excel** to better understand them by sorting and filtering and then **MS-SQL Server Management Studio** to merge all the files together and clean them.

First in excel I created two columns, ride-length, and day\_of\_the\_week, and I noticed that there were different values of ride\_lenght that were negatives, equal to 0 or with a short duration. Moreover, by creating a boxplot and filtering them I noticed some high values of duration that were probably to reflect a singular event. Finally, I discovered that many station names were missing.

I identified as outliers the ride\_lenght with a duration of more than 400 hours and less than 30 second and deleted them. I also didn’t consider the ride\_lenght with a negative time and delete them from the main table but before this I copied and pasted them in in another spreadsheet to investigate more the phenomenon and ask for explanations.

The queries that I used to clean the data can be found [here](https://github.com/Teoo98/Google.Data.Analytics.Project/blob/main/DATA_CLEANING.sql).

# Analysis

Now the data is cleaned, and we can take the next step and look closer into what insights can be extracted from it.

First, I did a descriptive analysis in excel calculating the mean of ride\_lenght, the max ride\_length and the mode of day\_of\_week, then I created a pivot table to quickly visualize the data, regarding the average ride\_lenght and number\_of\_rides for each day of the week and the total of the month. One example of the pivot table of one of the files can be see [here](https://github.com/Teoo98/Google.Data.Analytics.Project/blob/main/202202-divvy-tripdata%20(cleaned%2Banalyzed).xlsx).

I did this analysis for every file to gain the first insights and familiarize with the data.

Then I analyzed the full merged dataset in **MS-SQL Server Studio Management** calculating “the total number of rides”, “the number of rides in each day” and for each “rideable type”; this metrics has been calculated for both casual and members and for each month and the total period of the last 12 months. The SQL queries can be seen [here](https://github.com/Teoo98/Google.Data.Analytics.Project/upload/main).

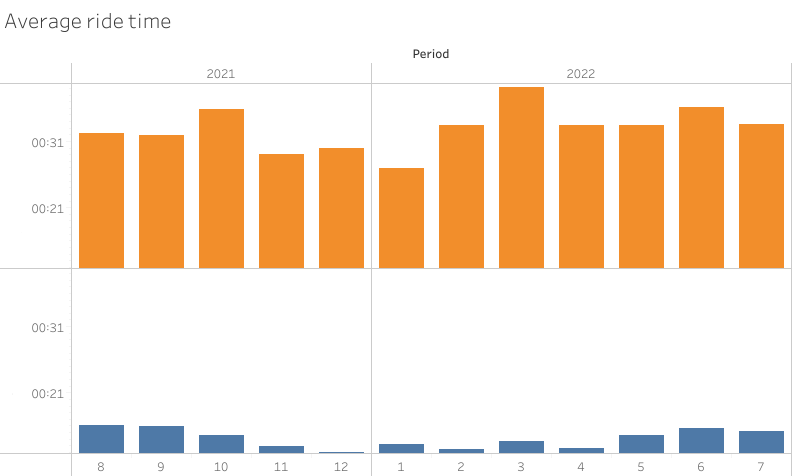
# Supporting Visualization and key findings

The results of the queries it’s been saved in a CSV format in order to update it in **Tableau** to create the Graphics and gain the insights.

|  |  |
| --- | --- |
| AVERAGE RIDE TIME | hh:mm:ss |
| MEMBER | 00:14:06 |
| CASUAL | 00:32:58 |

It’s possible to notice how the average ride time of the Casual customer is more than twice the one of the Member.

One possible explanation is that the members use the bike for short daily travels while the casual pay either for a single race or for a day pass so they want to exploit this service when they have the need to go to a more distant place, however this hypothesis needs to be confirmed gathering additional data.

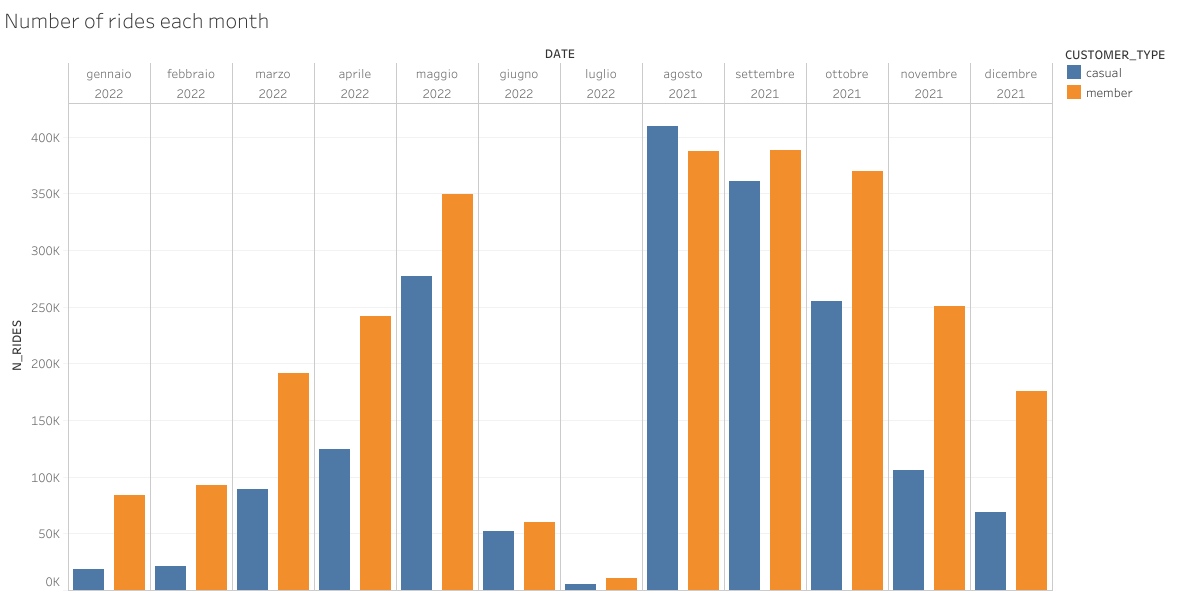


The difference of time between members and casual is constant in every month in the last year.

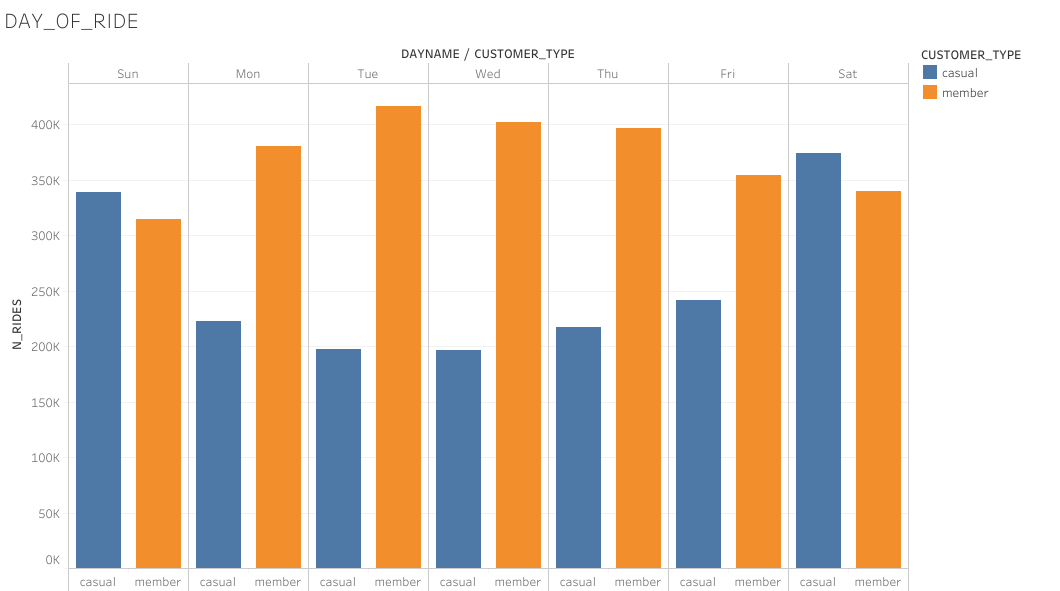
|  |  |
| --- | --- |
|  | The average distance is 1.68km for the members and 2.28km for the casual. So, the distance of the casual is 26% more than the one of the members. |

|  |  |
| --- | --- |
| Chart, pie chart  Description automatically generated | The total number of rides in the last 12 months is of 4 389 697 and those rides are done by neraly 60% by members and slightly more than 40% by casual. So a consistent number of rides is been generated by casual and a their conversion will enanche the business profit by a consistent amount |

Looking at the single months period is possible to notice that the mayority of the rides are in May,August,September and October. Moreover the gap between casual and member is bigger in the period from October to April and smaller in the one from April to September. In the Months of June,July,August and September the number is almost the same.



With regard to the day of ride, the casual customers ride more during the weekend with a consistent gap between Saturday and Sunday and the rest of the week while the member ride more during the week but with a slight gap between those days and the weekend.



The following pie chart represents the rideable type:

|  |  |
| --- | --- |
|  | The rides of the casual members are equally distributed between the three kinds of bike, while the members rides are made by nearly 50% by classic\_bike and 50% by electric\_bike. The members don’t use the docked bike. |

# Key Findings summary and Recommendations

**Findings summary:**

The casual members constitute the 40% of the total of the clients and a successful conversion strategy will bring many advantages to the company.

Casual members ride more during the week-ed and more during the summer period, especially between June and August.

A possible explanation can be that in those periods (weekends and summer) people hang out more since the weather is sunny and usually people don’t work during the weekend, however this hypothesis cannot be confirmed with the available data.

Members average ride time is twice the one of the casual, but the average distance is 25% higher, this may indicate a more propension of the casual to ride for leisure activities while the members for daily ones (like going to work).

Only the casual ride the docked bike while the members preferred one is the classic bike.

**Recommendations:**

Based on the actual available data, is not possible to make any conclusion and plan a conversion strategy, since the data is inconclusive because it allows us to gain insights about the main usage differences and not about the causes of that.

Therefore, is necessary to collect data about the value that drives the decisions of usage of the two group to understand the cause of the behavior and then act with a conversion plan who gives real value to the casual that will turn into members.