CASE STUDY REPORT: CYCLISTIC

Luis Felipe Santos Chagas

1. Summary

Cyclistic (fictional company) is a bike-sharing company in the city of Chicago. The company has two types of plans, casual (casual passengers) and annual (members). Financial analysts have concluded that annual members are more profitable than casual riders, and the marketing department believes the key to the company's growth would be maximizing membership by converting casual riders to annuals, rather than creating campaigns for new customers.

The purpose of this analysis is to create marketing strategies aimed at converting casual travelers into members. To do this, the team needs to better understand how both types of customers use the service, why casual travelers would want to purchase an annual plan, and what marketing strategy would be adopted. Bicycle journey histories will be analyzed to identify trends.

To this end, the following questions are initially asked:

- How do annual members and casual cyclists use bicycles?
- Why would casual travelers want to purchase annual plans?
- What marketing strategy can you adopt to influence casual travelers to become members?

These questions are intended to understand the data, in addition to answering the problem at hand: is it worth investing in converting casual travelers into members? If yes, what is the strategy to be adopted?

1. Prepare Stage

Historical route data is used to analyze and identify trends. As Cyclistic is a fictitious company, a set of public data will be used, provided by Motivate International Inc. under this license, being, therefore, valid data for the purpose of the case study. The data is available through this link.

An annual analysis will be carried out, starting in July/2021 until the month of June/2022. The data are available monthly, so they must be merged for the study of the indicated period.

2. Data Processing

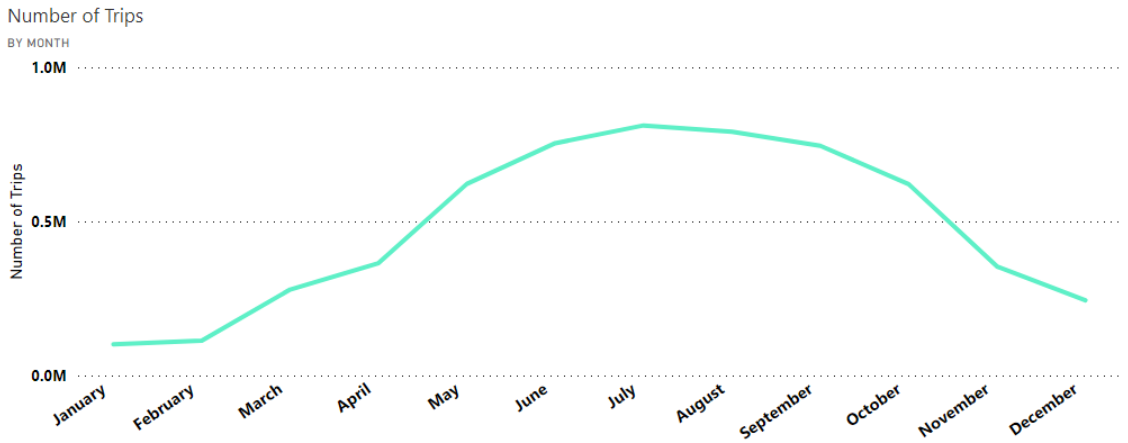
Due to the large volume of data, the R programming language (IDE RStudio) will be used for data manipulation. After the initial treatment and analysis, done in R, Power BI will be used to visualize the data. These processes are also documented in the R-Markdown script, available in the folder of this case study, called "R-Markdown_Cyclistic".

3. Data Analysis

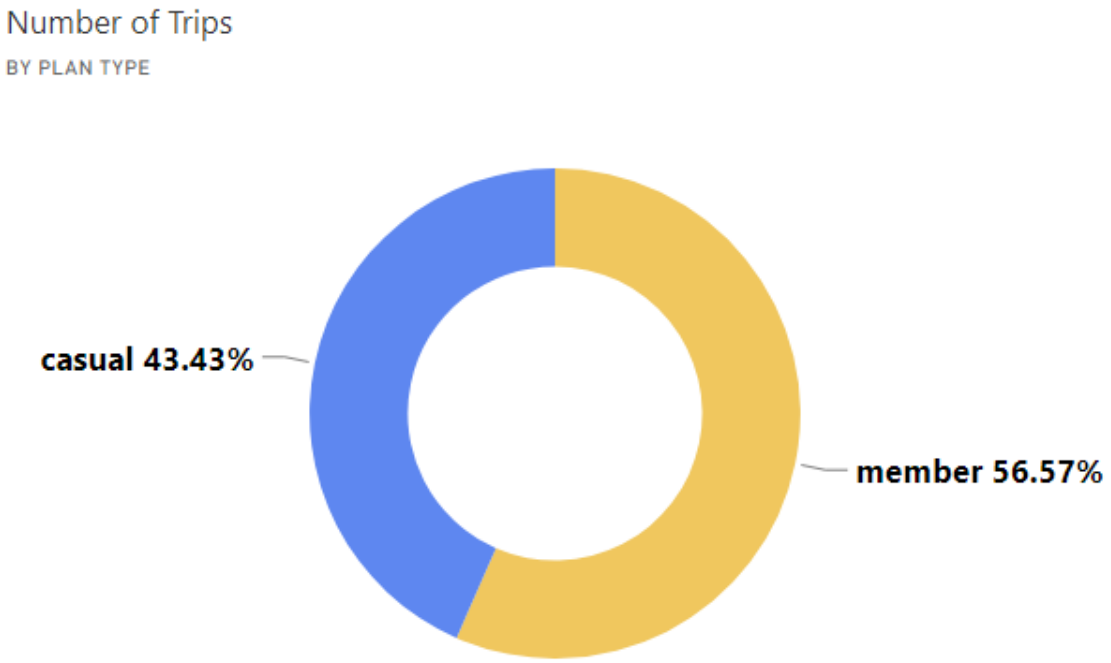
Three main analyzes were performed. The first in a general context, considering both types of clients. The second, an individual analysis of the members and casual passengers and, finally, an analysis of the ride times. These analyzes were performed in Power BI, in the file available in the folder of this case study, called "Dashboard_case_study_Cyclistic" (Google Drive link).

3.1. General Analysis

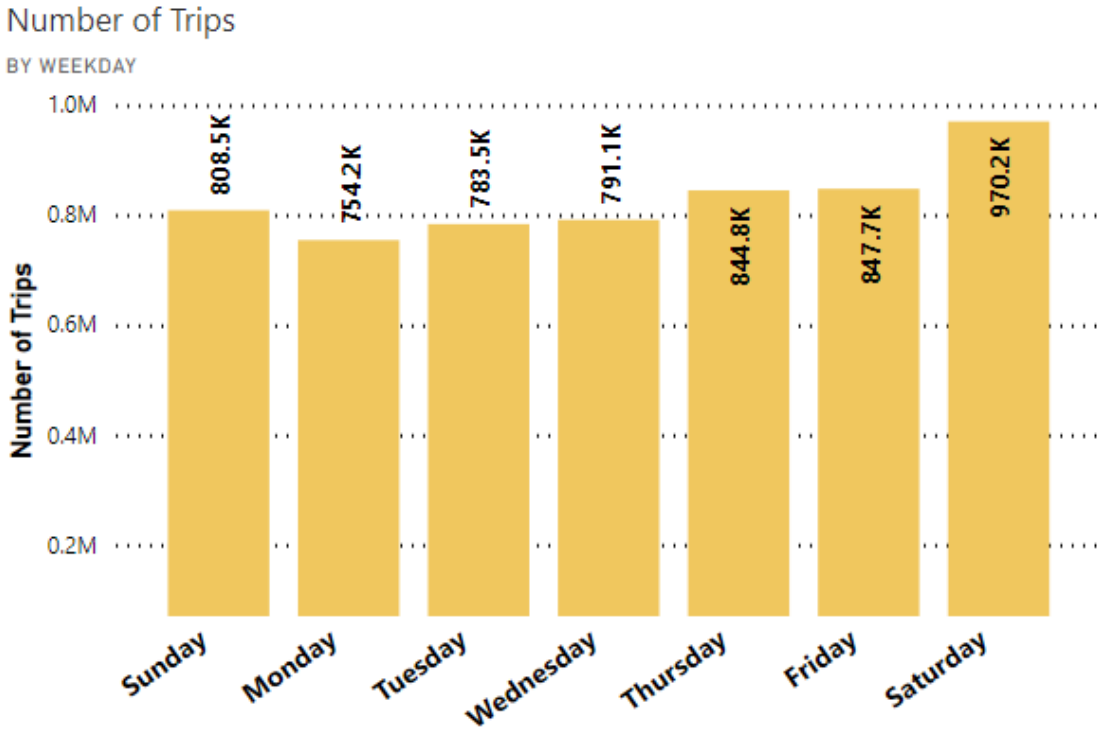
The general analysis shows a trend towards more rides in the months of May to September, representing the spring and summer months of the Northern Hemisphere. This can be seen in the **Number of Trips by Month** chart.



Another important point is the number of rides taken by members and casual passengers. As can be seen in the **Number of Trips by Plan Type** chart, there are more member rides, but there is not a big difference, with about 57% being members and 43% casual. This can mean that there is a good margin of casual passengers that can be converted into annual members.



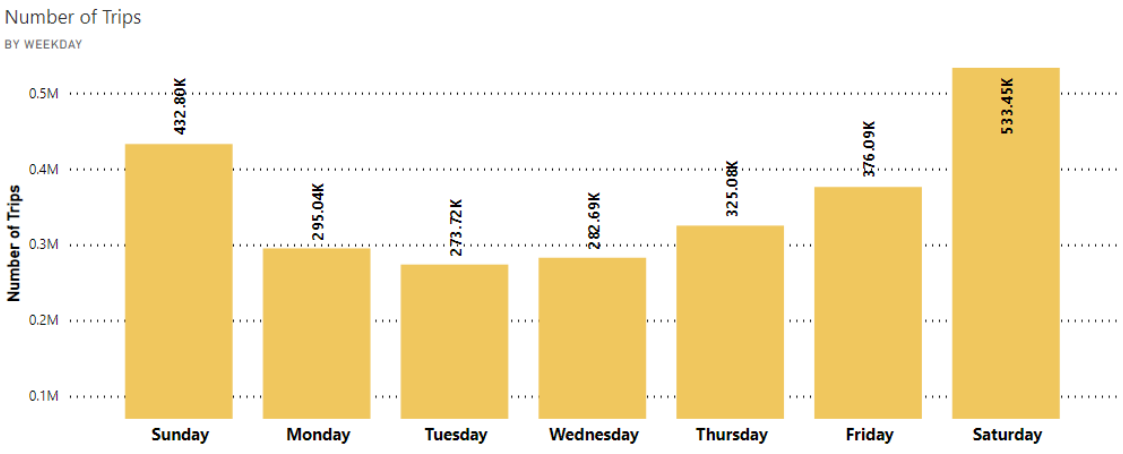
Another interesting analysis concerns the weekdays. There are more tours on the weekend, however, in general, the days are balanced, as can be seen in the **Number of Trips by Weekday** chart.



3.2. Analysis by plan

Now focusing on members, we have the same trend in the use of bicycles per month. Also, the number of rides is concentrated more in the middle of the week (Monday to Friday), a large part, probably, for commuting to work.

For casuals, there is the same trend of tours per month, with peak usage in July. Another point is that, although there are more rides on weekends, there are many rides from Monday to Friday. This represents casual passengers who have the same service usage pattern as annual members. This analysis can be seen in the **Number of Trips by Weekday** graph, filtered for casual passengers.



3.3. Analysis by ride time

The main objective of this analysis was to prove that many casual passengers have the same pattern of use of the service as annual members. In this analysis, for the members, more than 90% of the trips are less than 30 minutes on weekdays (which may indicate commuting from home to work and vice versa). And, for casuals, this same analysis returns a percentage of more than 80% of rides under 30 minutes.

4. Analysis Results

With this, it is possible to answer the questions initially asked:

- **How do annual members and casual cyclists use bicycles?** Members use it more from Monday to Friday, probably for commuting to work. On the other hand, casual passengers use it more during the weekend, but there are many casual passengers[1] who use the service from Monday to Friday, thus having the same usage pattern as annual members.
- **Why would casual travelers want to purchase annual plans?** As informed in the question above, there is a portion of casual passengers who use the service in the middle of the week, which is a pattern of use by members. These clients should be the focus of marketing actions.
- **What marketing strategy can you adopt to influence casual travelers to become members?** Demonstrate to those who have casual plans and use the service from Monday to Friday that switch to an annual plan is worthwhile, focusing on the benefits of being an annual member. We also have that, as the peak usage of casual passengers is in July, marketing could be implemented between the months of June and August.

5. Next Steps

Some more information could provide more realistic and accurate analyses. The dataset does not have passenger information (due to licensing), so the focus of this analysis was on rides. If this information were available, there could be a quantitative analysis of how many casual passengers could be converted into members.

Other valid information would be related to the plans. Knowing the value and benefits of each plan would be important for possible financial return analyses. In addition, this information could solve the doubt about the time limit of the rides. It was unclear if there was a time limit, and the presence of long rides in the dataset (over 5 hours) would be errors or if there really is such a possibility. This information is more detailed in the "R-Markdown_Cyclistic" file.

1 – This is an assumption as we are looking at the number of rides and not the number of customers. It is assumed that many tours equal many clients and not the same client taking rides multiple times.

Appendix A –Power BI Detailing

The Power BI file has 3 pages, namely:

1. **General:** analysis without data segmentation, where the amounts of rides are observed by month, by plan type and by day of the week.
2. **Plan Analysis:** page focused on comparing plans. It analyzes the number of tours by weekdays and by month, with data segmentation by plan type (member or casual).
3. **Ride Time:** page focused on the analysis of ride times. The times were separated into 4 categories (under 30 minutes, between 30 and 60 minutes, between 1 and 5 hours and over 5 hours[1]), to observe the pattern of use of each plan, also considering the days of the week. On this page there are two data segmentations, one for the plan type and other for the weekdays.

As informed in the detail file "R-Markdown_Cyclistic", the dataset was simplified, excluding columns that do not influence the analysis. Therefore, the ride identification columns (Id column), bicycle type and location columns (name and latitude/longitude of the start and end stations of the rides) were deleted.

With this strategy, it was possible to reduce the size of the file, facilitating its import and handling in Power BI, without loss of information relevant to the analysis in question. After importing the dataset "bikes_data_viz", three columns were created:

- Weekday_n: column that returns the day of the week (numeric, with 1 being Sunday and 7 being Saturday), considering the date informed in the Start Time column. This column has the main purpose of sorting the Weekday column (informed in the item below), based on numerical order (from Sunday to Saturday)
- Weekday: uses the information from the previous column, formatted to return the name of the day.
- Ride time: conditional column that classifies the ride time into 4 categories (under 30 minutes, between 30 and 60 minutes, between 1 and 5 hours and above 5 hours).

In addition to these columns, another table was also created, named "_measures" (no data), to enable the organization of measures. Three measures were created:

- **Number of rides:** count of the number of lines in the dataset, since each line is equivalent to one ride.
- **Average duration:** measure created to visualize (on a card) the average time of rides in minutes:seconds format.

1 – It is worth noting that, as the characteristics of the plans are not informed, we chose to include tour times longer than 5 hours, even without the certainty of this being a possibility, or an error in the data.

- Average duration seconds: complementary measure to the formatting process of the average duration.

Power BI has some limitations when working with the hours:minutes:seconds format. The first is that information with more than 24 hours is not counted, for example, if the ride is 26 hours long, Power BI would only inform 2 hours. Another limitation concerns descriptive analyses. If it is necessary to calculate the average time by ride, this average would not have the format hours:minutes:seconds.

Therefore, the formatting was done manually, through the measures "Average duration" and "Average duration seconds", considering the time of the tour, in seconds. The average duration of the rides is on the order of minutes (according to initial analyzes done in R), so the purpose of these measures is to write this average in minutes:seconds format.

The first measure concatenates the information of the integer part of the time (in seconds), divided by 60 (part of the minutes) with the information of the part of the seconds. This second information is made through the second measure. The “Average duration seconds” uses the rest of the time division by 60 (part of the seconds) and, if this remainder is less than 10 seconds, this value is concatenated with a zero (for example 08). If it is greater than 10, there is no need for this step.

It is worth noting that this process of concatenation with zero is not done for the minutes because at no time in the analysis were there averages below 10 minutes.