# Case Study: How does a bike-share navigate speedy success?

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

The role of this analysis was to answer the previous question in order to help the business task of Cyclist company to transform their casual riders in member riders, once these riders were more profitable.

# Data Description

The data was found online located, on a website page: Index of bucket "divvy-tripdata". Access: [Index of bucket "divvy-tripdata"](https://divvy-tripdata.s3.amazonaws.com/index.html). Provided by Motivate International Inc. under the license found here: [Index of bucket "divvy-tripdata"](https://divvy-tripdata.s3.amazonaws.com/index.html).

The data is organized monthly in zip files that we can download. After extracting it, we have csv files that contain tables in wide format with the following fields:

ride\_id: type: char; the id of the represented ride

rideable\_type: type: char; the type of bike: electric\_bike, docked\_bike or classic\_bike

started\_at: type: datetime; the time when the ride has begun

ended\_at: type: datetime; the time when the ride has ended

start\_station\_name: type: char; the name of the starting station of the ride

start\_station\_id: type: char; the id of the starting station of the ride

end\_station\_name: type: char; the name of the ending station of the ride

end\_station\_id: type: char; the id of the ending station of the ride

start\_lat: type: char; start station latitude

start\_lng: type: char; start station longitude

end\_lat: type: char; end station latitude

end\_lng: type: char; end station longitude

# Preparing and Processing

The data was processed on RStudio, because of Big Query free account storage limitations. The data was cleaned and had had its integrity checked. Each of its twelve files were briefly sorted and filtered on Excel spreadsheets, where the following has been observed: It has a lot of different formatting for id fields and station names, likewise mixing strings and numbers. As there has not been a solid resume of ids and stations’ names, there has not been any way to verify misspellings or inputting its names when it has been missing. This was the less reliable aspect of these datasets. But all the analysis that was meant to be done should not make use of them, so it can be said the it does not influence the subsequent analysis.

Furthermore, the datetime columns were well formatted as well as the rideable type column and those geographical coordinate columns, with exception of few missing compared with the datasets massive row number. All dataset has been merged to save time on dealing with missing values. These missing values have been dropped. Duplicates has been checked too and any was found. The cleaning process is documented on a RMarkdown report: “cyclist\_cleaning\_2024\_01\_08. Rmd”.

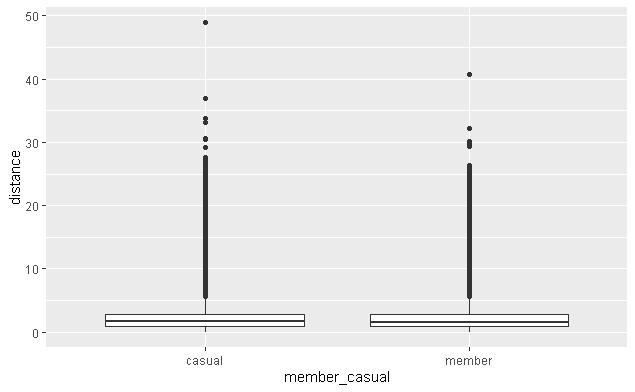
## Measures

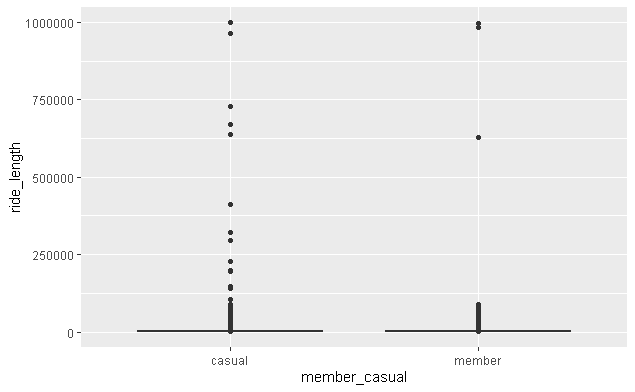
On the processing of data for analysis, some measures have been done, having in mind that the important thing for us was how this measure could define different rider profiles, casual and member. Some new datasets had had been generated as summaries, each one with its application. The measures performed have been:

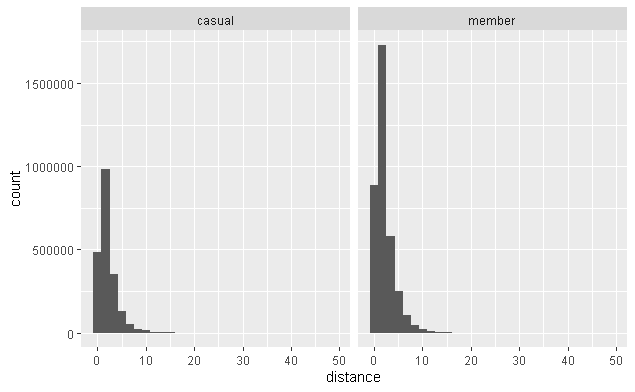
* ride\_length: the time length of each ride
* day\_of\_week: the day of week when each ride happened
* ride\_month: the month when each ride happened
* ride\_day: the day when each ride happened
* distance: the difference between the starting point coordinate (start\_lat; start\_lng) and the ending point coordinate (end\_lat; end\_lng). These measures are referred as radial distance on further dashboards
* ride\_hour: the hour when each ride happened

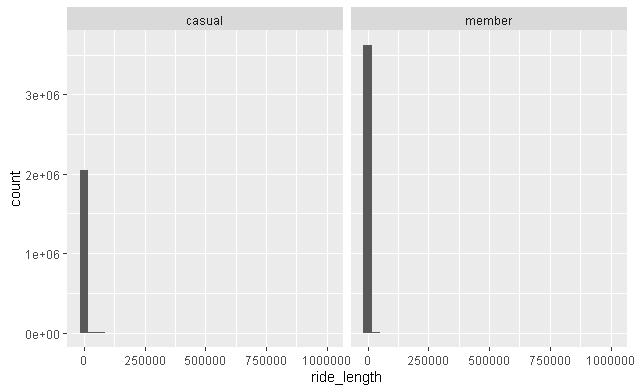
The dataset aggregated with these measures were saved, with its proper formatting on “data” directory by the name “df\_for\_hour\_analysis.csv”, on text file format.

## Summaries

For the analysis process, some resumes were performed and stored in tables as new text files. The first one consisted on some descriptive information about the data, like the number of ride of members and casual riders, average, minimum, maximum length and distance, as well as their standard deviations. These same measures have been performed for the overall data, which was idealized as a standard measure for comparing the other measures for each type of rider (member casual). A boxplot graph had had been generated, as well as histograms were plotted to see the member of rides distribution.







As it is possible to be seen, distance and length measures strongly converge to the mean and initial (small) values, which means that the mean behaviors observed for each ride profile should strongly be taken into account.

Furthermore, the data has been summarized by ride profile, day of week, month and day, following average, maximum, minimum values for length and distance and the number of rides by day. Then it was summarized for the third time by ride profile, rideable type, day of week, month and day, following the average length and distance measures, as well as the number of rides by day. Finally, it was summarized by ride customer profile, day of week and ride hour, measured by average length and the number of rides per hour.

All the resumes have been saved on “data” directory, on text files, by the respective names “summary\_data\_count\_2024\_01\_12.csv”, “summary\_data\_length\_2024\_01\_10.csv”, “summary\_data\_rideable\_type.csv”, and “summary\_hour.csv”.

# Analysis

The analysis has been taken on Excel and Tableau. The analysis document on excel has been named as “cyclist\_analysis\_2024\_01\_11.xlsx” and on Tableau as “cyclist\_analysis.twb”. The proceedings on both software are quietly similar, exception for fact that because of each software focus, on Excel the dashboards have more information in tables displayed, on Tableau the analysis has more focus on charts. The tables that have been used to generate the Excel dashboards are hidden and the tables and chart are protected under the password “1234” in order to avoid accidental editing, compromising the analysis. It will be followed on this report the Tableau analysis, given its focus on trends identification of casual and member riders’ behavior, but as the analysis on both software did not show any different insights. If it pleased to check exact measure values it is strongly recommended check the Excel table focused analysis.

## Descriptive analysis

Here we are going to focus our analysis on the “summary” worksheet on Tableau.

On average the radial distances do not show significant difference for both rider type, which means that even casual or members end its rides around 2.1 Km from the point where they started the ride. It might have been showing that both casual and members ride the same neighborhood.

The overall average length is around 910 seconds. Casual riders took more time on their rides, 1239 seconds against 724 seconds on member’s rides.

Their maximum and minimum distances are quietly the same. They are strange values and should be dealt in more detail in other moment, once there are many zero length values with corresponding positive distance (which means a displacement without time, that there is any physical meaning) and many extreme values like rides that ended 48 Km from the starting point and many rides taking almost 1000000 seconds. The standard deviations show that the distance varies around 2 Km more or less than the average for casual and members. The same measure for length shows a variation of around 3000 seconds above or below the average, while for members the average value varies 1500 seconds above or below of it.

As we see despite casual and member riders have been doing its rides on the same neighborhood, they have been wasting different length of time riding on average.

## Number of rides on Weekdays and Months

Here the analysis is going to be focused on the “Ride Number Analysis” dashboard on Tableau.

The number of rides by weekdays shows that casual riders reduce its rides on the working week and increases its rides on weekends, while member riders increase its number of rides during working week and reduces it on weekends. What we might be seeing is that members use bikes most of times as a mean of transport for work and casual riders uses it most of times for other purposes. The month chart analysis shows that the bike usage for both types of riders increases on summer and reduces on winter. A general information that could be noticed is that members perform more rides than casual riders. This behavior is maintained for most cases even when different weekdays and months are filtered. So, it’s a quietly robust result.

## Average Time Length Rides on Weekdays and Months

Here the following analysis took place on the Tableau dashboard “Ride Length Analysis”.

It is possible to see that the time spent on rides by casual riders are greater than that spent by members. While the time spent on rides by casual riders decrease on work week days the member time spent on rides remains almost the same, lightly varying positively on weekends. The grater variation is found when framing the analysis by month. The time spent on rides by casual riders increases considerably on summer months while member riders lightly increase its time spent on rides. These monthly variations that is observe probably is due to weather conditions of the region characterized by rigorous winter and snowfall which might turn bike rides harder than in summer.

## Average Radial Distance on Weekdays and Month

Here the following analysis took place on the Tableau dashboard “Ride Radial Distance Analysis”.

As it was observed on the descriptive analysis, during weekdays the distances remains almost the same for both casual and members. Now, when it comes to the monthly variation, the distance increases on summer months in almost the same rate, being the members distances slightly grater than for casual riders.

## Rideable Type Analysis

### Overall Rideable Type Ride Analysis

Here the following analysis was taken using the Tableau dashboard “Overall Rideable Type Ride Analysis”.

Looking to the charts it can be seen that there are more member rides than casual rides again. Casual riders prefer, in general, electric bikes to classic bikes. This preference remains true for member riders, but it is smaller for them. What really come to eyes is the fact that there are no rides on docked bikes done by members. Docked bikes are quietly curious and clumsy, which could do not fit so well for fast rides to work

### Rideable Type Daily and Month Analysis: number of rides

The following analysis is based on Tableau dashboard “Rideable Type Daily and Month Analysis”.

Over months and weekdays, the overall variations follow the same pattern for 3.2 session, which means that the overall number of rides for members increases over work week and casual rides follows an inverse pattern. The same analysis continues valid for the month variation, as the 3.2 session. As they maintain these patterns, the pattern for rideable type remains the same as has been shown on previous session for casual riders, but for members, this preference changes, and the classic bike begins to be the preferred bike for members’ rides, from July to November.

### Rideable Type Daily and Month Analysis: time length of rides

The analysis here is based on the Tableau dashboard “Rideable Type Daily and Monthly Length Analysis”.

The time length of rides when related to weekday and monthly variation remains almost the same, but as we should expect the time for casual rides greater than for members, being the time spent on classic bikes considerably grater than for electric bikes, in case its camos for casual riders. For members the time spent on classic bike rides are slightly greater then on electric. The huge variation comes when it is analyzed the docked bike rides. Through weekdays the time spent on docked bikes remains almost constant, but it hugely increases through months, and reaches its maximum on summer months. As we should expect the time spent on these bikes were much greater than on the other types, corroborating the hypothesis that these bikes are ever used for slower tours, as a leisure form.

### Rideable Type Daily and Monthly Distance Analysis

The analysis here is based on the Tableau dashboard “Rideable Type Daily and Monthly Distance Analysis”.

The general tendencies for the distance between the starting point and ending point of rides follows the same as analyzed on session [3.4](#_Average_Radial_Distance). But it is importance noticing that for both member and casual riders, the distances using classic bikes is less than for electric. This difference is slightly greater for members. Its is grater too, for month analysis.

## Overall Rides and Length by Hour

The analysis here is based on the Tableau dashboard “Overall Rides and Length by Hour” and “Rides and Length by Hour and Weekday”.

The number or rides by hour in general corroborate for the last time in this analysis the hypothesis stablished that the members have been using bikes for work transportation. It is seen that there is a huge peak on the number of ride coincident with the hush hour, around 8:00 a.m. and around 5:00 p.m. This same behavior happens to casual riders, but much less intense. This happens over the entire work week. On weekends both, casual and member riders converge its behavior, and the number of riders increases from around 5;00 p.m., reaching its peak about 3:00 p.m., and decreasing until 5:00 p.m.

When the analysis comes to length, the time spent on rides for members remains small than for casual riders. It remains almost the same all weekdays, but on work days it decreases a little in the early morning, on weekends it increases during the day and decreases over the night. The behavior is quietly the same for casual riders, but its variations are more aggressive. On work days the time spent on rides for casual riders strongly decreases in the early morning, and strongly increases over the day. On weekends this time strongly increases during the day and strongly decreases overnight. If bikes have been used for members as a way of transportation for work, its comprehensive to state that the use it objectively, wasting the less time possible on rides. So, it follows that the time increases on day hours once it is using by casual riders on leisure time.

# Suggestions, Conclusions and Acting Guidance

## The Need for More Analysis and Data

Despite the missing role here discussed, the role of extreme values that could took place on another more focused analysis, and the possibility of increase the data, including more detailed data about the kinds of neighborhood establishments, for trying to deduct with more detail the bikes has been used, it was possible to reach the perception of two different profile rider.

## Conclusions

The analysis here undertaken, leads to the conclusion that member and casual riders uses bikes for different purposes: the first, uses it primarily as a way of transportation for work, the second, as a way of leisure.

## Acting Suggestion

As the previous analysis was taken to guide the marketing team, it would be advised the advertisement should focused on showing for casual members why would be interesting for them using bike for going to work. For this would be interesting go in deeper analysis to understand more clearly why members like to go working on bike. This way of advertising would bring positive results.

It would be interesting that both groups prefer electric bikes so the possibility of focusing this kind of bike on advertising would be decisive for bringing more rides and why not converting casual and member riders.

Finally, would be interesting to understand why those who uses bike just for leisure purpose do not take the member. It would be interesting, but it would be work for another team, to think on a membership that would be profitable and fit the needs of those people that want to ride just for having fun. Established and understood this frame, it would be possible advertise this group focusing on health lifestyle what would be positive for the customers and for the company.