# Introduction

Welcome to the Cyclistic bike-share analysis case study! In this case study, you will perform many real-world tasks of a junior data analyst. You will work for a fictional company, Cyclistic, and meet different characters and team members. In order to answer the key business questions, you will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act. Along the way, the Case Study Roadmap tables — including guiding questions and key tasks — will help you stay on the right path. By the end of this lesson, you will have a portfolio-ready case study. Download the packet and reference the details of this case study anytime. Then, when you begin your job hunt, your case study will be a tangible way to demonstrate your knowledge and skills to potential employers.

**Scenario:** You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

# Step 1: Ask

**Problem Statement:** Identity the difference of Cyclistic bikes utilization between annual members and casual riders.

# Step 2: Prepare

**Data Description:** This data set is about the previous 12 months of Cyclistic trip data. The data is not organized in any order or form. The data is quite reliable and original. Besides, the data has a very all the sufficient information. To ensure the data is secured in terms of privacy and at the same time it is accessible for stakeholders, the spreadsheet is shared to the responsible persons to restrict the access. To verify the data integrity, I make sure to pre-process the data and then make sure to save a copy in order to have a backup file. The data should be able to answer the question and solve the statement, it has the information of both members and casual riders. However, the data does not show all the start station and end station, most likely because not every riders stop at the designated stop station but the coordinates of where the bicycles are situated are shown.

# Step 3: Process

**Documentation:** For this section, I use Python as my tool of choice because I am more familiar in Python and it has a package call Pandas, it is really good for data manipulation and processing. To ensure the original files are not being replaced, new copies are created produced after the cleaning process to make sure the original files are not destroyed. Firstly, the data is sorted based on their start date and time. Then, duplication is checked and discarded. Next, there is some data with starting time bigger than ending time, those are the inconsistent data and they are discarded as well. After that, the data is filtered and organized by the category of riders. Members are filtered in a group whereas casuals are also filtered in a group. This is to ensure better analysis for later in the process. Lastly, all the data is saved to a new csv file separated by the rider’s category. In the end, I imported the csv files to spreadsheet and found the difference of riding time and also the day of the week using spreadsheets formulas. With this, we can proceed to next step which is to analyze.

# Step 4: Analyze

**Data Analysis:** First and foremost, I used Google sheet to find the average and maximum length of ride via functions. Next, I also calculated the mode of the day of week. Then, I created a pivot table to calculate the average length of ride, maximum length of ride by day of week and the number of rides by day of week. After finished using spreadsheet, I moved on to using R for further analysis. Firstly, I imported the data into the Rstudio and concatenated both members’ and casuals’ data frame into a single data frame. Then, I calculated the summary statistic in seconds.

Screenshot 2022-08-28 194421

Figure 1: Summary Statistic of Length of Ride.

According to figure 1, we can see the minimum riding duration is 8 seconds whereas the maximum riding duration is 86,397 seconds. Besides, we can also see the average riding duration is 8,242 seconds and the median is 1,993 seconds.

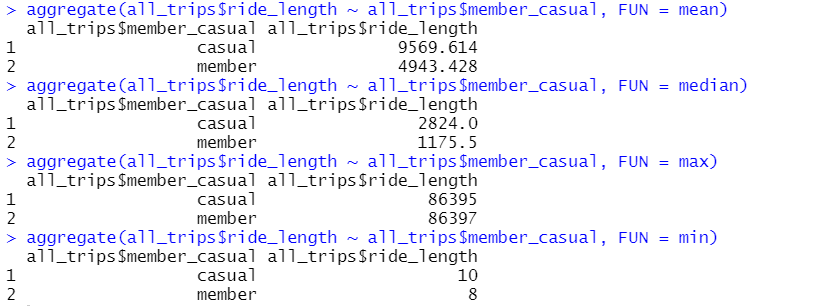


Figure 2: Comparison of Member and Casual Rider’s Summary.

Figure 2 shows up the comparison of member and casual riders’ mean, median, maximum and minimum. Averagely, casual rider has higher riding duration than member riders. Casual rider has higher median than member rider. Member rider has a higher maximum riding duration than casual rider but member rider has a lower minimum riding duration than casual rider. However, according to figure 3 below, we can see that casual riders are more than member riders.



Figure 3: Frequency of Casual Rider and Member Rider.

# Step 5: Share

**Data Visualization:** After all the analysis, I came out with some findings that might be useful for improving Cyclistic’s overall user’s environment.

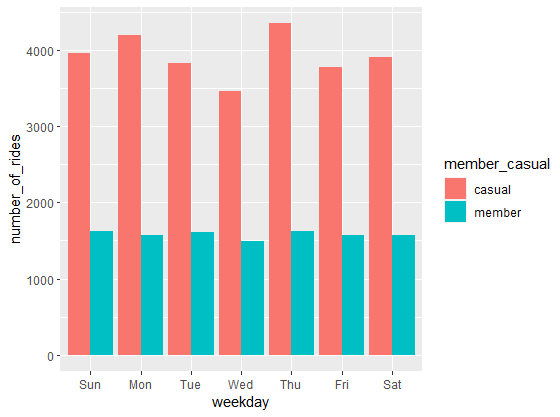


Figure 4: Number of Rides of Member And Casual Riders for Every Day.

As shown in figure 4, we can see that member riders are more consistent than casual riders when it comes to the usage frequency. Member riders are more consistent so one of the recommendations is we should focus on maintaining the member riders, give them some exclusive rewards to keep them active in using the Cyclistic apps. Besides, we should also convert those casual riders to members by attracting them with many good rewards and free gifts.

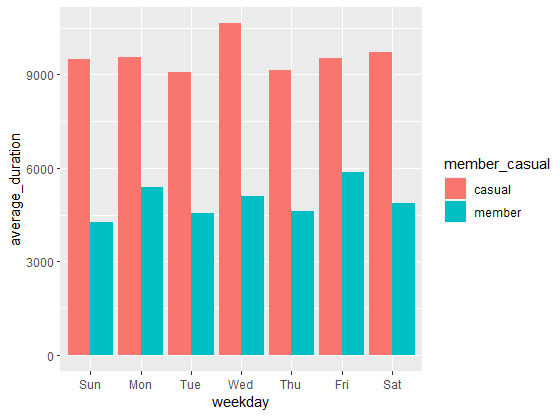


Figure 5: Average Riding Duration of Both Member and Casual Riders for Every day.

Figure 5 displays the average riding duration of every day by member and casual riders. As we can see, casual riders ride the most on Wednesday but the users highly use Cyclistic on daily basis. For member riders, they ride the most on Friday. Recommendation of focusing riders’ overall experience of using Cyclistic during the peak hour can improve their experience. This can make the users stay longer in using Cyclistic.

# Step 6: Act

**Recommendations:** From all the insights and findings, we can come out with some recommendations. Firstly, we can improve the overall experience of every users by enhancing and improving Cyclistic’s environment. Maintaining the bicycle is a good way to improve Cyclistic’s experience. Next, giving bunch of free gifts and rewards to member riders to keep them as members will be benefit to Cyclistic in long term. Besides member riders, giving some rewards to casual riders could also improve their experience and might convert them to join membership.