

Cyclistic Case Study - Final Report
Google Data Analytics Professional Certificate
Zamir Said
Sep. 15, 2023

Introduction

This is the final report for the case study project from the Google Data Analytics Professional Certificate Program. This report will be using the structure of the Data Analysis Process that was taught in this certification program. This means the main steps used in the analysis are Ask, Prepare, Process, Analyze, Share, and Act (A.P.P.A.S.A.).

The scenario for this case is study is a that I am working as a data analyst at a fictional bike-share company called Cyclistic. Cyclistic has two membership types: annual members and casual riders. The team wants to understand how annual members and casual riders use the bikes differently. They want to use these insights to try to convert casual riders into annual members.

Ask

This step in the data analysis process involves identifying the problem we are trying to solve or the question we are trying to answer. For this case study the question is: “How do annual members and casual riders use Cyclistic bikes differently?”.

Prepare

The prepare step involves finding the data you would like to use for the analysis, getting a sense of the general structure of the data, and determining the credibility of the data by questioning whether it is biased or not. For this case study, the data I will be using was made available by Motivate International Inc. and this data is being used because Cyclistic is not a real company.

The data is split by month, and for this analysis I will be using one month of data. The data has 14 columns with a mix of character and numeric variables. There is data on the start and end times of rides, the start and end stations, ID for the customers, and whether they are members or casual riders.

Process

This step is for cleaning the data before performing an analysis on it. That involves removing irrelevant columns, checking the data for errors, adding new columns made from data in existing columns, and transforming the data so that it can be analysed easily. In this step, there should also be a change log, which documents all the changes that were made to the data so that it can be easily recreated and if there is an error, it is easy to determine where this happened and correct it.

One of the errors that could occur in a dataset like this is that the end_time column has a time that is earlier than the start_time column. However, after checking for rows which have this issue, non appeared so this error does not exist in this dataset.

The next step is to transform the data to set it up for the analysis in the next step. I transformed the data in the following ways:

- Removed longitude and latitude columns because I will not need them for the analysis
- Using the date in the started_at column, I added columns for day, day of week, and date
- Using the time in the started_at and ended_at columns, I added a column for ride duration measured in seconds

Analyze

This step involves performing the analysis to find insights in the data that can help you answer the question that was posed at the beginning of the data analysis process.

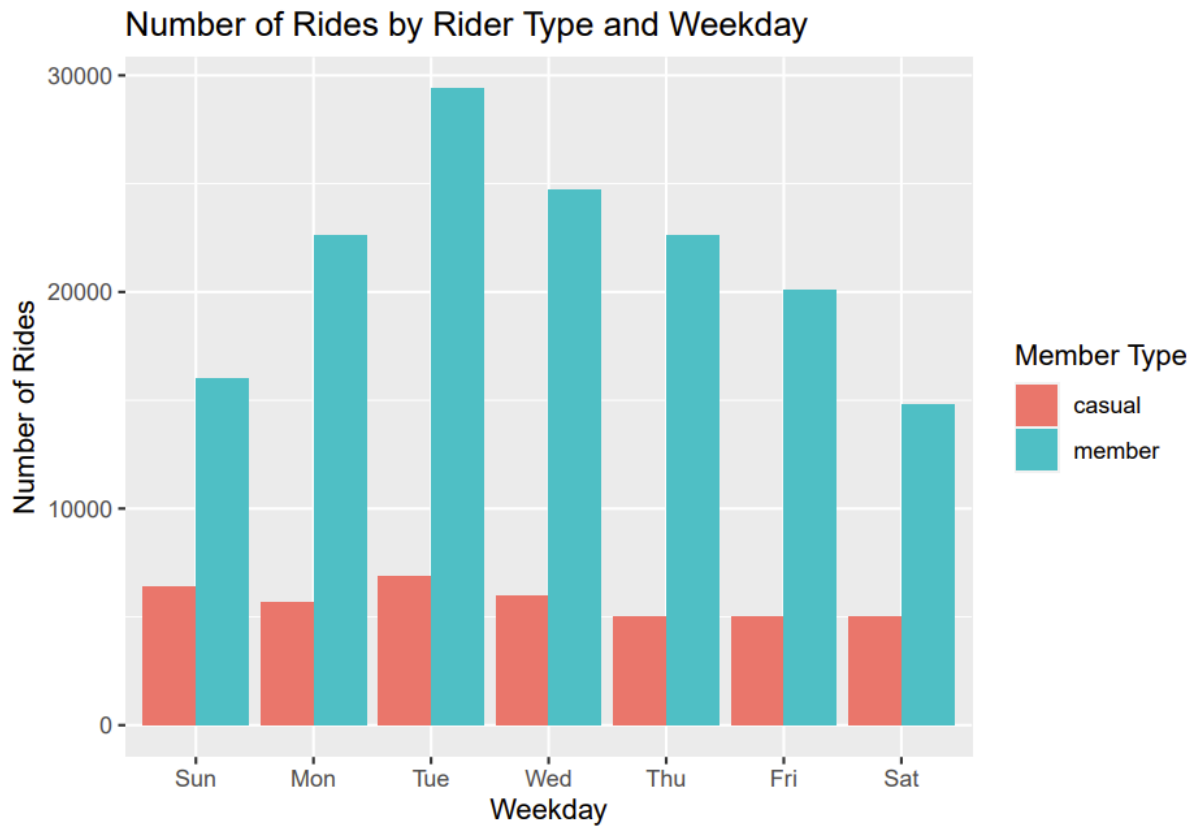
First, I checked the number of members compared to casual riders and found that there is approximately 4 times more members than casual riders. Next, I checked the number of rides across the week, and it looks like there are much more annual member rides than casual riders over all of the days of the week. Then, I used the descriptive statistics function in R to compare the minimum, maximum, mean, and median ride lengths for members and casual riders. It seems that the mean and maximum ride lengths for casual riders is much higher than that of annual members. Finally, I found the average ride length for members and casual riders split by the days of the week. This showed that the average ride length of members is relatively consistent across the week whereas the average ride length for casual riders is much greater in the weekends than it is during the week. It also showed that across the whole week, casual riders have higher average ride durations compared to annual members.

The full analysis with the R code is available as a PDF file in the same folder as this report.

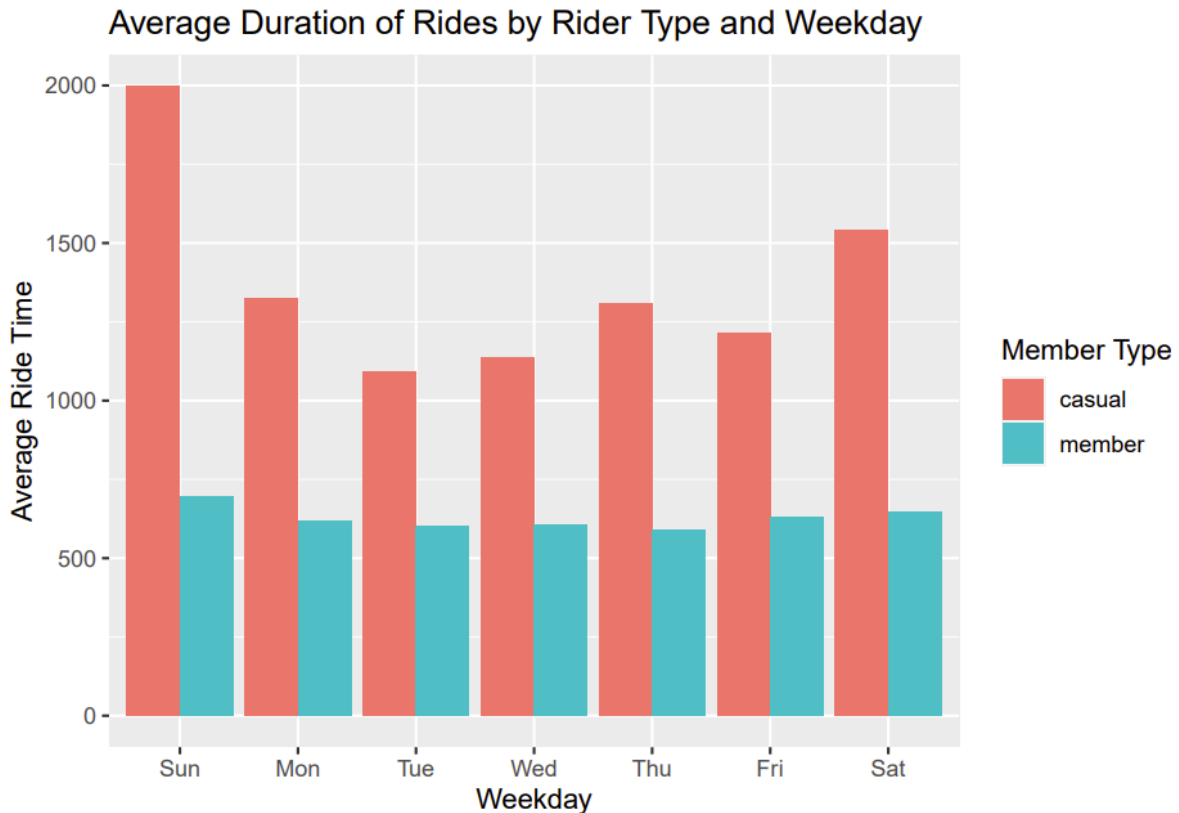
Share

This step involves making visualizations to convey the insights you have found from the data analysis to an audience. The audience may or may not have a technical background so the visualizations should not be too technical. They should be simple but still effective in showing the message you are trying to communicate.

The following visualization shows the number of riders across the week, separated by membership type. It shows that there is a higher number of annual members across all the days of the week compared to casual riders.



The following visualization shows the average ride duration across the week, separated by membership type. This visualization shows that the average ride duration of casual riders is much higher than members across all days of the week. It also shows that the average ride duration of members stays mostly consistent throughout the week whereas the average ride duration of casual riders increases on the weekends.



Act

This step involves applying the insights from the analysis to propose a course of actions that the stakeholders could take based on the analysis.

In this case, the original question was “How do annual members and casual riders use Cyclistic bikes differently?” and the conclusions from the analysis were that members use the bikes for consistent ride durations throughout the entire week whereas casual riders use the bikes for longer ride durations during the weekends and shorter ride durations during the week.

One of the steps the company can take is to supply more bikes during the weekend since bikes are going to be gone for a longer period of time. Since the stakeholders are looking for ways to convert casual riders to annual members, they could cap the amount of time that casual riders can use the bikes on the weekends. This could encourage casual riders to get the membership so they can ride for a longer duration on the weekends.

Conclusion

This concludes the final report for the case study project for the Google Data Analytics Professional Certificate. This project is an exercise for using the Data Analysis Process which involves the steps: Ask, Prepare, Process, Analyse, Share, and Act. Also, this is practice for using

R code for data analysis and data visualizations. Finally, this project is also a good exercise for finding insights from data and translating those insights into steps that the stakeholders can take to apply the findings of the analysis.

If I were to do this project again in the future, there would be a few things that I would like to do differently to get a better analysis of the data. First, I would combine all the months of data from 2023 to get one large data set for the entire year. This would ensure that the results from the analysis are more accurate, and it can be applied to the entire year instead of just January. I could also separate the data by seasons to get different insights for different seasons. I would also perform an analysis on more variables than just the ride duration and days of the week. One example of that could be to compare the starting and ending locations for annual members and casual riders. For this, I could also create a geographic visualization using a software such as Tableau. Finally, I would ask more questions that may be able to be answered from these data sets.