Capston Project: How Does a Bike-Share Navigate Speedy Success?

Zemelak Goraga

2023-02-21

Introduction

This case study focused on a fictional company called 'Cyclistic', which is a bike-share company. Key questions related to the company's business activities were addressed following the Ask, Prepare, Process, Analyze, Share and Act steps of the data analysis process.The previous 12 months of Cyclistic's historical trip data were used to analyze and identify trends and draw key insights.

Scenario

In this case study, it was assumed that the director of the marketing team believes that the company's future success depends on maximizing the number of annual memberships. Therefore, in this case study, the interest of data analysis was to understand how casual riders and annual members use Cyclistic bikes differently. So, insights and professional data visualizations obtained from this study will be helpful in designing a new marketing strategy to convert casual riders into annual members.

Business task

Analyze the Cyclistic's historical trip data of the previous 12 months (Feb 2022 - Jan 2023) in order to draw trends and key insights that can be used as an input for designing a new marketing strategy to convert casual riders into annual members.

Business Question

How casual riders and annual members use Cyclistic bikes differently?

Based on the SMART approach, the business question was brokendown into pieces as follows:

i) How many bike rents did Cyclistic made over the previous 12 months period? What percentages of the total bike rents made in the previous 12 months were accounted by the causal-riders and annual members, respectively?

ii) How is the trend in ride-length between causal-riders and annual members over the previous 12 months period? Does it differe between causal-riders andannual members? Which of the two rider groups used the Cyclistic's bikes more intensively?

iii) Do Cyclistic's riders prefer specific bike-type? Which bike-types are mostly used by causal-riders and annual members, respectively?

iv) In which months and days of the week, there was a higher use of Cyclistic's bikes? Does it differe between causal-riders and annual members?

Methodology

Data Source and licence agreement:

The company's historical trip data was made available by Motivate International Inc. Please visit the following link for further information: https://ride.divvybikes.com/data-license-agreement

About the Data

The data used in this case study were accessed from the source in Zip file containing twelve Tables formated as .CSV files. The data used for analysis had a total of 13 attributes(columns) and 5,754,254 observations (raws). The 13 attributes are described as follows:

Data Attributes:

- ride_id (the Id of a specific ride, STRING)
- rideable_type (type of the bike, STRING)
- started_at (started time of a specific ride, TIMESTAMP)
- ended_at (end time of a specific ride, TIMESTAMP)
- start_station_name (start station name of a specific ride, NULLABLE)
- start_station_id (start station ID of a specific ride, NULLABLE)
- end_station_name (end station name of a specific ride, NULLABLE)
- end_station_id (end station ID of a specific ride,STRING)
- start_lat (start latitude of a specific ride, FLOAT)
- start_lng (start longitude of a specific ride,FLOAT)
- end_lat (end latitude of a specific ride, FLOAT)
- end_lng (end longitude of a specific ride, FLOAT)
- member_casual (type of riders, STRING)

Analytical tools

In this case study, different analytic tools such as SQL, Spreadsheets and R programming were used to perform one or more components of data importation, data cleaning, data transformation, data analysis and visualization.

Data Importation procedure

In this case study, Cyclistic's historical trip data representing the previous 12 months period (February 2022 to January 2023) were downloaded from its source into a folder located in my computer. Afterwards, the data were imported from my computer into Google Bigquery. Detail steps are presented as follows:

1. First, a new folder called 'capstone' was created in my personal computer.
2. The data in a Zip folder containing several .csv files was downloaded from its source and saved into the newly created 'capstone' folder in my computer
3. The .csv files in the zip folder were extracted into a sub_folder within the 'capstone' folder.
4. A total of 12 separate tables representing the previous 12 months period were selected to be imported into the BigQuery.
5. To import those twelve tables from my computer into bigQuery,first, a new project with project ID (capstone-378215) was created in SQL workspace and a new dataset "Capstone_BikeShare" was created within the project.
6. Afterwards, separate new Tables were created within the "Capstone_BikeShare" dataset for each of the uploaded tables. As some of those tables were contining big data, the google cloud system didn't allow me to import them into BigQuery. To solve this, those tables with big data were each splited into two tables using 'Split' software. For instance, one of the table with big data was the table containing the

data representing the month of February 2022. So. the data in this table was entered into BigQuery as two separate tables. Afterwards, merged as one table using 'union all' function of SQL as follows:

- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_feb1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_feb2`
- where ride_id is not NULL

So, all the different tables were imported into BigQuery as follows:

- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_feb`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_mar`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_apr`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_may1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_may2`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_jun1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_jun2`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_jul1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_jul2`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_aug1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_aug2`

- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_sep1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_sep2`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_oct1`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_oct2`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_nov`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_dec`
- where ride_id is not NULL
- union all
- SELECT distinct *
- FROM `capstone-378215.Capstone_BikeShare.table_jan`
- where ride_id is not NULL

Once, a single table containing all the data of the 12 separate tables was created, it was saved as 'capston_bike_share_data' using the 'SAVE RESULTS' and then 'BigQuery table' options:

Afterwards, this 'capston_bike_share_data' was used for further data manipulation and analysis steps:

SQL queries used for extracting year, month, day and time attributes from Cyclistic's ride-share data:

- SELECT *
- , EXTRACT(YEAR FROM started_at) As year_of_week_started
- , EXTRACT (month from started_at) As month_of_week_started
- , EXTRACT(date FROM started_at) As day_of_week_started
- , EXTRACT(time FROM started_at) As time_of_day_started
- , EXTRACT(YEAR FROM ended_at) As year_of_week_ended
- , EXTRACT (month from ended_at) As month_of_week_ended
- , EXTRACT(date FROM ended_at) As day_of_week_ended
- , EXTRACT(time FROM ended_at) As time_of_day_ended
- FROM `capstone-378215.Capstone_BikeShare.capston_bike_share_data`

Finally, the 'capston_bike_share_data' file was exported into ' Google Sheet' using the 'Open with' followed by the 'Connected Sheets' options.

Data Analysis

SQL query for calculating 'ride_length':

- SELECT distinct *
- , started_at - ended_at As ride_length
- FROM `capstone-378215.Capstone_BikeShare.capston_bike_share_data`
- where started_at is not NULL and ended_at is not NULL

SQL query for descriptive statistics:

Descriptive statistics grouped by a single variable: member_casual type

- SELECT distinct
- Sum(ride_length) As sum_ride_length
- , Max(ride_length) As max_ride_length
- , Min(ride_length) As min_ride_length
- , Var(ride_length) As var_ride_length
- FROM `capstone-378215.Capstone_BikeShare.capston_bike_share_data`
- where ride_length is not NULL
- group by member_casual

Descriptive statistics grouped by two variables: 'month_of_week' and 'member_casual type'

SQL query for descriptive statistics:

- SELECT distinct
- Sum(ride_length) As sum_ride_length
- , Max(ride_length) As max_ride_length
- , Min(ride_length) As min_ride_length
- , Var(ride_length) As var_ride_length
- FROM `capstone-378215.Capstone_BikeShare.capston_bike_share_data`
- where ride_length is not NULL
- group by month_of_week, member_casual

Data Visualization

The 'capston_bike_share_data', which was exported from SQL workspace into ' Google Sheet,' was used to make different charts using the 'Pivot table' tool as follows:

for instance: i) to map a line graph showing trend of ride length over the previous 12 months (see below Figure_8), - first, a new table with few required variables was extracted from the main table 'capston_bike_share_data' using the 'extract' function in the google sheet.

- Next, by clicking on ' pivot table', the variable 'ride_length' was selected in the 'Add' button of the 'Value' option and then 'sum' option was selected for this variable.
- the variable 'month_of_week' was selected in the 'Add' button of the 'Raw' option and the variable 'causal_member' was selected in the 'Add' button of the 'Column' option

Finally, after clicking on the 'Apply' button, the 'chart' was selected in the 'insert' option of the menu bar. Afterwards, a line graph was selected in the chart options. The title, subtitle, axis and legend of the graph were labelled with appropriate naming using the 'setup' option of graph editing.

Similar approaches were applied to make all the remaining charts presented below. Once the charts were made in google sheet, they were downloaded into a folder located on my computer and named as Figure_1 to Figure_8. Afterwards, a new folder called 'Cap' was created in the 'File' pane of the R cloud Studio and those figures(charts) were imported to the folder using the 'import' option located in the 'File' pane. Finally, the following chunk codes were applied to open those figures in this Rmarkdown document:

```r
knitr::include_graphics("Cap/Figure_1.png")
```

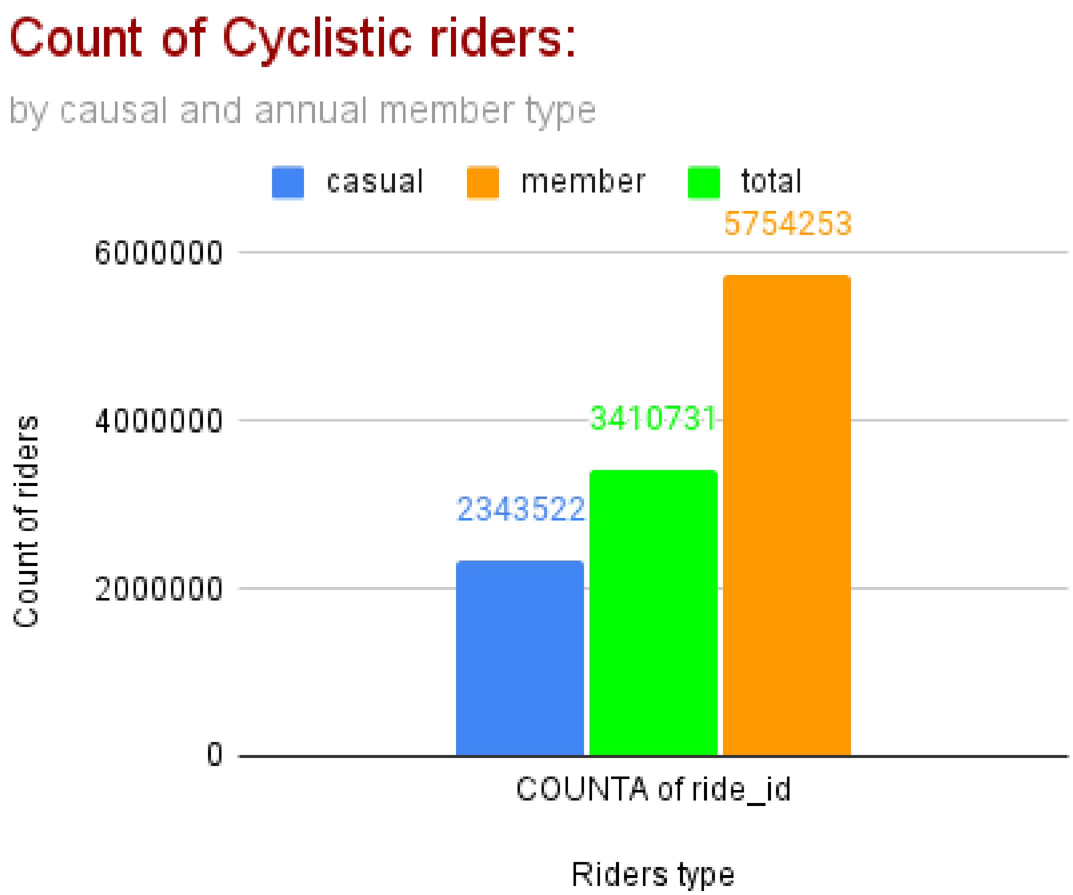

Figure_1: Narrative summary:

The figure shows that the Cyclistic bike-share company rented its bikes 5,754,253 times in the previous twelve months period (February 2022 to January 2023). From those rents, a total of 3410731 were contracted by Annual members; while, 2343522 were contracted by causal riders. This shows that although both rider groups contracted higher share of the total rents of the company, the annual members were contracting much higher proportion of the rents than the causal riders.

```r
knitr::include_graphics("Cap/Figure_2.png")
```

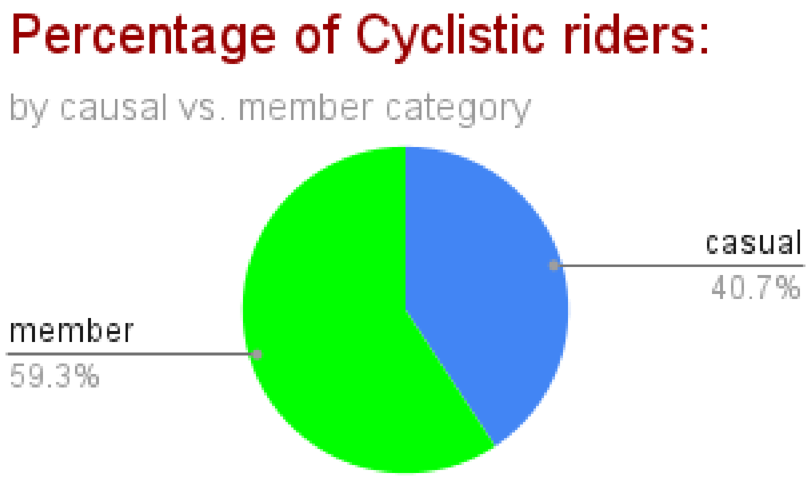

Figure_2: Narrative summary:

When we see the percentage of the total bike rents made to those two riders groups of the company,in the period between February 2022 to January 2023, the Annual members got the majority (59.3%) of the rents.

```
knitr::include_graphics("Cap/Figure_3.png")
```

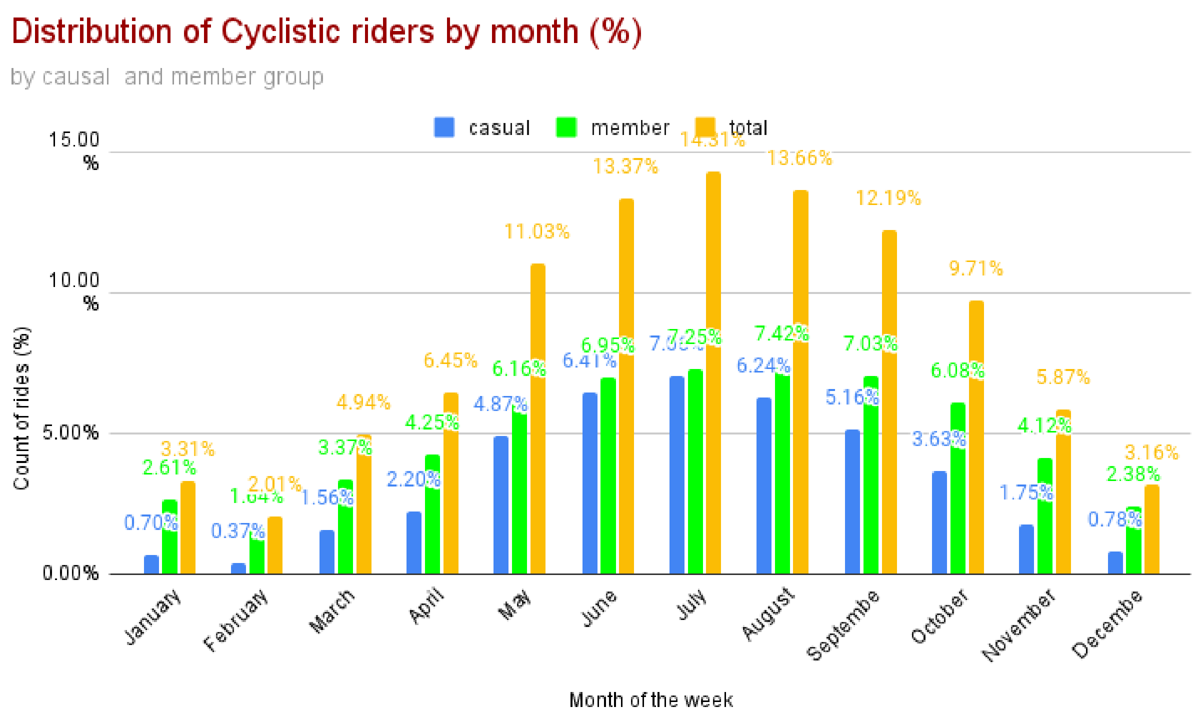

Figure_3: Narrative summary:

When we see the monthly distribution of Cyclistic's historical trip data, the minimum (2.01%) and maximum (14.31%) trip history were observed in the month of February and July, respectively. As can be understood from the figure, the company had better bike renting business in the months between May to September representing above 10% of the annual rents. When we compare the historical bike trip trend between the causal and annual members, both groups showed similar increasing and decreasing trends, but the later had relatively higher usage of the Cyclistic's bikes as compared with the former group.

```r
knitr::include_graphics("Cap/Figure_4.png")
```

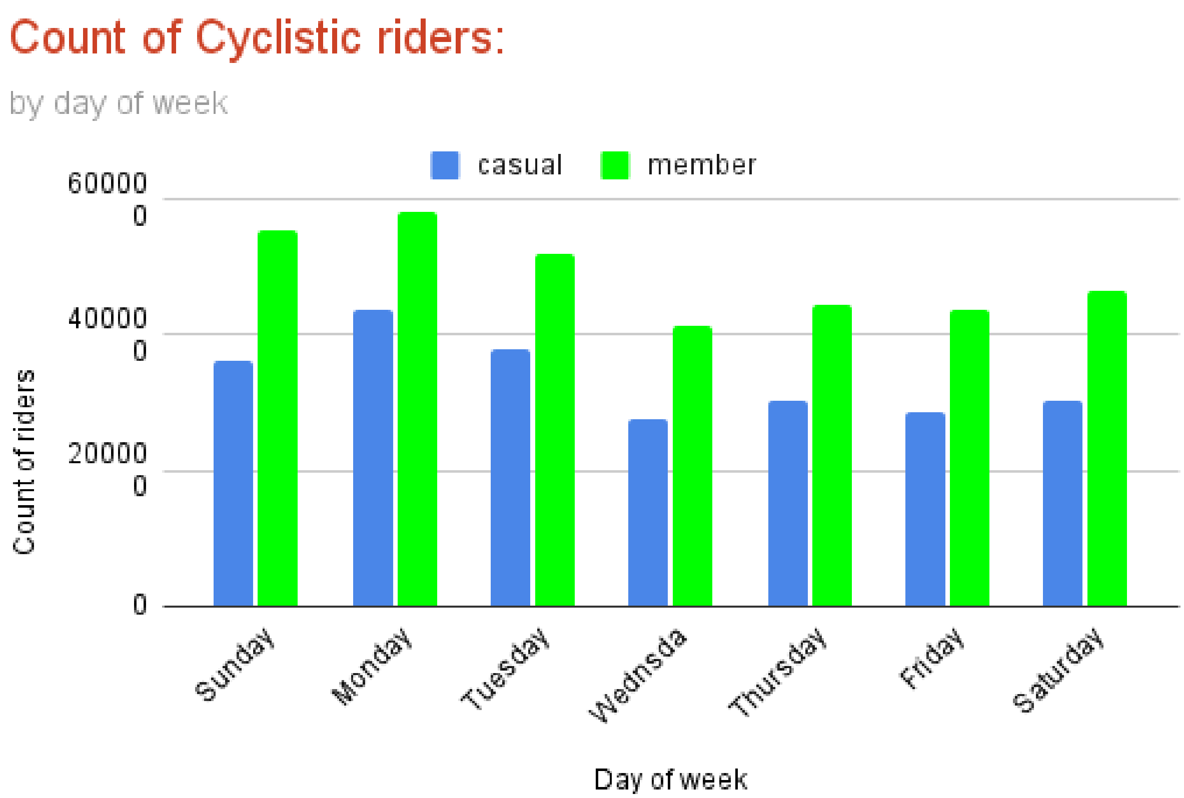

Figure_4: Narrative summary:

When we see the distribution of Cyclistic's historical trip data by day of week, the company had bike renting business almost on all days of the week with minimum and maximum trip history on Wednesday and Monday, respectively. When we compare the daily historical bike trip trend between the causal and annual members, in all cases the annual members had contracted relatively much higher rents than the causual group.

```r
knitr::include_graphics("Cap/Figure_5.png")
```

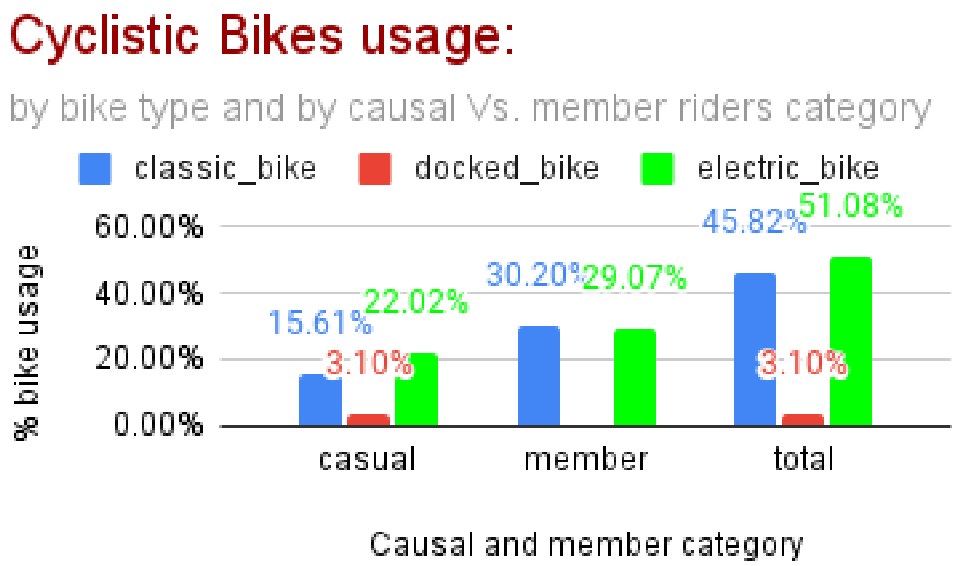

Figure_5: Narrative summary:

The other important insight that can be draw from this case study is the behavior of the rider groups towards the different types of the Cyclistic's bikes. Among the three types of bikes owned by the company, the electric bike followed by the classic bikes were the most preferred types; while, the docked bike had the lowest preference and rented by only few causual riders.

```
knitr::include_graphics("Cap/Figure_6.png")
```

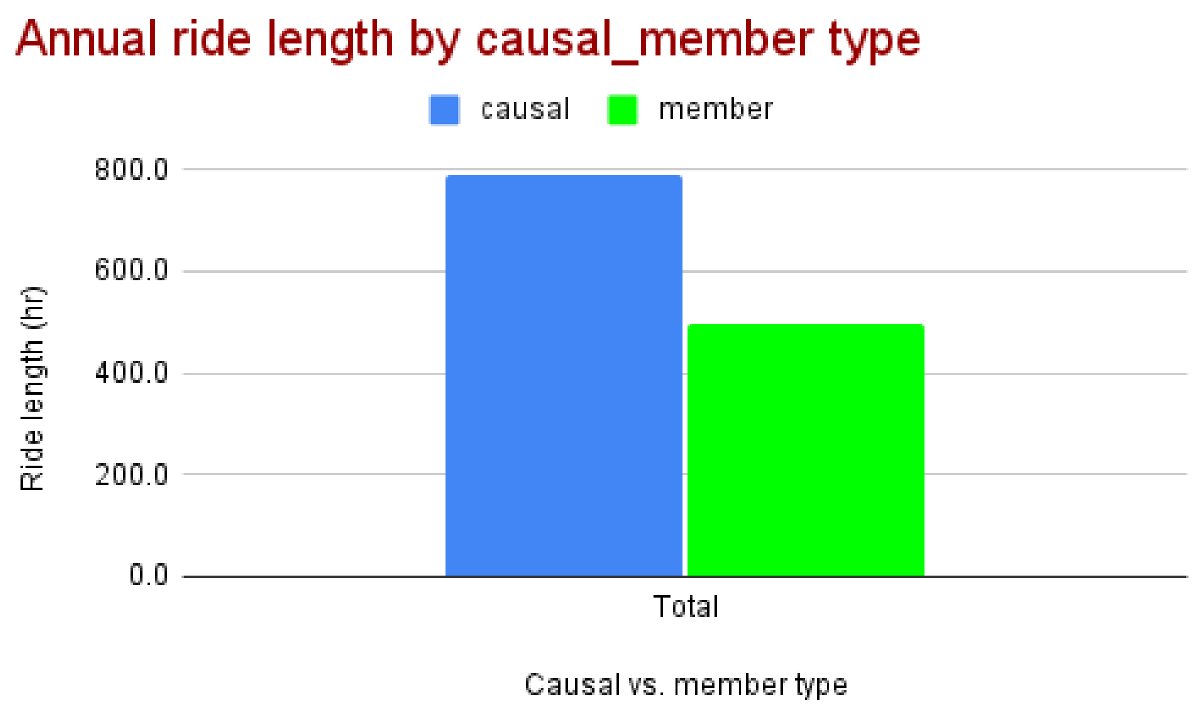

Figure_6: Narrative summary:

This figure presents the sum of ride length (in hours) commuted in the previous twelve months period. For instance, the causual riders commuted closer to 800 hours in the previous twelve months period with the average monthly commute of 66.7 hours or 3 days per month. On the other hand, the Annual members commuted closer to 450 hours in the previous twelve months period with the average monthly commute of 37.5 hours or 1 and 1/2 days per month. So, the data shows that the causal riders did ride for longer hours than Annual members using the Cyclistic bikes in the past twelve months period.

```
knitr::include_graphics("Cap/Figure_7.png")
```

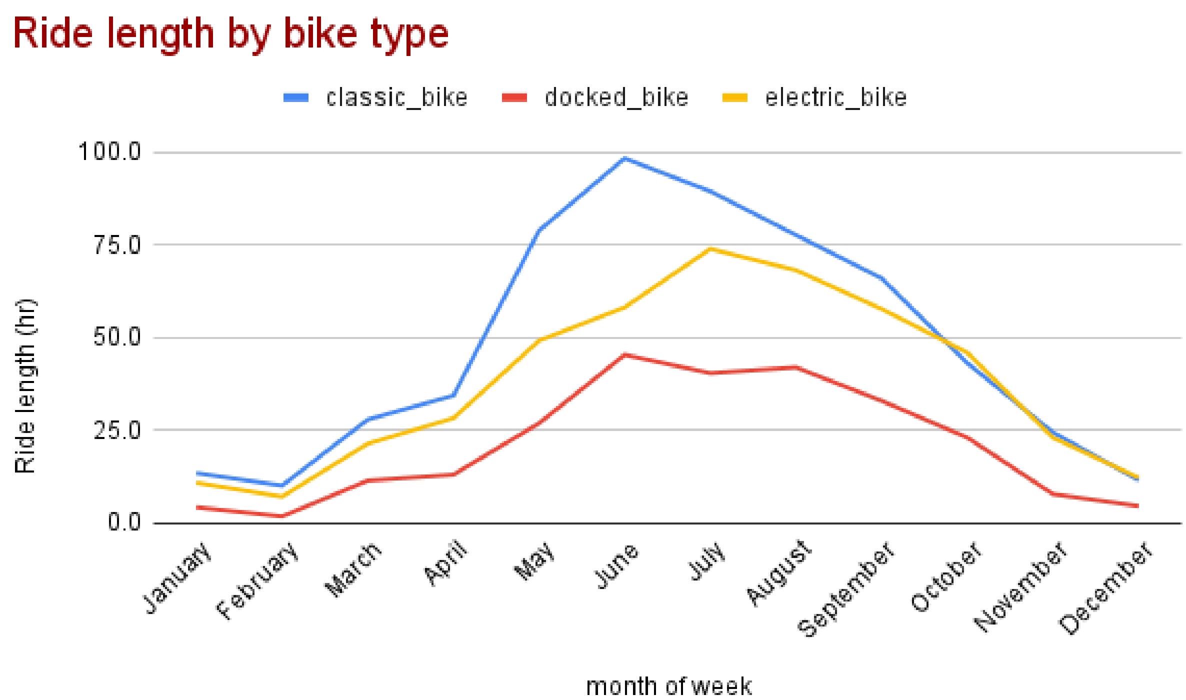

Figure_7: Narrative summary:

The figure compares the trend in ride length (in hours) between the causal riders and Annual members over the previous twelve months. The data shows that there is a clear difference between the two riders groups. In other words, the causal riders did intensively used the bikes as compared with the Annual members. When we see the length of hours of commute, relatively higher commutes were observed in both groups in between May to August with a peak commute in June and July months.

```
knitr::include_graphics("Cap/Figure_8.png")
```

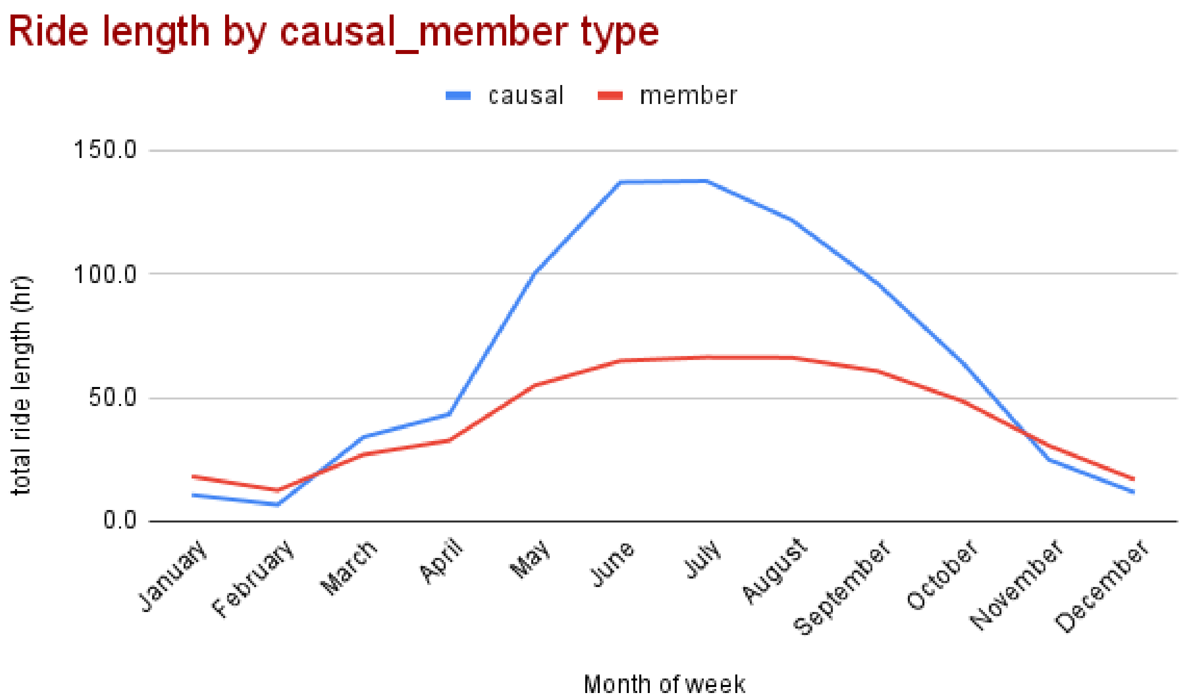

Figure_8: Narrative summary:

The figure shows the ride length (in hours) commuted using the different bike types over the previous twelve months period. As can be seen from the figure, despite the different degree of usage, all the three bike types were used by Cyclistic's riders in all months of the year. Although usage of all those three bike types had similar increasing and decreasing trends over the previous twelve months period, much higher ride-length was commuted using classic-bikes followed by electric-bikes.

Conclusions

i) The case study revealed that 59.3% of the Cyclistic's annual bike rents were the annual memebrs. This may indicate that annual members are the major sources of income for the company.

ii) The data showed that Cyclistic had better bike renting business in the months between May to September of the year where each month representing above 10% of the annual sales or number of rents. In this time interval, both the causal riders and annual members had similar increasing and decreasing bike renting trends.This might implies that the time interval between May to September is a peak ride-share business time and it is a good season to make market promotion campaigns for improving the ride-share business.

iii) Among the three types of bikes owned by the company, the electric_bikes were the most demanded ones in the previous twelve months period. Comparing the causal riders with annual members by their most prefered bike category, the former had more preference for electric bikes; while, the later had more demand for classic-bike type. Those insights obtained from this case study might indicate that electric-bikes are the bikes of choice for most riders.Thus, it is important that the company needs to invest more on electric_bikes inorder to attract more customers and bring the causal riders into annual membership.

iV) The case study revealed a clear difference between the two riders groups in terms of the length of hours they ride the Cyclistic bikes over the previous twelve months period. Comparing the two groups of riders, the causal riders did ride the company's bikes for more hours as compared with the Annual members. The interval between May to August was the time when both groups of reders did commute for longer hours than the remaining months of the year. Interestngly, this time interval also aligns with the peak ride-share renting period of the company. This further supports the idea of planning marketing promotion activities for attracting new customers and also bringing the causual riders to Annual membership in this specific periods of the year.

V) Data analysis further point out the clear difference among the three bike-types in terms of their usage by riders to commute for long hours.Comparing those three bike-types, the classic-bikes were the most bike groups that were used for commuting longer hours. This might indicate that although there is a tendency of increasing demand for electric-bikes, the importance of classical-bikes to commute longer distance make them impotant bike-types for improving and maintaining Cyclistic's ride-share business.

Recommendations

This case study was very helpful to put into practice all the knowledge and skills I gained through google data analytics certificate program. The important insights obtained from the study can help the cyclistic bike-share company to make data-driven marketing decisions for attracting new customers and also bringing causual riders into annual memebrship. Based on the insights obtained from the study, it is recommended to plan such marketing promotions in Cyclistic's peak ride-share business period which is between May to August of the year. In this specific case study, the financial parameters like 'how much the causal-riders and annual members pay' for using Cyclistic's bikes, was not considered. So, it was not possible to draw additional key insights, for instance, which group pay more? and 'how much the Cyclistic's could benefit by converting riders from causal to annual membership.' Therefore, I recommend further investigation on additional parameters including financial attributes in order to draw additional key insights that can be used as an input for developing a successful marketing stratagy to boost Cyclistic's ride-share business.

Acknowledgements

First of all I would like to thank Merit Ameria for accepting me for Google Data Analytics Certificate Program and sponsoring my study. Special thanks to my Squad Coach Mrs. Sherri Harvin, Job Placement Success Coach, Merit America, USA, for her overall advice, encuragement and professional inputs towards the successful completion of this Certificate Program. Last but not list, I highly appreciate all course providers and the tech-collab session organizers of Merit America for sharing me their knowledge and skills in Google data analytics.