Cyclistic Case Study Analysis

This file contains the analysis process of the Cyclistic case study, a study I did as my Capstone Project for my Google Data Analytics Professional Certificate course (Coursera).

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.

This report includes;

1. Clear summary of the business task

2. A description of all data sources used

3. Documentation of any cleaning or manipulation of data

4. A summary of your analysis

5. Supporting visualizations and key findings

6. Your top high-level content recommendations based on your analysis

Phase 1 - Ask

**Business task**

The business task is to design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics.

Phase 2 - Prepare

**Origin and data structure**

Cyclistic’s data is publicly available and it is stored on an AWS page: https://divvy-tripdata.s3.amazonaws.com/index.html. I used data from the last 12 months as requested in the assignment. I therefore only downloaded files (csv format) containing data from June 2021 until May 2022, one for each month.

Each file contains a table with 13 columns and tens of thousands of observations. Each column represents a field, a type of data that defines the use of the bike service provided by Cyclistic:

* ride\_id
* rideable\_type
* started\_at
* ended\_at
* start\_station\_name
* start\_station\_id
* end\_station\_name
* end\_station\_id
* start\_lat
* start\_lng
* end\_lat
* end\_lng
* member\_casual

Phase 3 - Process

**Documentation of any cleaning or manipulation of data**

1. Downloaded files.
2. Saved them on pc in xlsx format.
3. I decided to use SQL for my analysis, so I uploaded the files into Microsoft SQL server management studio.
4. I merged the files of each of the 12 months into one file(table)
5. Added ‘ride\_length’ , ‘day\_of\_week’ and ‘month\_of\_year’ columns as requested in the assignment briefing.
6. The “started\_at” and “ended\_at” columns contained the date and time together, so I separated the date and time in both columns
7. I started the cleaning process by checking and removing duplicate rows
8. Then removed rows where “ended\_at” is less then “started\_at”

Phase 4 - Analyse

**Analyses performed**

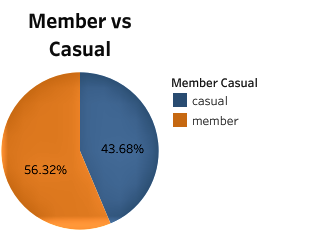
* Selected the more relevant columns, which are ride\_id, rideable\_type, started\_at, ended\_at, member\_casual, ride\_length, date\_started, date\_ended, time\_started, time\_ended, day\_of\_week, month\_of\_year.
* I did a distinct count of all ride\_id columns, this is to know the total number of rides for the those 12 months. I found 5,860,613 distinct rides
* I took a users count, of all the users that rode bikes during those 12 months 56.32% were member while 43.68% were casual users.
* Identified the various rideables used by the riders, I found 3; Electric bike, classic bike, docked bike
* I check for the most preferred rideables between the members and the casual users, I found that though a lower percent of the casual uses the docked bike, the members don’t the use the docked bike at all
* I also took a look at how the riders use the bikes on the weekdays and the weekends, I found the members use the bikes mostly on the weekdays, and casual users use the bikes mainly on weekends
* Taking a look at the average ride length by two categories of users, The casual users have a higher average ride length than the members.

**Phase 5 - Share**

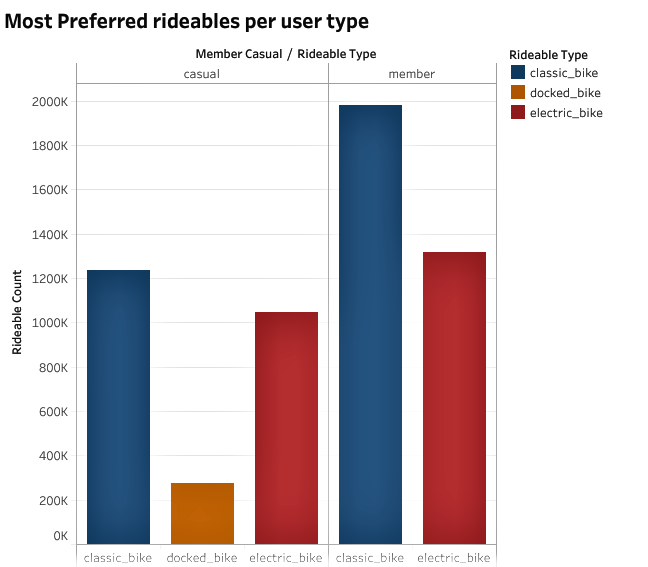
**Introduction**

The following visualisations demonstrate the differences between Cyclistic’s casual riders and annual members in terms of user behaviour.

**Member type percentage distribution**



Here, we see that the casual riders is very well represented, this will definitely increase the Cyclistic revenue if casual rider became annual member.



**Ride count of each user category per day of the week**

