Cyclistic Bike Share Analysis Project

Sebastian Thomas

2022-07-07

Scenario

Cyclistic is a bike-share company based in Chicago whose bikes are used by customers to ride for leisure (casual riders) and also to commute to work each day (annual members). The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, Cyclists' team want to conduct a thorough analysis to understand how casual and annual riders differ. From the insights they derive, they plan to implement a new marketing strategy to convert casual riders into annual members.

Business Task

The task is to spot differences in how casual riders and annual members use Cyclistic's services. This analysis would be useful for Cyclistic in determining whether converting casual riders into annual members would contribute towards its future success.

Key Stakeholders

There are many stakeholders involved In this project, including: * Cyclistic's executive team * Lily Moreno * Cyclistic's Marketing Analytics team * Cyclistic customers who have either single-ride passes, full-day passes or annual memberships

Overview of the data used

The data sources used for this task come from a public data repository, https://divvy-tripdata.s3.amazonaws.com/index.html. For the case study, I have chosen to use the historical data between January 2021 and December 2021. It contains the following field names:

- Ride_id – unique identifier for the trip
- Rideable_type – The type of bike that was used
- Started_at – The time and date the rider took the bike for the trip
- Ended_at - The time and date the rider finished the trip with the bike
- Start_station_name – The station from which the bike was picked up from
- Start_station_id – unique identifier for the start station
- end_station_name - The station from which the bike was kept after the trip
- End_station_id – unique identifier for the end station
- Start_lat – latitude of the start station

- Start longitude – longitude of the start station
- End_lat – latitude of the end station
- End_longitude – longitude of the end station
- Member_casual – field to identify is the rider for a trip is a casual rider or part of the annual membership program
- Ride_Length – Amount of time a Cyclistic bike was used for the ride
- Day_of_Week – day of the week on which the ride was taken

This dataset has over 90,000 records on rides for 12 months, which means we have enough data to deem this dataset reliable. A third party, Motivate International Inc, has put together this dataset. For this reason, it would have a low level of originality. Many parameters are included, and since most of them are self-explanatory, the dataset seems to be quite comprehensive. It is also relatively current as it has been collected within 2021 but is not cited well as it is from a third-party source for whom minimal information is available. In regards to addressing licensing, this dataset has been provided with royalty-free and non-exclusive access Despite its drawbacks from being from a third party, I believe its reliability, comprehensivity, and currency outweigh the disadvantages and make it a recommendable source.

Tools used to conduct analysis on the data

The tools that will be used to process and analyse the data include the following:

- Microsoft Excel
  - This would be useful to to take a quick glance at the data and understand any obvious trends. It would also provide an easy to use user interface to identify any errors within the data.
- RStudio & R
- Tableu