Cyclistic Data Analysis - Business Problem

Gayatri Paul

2023-01-06

1. Introduction

Cyclistic is a bike share company launched in 2016. It has more than 5000 bicycles locked into more than 600 stations across Chicago. Bikes can be unlocked from one station and returned to any station in the system. Cyclistic has flexible pricing plans: single ride passes, full-day passes and annual memberships. Customers with annual memberships are called Cyclistic members and others with passes are called casual riders.

Cyclistic's finance analysts have concluded that annual members are more profitable and maximizing the number of annual memberships will be key to growth. Mareno- the director of marketing believes that rather than creating a campaign to target new customers, we can try to convert existing casual riders into members as they are already aware of the program and have chosen Cyclistic for their mobility needs.

The goal here is to design marketing strategies aimed at converting casual riders into annual members. For that, we need to understand how members and casual riders differ, why would they opt for an annual membership and how digital media be used to influence their decision?

Pricing:

Classic Bike:

Single Ride: $3.30 - Ride for 30 mins - Delay charges: $0.15/minute Day Pass: $15 - Ride for 180 mins - Delay charges: $0.16/minute Annual Membership: $108 per year, 45 mins each ride - Delay charges: $0.15/minute

Electric Bike:

Casual riders: $1 to unlock + $0.39 per minute Members: $0 to unlock + $0.16 per minute

In this case study, we will analyze the key differences between casual riders and annual members and provide top 3 recommendations based on our analysis. The recommendations will be then shared with the stakeholders- Lily Moreno, the director of marketing and Cyclistic executive team.

Data Source:

The datasets have been collected from following website:

https://divvy-tripdata.s3.amazonaws.com/index.html (https://divvy-tripdata.s3.amazonaws.com/index.html)

The data has been made available by Motivate International Inc. under this license:

https://ride.divvybikes.com/data-license-agreement (https://ride.divvybikes.com/data-license-agreement)

We are using the data from October 2021 to October 2022.

2. Prepare

● How is the data organized?

The data has been arranged in Wide format. With following columns:

Column Name Description

ride_id:	Ride ID (Alphanumeric of varying lengths)
rideable_type:	Bike Type (Classic / Dock/ Electric)
started_at:	Date and time @ the start of the ride (format: mm/dd/yyyy hh:mm)
ended_at:	Date and time @ the end of the ride (format: mm/dd/yyyy hh:mm)
start_station_name:	Name of the station where ride started
start_station_id:	ID of the station where ride started
end_station_name:	Name of the station where ride ended
end_station_id:	ID of the station where ride ended
start_lat:	Latitude where the ride started
start_lng:	Longitude where the ride started
end_lat:	Latitude where the ride ended
end_lng:	Longitude where the ride ended
member_casual:	Rider type. 1-Casual, 2-Member

● Are there issues with bias or credibility in this data?

As the data has been directly downloaded from the website and the records are directly taken from tracked rides, there are no bias or credibility issues in the data. The data is ROCCC: Reliable, Original, Comprehensive, Current and Cited.

● How did you verify the data's integrity?

The data source and license to use the data for non-commercial purposes has been provided and linked in the document. The data does not include any personal information such as the rider's name, address or their credit card details. Hence it abides by privacy and security regulations.

● How does it help you answer the question?

The data contains information regarding type of rider and number of trips. We know the prices of each type of pass; hence we can utilize this information for our analysis and gain valuable insights.

● Are there any problems with the data?

1. Some columns of the data sets contain missing information.
2. The data does not contain any information which can help us understand how many times the same person used the bike service. This would have been beneficial for the analysis.
3. There are a few rides for which recorded bike return time is before the start time, we will have to remove those rows.
4. There is two options available for casual riders. Full day pass and single ride pass. There is no data which can tell us which pass was used for the ride. This data is crucial as it will tell us how much money the rider spent on the ride.