CYCLISTIC LTD. BIKE-SHARING STUDYCASE

By: Victor Anim-Opare
(07/02/2022)

INTRODUCTION

This case study is my capstone project in the Google Data Analytics Professional Certificate course. I will follow the six data analysis steps discussed in the course:

- Ask
- Prepare
- Process
- Analyze
- Share
- Act

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.

==========

STAGE 1: ASK

==========

Identification of Problem:
After Cyclistic Ltd. introduced its marketing strategy which focused on building awareness and appealing to broad consumer segments the firm is now facing the issue of not maximizing its profit due to the composition of its consumer population.

Business Task
The main objective is to develop a new marketing strategy aimed at converting casual riders to annual members by understanding the differences between these two groups

Key Stakeholders
1.Director of marketing (Lily Moreno):
Responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels

2.Cyclistic marketing analytics team:
Responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.

3.Cyclistic executive team:
Decides whether to approve the recommended marketing program.

Questions

Three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

==========

STAGE 2: PREPARE

==========

Description of data

In this study I will be using 12 different dataset representing trip data for the Cyclistic Ltd. in the last 12 months (June,2021 -May,2022). Each dataset has 13 fields namely *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_lat* and *member_casual*. The data records daily ride details.

Source Of Data

The data has been made available by Motivate International Inc. under the license found below) #https://ride.divvybikes.com/data-license-agreement

Credibility of data

The data provided is original, appropriate, and current. Given that the data has information on the type of customer, start time and end time on each ride, this will help me to answer the business questions. However, due to data-privacy issues i cannot use any rider's personally identifiable information. This means that i won't be able to connect pass purchases to credit card numbers to determine if casual riders live in the Cyclistic service area or if they have purchased multiple single passes.

Tools

Due to the large size of the data, I will use SQL for data cleaning and analysis. For visualizations I will use PowerBI.

==========

STAGE 3: PROCESS

==========

Data Cleaning Process

The 12 datasets were originally saved as CSV. Even though CSV consumes less memory, I will save the file as an Excel Workbook File because it is easier to read large files this way. Then I will export each dataset into SQL. Before starting any actual data analysis, I combined the 12 different datasets into one database in SQL. It is important to note that, to join the datasets, each dataset must have the same number of columns and the fields must be of the same data type. Start_station_id for July 2021, November 2021 and April2022 must be converted from float to nvarchar data type. The same must be done for end_station_id. Once this has been done, we can now combine the datasets. The resulting database has 5,860,776 records. I then proceed to explore the data to identify problems with the data. One major issue was missing data. Please see table below a summary of the missing data.

Table 1

VARIABLE	NUMBER OF MISSING ITEMS
start_station_name	823,145
start_station_id	1,549,901
both start_station_name and start_station_id	823,142
end_station_name	877,927
end_station_id	1,278,032
both end_station_name and end_station_id	877,927
end_lat	5,036
end_lng	5,036

Data Validation

To confirm if the dates in the data were correct, I ran a query to identify instances when the start date recorded occurred after the end dates. I also checked rides where the start and end times were the same. Table 2 gives the summary of the results

Table 2

CASE	NUMBER OF CASES
started_at = ended_at	507
started_at > ended_at	139
*ride_length=0**	37284

To deal with the issue of missing data and to ensure that the data to be used for the analysis was valid I removed all records with missing data. I also deleted records that met any of the three conditions in Table 2. The third condition was added to table2 because even though the ride length was calculated as 0, the start and end times were different. Those cases occurred because

unit for the *ride length* variable is 'minute'. Hence any ride length less than a minute will show as 0. The final table to be used for analysis has 3,563,153 records.

One major factor that will distinguish casual riders from members is the length of ride. Hence, I will create a new field labelled *ride_length* which will be calculated by subtracting the start time from the end time. An outlier of 55944 minutes of ride length for a casual rider

To help us understand the two groups better, I will create three other fields that tells when the ride was taken. These are *Day*, *Month* and *Season*. In this study, there are four different seasons namely Winter, Spring, Summer and Fall.

==========

STAGE 4: ANALYSIS

==========

In this stage, I will perform various descriptive analysis in order to understand the two customer types.

4.1. Minimum, Maximum And Average Ride Length Between Members And Casual Riders

Customer Type	Minimum Ride	Maximum Ride	Average Ride
	(in minutes)		
Member	1	1495	12
Casual	1	55692	28

Table 1

4.2. Total Ride Length Between Members And Casual Riders

Customer Type	Total Ride Length (in hours)
Casual	692775
Member	426955

Table 2

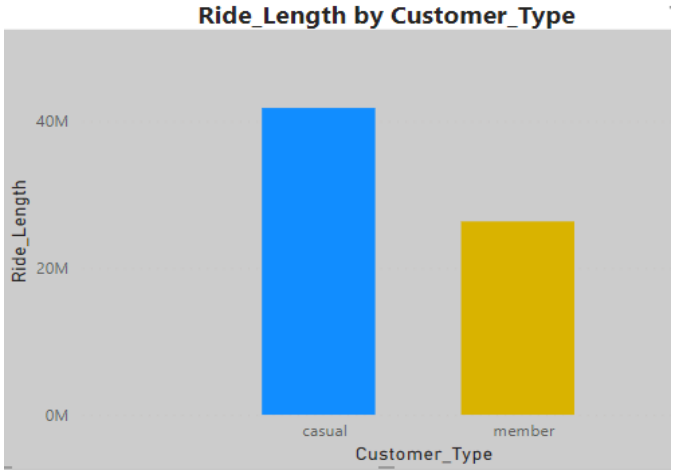

Figure 2.1

Bar Chart Showing Total Ride Length Between Members And Casual Riders

3.Average Ride Length (By Days Of Week) Between Casual Riders and Members

	Day	Customer_Type	Average_Ride_Minute_Days_of_Week
1	Sunday	casual	32
2	Saturday	casual	31
3	Monday	casual	27
4	Friday	casual	27
5	Wednesday	casual	24
6	Tuesday	casual	24
7	Thursday	casual	24
8	Sunday	member	14
9	Saturday	member	14
10	Monday	member	12
11	Friday	member	12
12	Wednesday	member	11
13	Tuesday	member	11
14	Thursday	member	11

Table 3

4.Total Ride Length (By Days of Week) Between Casual Riders and Members

	Day	Customer_Type	Total_Ride_Length_hours
1	Saturday	casual	168756
2	Sunday	casual	153690
3	Friday	casual	91901
4	Monday	casual	76981
5	Thursday	casual	70934
6	Wednesday	casual	66524
7	Wednesday	member	64390
8	Tuesday	casual	63989
9	Tuesday	member	63662
10	Saturday	member	63543
11	Thursday	member	61150
12	Sunday	member	59848
13	Monday	member	57880
14	Friday	member	56482

Table 4

Table Showing The Total Ride Length (By Days of Week) Between Casual Riders and Members

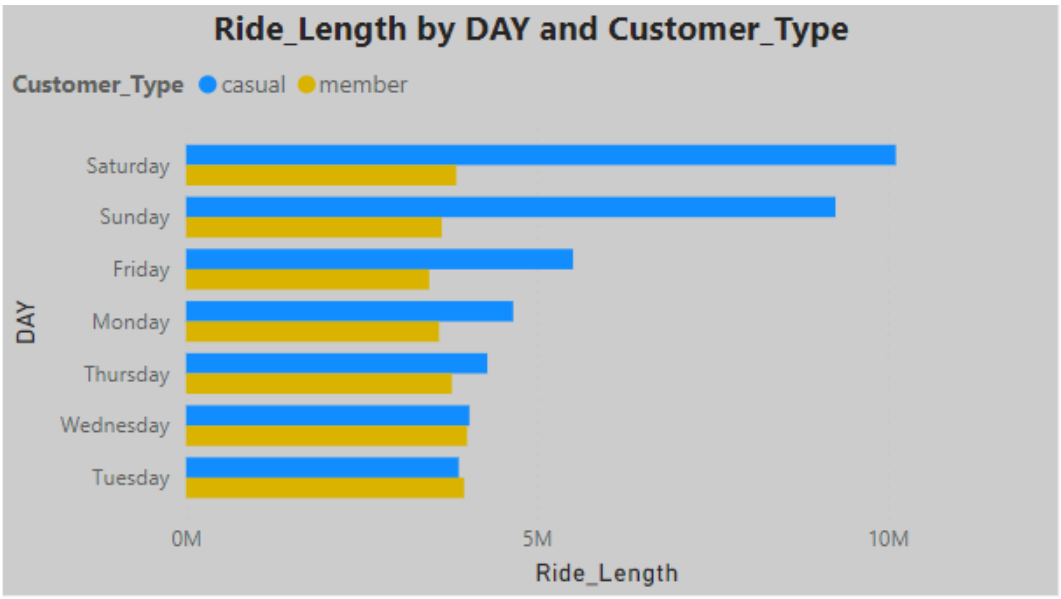

Figure 4.1

Clustered Bar Chart Showing The Total Ride Length (By Days of Week) Between Casual Riders and Members

5.Total Monthly Ride Length Between Casual Riders And Members

	Month	Customer_Type	Total_Ride_Length_hours
1	June	casual	160505
2	August	casual	135518
3	September	casual	116884
4	May	casual	83777
5	October	casual	71848
6	August	member	69604
7	June	member	66583
8	September	member	65984
9	May	member	57809
10	October	member	53658
11	July	casual	43537
12	March	member	27341
13	March	casual	26227
14	December	member	21486
15	April	casual	16762
16	December	casual	16124
17	April	member	14273
18	July	member	13626
19	November	member	13625
20	February	member	12257
21	November	casual	11382
22	January	member	10709
23	February	casual	5155
24	January	casual	5056

Table 5

6. Total Seasonal Ride Length Between Casual Riders and Members

	Season	Customer_Type	Total_Ride_Length_hours
1	Summer	casual	339560
2	Fall	casual	200114
3	Summer	member	149813
4	Fall	member	133267
5	Spring	casual	126766
6	Spring	member	99423
7	Winter	member	44452
8	Winter	casual	26335

Table 6

Table Showing Total Seasonal Ride Length Between Casual Riders and Members

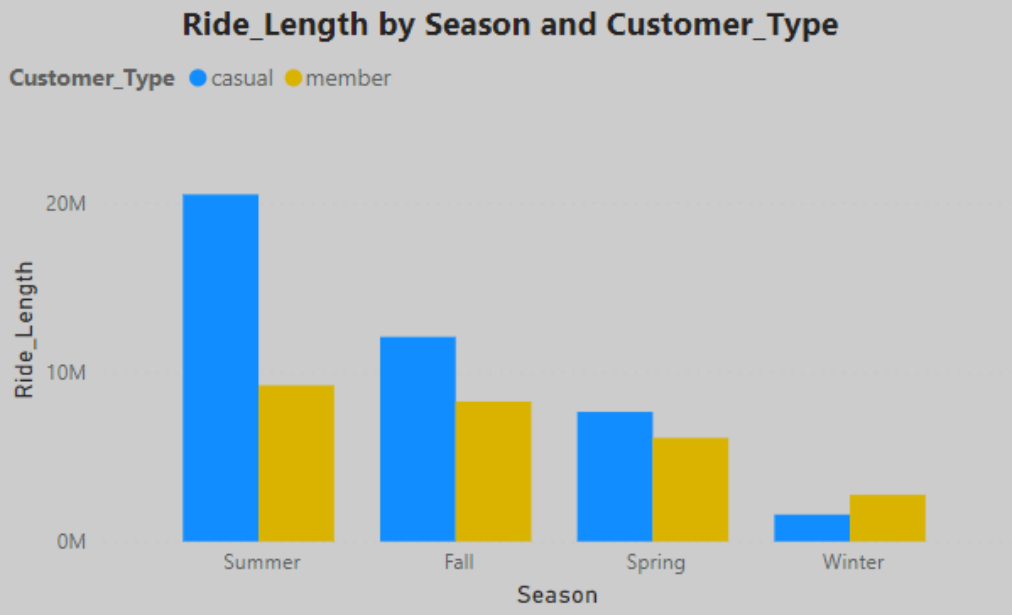

Figure 6.1

Clustered Bar Graph Showing Total Seasonal Ride Length Between Casual Riders and Members

==========

STAGE 5: SHARE

==========

From the above analysis the following conclusions can be made:

1. Casual riders had a higher total ride length than members

2.The average ride length for casual riders is more than twice that of members.

3. The highest average ride length for casual riders occurs on Sunday and followed by Saturday. This is similar for members where both Saturday and Sunday are ranked as the days with the highest ride length.

4.For casual riders, the total length of ride time is high on Saturdays. For members, it occurs on Wednesday.

5.The first three months with the highest total ride length for casual riders are June, August, and May respectively. For members the first three months are August, June, and September respectively.

6.The summer season recorded the highest total ride length for both casual riders and members.

SHARE

This case study will be available on my GitHub profile.

==========
STAGE 6: ACT
==========

RECOMMENDATIONS

1.Since there is fairly an inelastic and high demand for the bikes by casual riders during weekends, Cyclistic Company should increase the price for casual riders on Saturdays and Sundays.

2. Management should introduce a discount for members during summer seasons. During this period the company records high ride length for both customer types. Casual riders will be attracted by the sales and convert into members.

3. Increase the number of electric bikes. It was observed that majority of the casual riders used electric bikes. If the company increases the number of electric bikes, it may win the loyalty of some casual riders who will eventually convert to be members.