Case Study: How Does a Bike-Share Navigate Speedy Success?

Drenusha Gllavica *Université Lumière Lyon 2*

This document provides an analysis of a Case Study, which was done as part of Google Data Analytics Certificate. For the section Data cleaning and manipulation & A summary of key findings was used SQL. Creating visualisations was done via Tableau and this rapport was written via R Markdown.

Keywords: Google Data Analytics Certificate, R Markdown, SQL, Tableau, Cyclistic

---

INTRODUCTION

Cyclistic is a bike-share company based in Chicago, operating since 2016. This company has more than 5,824 bicycles in 692 docking stations. What makes Cyclistic unique is the fact that besides traditional bikes, it offers recycling bikes, hand tricycles and cargo-bikes. Making it attractive for individuals who cannot use traditional bicycles.

Cyclistics pricing plans:

- Members – customers who buy annual memberships.
- Casual riders – customers who buy single-ride passes and full-day passes.

STATEMENT OF THE BUSINESS TASK

Cyclistic's finance analysts determined that members generate more revenue than casual riders. Consequently, they believe that the goal to be achieved in order to boost profits is to increase the number of members.

The goal – to design marketing strategies intended to convert casual riders to members.

Business task – to annalyze how casual riders and annual members use Cyclistic bikes differently and to offer recommendations on how to convert casual customers to annual members.

Stakeholders – *Lily Moreno*, the director of marketing and my manager, *Cyclistic marketing analytics team*, *Cyclistic executive team*.

In order to complete the business task I am going to find:

1. Which is the busiest month and season for Cyclistic?
2. Which stations are used the most, by members and by non-members?
3. Who uses bikes more, members or non-members and which type of bikes are used the most?
4. How do annual members and casual riders use Cyclistic bikes differently during the week?
5. Is there a difference between the length of rides between members and non-members?

DATA SOURCES

In this analysis we used data from Cyclistic's historical trip data, for 2020, to analyze and identify trends. The data has been made available by Motivate International Inc who operates the City of Chicago's Divvy bicycle sharing service. Data has been downloaded from https://divvy-tripdata.s3.amazonaws.com/index.html, precisely 202004-divvy-tripdata.zip, 202005-divvy-tripdata.zip, 202006-divvy-tripdata.zip, 202007-divvy-tripdata.zip, 202008-divvy-tripdata.zip, 202009- divvy-tripdata.zip, 202010-divvy-tripdata.zip, 202011-divvy-tripdata.zip, 202012-divvy-tripdata.zip.

Variables			
member_casual	start_station_name	start_lat	ride_id
rideable_type	start_station_id	start_lng	
started_at	end_station_name	end_lat	
ended_at	end_station_id	end_lng	

DATA CLEANING AND MANIPULATION

In order to clean the data, I used SQL via Miscrosoft SQL Server Management Studio.

- First, I added 10 different databases in SSMS as tables.
- Next, I checked NULL values in each table, and I deleted rows that had NULL values for both: name of the start station & start station ID, as well as name of the end station & end station ID.

```
DELETE FROM capstone..trips_7
WHERE start_station_id IS NULL AND start_station name IS NULL
DELETE FROM capstone..trips_7
WHERE end_station_id IS NULL AND end_station_name IS NULL
```

- Following, I used **UNION ALL** to bind together all the datasets.
- I checked for duplicate rows – there were none.
- I used **LEN** to check for the length of letters in string columns. An example is shown below:

```
SELECT LEN (ride_id)
FROM finaldata..datacyc
```

Since the majority of rows had 16 letters, I checked if there was a row with a different length:

```
SELECT ride_id
FROM finaldata..datacyc
WHERE LEN(ride_id) > 16
```

Results | Messages

	ride_id	ride_id_length
1	EACB19130B0CDA4A	16
2	8FED874C809DC021	16
3	789F3C21E472CA96	16
4	C9A388DAC6ABF313	16
5	943BC3CBECCFD662	16

I did the same action for all string variables. There were not any errors regarding the length of letters in string columns. There was not a need to TRIM any of the observations either.

A SUMMARY OF KEY FINDINGS

In order to continue with data analysis, I created a new column, in each dataset called *length* which shows the difference between when the trip started and the trip ended.

```
SELECT *
, DATEDIFF(MI, started_at, ended_at) AS length
FROM dbo.trips_12
```

Next, I extracted the day out of the column started_at, in order to find out which day of the week was the busiest regarding bike renting.

```
SELECT *
, DATENAME(Weekday, started_at) AS start_station_day
FROM dbo.trips_q1
```

Additionally, I created a table to measure the rides on each month.

Also, I used *COUNT* to check on which type of bikes are most used by members and non-members.

```
SELECT COUNT(member_casual) as number_of_rides
FROM cyclisticdata
WHERE rideable_type = 'classic_bike' AND member_casual = 'member'
GROUP BY member_casual, rideable_type
```

According to the data the most used stations by members and non-members are:

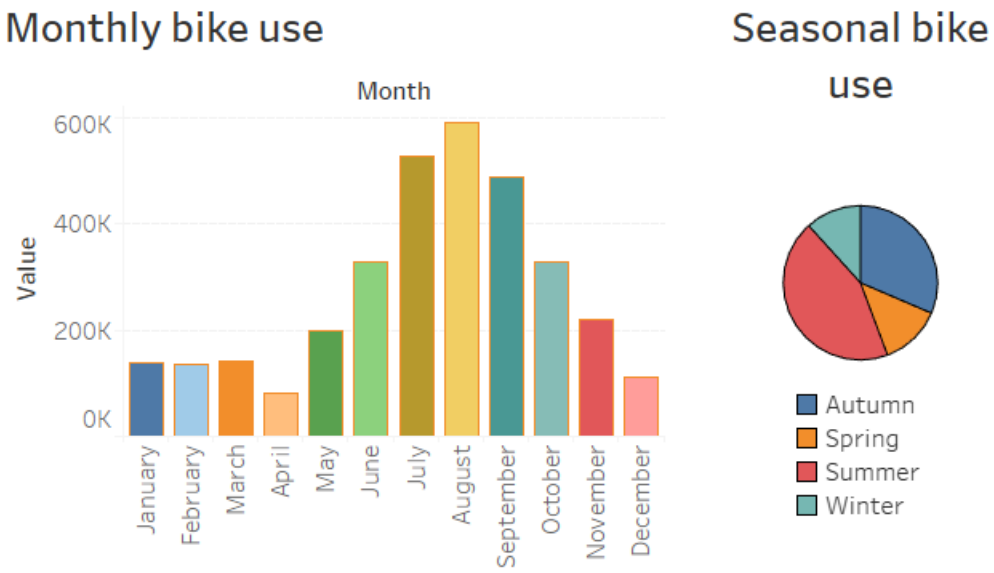
Starting station	Ending station
Non-Members(casuals)	
• Streeter Dr & Grand Ave • Lake Shore Dr & Monroe St • Millennium Park • Theater on the Lake • Michigan Ave & Oak St	• Streeter Dr & Grand Ave • Millennium Park • St. Clair St & Erie St • Dearborn St & Erie St • Broadway & Barry Ave
Members	
• Clark St & Elm St • Kingsbury St & Kinzie St • St. Clair St & Erie St • Dearborn St & Erie St • Broadway & Barry Ave	• Clark St & Elm St • St. Clair St & Erie St • Kingsbury St & Kinzie St • Dearborn St & Erie St • Broadway & Barry Ave

Results show that average length of trips is different for members and casuals:

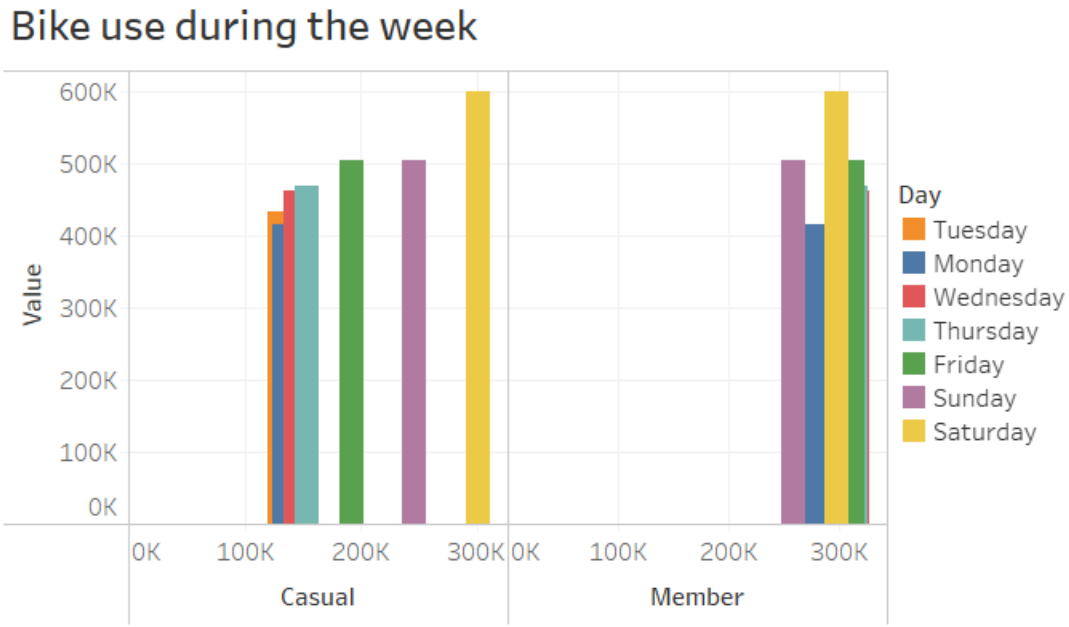
- Members – 12 minutes
- Casuals – 46 minutes

VISUALISATIONS

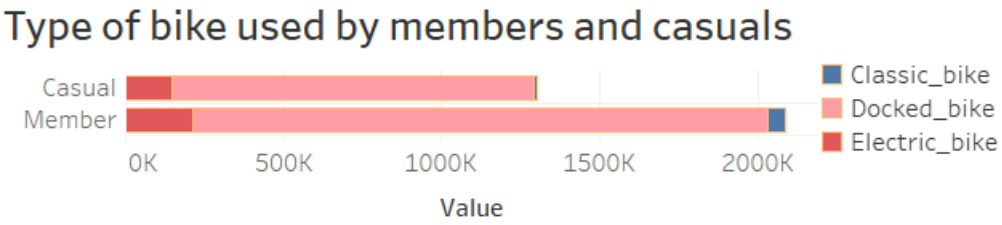
Now that I have gained some insight into the data, I am going to share findings through visualizations done via Tableau.



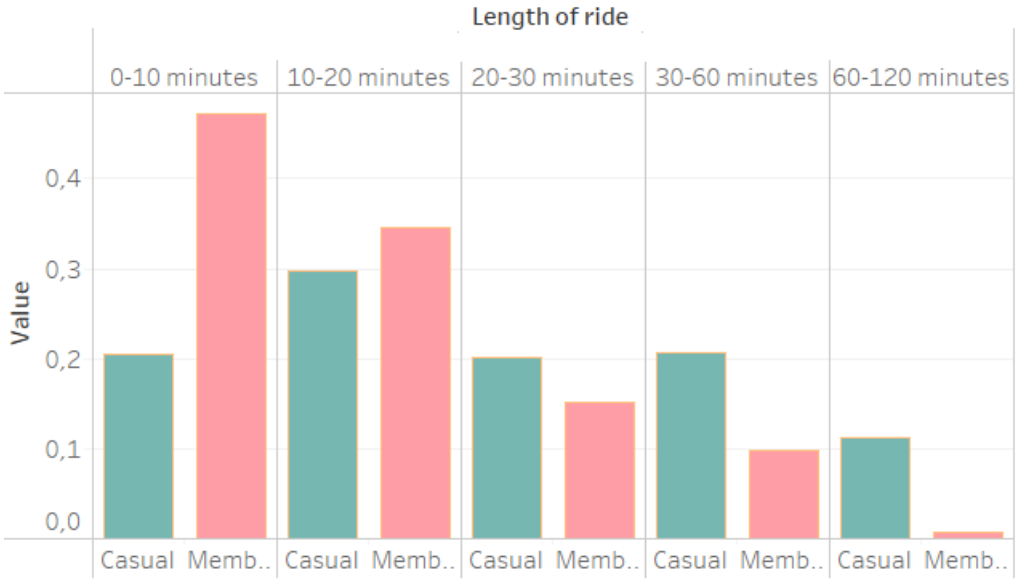
This graph shows that Cyclistic was the busiest during summer months. We can see a more clear picture in the first graph where we see that during June, July, August, September and October Cyclistics bikes were used the most.



Next, we can see the biking differences between members and casual riders. We can see that casual riders prefer Friday, Saturday and Sunday to rent bikes, in comparison to other weekdays.



We can see that concerning the number of clients, members use bikes more than casual riders. Additionally, docked bikes seem the preferred bike among members and casual riders.



Finally, this last graph shows that members generally like to use bikes between 0 to 20 minutes. Whereas, casual riders seem to like to bike from 0 to 120 minutes, with a spike for 30% of them who like to bike for 10-20 minutes.

RECOMMENDATIONS

Based on my analysis, members are more profitable than casual riders, therefore:

1. Cyclistic can offer a discount for casual riders to register as members. This promotion can be offered during summer months, when Cyclistic is used the most. Or on Fridays, since casual riders seem to prefer Friday to Sunday to bike more.

2. Since casual riders use bikes for longer rides, as we can see by the average length of trip duration – 46 minutes, we can create a point system. In this point system, we can offer 1 point for each minute passed. Afterward, points can be accumulated to get a free month of biking or other gifts.

3. For both of these recommendations, the marketing team can focus on advertising close to the stations that are used the most by casual riders. In particular, they can use geotargeted advertising on social media in order to reach consumers that are situated near the most used stations.