Google Data Analytics Capstone Project: Cyclistic Case Study

Introduction:

For the past few months, I have been completing the Google Data Analytics Professional Certificate offered on **Coursera.** I have used the six steps of the data analysis process that I learned through out course to answer key business questions .

Ask, **prepare**, **process**, **analyze**, **share**, and **act** and technical tools **(Excel ,SQL, Google data studio)** . To finish the course, each student needs to creates their own Capstone Project to highlight the technical skills they learned during the course and to showcase their understanding of each step of the data analysis process. I have completed my Capstone Project on the **Cyclistic Case Study.**

Scenario

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. You joined this team six months ago. Lily Moreno, (my manager) and 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.

Ask:

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.

Moreno has assigned me the first question to answer:

How do annual members and casual riders use Cyclistic bikes differently?

In this first phase of analysing data, the questions like:

- What is the problem you are trying to solve?
- How can your insights drive business decisions?

will guide me.

Prepare:

To answer this question, I will be analysing historical Cyclistic bike trip data which is provided by google.

It is public data on the fictional company called Cyclistic provided by Motivate International Inc. under the license.

The data is reliable, free of any bias, and has been collected by Cyclistic and stored on the company's database separated by month in CSV format.

I downloaded the data of last 12 months and stored it in my PC in an appropriate manner.

Problems with the Data: Too many blank or null values

Note: The Cyclistic dataset can be found **here**.

Process:

In this process stage I need to combine and clean the data.

For this I use Excel and SQL on Google's **Big query** platform.

To import data files into bigquery, I need to first delete unwanted data in Excel like longitude, latitude column. This helped reduce the

size . **(Before starting any process, I Created subfolders for the .CSV file so that you have a copy of the original data.)**

I added 2 columns as well ride_length and days _of_week.

Then I checked for any duplicate ride id.

Once in BigQuery, I appended all the datasets, and it has 5,723,445 rows in total.

Then I Check how the rides data is distributed among casual and annual subscribers and see if the number of one group differs from the other.

4. ANALYSE AND SHARE

- After running the following query, I got

```
SELECT member_casual, COUNT(member_casual) ROWS_CNT
FROM `testing-344111.Cyclistic_trip_data.Clean_data`
GROUP BY member_casual
```

- Trips by casual group (2546501) account for 45% of the total number of trips, which is slightly lesser than that of member groups (3176944), 55%

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	member_casual	ROWS_CNT
1	casual	2546501
2	member	3176944

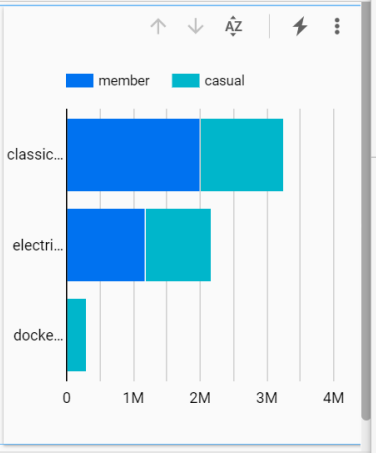
-

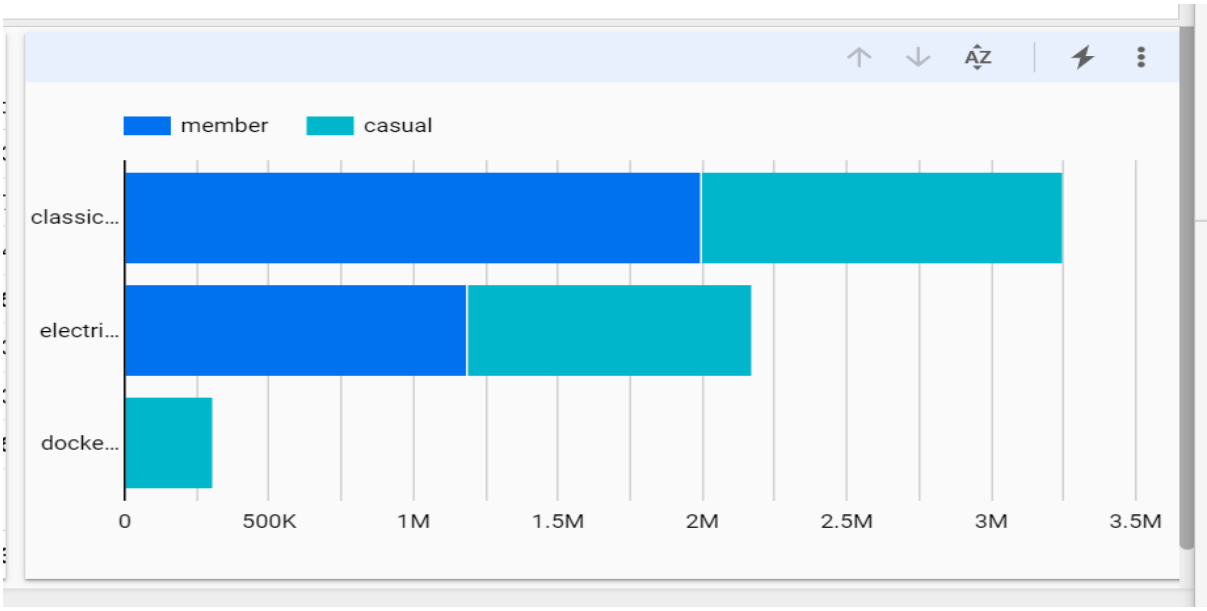
Then I export data to make Pivot table

member_casu...	rideable_type	ride_length	Record Count
casual	docked_bike	1,385	303,984
	classic_bike	1,273	1,257,630
	electric_bike	436	984,887
	Total	**1,434**	**2,546,501**
member	classic_bike	1,002	1,993,023
	electric_bike	444	1,183,921
	Total	**1,023**	**3,176,944**
Grand total		**1,436**	**5,723,445**

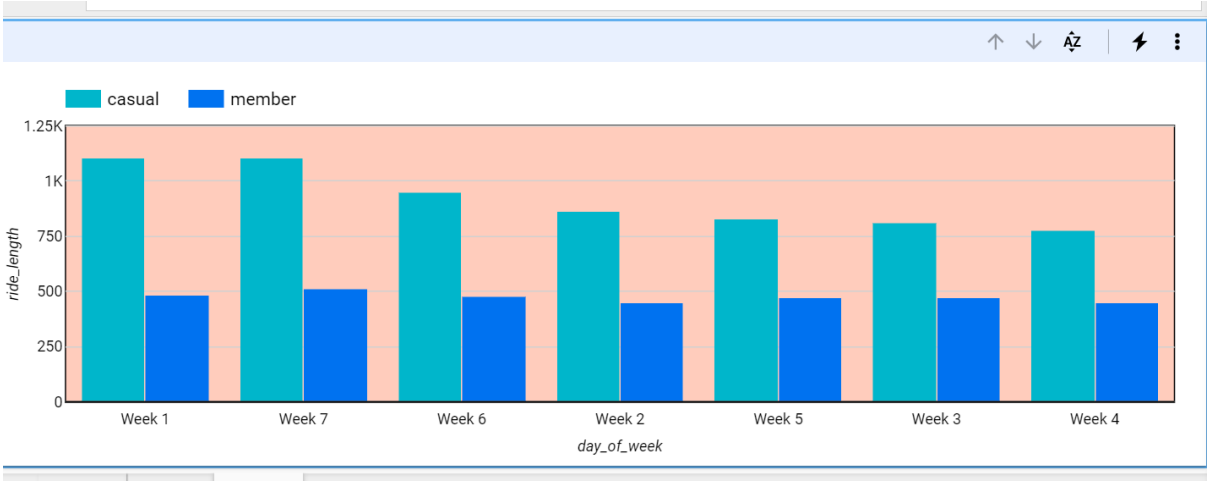
Then I create a chart to know exact numbers through viz

member_ca...	rideable_type	ride_length	Record Count
casual	docked_bike	1,385	303,984
	classic_bike	1,273	1,257,630
	electric_bike	436	984,887
	Total	**1,434**	**2,546,501**
member	classic_bike	1,002	1,993,023
	electric_bike	444	1,183,921
	Total	**1,023**	**3,176,944**
Grand total		**1,436**	**5,723,445**





Average trip duration by the day of the week



Act

In this final step, my task was to state and act on key findings by providing recommendations that will help the marketing director (my manager) maximize profits.

Key Findings

1. Casual riders were found to take longer trips or rides on average than annual Members. This could mean Casual riders used Cyclistic bikes for leisure.

2. Casual riders often ride on weekends whereas annual Members use the program more over the week than on weekends. This could indicate that annual Members are using the bikes to commute to work.

3. Casual riders preferred using electric bikes to other ride types like classic and docked bikes. Annual Members used classic bikes more than other types.

(Recommendations)

1. Users that take long rides could be offered a discount on renting bikes or better yet annual Members that travel more than a certain distance say for instance 10 miles should be offered a discount.

2. Cyclistic bike-share should introduce a weekend only membership that costs less than the current 7-day membership.

3. A rebate program in form of refunds to annual Members that use a certain ride type more others should be introduced. It could be a small percentage of the cost.