

Google Data Analytics Professional Certificate Final Project:

How Does a Bike-Share Navigate Speedy Success?

by

Birute Lauzonyte

March 27th, 2022

I. Introduction

Cyclistic is a bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

The director of marketing believes the company's future success depends on maximizing the number of annual memberships.

The objective is to identify how casual riders and annual members use Cyclistic bikes differently.

Task is to identify how casual riders and annual members use Cyclistic bikes differently.

The data has been made available to download by Motivate International Inc. [here](#) under this [licence](#).

II. Outline

This project will be focusing on two main tools to achieve the above objective, SQL (BigQuery) for data manipulation and Tableau for data visualization. Following the steps below:

1. Download the last 12 months ".csv" data files.
2. Upload files into BigQuery to be combined into one table for all months.
3. Start data exploration by filtering, sorting and manipulating data as needed to get insights.
4. Export the full dataset to Tableau for additional insights using visualization.
5. Propose recommendations based on the data results.

III. Data preparation and manipulation

In order to upload “.csv” files into BigQuery, each individual file needs to be no more than 100 Mb. After examining the monthly “.csv” file sizes, I found out that months from June till October all exceed the size limit. In order to get around that issue, I will use “Split CSV” online program to split each one of those months into 2 files, that way files can be successfully uploaded into BigQuery without any issues.

Once all files are of the allowable sizes for upload, they are uploaded into BigQuery using Auto-Detect schema to be combined into one dataset representing the riders data of the last 12 months. Below is the SQL code used to manipulate & combine the data as needed to create the full dataset for one year.

```
SELECT *
FROM `sql-project-339918.Cyclistic.case_study`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study202110_1`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202105`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202106_1`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202106_2`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202107_1`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202107_2`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202108_1`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202108_2`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202109_1`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202109_2`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202110_2`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202111`
union all
```

```

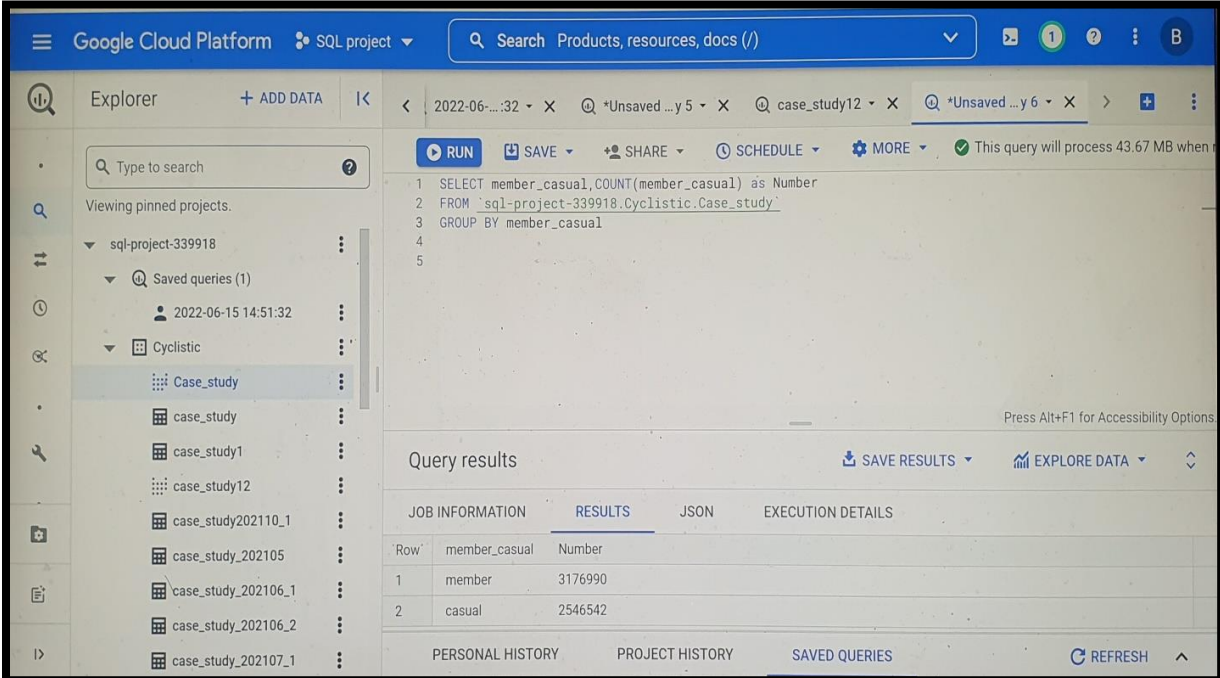
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202112`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202201`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202202`
union all
SELECT *
FROM `sql-project-339918.Cyclistic.case_study_202203`

```

IV. Data exploration using SQL

After combining the last 12 months data, exploring the full dataset using SQL shows that the table has 337,230 observations.

Below you can see the SQL code and the results to see the total number of Cyclistic annual memberships and casual riders in the last 12 months. The query results below (Picture 1) shows that there are more than 3 million members and around 2.5 million casual users.



The screenshot shows the Google Cloud Platform SQL interface. The left sidebar displays the project structure for 'sql-project-339918', including a 'Cyclistic' dataset and various 'case_study' tables. The main panel shows a SQL query being executed:

```

1 SELECT member_casual,COUNT(member_casual) as Number
2 FROM `sql-project-339918.Cyclistic.Case_study`
3 GROUP BY member_casual
4
5

```

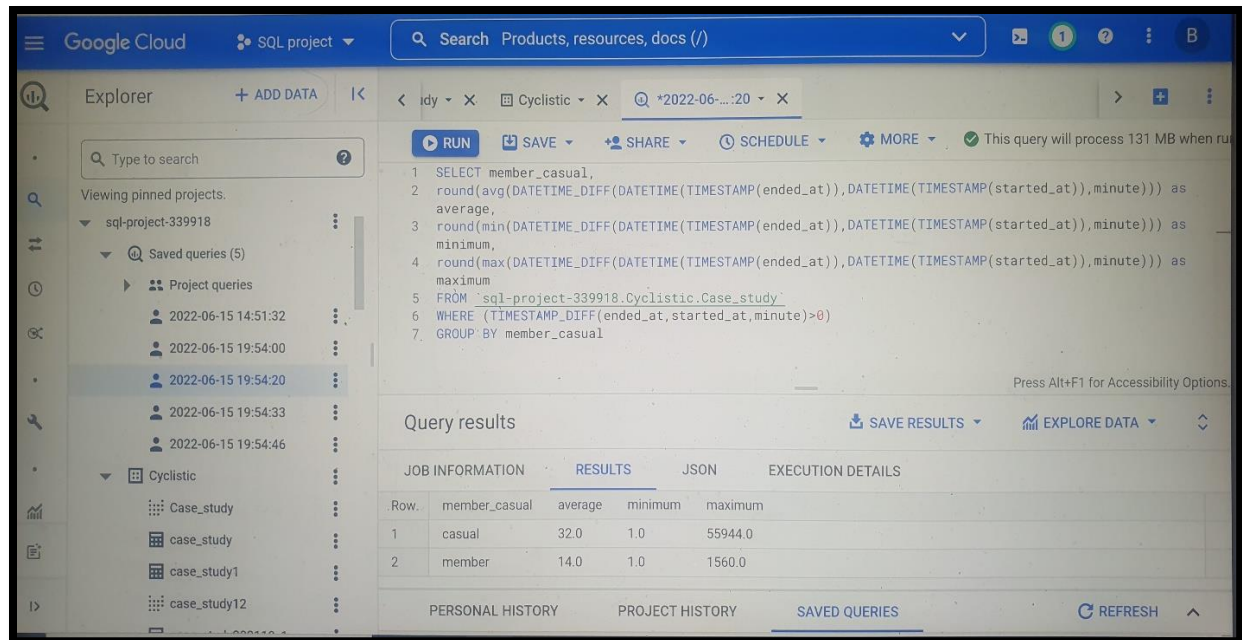
Below the query, the 'Query results' section is visible, showing a table with two rows of data:

Row	member_casual	Number
1	member	3176990
2	casual	2546542

The interface also includes tabs for 'JOB INFORMATION', 'RESULTS', 'JSON', and 'EXECUTION DETAILS'. At the bottom, there are tabs for 'PERSONAL HISTORY', 'PROJECT HISTORY', and 'SAVED QUERIES', along with a 'REFRESH' button.

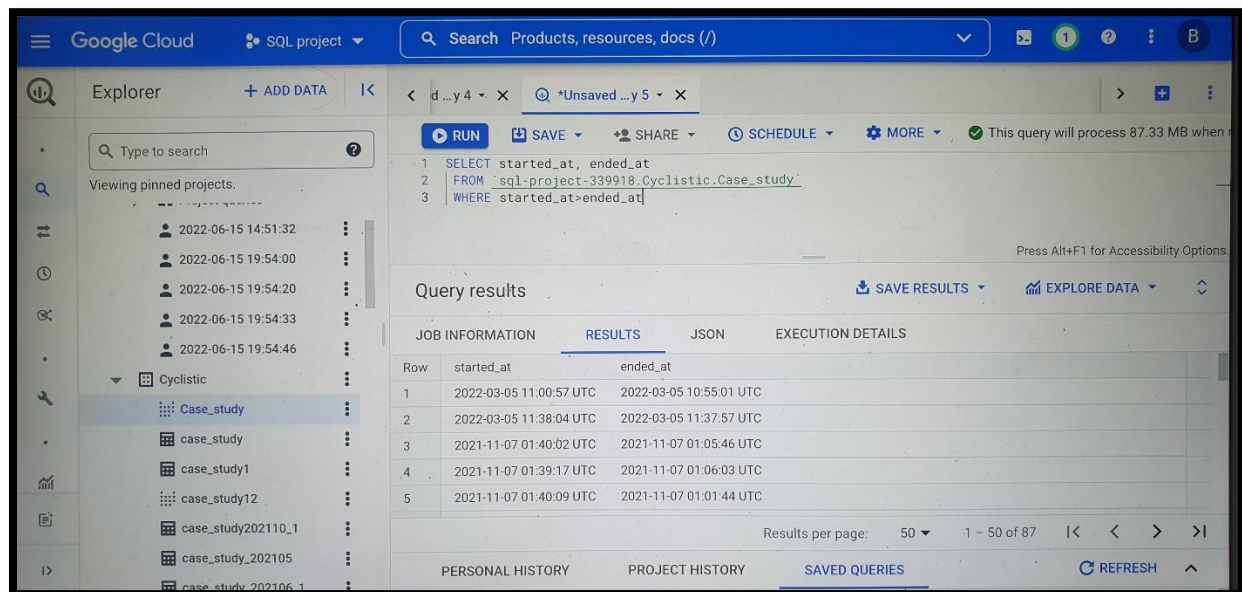
Picture 1. Total number of annual memberships and casual riders.

When exploring the average, the minimum and the maximum trip times for each user group in minutes, the query results below (Picture 2) are very interesting. The results show that casual users are using bikes for longer times than annual members. With the average trip time for members being 14 minutes, while the average trip time for casual riders are 32 minutes, that is more than double comparing to annual members.



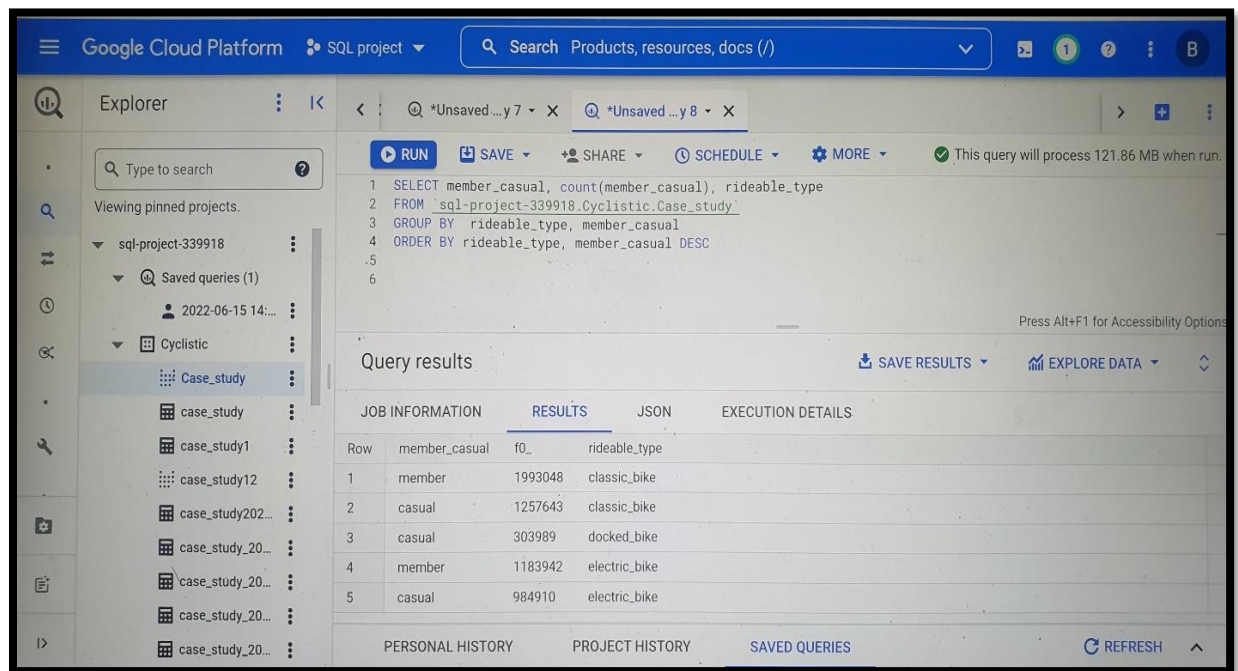
Picture 2. The average, minimum and maximum trip times in minutes for each user group.

The results above show that there is an issue with the minimum trip times in the dataset. When trying to understand the reason for that, I wrote a query that checks the entries in which the trip end time was before the trip start time. Obviously, those entries were due to an error, but the results below (Picture 3) show that there are 87 entries in the whole dataset that meets this condition and this is why the minimum trip times are not accurate.



Picture 3. Times when trip end time was written before trip start time.

Finally, when exploring which bike types (Picture 4) Cyclistic users use the most by different user group in general classic bikes are the most popular choice followed by electric bikes. Also, docked bikes are used just by casual riders.

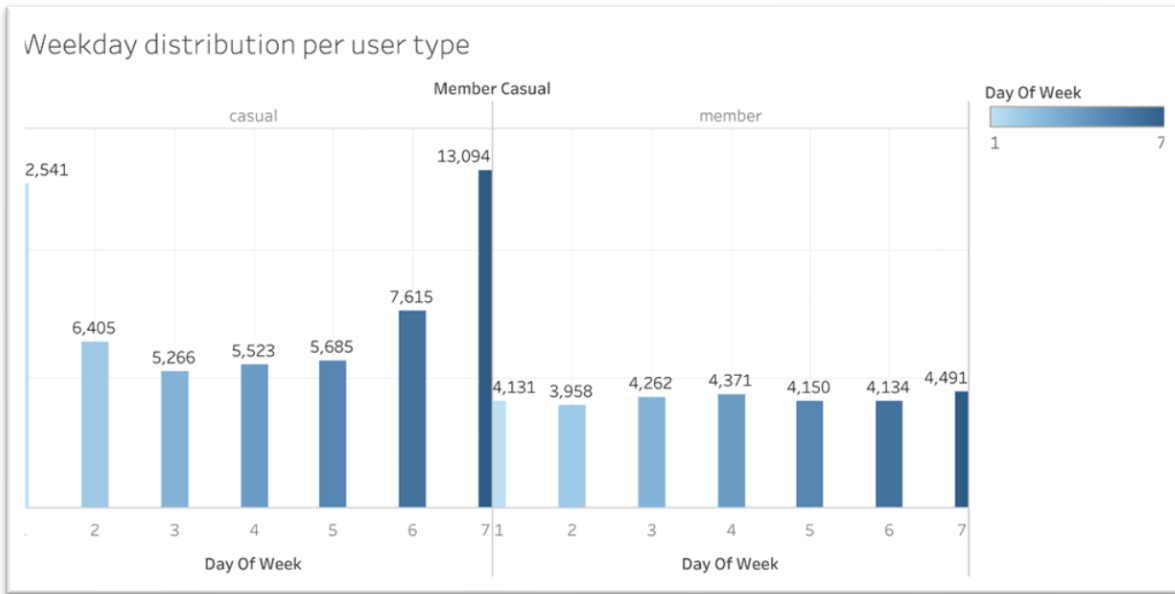


Picture 4. The most used bike types by different user group.

V. Data exploration using Tableau

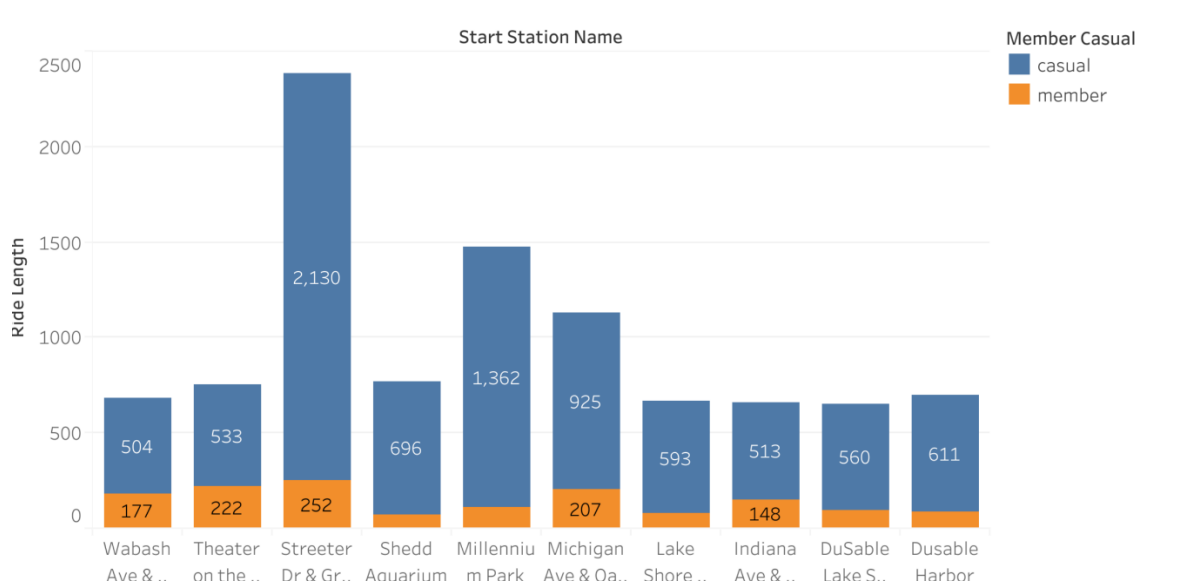
For results visualization and additional insights, I exported data to Tableau. The first visualization below (graph 1) will illustrate how casual riders and annual members use bikes during the week, counting that 1 = Sunday and 7 = Saturday.

The results below show that for casual riders weekends are the busiest days of the week, while annual members use bikes the most on Sunday and Wednesday but with not such a big difference as casual riders.



Graph 1. Weekday distribution per user type considering that 1 = Sunday and 7 = Saturday.

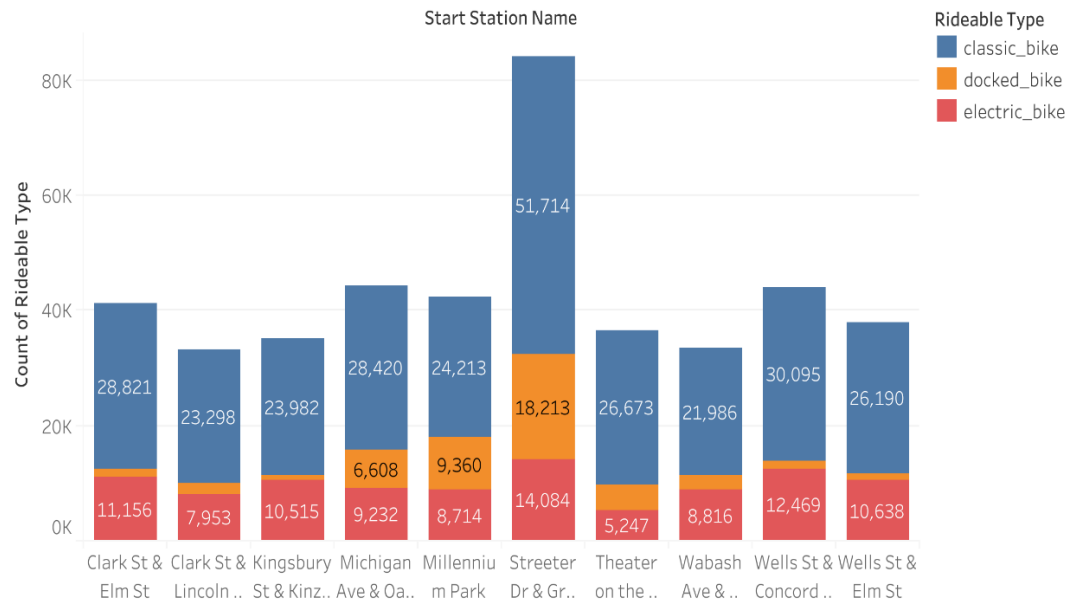
The next bar chart (graph 2) will illustrate the top 10 most popular stations from which trip starts that shows the number of each user type per station.



Graph 2. Top 10 busiest stations per user type.

Following the same method of visualization, I explored the most popular bike types for the top 10 busiest stations. Below you can see a stacked bar graph (graph 3) that shows the number of each bike type rented in the most popular top 10 stations.

Bike type distribution for the top 10 stations

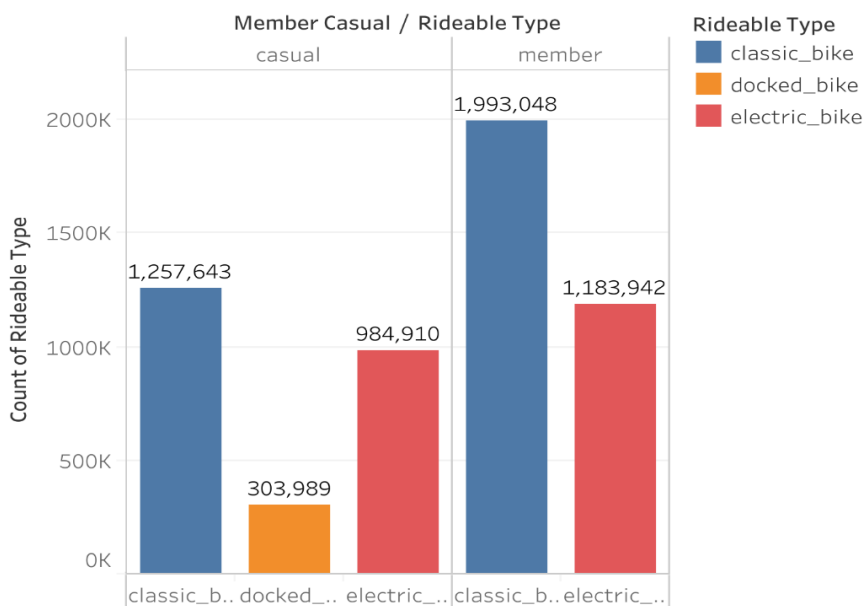


Graph 3. Bike type distribution for the top 10 stations.

Below there is the last stacked bar graph (Graph 4) that illustrates bike type distribution per each user group.

The results match the outcome of the SQL queries and shows that the most popular are classic bikes followed by electric ones, and docked bikes are exclusively used only by casual riders.

Bike type distribution per user type



Graph 4. Bike type distribution per user group.

VI. Insights and recommendations

I discovered several important insights through the data exploration that I performed above. The results of data were the key in determining the right recommendations to achieve the objective of this project and help the Cyclistic company to achieve their goal and maximize the number of annual memberships.

The most important insights are:

- The number of casual riders almost matches the number of annual memberships. That means that there is a big target audience to make the marketing campaigns and a possibility to almost double the number of current annual memberships.
- Casual riders trip times are more than double than current members.
- Annual members use bikes almost same on weekends as on working weekdays, but casual riders use bikes much more often on weekends than on labor days.
- Docked bikes are used just by casual riders.

After studying those insights, my recommendations are:

- Make marketing campaigns and encourage casual users to apply for memberships by creating special offers on rides for weekends and that are longer than 15 or 20 minutes.
- Encourage casual riders to apply for memberships by creating special offers for docked bikes.
- Concentrate on the stations that casual riders use the most:
 1. Streeter Dr.
 2. Millennium Park
 3. Michigan Ave