**BACKGROUND:** This analysis is based on case study number one (1) of the Google Data Analytics Certificate, in which I am assuming the position of a ‘Jr. Data Analyst’ at Cyclistic, a fictitious bike-share company based in Chicago. The aim of the analysis is to understand the behaviors of Cyclistic’s two main users: members and non-members(casual). The reason why the company would like to understand the behaviors between the two is that the company’s membership numbers have been stagnant and the management team does not have strategies to help increase the membership. Analyzing the behaviors between the two would create insights for the marketing team that would help create strategies to promote their membership plan.

**TOOLS:**

The data will be cleaned and analyzed using SQL Server Management Studio (SSMS) and Power BI and will follow Google’s six-step process of processing data: ask, prepare, clean and process, analyze, share, and act.

**Ask**

It is important to understand what are the major questions that drive the analysis. Given that the company has a problem with increasing its membership numbers and it has data about its users, hence the question: “**How do annual members and casual riders use Cyclistic bikes differently?**”

**Prepare**

The data for this analysis was collected from [here](https://divvy-tripdata.s3.amazonaws.com/index.html). It was decided to use data for January, 2022 to December, 2022, downloaded as .csv’s. All of the twelve separate files/tables were combined as one table as illustrated in SQL code. This was easy because all of the tables had the same attributes.

The length of ride\_id’s (primary key) was checked to make sure that I know what the primary key looks like. Verification of any null and duplicate values per row was also done. These have been elaborated in the SQL code.

## Clean and Process

It was established the ride\_id column had no NULL and Duplicate values. The columns that had NULL values were highlighted as per SQL code also uploaded. The next step was to run code that would eliminate these NULL values and generate a final table/code, which was uploaded to Power BI by running it in the Power BI SQL connector or further Visualization.

After there were no NULL values and duplicate values, it’s time to add new rows for analysis purposes as below;

* Day of the week AS day\_of\_week
* Month AS month
* Day AS day
* Year AS year
* ride length as ride\_length

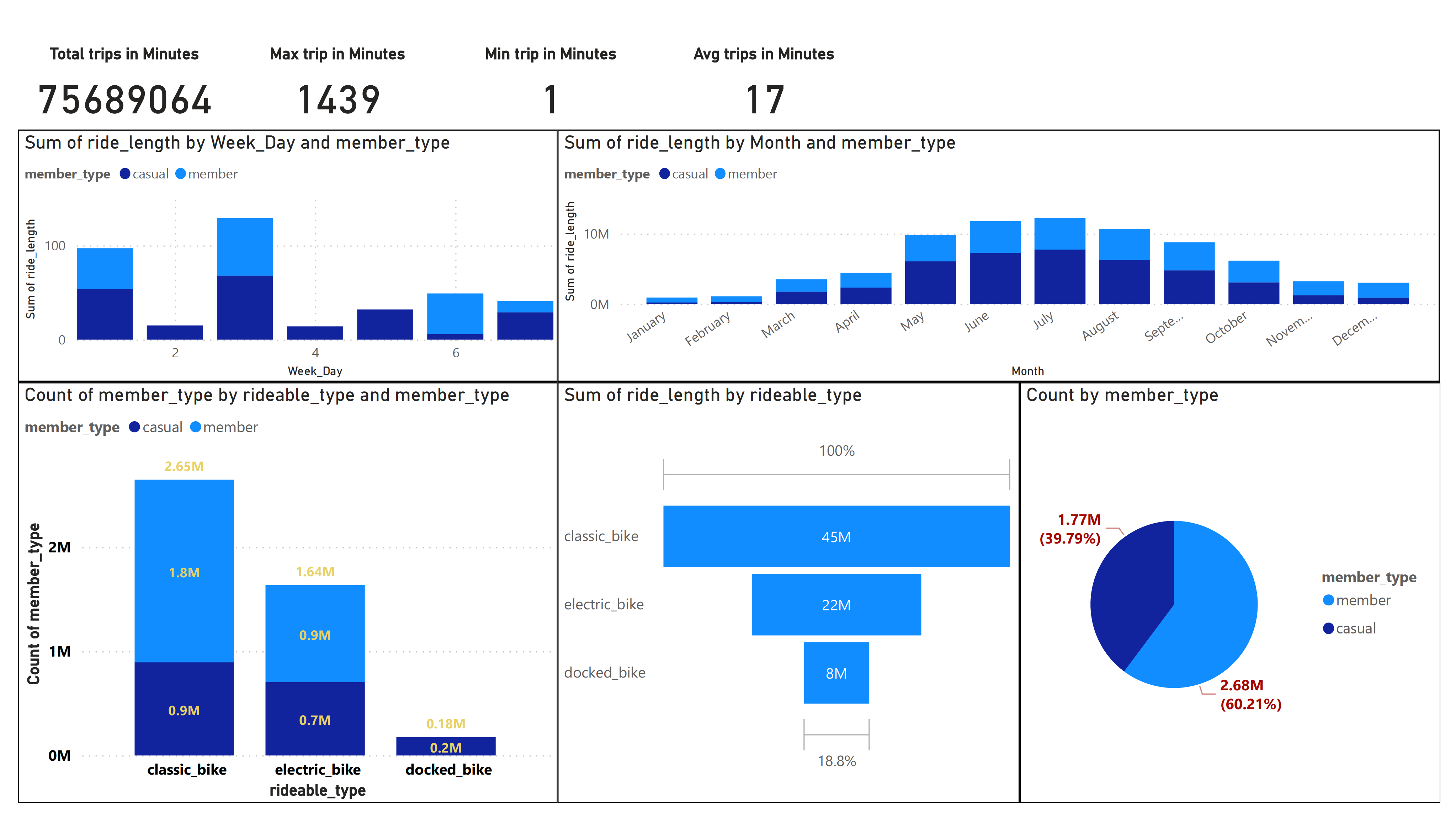
We included data that had a valid time length (more than 1 minute and less than 24 hours).

## Analyze

Once the data was cleaned, analyzing the data to answer the question; “**How do annual members and casual riders use Cyclistic bikes differently?”,**which can be answered in two main categories: behaviors in the difference between time spent on the bicycle and the routes each users take. The analysis on the time spent on the bicycle were made through many queries because we would like to know the length of the ride in terms of minutes and hours while also knowing the trends when people ride bikes in days and months. The queries are elaborated in the uploaded SQL code.

## Share

Here we share the insights derived from the analysis through visualizations using Power BI:



**ACT:**

Members use more of the classic bike as well as the electric bike and the docked bike is used more by casual users. The pie chart also shows Cyclistic to have more members (~60%) as opposed to casual at (~39%). We also see that the classic bike has more ride time followed by the electric bike then the docked bike. We also notice that through the week, casual members are seen to ride more with members’ numbers only high on Saturdays. Through the year, we see un upsurge of casual riders around April to September. Classic bikes as well as electric appear to be more used by members. It could also be concluded from the weekly usage that most members may be using the bikes to go for work and back.