**Cylistic Marketing Strategy 2020**

Cyclistic Bikeshare Analysis - Case Study

Google Data Analytics Professional Certificate

This case study is a **capstone project** for the Google Data Analytics Professional Certificate. For the purpose of this case study, I am a **junior data analyst in the marketing analyst team** at Cyclistic - a fictional bike-share company - in Chicago.

**Ask**

**Background**

Cyclistic offers a bike-share program that features more than 5,800 bicycles and 600 docking stations. They offer reclining bikes, hand tricycles and cargo bikes making bike-share more inclusive to people with disabilities and riders who cannot use a standard two-wheeled bike. The company has flexible pricing plans: single-day pass, full-day pass and annual memberships. People who opt for the single and full day pass are considered ***casual riders*** while those with annual membership are ***members***.

**Problem Statement**

The cyclistic finance analysts have come to a conclusion that annual members are much more profitable than casual riders. The marketing team has been tasked with **designing marketing strategies aimed at converting casual riders into annual members**.

Lily Moreno, the director of marketing believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

In order to do that, 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 marketing tactics.

**Key Stakeholders**

**Lily Moreno**: The director of marketing. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.

**Cyclistic marketing analytics team**: These data analysts are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.

**Cyclistic executive team**: This team is made up of detail oriented members who will decide whether to approve the recommended marketing program.

**Prepare**

The data analyzed in this case study is the [Divvy Trip Data](https://divvy-tripdata.s3.amazonaws.com/index.html) of the last 12 months - December 2020 to November 2021. The dataset has a different name because Cyclistic is a fictional company. However, for the purpose of this case study, the dataset is appropriate. The data has been made available by the Motivate International Inc. under this [license](https://www.divvybikes.com/data-license-agreement). The data is structured - arranged in rows and columns. The data is Reliable, Original, Comprehensive, Current and Cited.

The dataset includes twelve(12) csv files each arranged in rows and columns with trip information such as ride Id, date, ride type, longitude, latitude, start station details and end station.

**Process**

Due to the large data, two analytical packages were used - **Microsoft Excel** and **R**. Ms Excel was used to clean the data to an extent: *remove unwanted columns E - L, filter for nulls, calculate trip\_duration and extract day\_of\_week for each entry* then R was used for further cleaning and visualization.

Cleaning Process

* Deleted the columns listed below:

Start Station Name

Start Station ID

End Station Name

End Station ID

Start Lat

Start Lng

End Lat

End Lng

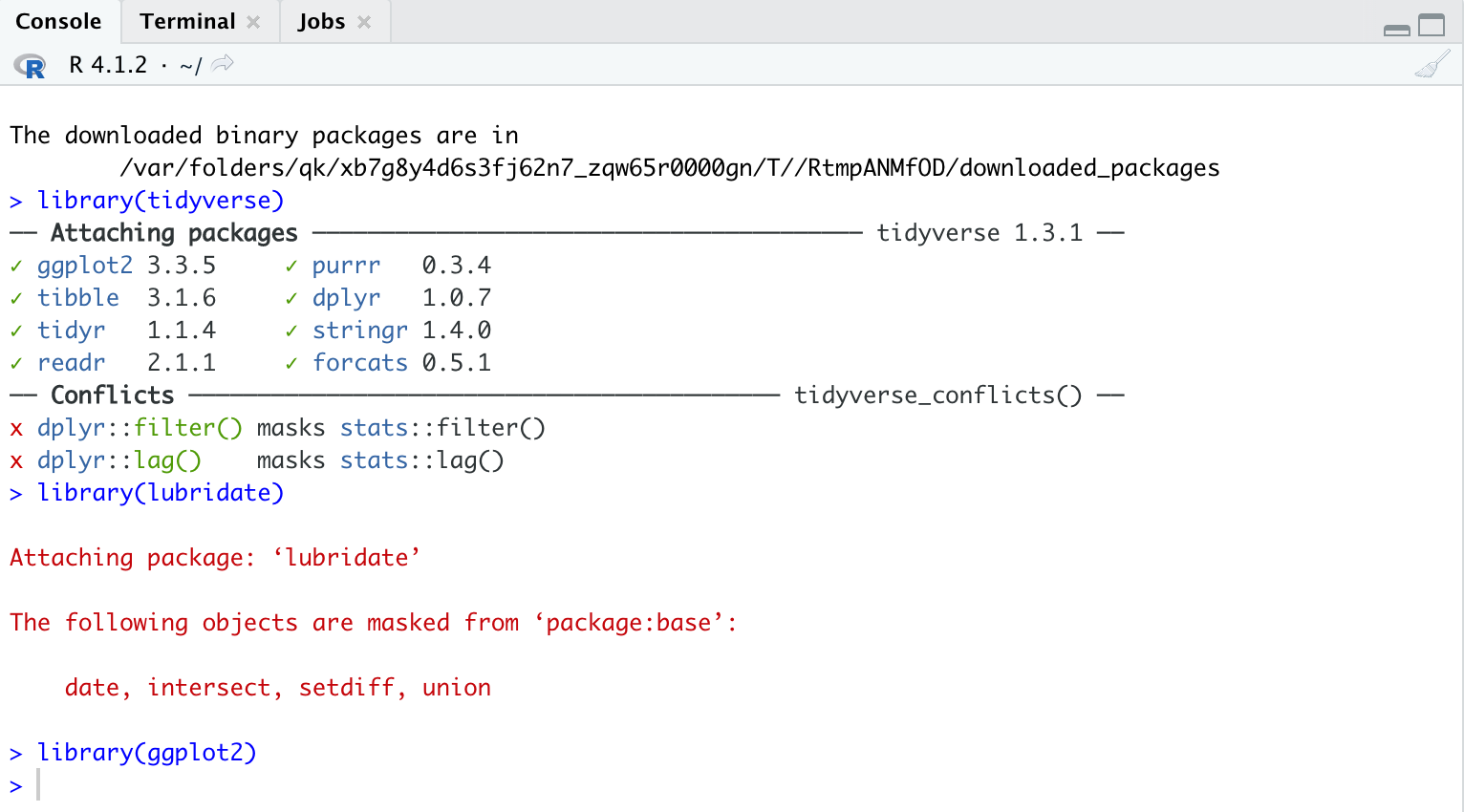
* Deleted all blank cells in the spreadsheet
* Created a new column - **day\_of\_week** - which contains the numerical representation of the weekday.
* Created a new column - **ride length** - which contains the difference between start time and end time in minutes. In the new column, some rows contained **negative values**. Such errors happened because the ending time was earlier than the starting time in their respective rows. All such rows are deleted.

After the cleaning process in EXCEL, R is used for further cleaning and analysis.

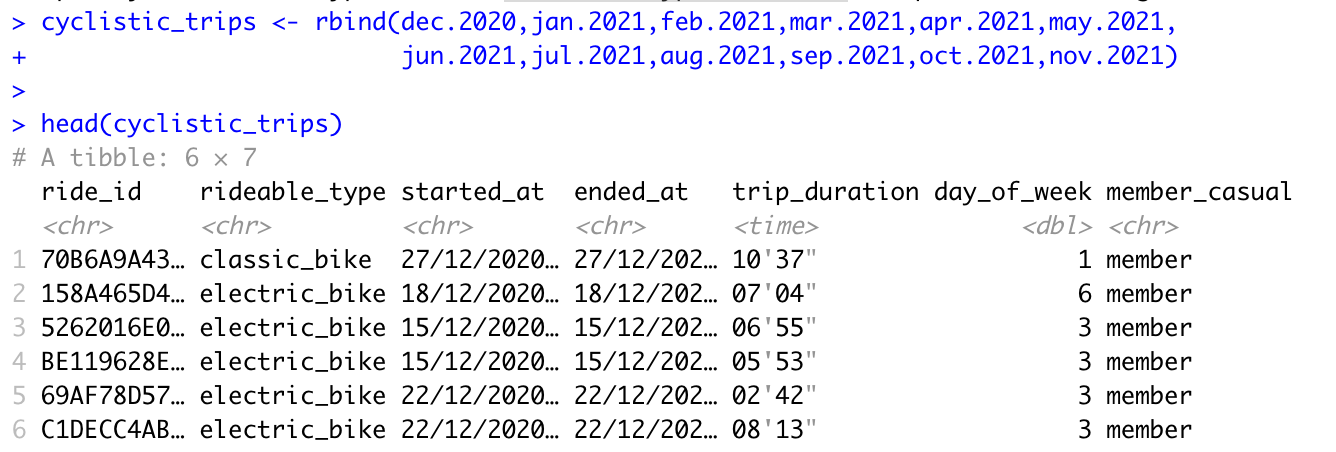
**Analyze**

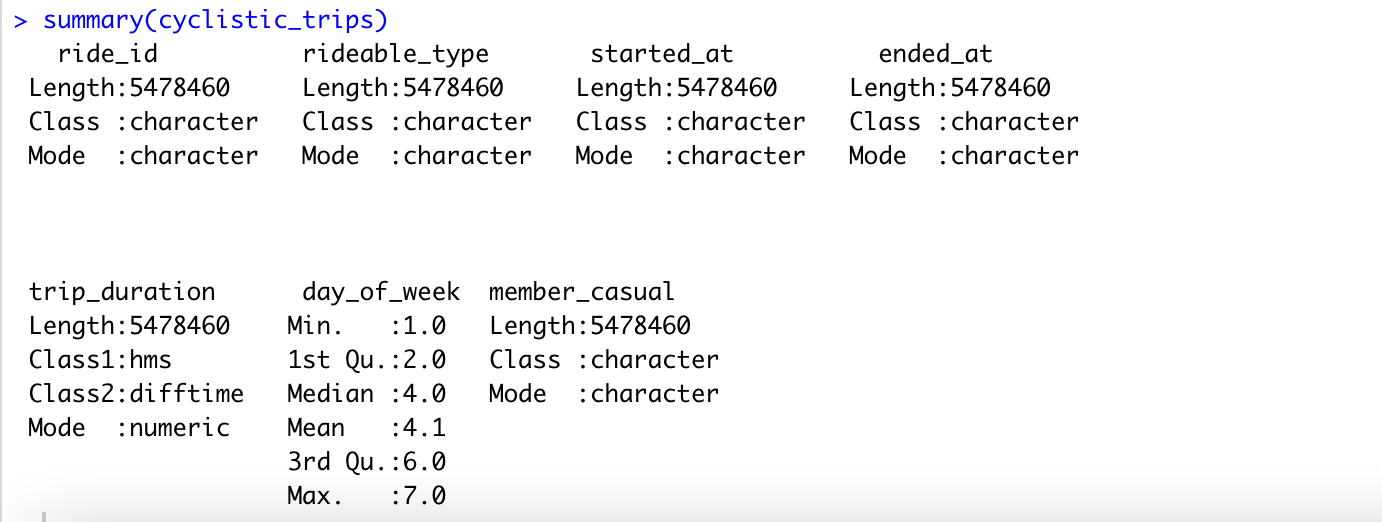
For the first step in R, the *tidyverse* and *lubridate* packages were loaded. The working directory set and data for the last 12 months downloaded using the read\_csv() command.

The data for each dataset was then verified to see that the columns were all in the same format.



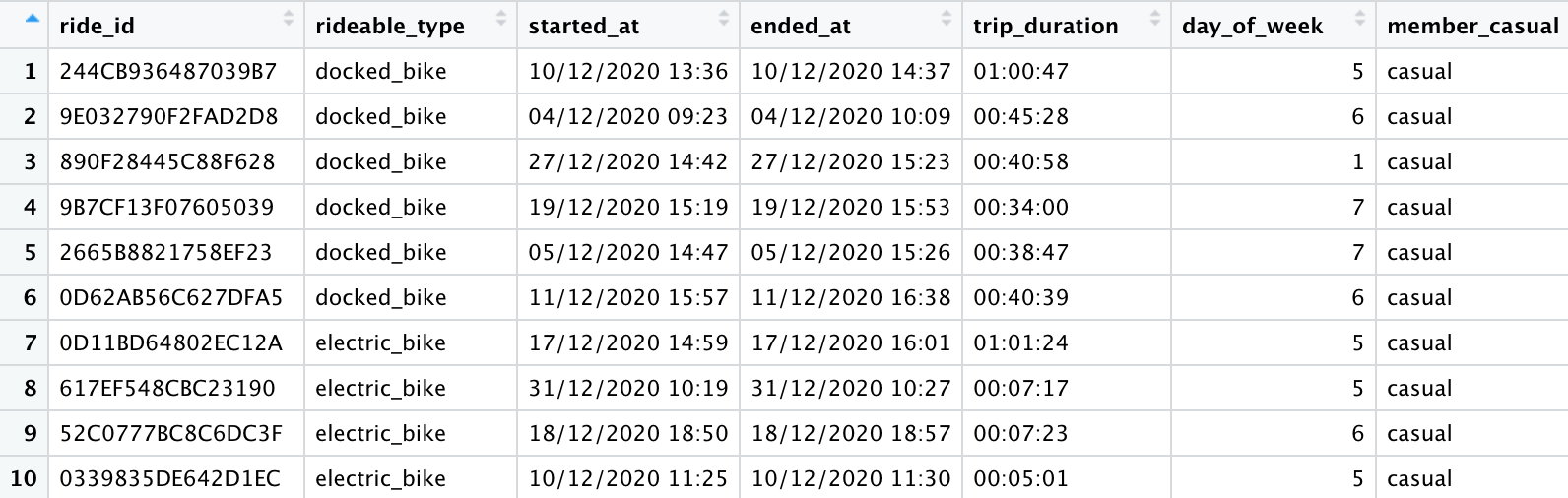
Binding all 12 files into one single file using the rbind function

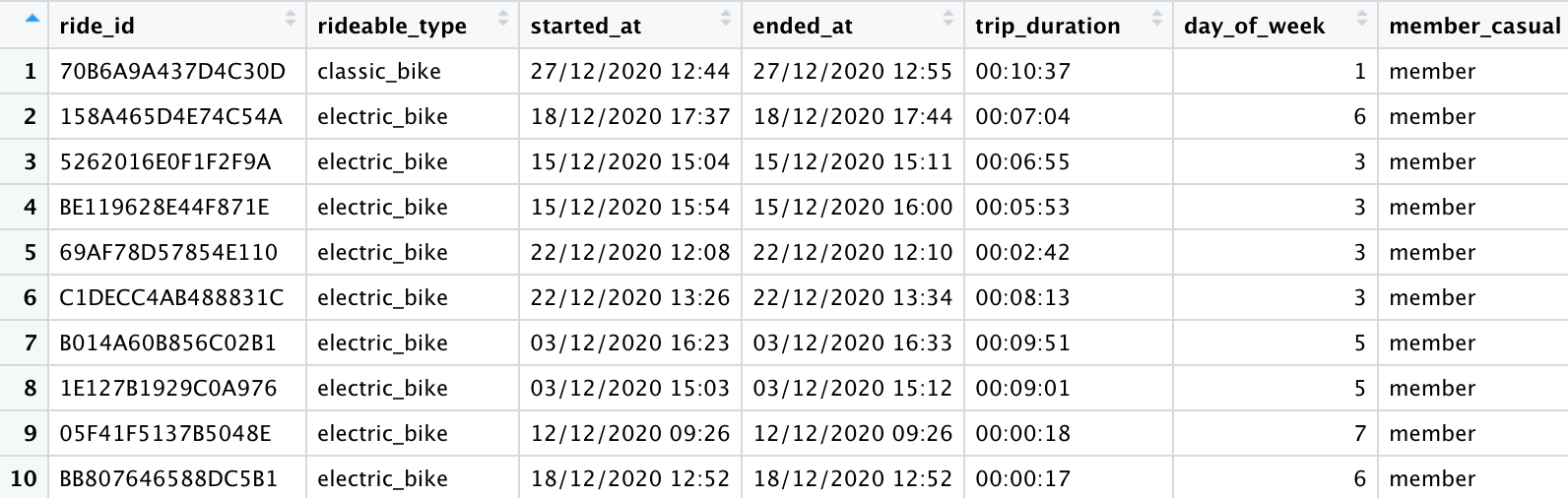




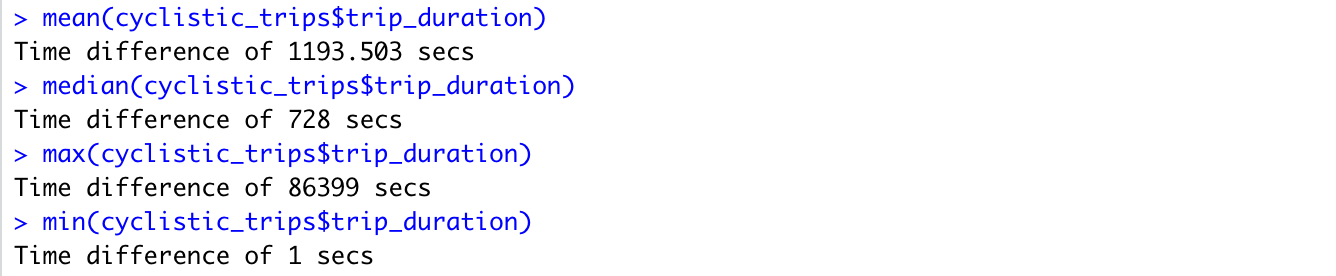
The table above shows the data was successfully bound as a single table containing 5,478,460 rows..

Next, the member and casuals are filtered into separate tables:

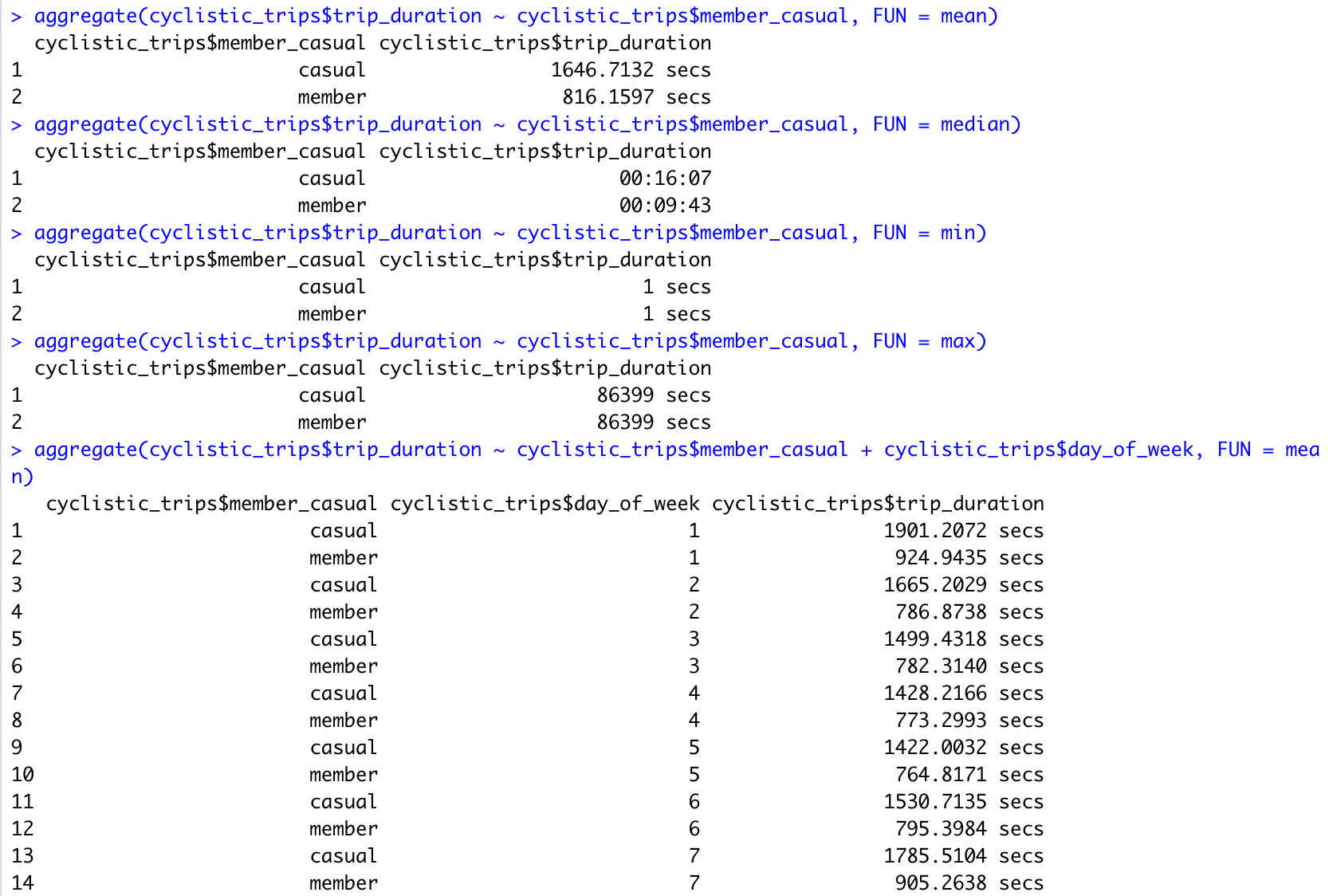




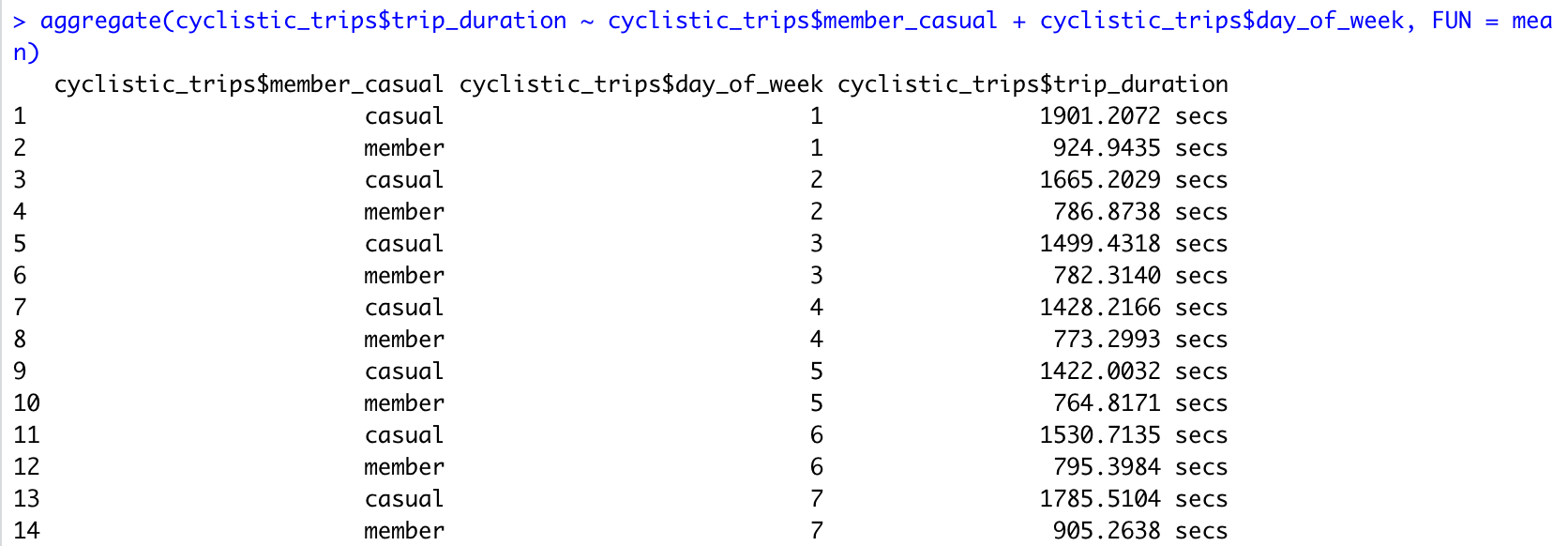
Some basic statistics of the entire cyclistics trips were calculated below:



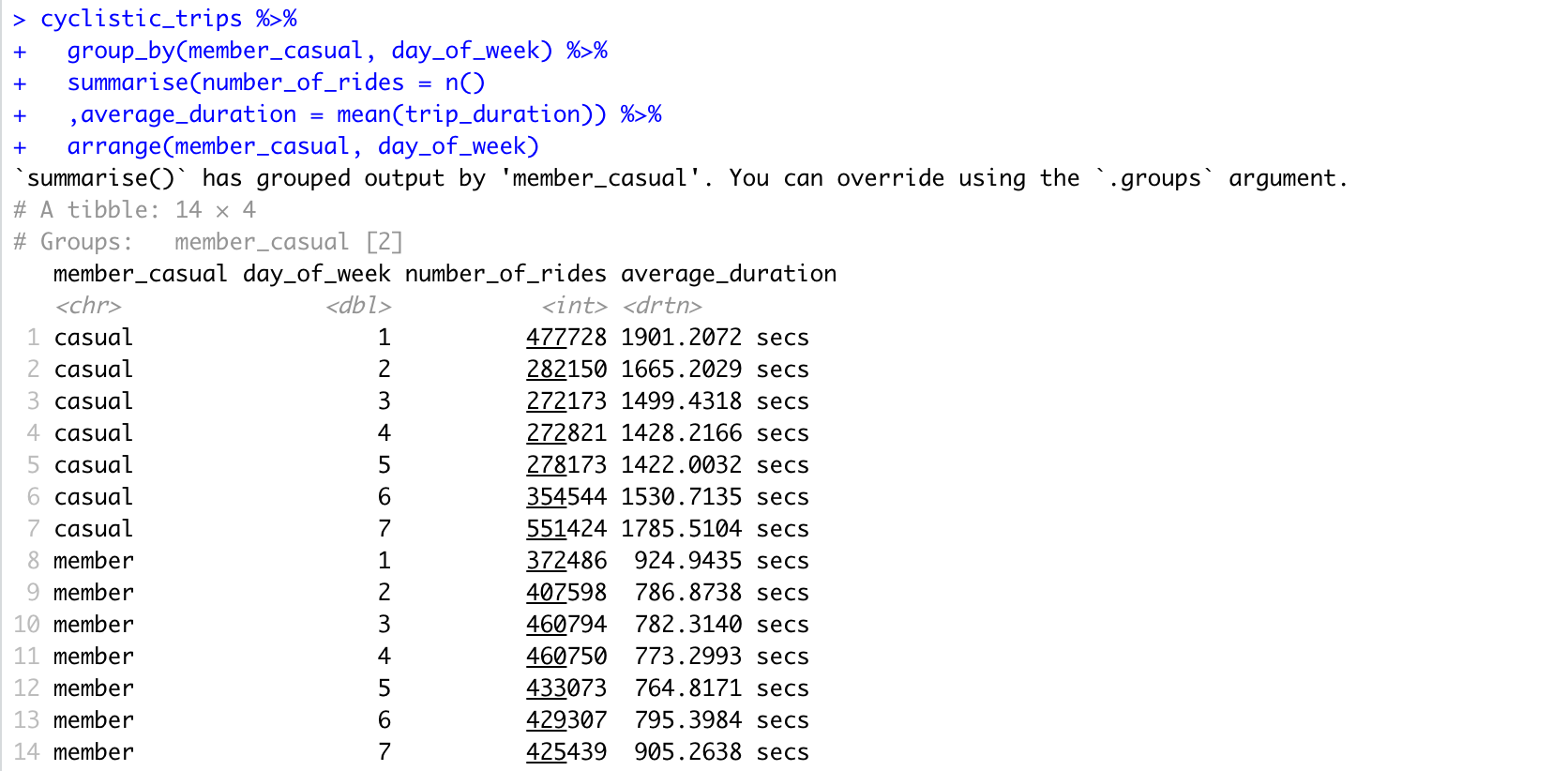
To do some descriptive analysis, we compare the annual members and casual users riding time with reference to mean, median, max and min values.



The code below compares the average riding time for annual members against casual users for each day, **1 being Sunday and 7 Monday**

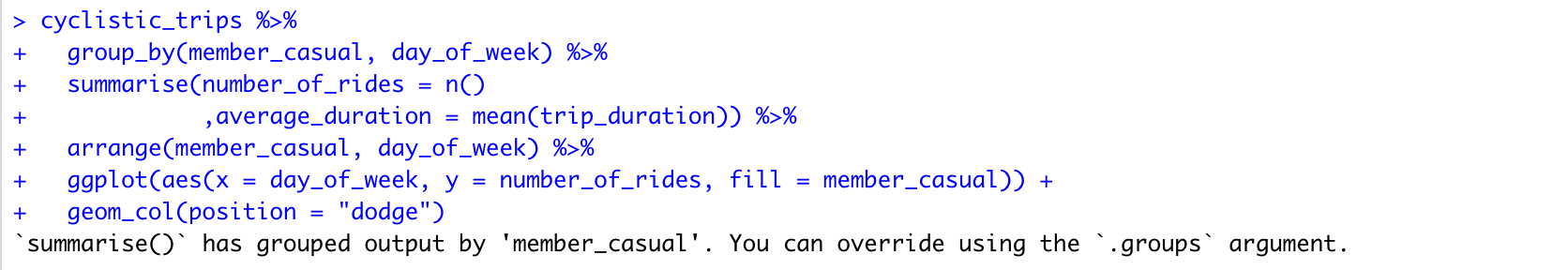


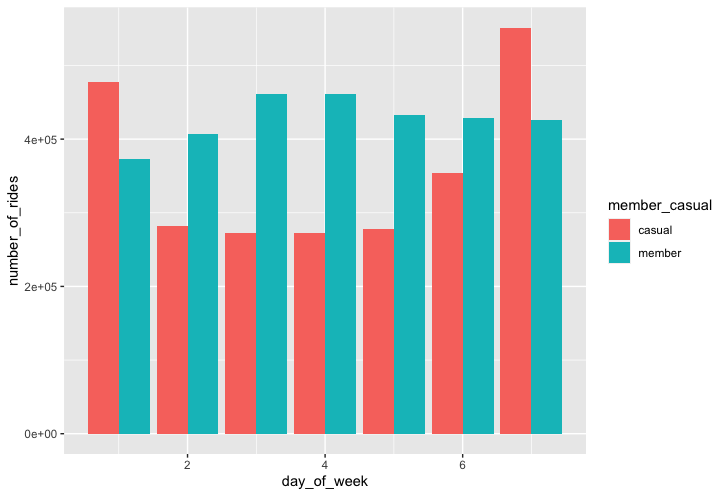
Next, we calculate the number of rides each weekday for the casual users and annual members



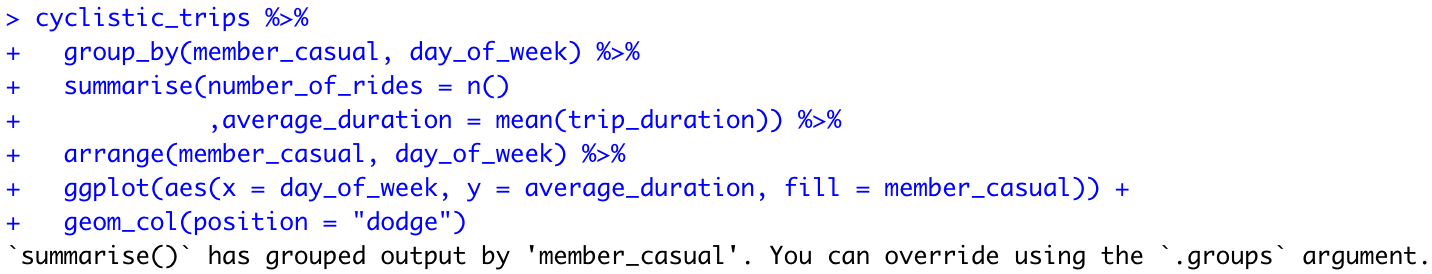
**Share**

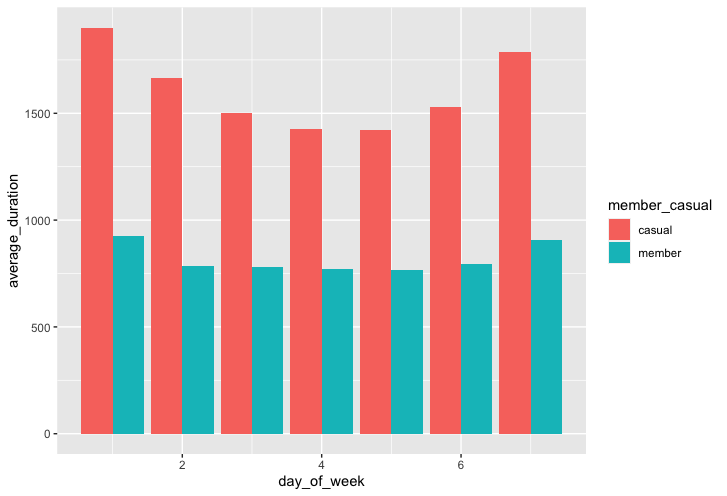
Next, we will show a visualization of the number of rides taken by annual members and casual users





Next we will show a visualization of the average ride duration in seconds by annual members and casual riders





**Act**

Based on the dataset analysed in this case study, the following are observations and recommendations for the Cyclistic marketing team.

**Observations**:

* From Monday to Friday, annual members took more trips than casual members. This implies that members mostly use the bikes for work commutes.
* The trips of members are shorter than those of casual riders.
* Over the weekend casual riders take more rides than members
* The average trip duration of casual riders is more than that of members
* Annual members mostly use their bikes for work commutes and their journeys are shorter than those of casual riders who ride for leisure. By promoting the benefits of riding to work and home we may convert more casual riders to annual members

**Recommendations**

* **Discount** the **membership fee** for casual riders who ride more than three(3) times a week.
* Introduce a **weekend only membership** package.
* Create a **loyalty program** for casual riders with an average duration of 1500 seconds which converts them to members when their points reach an agreed benchmark.