Introduction

This is my Capstone Case Study Project of Google Data Analytics Professional Certificate, where I perform real-world tasks of a junior data analyst at a fictional company, Cyclistic. In order to answer the key business questions, I will follow the steps of the data analysis process that I previous learned: Ask, Prepare, Process, Analyse, Share, and Act and using technical tools **(R and Tableau)** in each stage to ultimately gather insights from large amounts of data.

Background

In this case study, I am the Junior Data Analyst at Cyclistic, a bike-share company based in Chicago. Cyclistic is a bike-share program that features more than 5,800 bicycles and 600 docking stations. around Chicago. They have two customer segments: **Casuals and Members**. **Casuals** are classified as customers who either buy a single-trip pass or a day pass, whereas **Members** are classified as customers that buy annual memberships.

Ask

The finance department at Cyclistic has determined that Members are the more profitable customer segment and have tasked the marketing department with creating a campaign aimed at converting Casuals to Annual Members. To help them complete this business task, the marketing team has assigned me with answering the following question **“How do annual members and casual riders use Cyclistic bikes differently?”**

Prepare

To answer this question I will be analysing historical Cyclistic bike trip data for all 12 months from February 2023 to February 2024. 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(link). For the purpose of this project, I have saved the 12 relevant CSV files in my local drive.

The data collection team at Cyclistic have outlined some key facts and constraints about the data:

- All personal customer information has been removed for privacy issues.
- The data should have no trips shorter than 1 minute or longer than one day. Any data that does not fit these restraints should be removed as it is a maintenance trip carried out by the Cyclistic team.

Process

To combine and clean the data I used **R** on **Visual Studio Code** application. Below is an outline of my process:

Data Combination

- Created a folder called ‘CaseStudy_Cyclistic’ and imported the 12 CSV files as 12 separate monthly tables in the dataset.
- Compare columns names of all tables before, stack individual data frames into one big data frame, containing all bike trip data from 02/2023 to 02/2024.

Cleaning the Data

My data cleaning code can be viewed **here** on GitHub.

A summary of the cleaning steps I took:

- Created day of the week and ride time length columns to bolster analysis.
- Removed geographical columns: *start_lat, start_lng, end_lat, end_lng.* Stations name are provided.
- Created a new version of dataframe as I removed trips where the ride time length was less/equal than 1 minute or greater/equal to 1 day, and trips that were duplicated.
- In total, I removed 2,891,164 rows to be left with a clean combined table with 3,006,449 rows.

Analyse/Share

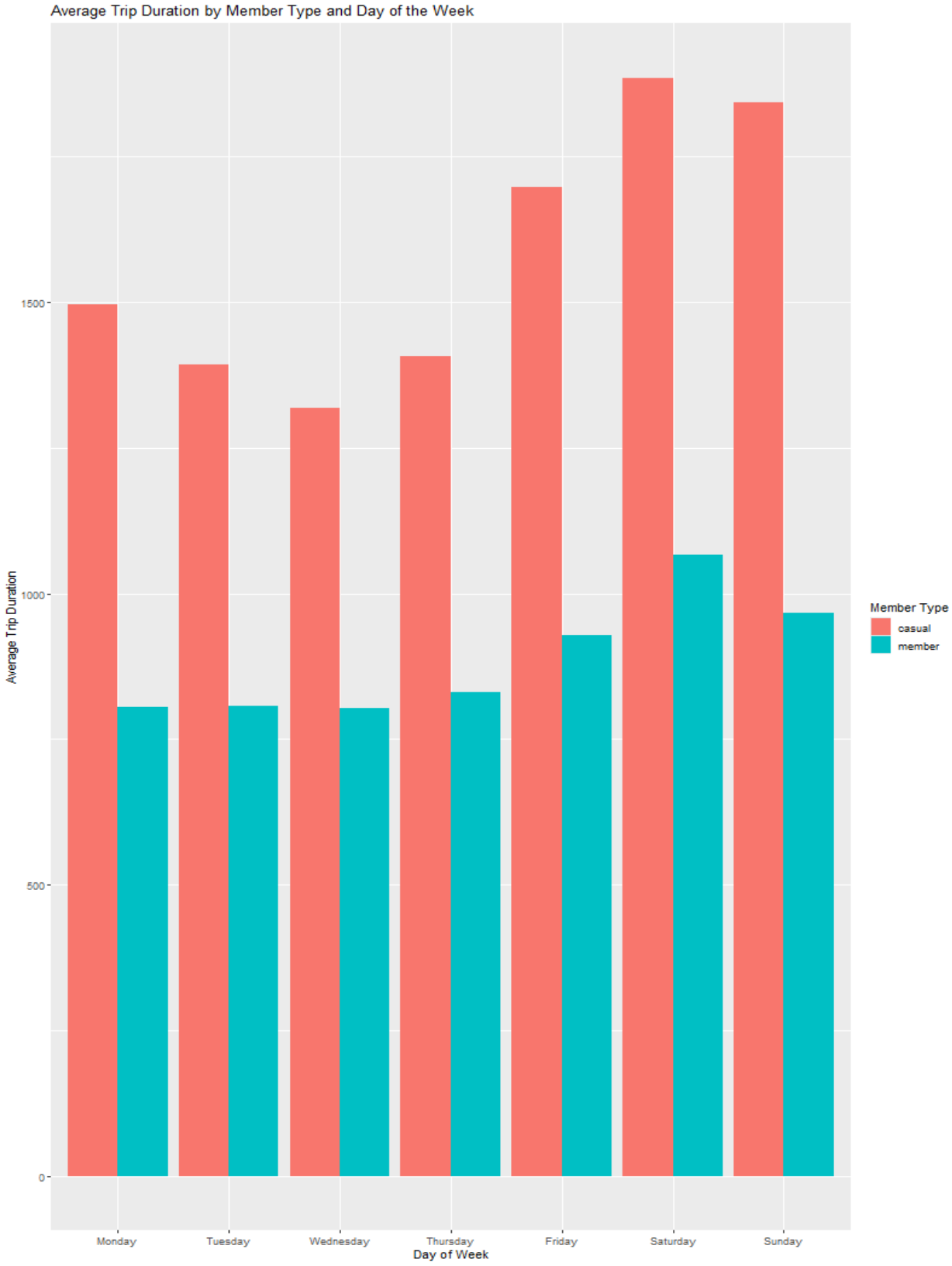
To analyse my data I used **R** to sort, filter, aggregate and visualize my data before importing it into **Tableau** for further analysis. My full R analysis code can be viewed here and two different versions of my Tableau visualizations can be found here(v1 and v2).

In **R**:

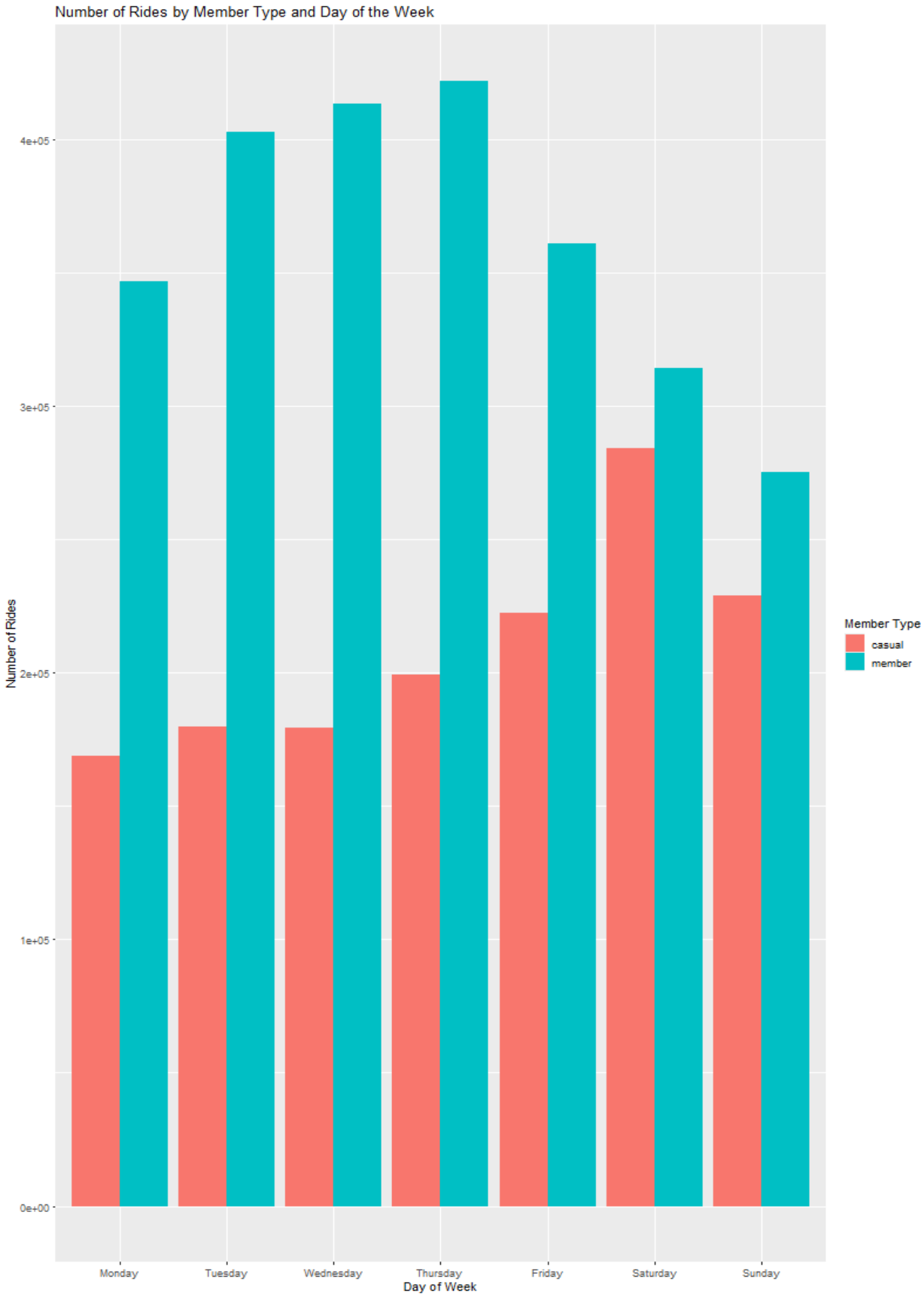
- The average of the ride's length in between casual members and annual members by day of the week:

```
# Groups:   member_casual [2]
   member_casual day_of_week number_rides average_duration
   <chr>         <ord>              <int>            <dbl>
 1 casual        Monday            168719            1497.
 2 casual        Tuesday           179645            1393.
 3 casual        Wednesday         179219            1318.
 4 casual        Thursday          199200            1407.
 5 casual        Friday            222448            1697.
 6 casual        Saturday          284263            1885.
 7 casual        Sunday            229047            1843.
 8 member        Monday            346786             805.
 9 member        Tuesday           402901             807.
10 member        Wednesday         413271             803.
11 member        Thursday          421775             830.
12 member        Friday            360800             928.
13 member        Saturday          314385            1066.
14 member        Sunday            275207             966.
> 
```

- The average trip duration by Member type and Day of the week:

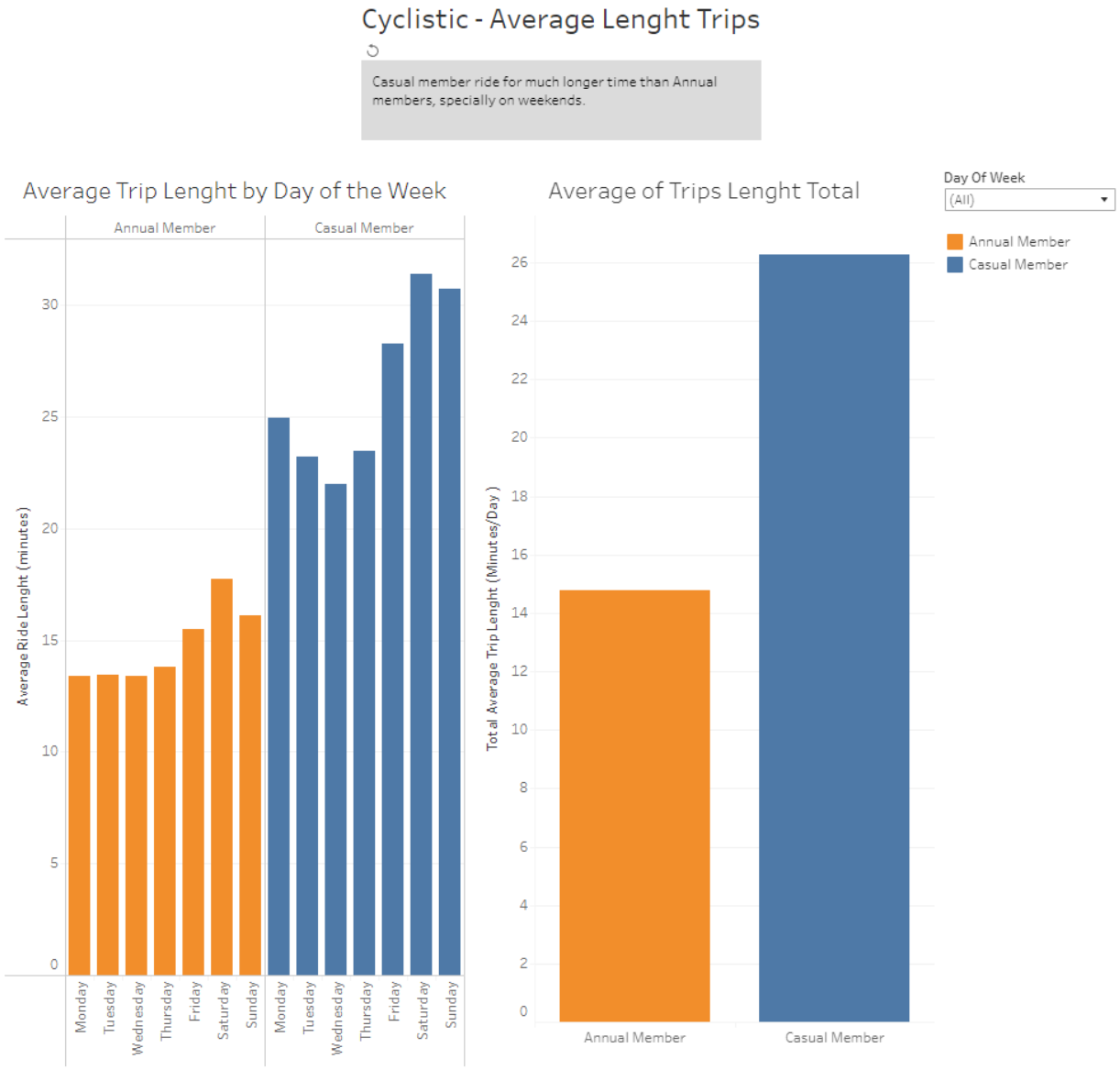


- The number of rides on each day of the week by each type of member:

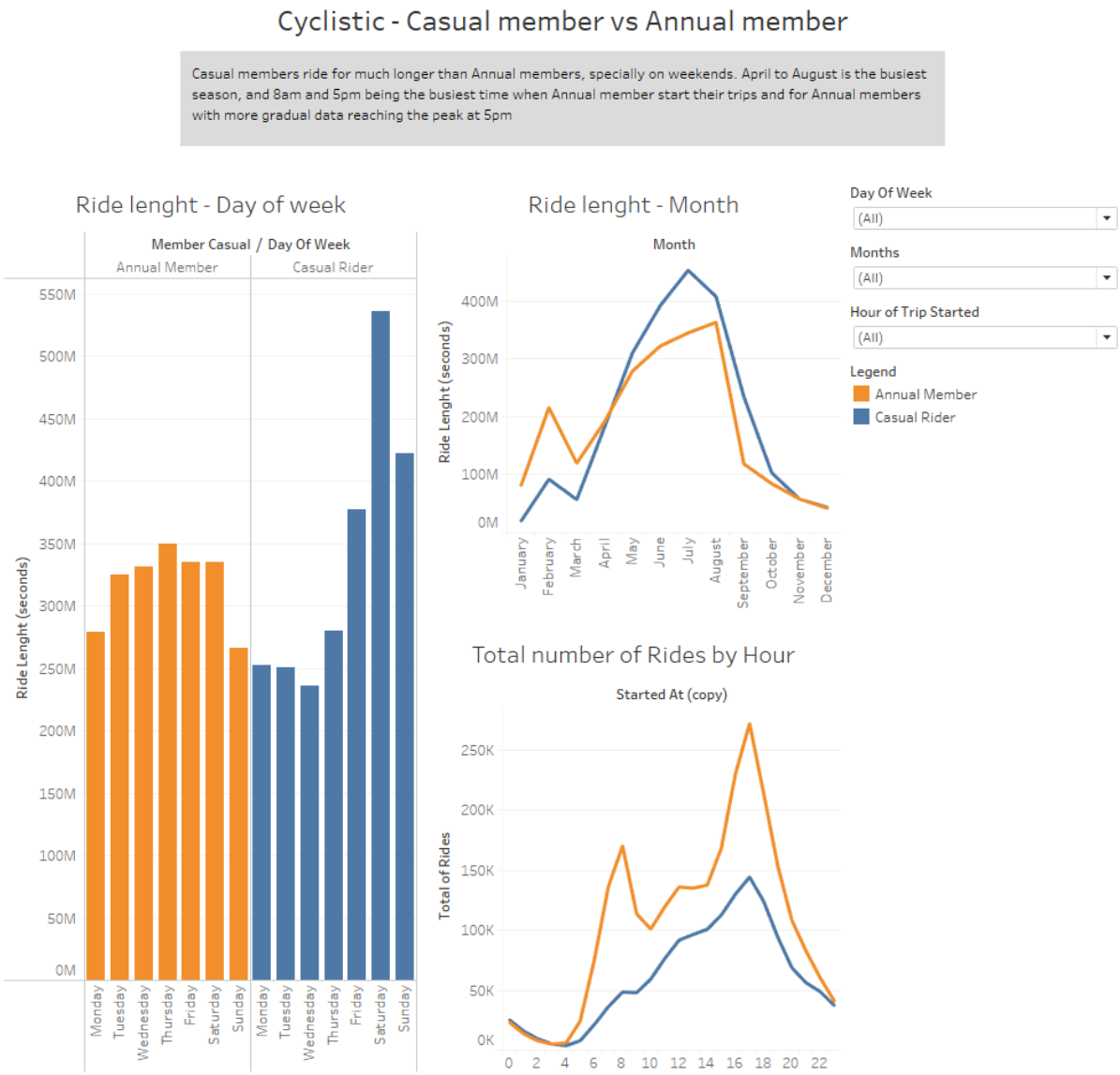


In **Tableau**:

- Version 1 - Average Trip Length in bar charts:



- Version 2 – For deeper analysis I created a second Tableau story which contains: Ride length by month, day of week and number of trips started per day by hour.



Act

I can conclude from my analysis that we can see that the biggest differences in between Annual members and Casual member are:

- Casual members tend to ride longer, almost 2 times more per trip (average annual members = 14.77min and casual members = 26.28min).
- Casual members ride with more frequency on weekends while annual members use it more often on commute hours during the weekdays.

As suggestions in how to convert more Casual members to Annual members I present the following:

1. Offer new riding rates on annual members plan that includes discounts on weekends.
2. New membership plans, of 6 months and 3 months. Where the Annual membership is the best value for price. Including bigger discounts for winter season, as the usage drops drastically compared to summer season.
3. Invest in marketing, with limited time discount deals, announcing them on strategic times of the day, when there is more traffic of riders.