Background

Cyclistic future success depends on its ability to change casual members who are riders who purchase a single ride pass or a full day pass to an annual membership which has been determined to be more profitable. In light of this I will use data insights on how cyclistic customers that is, casual and annual members differ to inform marketing strategy decisions targeted towards causing casual members to convert to an annual membership.

Data Sources

The data used in this analysis is Cyclistic’s previous 12 months of data starting from June 2022 to May 2023. The data was acquired from Bikeshare, the original source and contains information dates, times, locations and the types of bikes for both casual and annual members which is critical information that will assist in finding the differences between casual and annual members.

No bias issues were observed as the data is collected electronically using logs thus leaving little room for observer, interpretation and confirmation bias and there was no sampling bias as the data contains information on all rides that occurred monthly. However, there was no assurance made by bikeshare that the data is completely accurate.

Data Cleaning Process

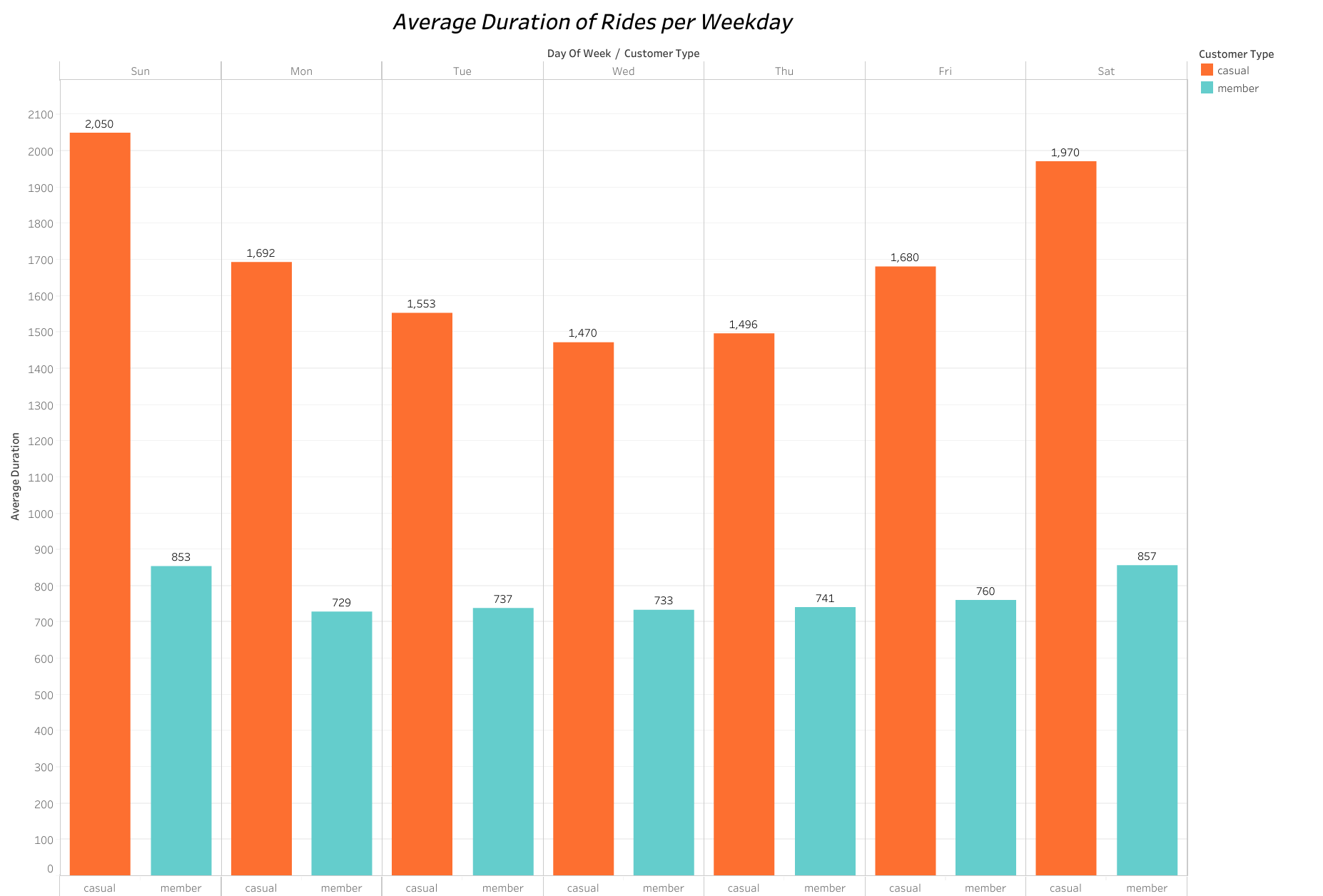
I made use of Rstudio to clean data as I was not able to import data into BigQuery Sandbox and also because of the excel limitation in number of rows which made it impossible to combine all 12 months of data into one worksheet.

1. Imported the 12 csv files into R Studio for each from May 2022 to April 2023.
2. Combined the 12 data frames into one data frame.
3. Renamed the column member casual to customer type.
4. Created new column ‘date’ which derives only the date from the datetime column ‘started at’ with the format YYYY-MM-DD.
5. Created a new column ‘month’ which gives the name of the month when the ride started.
6. Created a new column ‘day’ which indicates the day in which the ride started for instance ‘01’ for the first day of the month.
7. Created a new column ‘year’ which indicates the year in which the ride started with format ‘YYYY’.
8. Created a new column ‘day of week’ which indicates the weekday on which the ride began for instance ‘Sun’ for Sunday.
9. Created a new column ‘start time’ which indicates the time of start for the ride in the format ‘HH:MM’.
10. Added a new column ‘ride length’ which indicates the duration of each ride by finding the difference in time between the columns ‘started at’ and ‘ended at’ given in seconds.
11. Converted the ‘ride length’ column to numeric in order to perform calculations on it.
12. Removed rows where ride length is less than 0 which is erroneous and less than 60 seconds that were potentially false starts.

Analysis

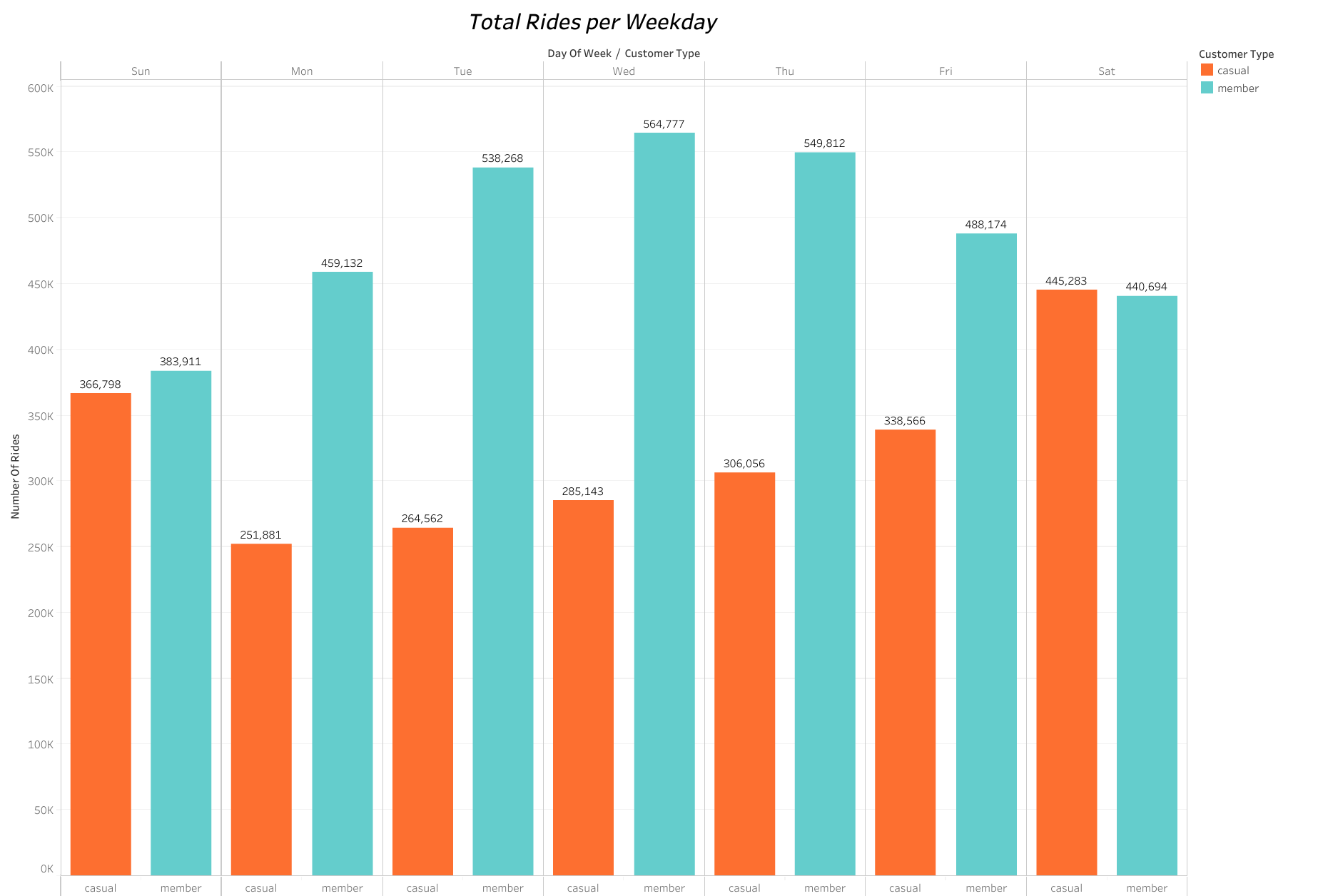
I performed analysis using R and made use of Tableau for the visualisations. Upon performing preliminary analysis, I observed that casual users ride the bikes for significantly longer than members as casual user’s average ride length was 1732.2 seconds compared to that of members which is 767.7 seconds which is 945.5 seconds lower than that of casual riders. However, members have more frequent rides as they totaled rides of 3424768 rides in the last 12 months against the 2258289 total rides of casual riders which is 1166479 rides lower.

Days of the week



From the visualisation above we can note that both casual riders and members ride for longer during the weekends as Saturday and Sunday have the tallest bars. For casual riders after the weekend the rides progressively become shorter until Wednesday, the day with the shortest rides, after which they begin to increase again to peak on Sundays, the day with the longest rides. Members ride more consistently during the week with regards to time and there is a slight increase on weekends. Casual rider’s average duration has a standard deviation of 16772.477 whereas members have a standard deviation of 1771.795 which indicates that casual rides are significantly more variable throughout the week than members’ rides.

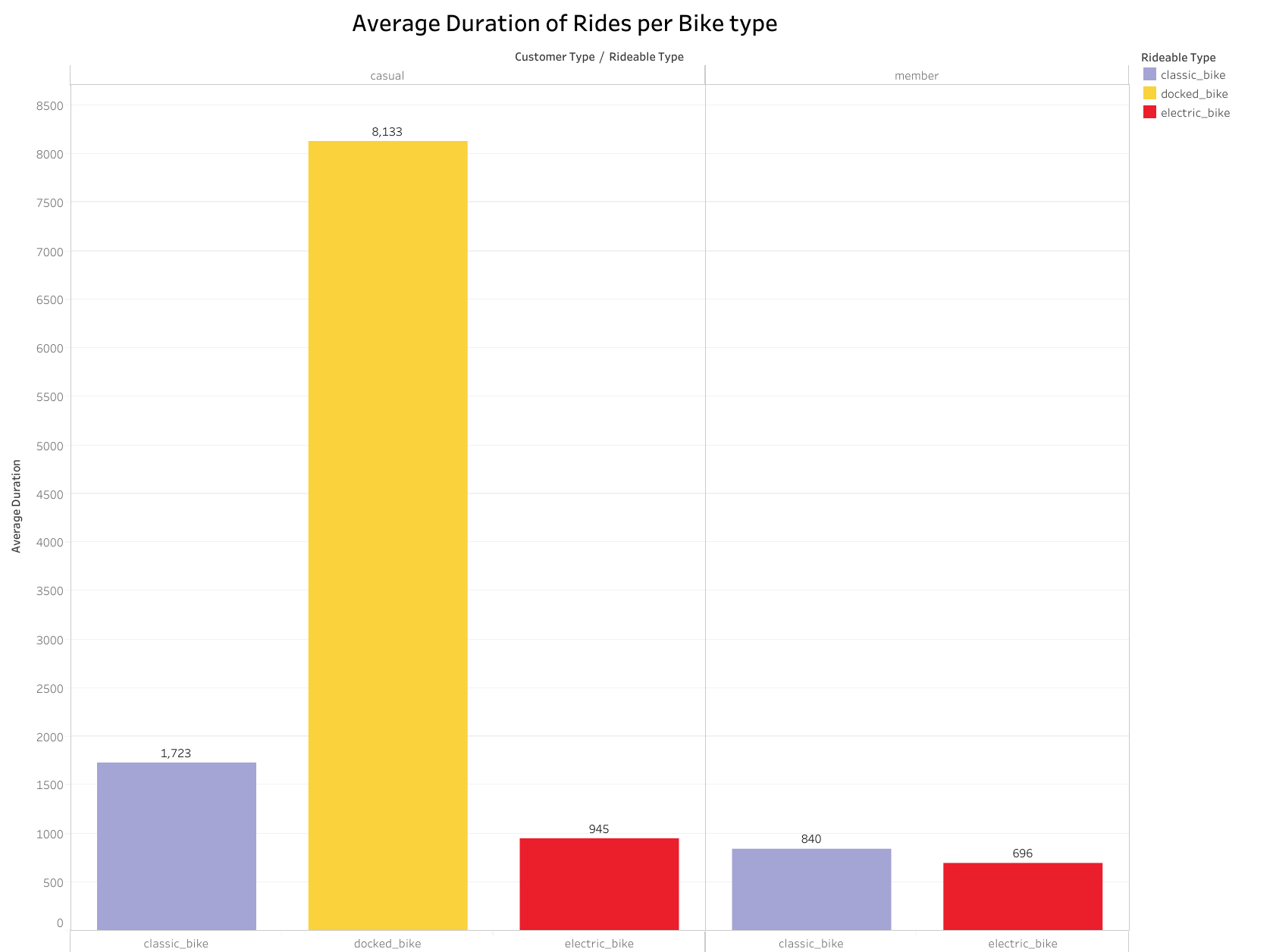
It can be inferred therefore that, members are more routine in the way they use the bikes which might be for daily activities like commuting to work or for fitness purposes as part of a routine workout. Casual riders seem to ride more during their free time that is, on weekends where most do not go to work.



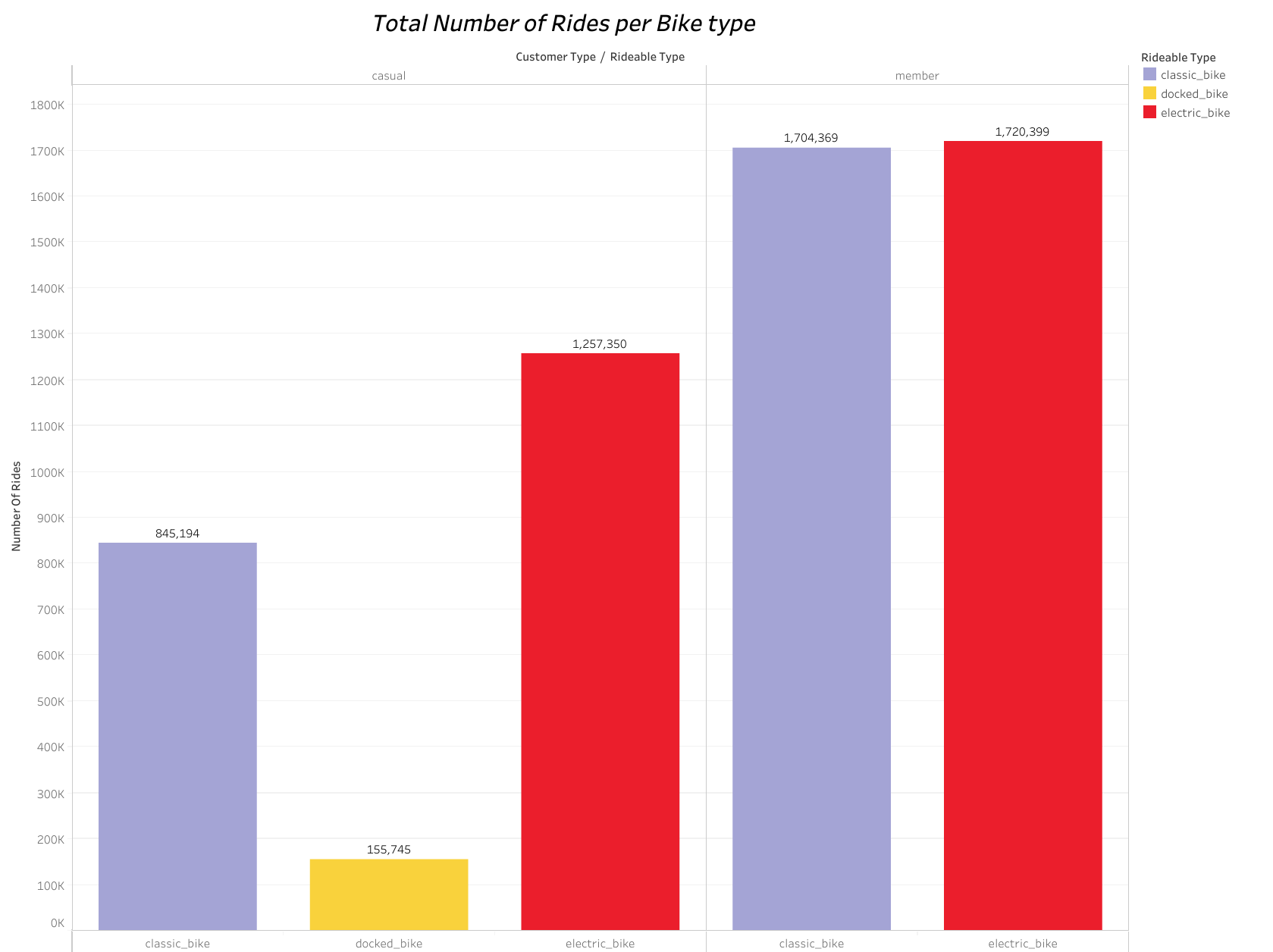
Casual riders ride less frequently as compared to annual members on most days of the week except on Saturdays, which is the day that has the most casual rides. It is also observed that casual riders ride the most on weekends. After the weekend the number of rides plummets on Monday which is the least ridden day for casual riders from then the rides gradually increase throughout the week to peak on Saturday. However, members gradually increase their rides from Sunday to peak on Wednesday and then gradually decreases to have the lowest rides on weekends. It is noted that we are unable to tell if riders rode more than once in a day.

Casual riders seem to ride more frequently on weekends when most people are not at work and this may be because they ride more for leisure purposes and many may not be using the bikes for commuting to work as rides are significantly less frequent during the week. Members ride more during the week possibly because they use them throughout the day for most transportation purposes.

Type of Bike

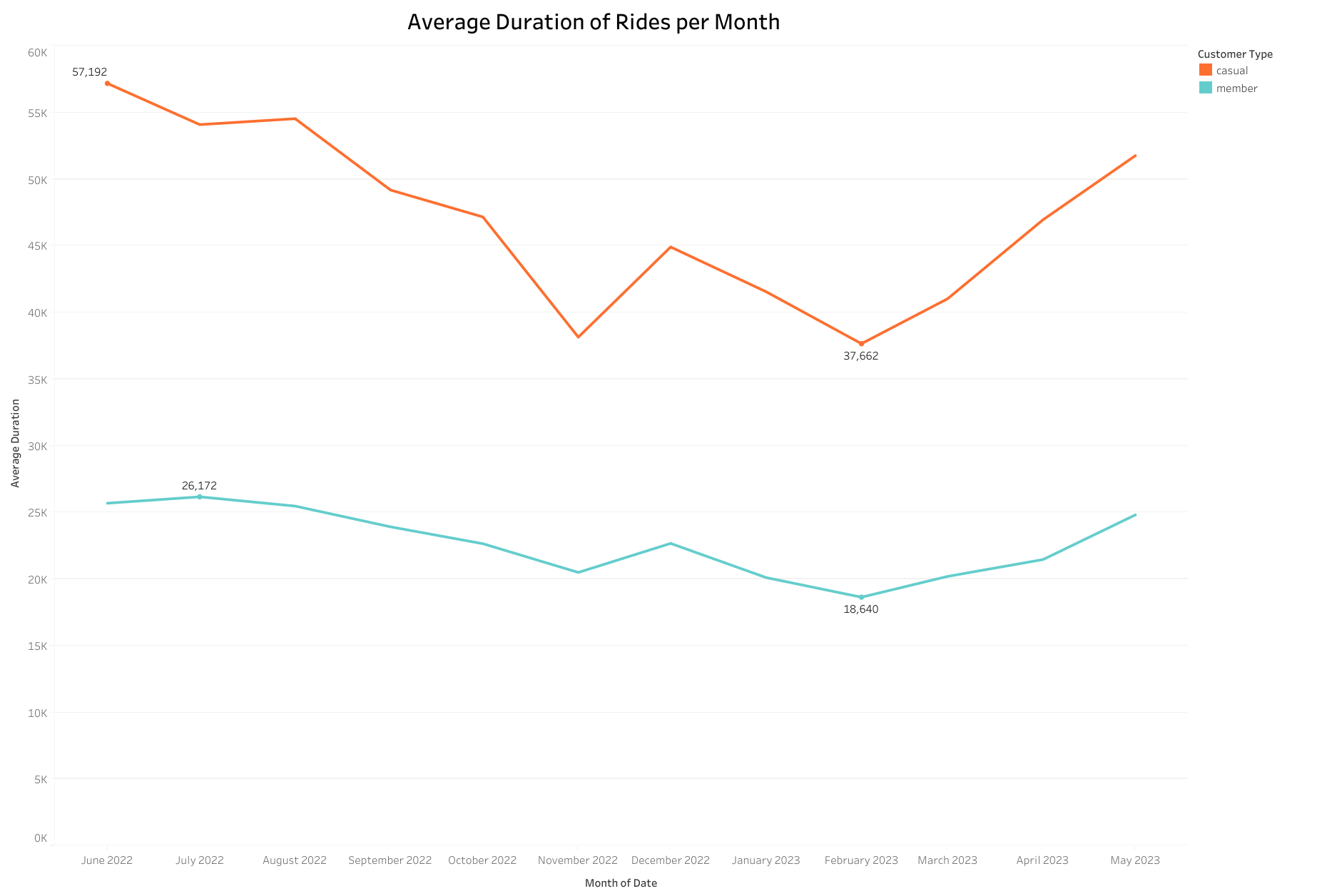


It is evident from the graph that the docked bike is ridden the longest between the three bike types and they are only ridden by casual riders. In addition, the classic bike is ridden for much longer than electric bikes by casual riders than by members, that is 45% more in comparison to only 17% more for members.



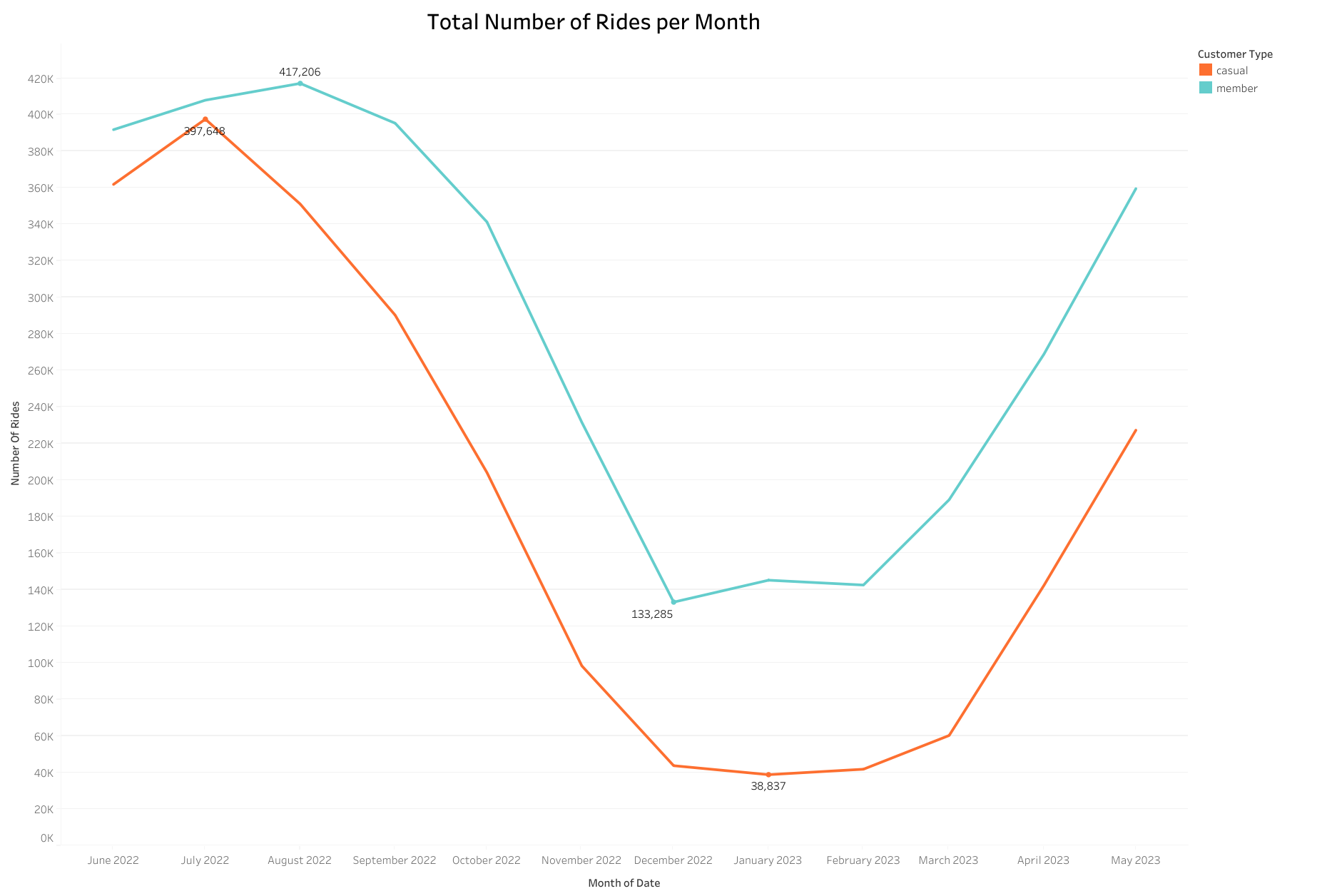
From the above visualisation, it is observed that casual riders ride the electric bike significantly more frequently than the classic bike in comparison to the members who are more consistent between the two bike types. Members do not ride the docked bikes at all but casual riders do.

Rides per Month



This is a visualisation showing average duration of rides per month for casual riders and members. Casual riders have their longest rides in the month of June after which there is a decline in the duration until it reaches November, there is an increase in duration of rides in December then the rides become shorter until February which is the month with the shortest ride duration. Members ride more consistently than casual riders as it is less variable with a standard deviation of 81.222 for members and 220.874 for casual riders.

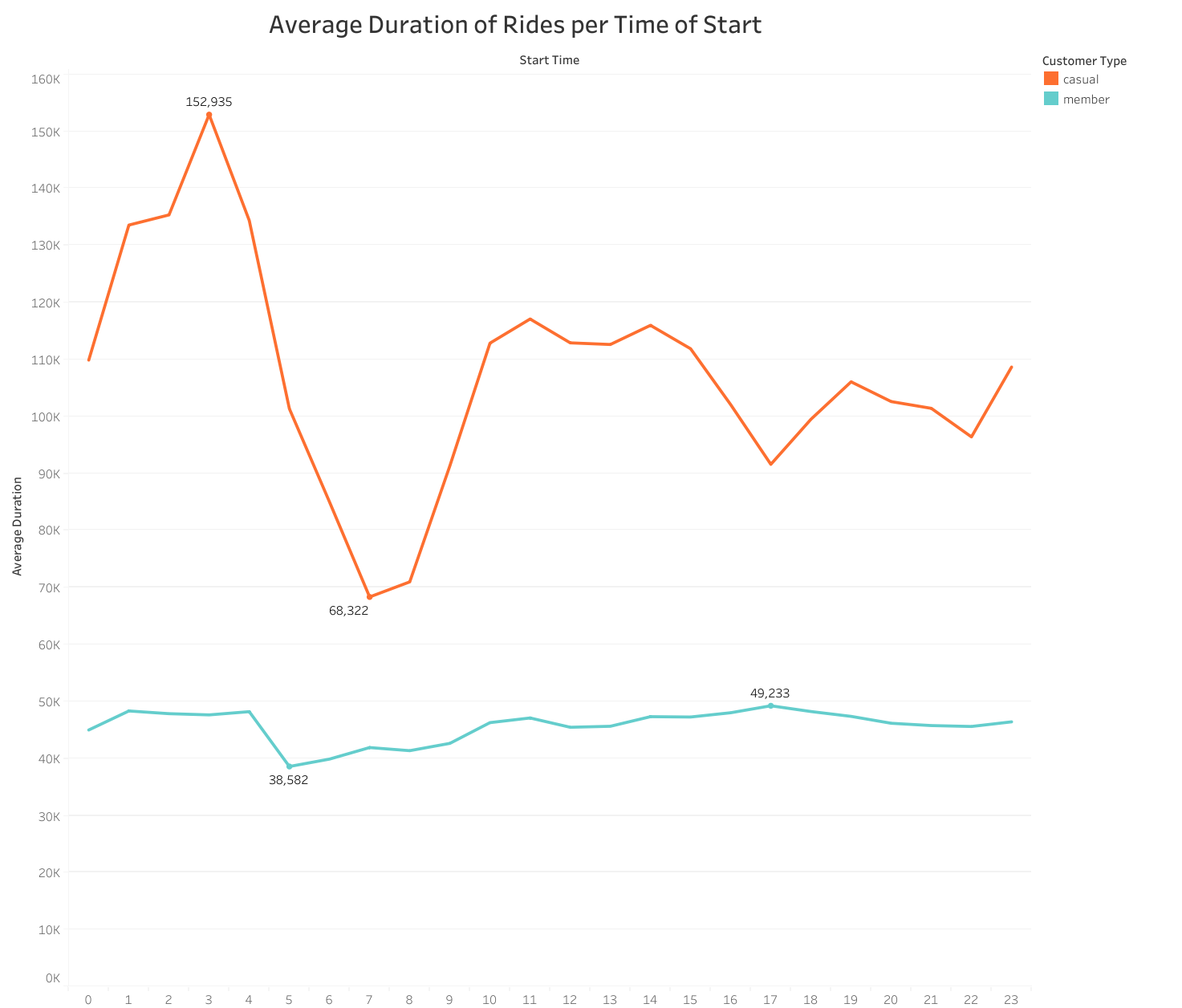
It is observed that duration of rides is affected by the season the month is in. In spring, that is between the months of March to May ride duration steadily increase for both casual riders and members with there being a greater decrease for casual riders. In summer, between the months of June to August is where we observe the longest rides for both casual rides and members. Ride durations begin to decrease in Autumn, from September to November and the shortest rides occur in winter for both casual riders and annual members.



At first glance there is a similar shape for both casual riders and members but upon closer inspection it is observed that casual riders ride the most in the month of July whereas members ride the most in August. Casual rides drop until they reach January which has the least rides after which ride frequency rises. Members rides drop until they reach December when they ride the cyclistic bikes the least.

The changes detected during the months of the year can be attributed to changes in seasons. There most rides are in summer, least rides are in winter, rides steadily decline in autumn and in spring there is a rise in ride frequency for both groups. Casual rides significantly lower in winter in comparison to member’s rides. Most casual riders do not ride when it is cold.

Time of Start

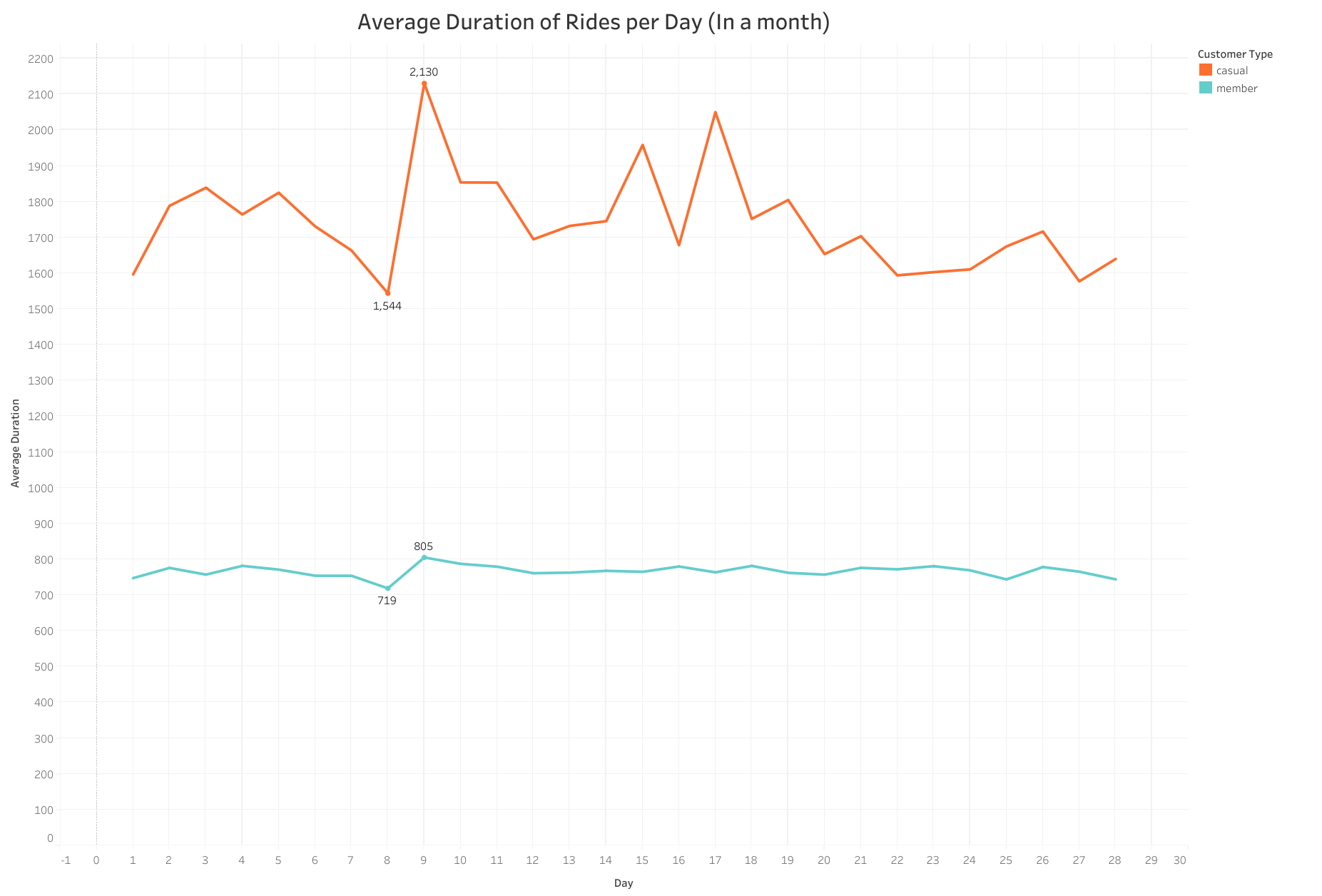


As seen from above members are more consistent with the duration of the rides throughout the day. Casual riders start their longest rides at 0300hrs after which there is a sharp decline in lengths of the rides until 0700HRS when the shortest rides start. Rides begin to increase again in duration until 1100HRS. Members have their longest rides on average at 1700HRS and their shortest at 0500HRS.

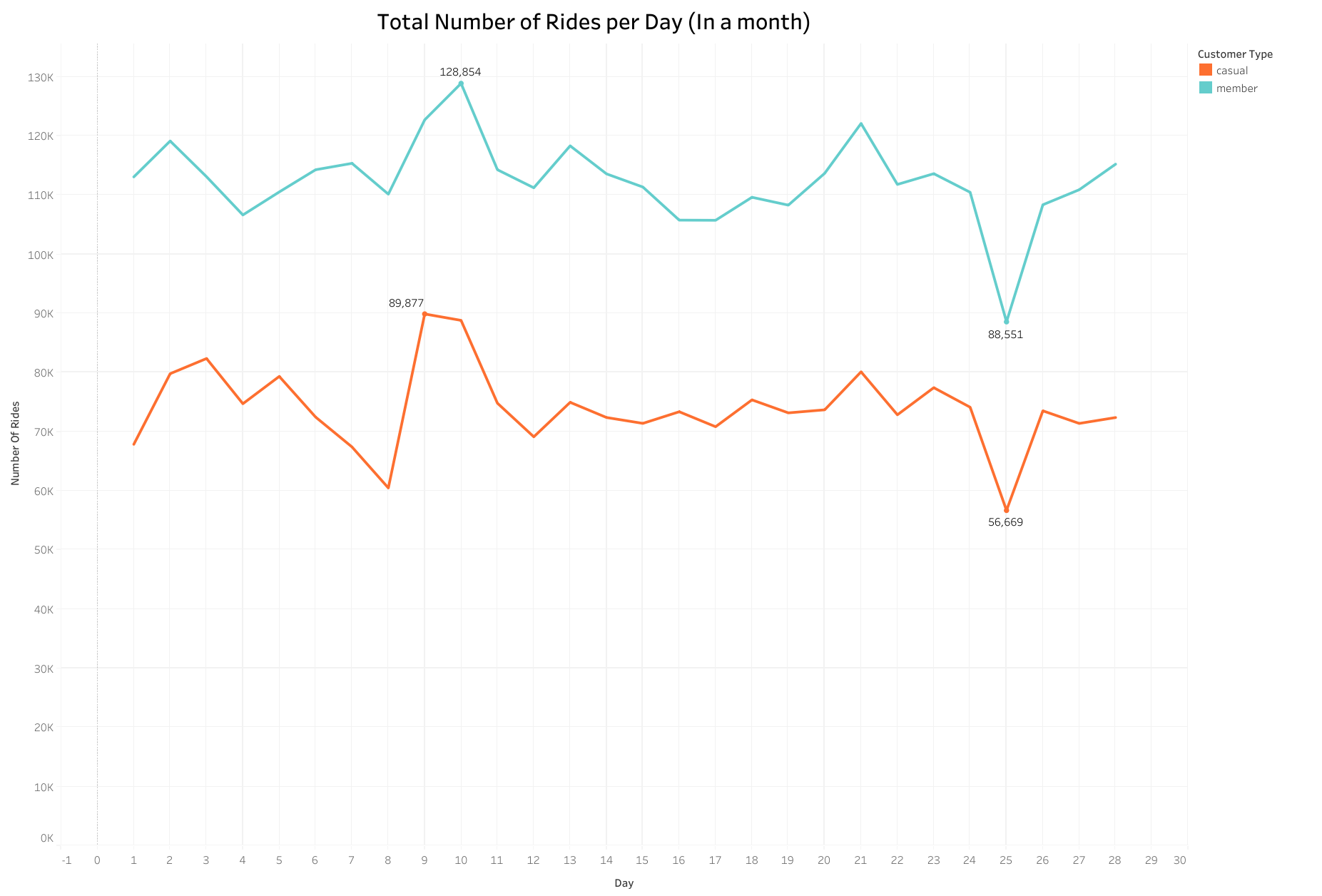


Most rides for members and casual riders start at 1700HRS and throughout the day members ride more frequently than members except between 0000HRS and 0300HRS. For casual riders after 0300HRS there is a gradual increase in number of rides per hour of start which peaks at 1700HRS after which there is a sharp decline until 2300HRS. During the morning hours members peak at 0700HRS, the ride frequency drops between 0800HRS and 0900HRS after which there is a gradual increase which sharpens between 1400HRS and 1700HRS then the rides drop.

Rides per Day in Month



The above visualisation explores the average duration of rides starting on each day of the month. The notable difference we already see is members have more consistent ride lengths throughout the month as casual riders have more variation. The longest casual rides are started on the 9th of the month and the shortest on the 8th for both types of riders.



Casual riders and members are similar in the daily number of rides consistency with members riding more frequently throughout the month. The only notable differences are casual riders ride less frequently than members and casual rides peak on the 9th of the month whereas member’s rides peak on the 10th.

Recommendations

* Membership for those who take longer rides, casual riders may not be encouraged to sign up for a membership as they may have fewer longer rides.
* Discounted or free trial period
* Winter sign up promotion, casual rides significantly drop in winter thus proving winter biking gear on sign up may encourage casual riders to convert to an annual membership.
* Introduce bikes for times when there is snow exclusive to annual members.