Background

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Customers who purchase single-ride or full-day passes are known as casual riders. Customers who purchase annual memberships are members.

Scenario

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Ask Phase

Objective: Increase membership and revenue by converting causal rider into annual members.

Task: How casual riders and annual members use Cyclistic bikes differently?

Stakeholders: Lily Moreno, Cyclistic marketing analytics team and Cyclistic executive team.

Prepare Phase

Data located: The Cyclistic's historical trip data is available online divvy-tripdata. The data has been gathered by Motivate International Inc.

Data organize: The data is available in csv file format. It contains latest Cyclistic trip data of 12 months from Jan to Dec 2022. It contains attributes ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng and member_casual.

Data Credibility/bias: The data is reliable and directly from the source. Data has all the important information which are required for this project. It contains the latest data of 2022.

Licensing, Privacy, Security, and Accessibility: The data has been made available to the public by Motivate International Inc. under the Data License Agreement. The data is privacy is protected as all the Cyclistic trip data doesn't contain name, phone number or address of the bike users.

Data answering the question: The data helps to answer the question as the dataset contain the information of both member types causal riders and member riders .Causal riders are the one who purchase single-ride or full-day passes and Member riders are the one who purchase annually memberships.

Problems with data: There are problems with the dataset it contain dirty data like inconsistent data, incorrect data and incomplete data. There was no duplicate data.

Process Phase

What tools are you choosing and why?

I choose three tools Excel, R and Tableau. It is because I didn't wanted to limit myself to one tool instead I wanted to do data cleaning using excel, making new columns/checking tables using r and visualization/analysis using tableau.

Microsoft Excel: I download the divvy-tripdata 12 months dataset From Jan 2022 till Dec 2022. I thought to combine all the 12 months of data into one but I thought it would take lots of time and CPU power to do so I decided to clean my data one by one. I used the excel built in function filter to remove all the empty rows and columns after that I used the remove duplicate function to remove the duplicates if any.

After cleaning the data I added two new column in all the 12 months dataset files. The columns I added was ride_length and day_of_week.

- ride_length: Calculated by using if statement on started_at column and ended_at column.

```
=IF(D2>C2, D2-C2, 1-C2+D2)
```

  After that Format of column was set to HH:MM:SS.

- <u>day_of_week</u> : Calculated by using WEEKDAY command on started_at column.

```
=WEEKDAY(C2,1))
```

R

I loaded all the required libraries

```
library("tidyverse")
library("lubridate")
library("hms")
library("hydroTSM")
library("data.table")
```

I loaded 12 months data into r using read.csv and made separate data frames for each month.

```
jan01_df <- read_csv("202201-divvy-tripdata.csv")
feb02_df <- read_csv("202202-divvy-tripdata.csv")
mar03_df <- read_csv("202203-divvy-tripdata.csv")
apr04_df <- read_csv("202204-divvy-tripdata.csv")
may05_df <- read_csv("202205-divvy-tripdata.csv")
jun06_df <- read_csv("202206-divvy-tripdata.csv")
jul07_df <- read_csv("202207-divvy-tripdata.csv")
aug08_df <- read_csv("202208-divvy-tripdata.csv")
sep09_df <- read_csv("202209-divvy-publictripdata.csv")
oct10_df <- read_csv("202210-divvy-tripdata.csv")
nov11_df <- read_csv("202211-divvy-tripdata.csv")
dec12_df <- read_csv("202212-divvy-tripdata.csv")
```

I used rbind to combine all the data frame into one data frame named cyclistic_all_months.

```
cyclistic_all_months <- rbind(jan01_df,feb02_df,mar03_df,apr04_df,
                              may05_df,jun06_df,jul07_df,aug08_df,
                              sep09_df,oct10_df,nov11_df,dec12_df)
```

I removed the created data frame as the data frame has already combine into one in previous step.

```
remove(jan01_df,feb02_df,mar03_df,apr04_df,
       may05_df,jun06_df,jul07_df,aug08_df,
       sep09_df,oct10_df,nov11_df,dec12_df)
```

Then I view dataset using different r functions like str, head, colnames, glimpse and summary.I decided to remove all the ride length where value is either zero or less than zero.

```
cyclistic_all_months<- cyclistic_all_months[!(cyclistic_all_months$ride_length <=0),]
```

I created new columns day_name, date, month_date, month_name, started_time, ended_time, started hour, ended hour and seasons.

```r
#Converting STARTED AT & END AT INTO POSIXct
cyclistic_all_months$started_at <- format(as.POSIXct(cyclistic_all_months$started_at,format="%m/%d/%Y %H:%M"))
cyclistic_all_months$ended_at <- format(as.POSIXct(cyclistic_all_months$ended_at,format="%m/%d/%Y %H:%M"))
#day name
cyclistic_all_months$day_name <- format(as.Date(cyclistic_all_months$date), "%A")
#date
cyclistic_all_months$date <- as.Date(cyclistic_all_months$started_at)
#month date
cyclistic_all_months$month_date <- format(as.Date(cyclistic_all_months$date), "%d")
#month
cyclistic_all_months$month_name <-month(ymd(cyclistic_all_months$date), label = TRUE, abbr = FALSE)
#other way
cyclistic_all_months$month_name <- format(as.Date(cyclistic_all_months$date), "%m")
#time
cyclistic_all_months$started_time <-as_hms(as.POSIXct(cyclistic_all_months$started_at))
cyclistic_all_months$ended_time <-as_hms(as.POSIXct(cyclistic_all_months$ended_at))
#hour
cyclistic_all_months$start_hour <- (hour(cyclistic_all_months$started_at))
cyclistic_all_months$end_hour <- (hour(cyclistic_all_months$ended_at))
#seasons
cyclistic_all_months$seasons <- time2season(cyclistic_all_months$date,out.fmt = "seasons")
```

Analyze and Share Phase

After cleaning and new columns were made I did some calculations with my dataset which include the following:

- Max, Min, Median ride length
- Total count for member type/ridership type, member type and ridership type together
- Average ride length for member type/ridership type, member type and ridership type together.
- Average ride length for member/ridership type by day/month/month date/seasons
- Total no of rides for member/ridership type by day/month/month date/seasons
- Average ride length for member type by hour
- Total no of rides of member type by hour

```
fwrite(cyclistic_all_months,"cyclistic_data_combine.csv")
```

From R I created the cv file with new columns using the fwrite function so that I could use the merge for data visualization in tableau. To see full r code it is available here.

Tableau

Merge csv file was imported to tableau. When I imported the merge dataset I realize that ride_length column is in seconds instead of minutes. I decided to make a new calculate field in which I divide the ride_length by 60 and change the format to nn:ss. I made a few graphs related to the task given and a dashboard combining all graphs.
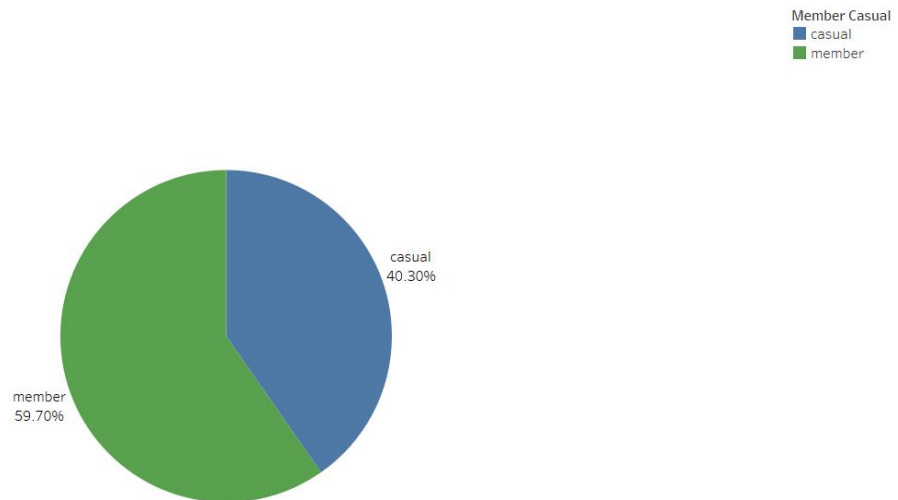
Total Rides of Member Type

casual	member	Total Count
1,742,909	2,581,902	4,324,811

From the table we can see that the total no of rides is 4,324,811 in which member rider has 2,581,902 total ride that is more than the casual rider which has 1,742,909 total rides.

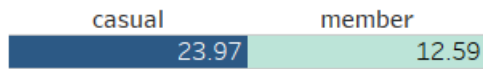## Total Rides of Member Type in Percentage

\<Total Rides of Member Type in Percentage\>

Member Casual
- casual
- member

casual
40.30%

member
59.70%

In the pie chart it is clearly showing us that member rider has around 19 % more rides than the casual rider.

Average Ride Length of Member Type

<Avg Ride Length of Member Type>

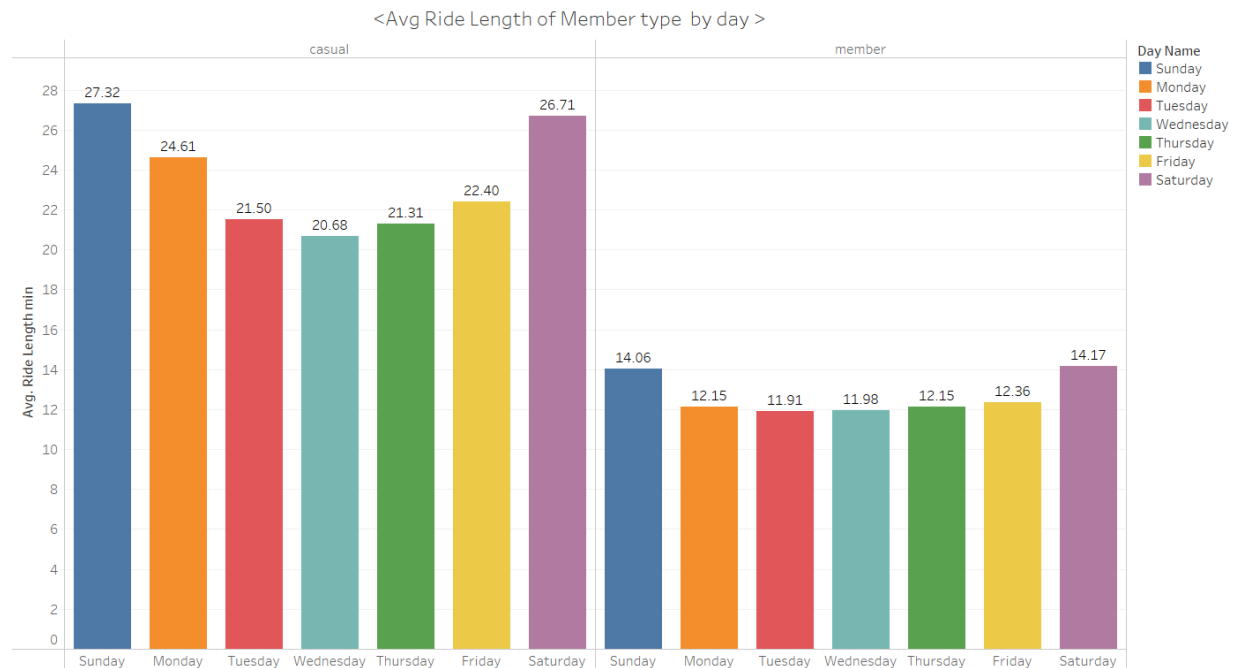| casual	member
23.97	12.59

From the table we can see that casual rider has ride of 23.97 min on average which is longer compared to member rider which has ride of 12.59 min on average.

Average Ride Length of Member type by day

<Avg Ride Length of Member type by day >

From the bar chart we can see that casual rider has better average ride length than the member rider whether it is Sunday or Monday or any day. Member rider is nowhere close in term of average ride length. Sunday being the day for highest average ride length of 27.32 min for casual rider whereas Saturday being the day for highest average ride length of 14.17min for member.

Total Rides of Member Type in different Seasons

Month Name

Seasons
- autumm
- spring
- summer
- winter
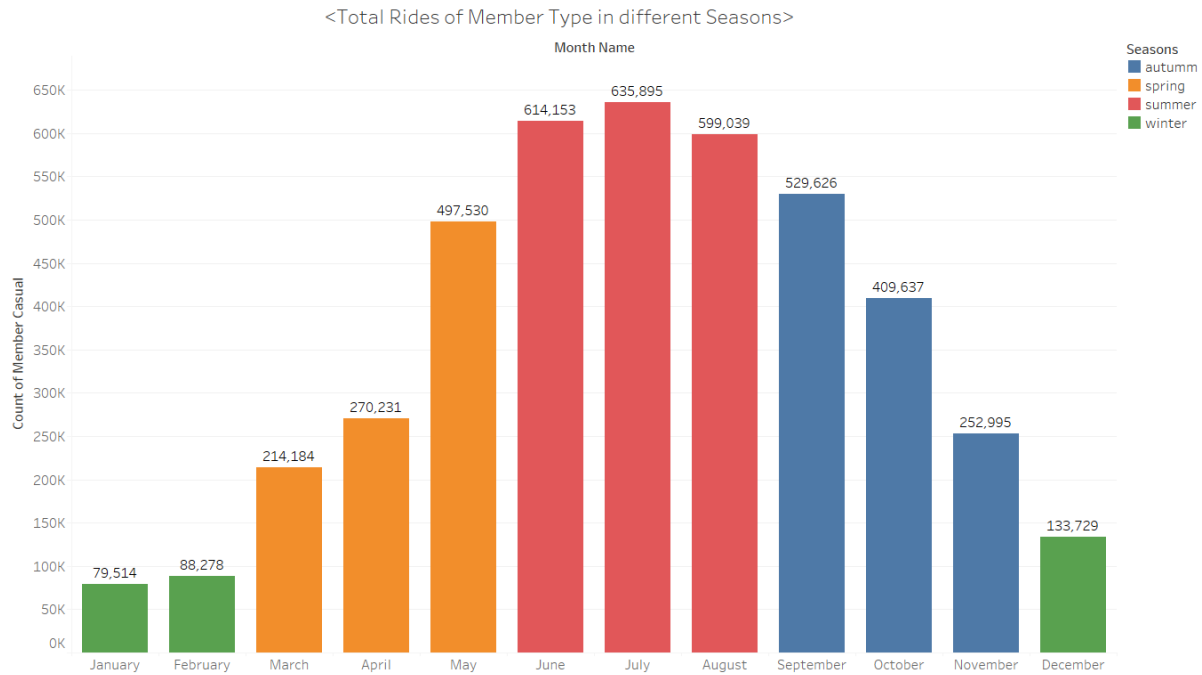
From the bar chart we can see that total no of rides are the highest in the month of Summer July that is 635,895. Winter January has the lowest no of rides that is 79,514. It shows us that July is the busiest season in total no of rides whereas January is the least busiest season in total no of rides.

Total Rides of Member Type by Year
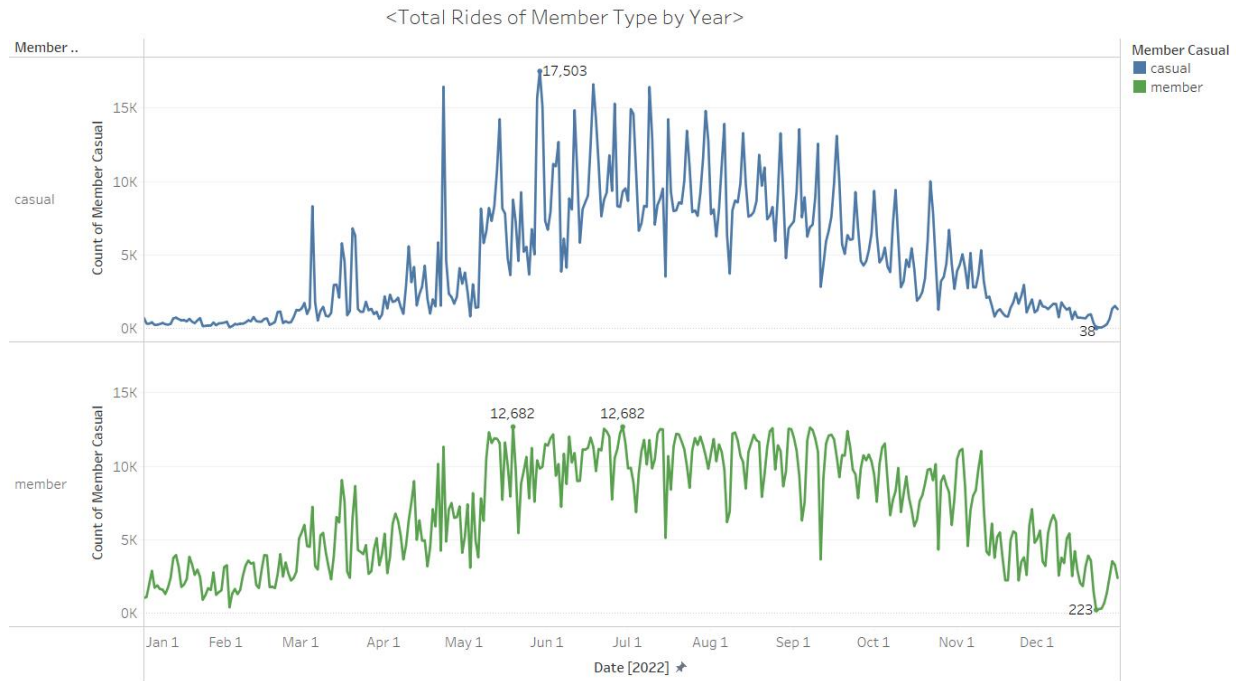
From the line graph we can see that causal rider has highest ride of 17,503 that is more than member ride which have 12,682 no of rides as the highest ride in the year. Both casual and member rider has up and down in no of rides throughout the year but for casual rider the rider was quite low during the start and end of the year compared to member rider.

Total Rides of Member Type by Ridership

<Total Rides of Member Type by Ridership >
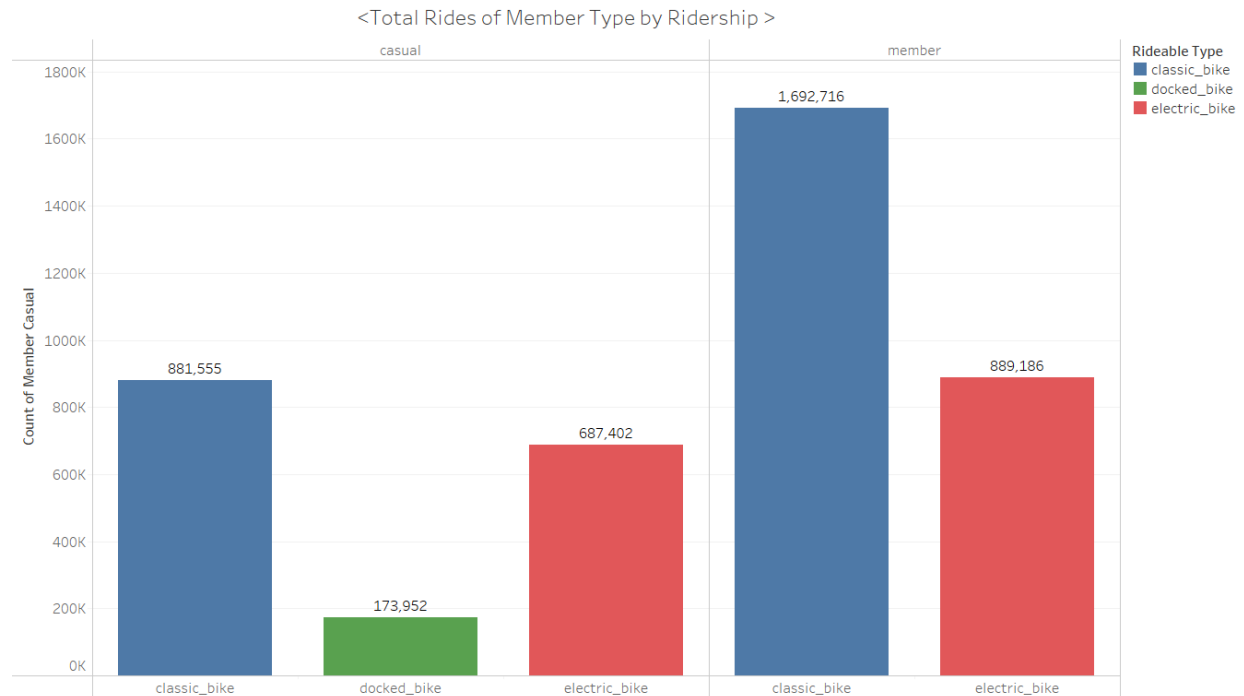
From the bar graph we can see that member rider has 1,692,716 no of rides in classic bike that is more than casual rider which have 881,555 no of rides. For Member rider the classic bike ridership is almost double compared to casual rider. Causal rider has extra advantage of 173,952 docked bike type compared to member rider.

Total Rides of Member type by hour
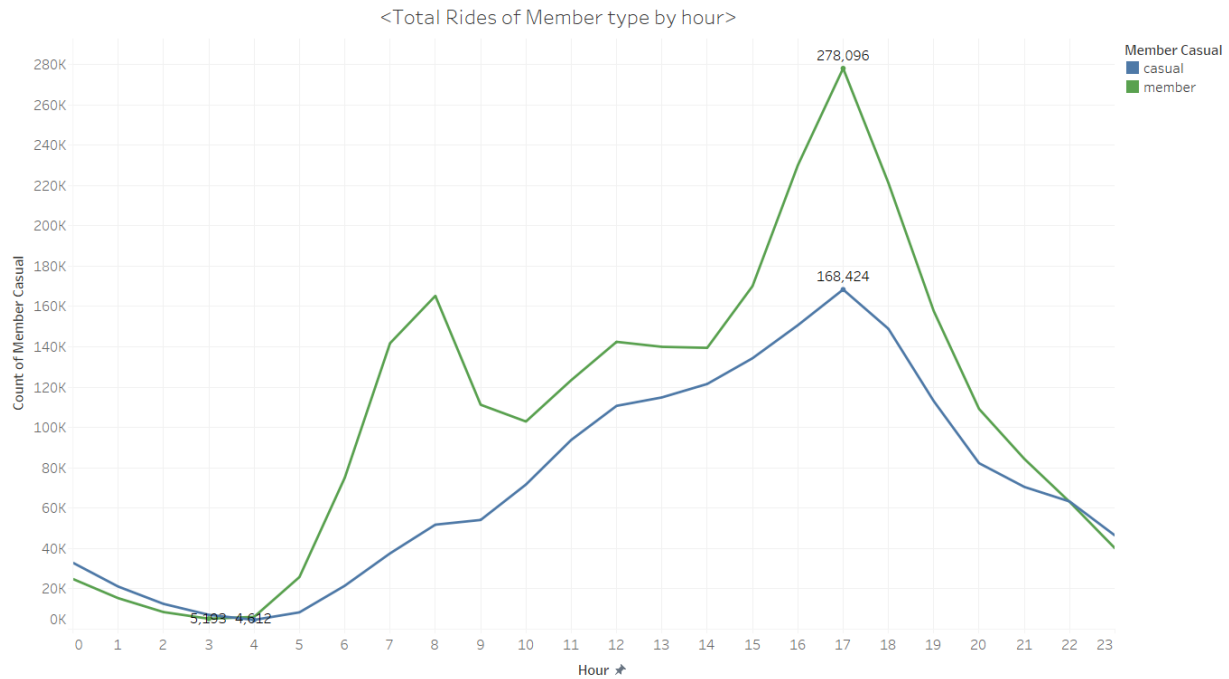
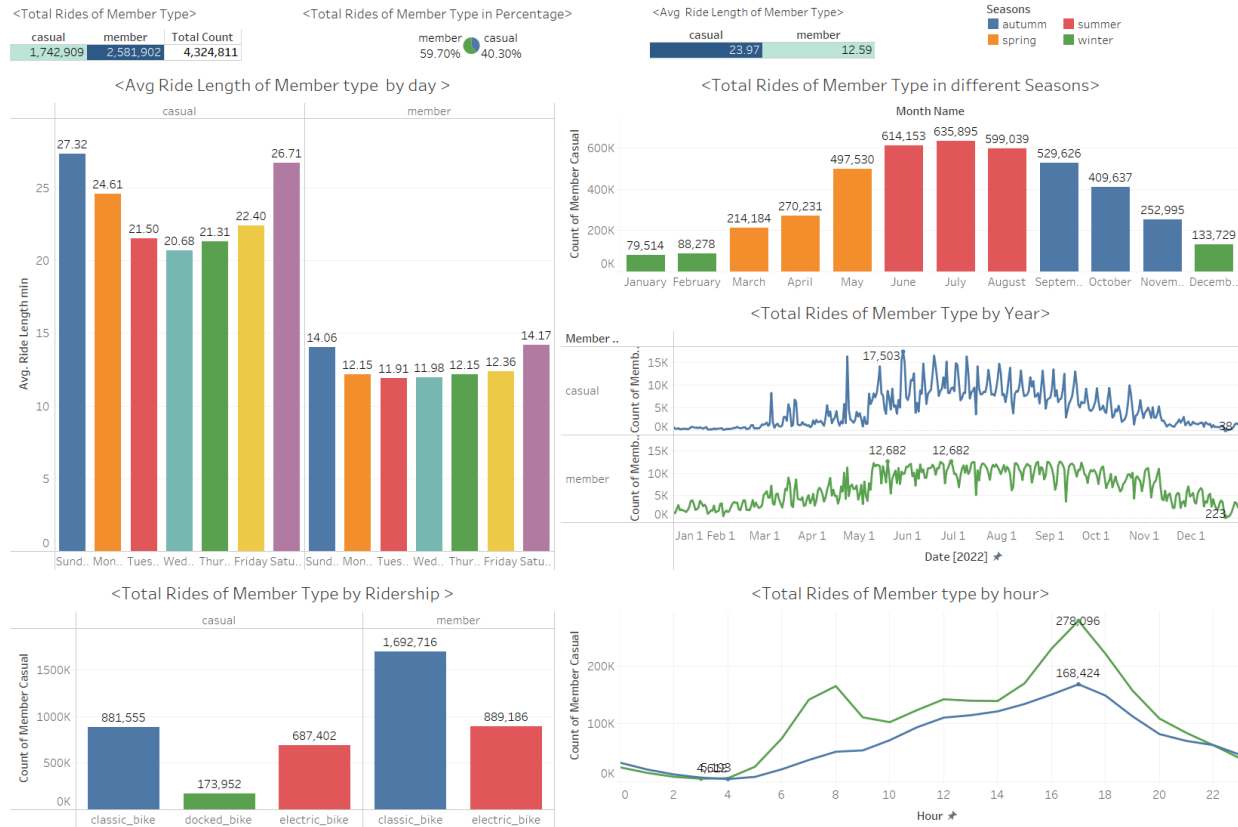From the line chart we can see that from 15 till 18 hour was the most busiest hour for both causal rider as well as member rider. Casual ridership recorded 278,096 no of rides in the 17th hour whereas member rider recorded 168,424 no of rides in the 17th hour.

Final Dashboard

\<Google Data Analytics Capstone | Cyclistic Bikes\>

\<Total Rides of Member Type\>

casual	member	Total Count
1,742,909	2,581,902	4,324,811

\<Total Rides of Member Type in Percentage\>

member 59.70%   casual 40.30%

\<Avg Ride Length of Member Type\>

casual	member
23.97	12.59

Seasons
- autumn
- summer
- spring
- winter

\<Avg Ride Length of Member type by day \>

casual: Sunday 27.32, Monday 24.61, Tuesday 21.50, Wednesday 20.68, Thursday 21.31, Friday 22.40, Saturday 26.71

member: Sunday 14.06, Monday 12.15, Tuesday 11.91, Wednesday 11.98, Thursday 12.15, Friday 12.36, Saturday 14.17

\<Total Rides of Member Type in different Seasons\>

January 79,514, February 88,278, March 214,184, April 270,231, May 497,530, June 614,153, July 635,895, August 599,039, September 529,626, October 409,637, November 252,995, December 133,729

\<Total Rides of Member Type by Year\>

casual: 17,503

member: 12,682, 12,682, 223

\<Total Rides of Member Type by Ridership \>

casual: classic_bike 881,555, docked_bike 173,952, electric_bike 687,402

member: classic_bike 1,692,716, electric_bike 889,186

\<Total Rides of Member type by hour\>

278,096, 168,424

To see my visualization, it is available on [Tableau](#).

Act Phase

Recommendation

- Offer extra month of free membership when the casual rider purchase the annual membership.
- Focus on digital marketing like ads, signboards to show causal riders how much money can be saved using the annual membership.
- Introduce a point based system in annual membership to be eligible for coupons and discounts in the future.
- Give discount to casual rider in winters to increase ridership and memberships as well.