

# **Google Data Analytics Capstone Project- Cyclistic Bike-Share Analysis.**

**Sunooja Suhara**

8-Aug-2022

## **Introduction**

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.

The goal of this project is to analyze the trends in data, how casual riders and annual members use Cyclistic bikes differently and how to convert casual riders into annual members.

## **Project Task**

The task of this project is to take actionable measures to convert casual members to permanent members. The analysis will help the stakeholder better understand business growth and can take necessary action to make it live.

## **Dataset**

The data is provided from the Cyclist company - dataset. Considered data of 10 months from August,2021 to May,2022.

## **Phases Of Analysis**

### **Ask Phase:**

In this phase the stakeholder has provided the clear picture of their questions in the document:

- How do annual members and Casual riders use Cyclist bikes differently?
- Why would casual riders buy Cyclist Annual membership?
- How can a Cyclist group use Digital media to influence casual riders to become annual members?

## Prepare Phase:

- Download the data provided by the stakeholder and store it appropriately.
- Import the data to R and identify the way the data is organized, column names, their data types, and give a rough calculation for the data that will lead the analysis. At the initial stage of analysis, there were 13 columns.

```

rows: 4,308,771
columns: 13
; ride_id          <chr> "99103BB87CC6C1BB", "EAFCCCFB0A3FC5A1", "9EF4F46C57AD234D", "5834D3208BFAF1DA", "CD825CB87ED1D096", "612F12C94A964F3E", "C7435946...
; rideable_type    <chr> "electric_bike", "electric_bike", "electric_bike", "electric_bike", "electric_bike", "electric_bike", "electric_bike", "electric_...
; started_at       <chr> "2021-08-10 17:15:49", "2021-08-10 17:23:14", "2021-08-21 02:34:23", "2021-08-21 06:52:55", "2021-08-19 11:55:29", "2021-08-19 12...
; ended_at         <chr> "2021-08-10 17:22:44", "2021-08-10 17:39:24", "2021-08-21 02:50:36", "2021-08-21 07:08:13", "2021-08-19 12:04:11", "2021-08-19 12...
; start_station_name <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
; start_station_id  <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
; end_station_name  <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
; end_station_id    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", ...
; start_lat         <dbl> 41.77000, 41.77000, 41.95000, 41.97000, 41.79000, 41.81000, 41.77000, 41.94000, 41.92000, 41.74000, 41.95000, 41.97000, 41.96000, ...
; start_lng         <dbl> -87.68000, -87.68000, -87.65000, -87.67000, -87.60000, -87.61000, -87.62000, -87.64000, -87.66000, -87.53000, -87.71000, -87.7100...
; end_lat           <dbl> 41.77000, 41.77000, 41.97000, 41.95000, 41.77000, 41.80000, 41.81000, 41.95087, 41.92000, 41.74000, 41.97000, 41.96000, 41.89000, ...
; end_lng           <dbl> -87.68000, -87.63000, -87.66000, -87.65000, -87.62000, -87.60000, -87.61000, -87.65915, -87.66000, -87.53000, -87.71000, -87.7100...
; member_casual    <chr> "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", "member", ...

```

- Check for the data type of each column and change to the appropriate data type, which makes the analysis easier. Here in Bike share, the data type of 'started\_at' and 'ended\_at' changed to the datatype 'Date and Time'.

```

#renaming existing columns
colnames(data_bike_share)[colnames(data_bike_share) == "rideable_type"] <- "bike_type"
colnames(data_bike_share)[colnames(data_bike_share) == "member_casual"] <- "user_type"

```

```

#casting started_at and ended_at to time data type
data_bike_share$started_at <- strptime(data_bike_share$started_at, format = "%Y-%m-%d %H:%M:%S")
data_bike_share$ended_at <- strptime(data_bike_share$ended_at, format = "%Y-%m-%d %H:%M:%S")

```

- Next, add the columns which will be very crucial for analysis and rename the columns to ease analysis while coding. After adding the calculated fields, the data contains 18 columns:

To calculate the trip duration, hour, week, month, and year of travel.



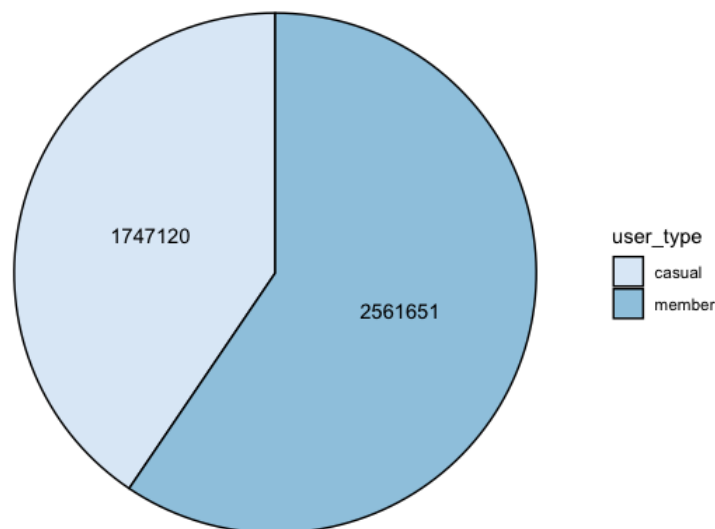
'start\_hour' - This column gives us the info about when the trip started in a day, by which analysis can be calculated based on the timing.

```
#Adding columns for analysis
data_bike_share$trip_duration <- as.numeric(with(data_bike_share, ended_at - started_at))
data_bike_share$weekday <- with(data_bike_share, weekdays(started_at))
data_bike_share$ride_month <- format(data_bike_share$started_at, format = "%b")
data_bike_share$ride_year <- format(data_bike_share$started_at, format = "%Y")
#Calculating the time of travel(morning, Afternoon, Eve, night)
data_bike_share$start_hour <- hour(data_bike_share$started_at)
data_bike_share$start_hour <- ifelse(data_bike_share$start_hour >= 5 & data_bike_share$start_hour < 11, "morning",
                                     ifelse(data_bike_share$start_hour >= 11 & data_bike_share$start_hour < 17, "Noon",
                                             ifelse(data_bike_share$start_hour >= 17 & data_bike_share$start_hour < 22, "Evening", "Night"))) )
```

Creating charts with the values that are in the dataset and with the calculated fields.

- Comparing casual and paid member users

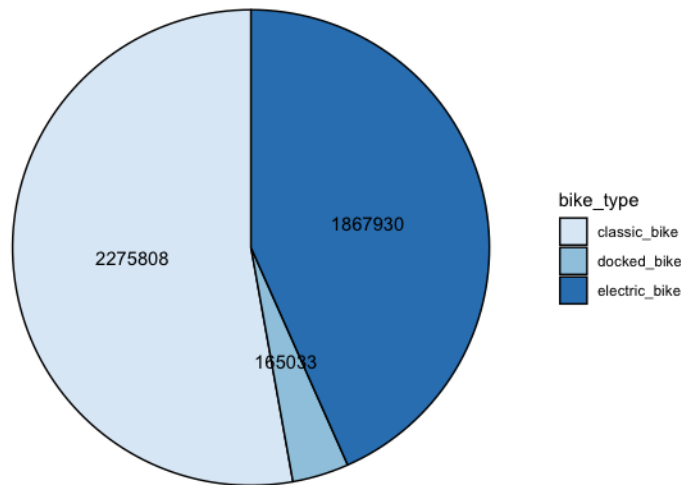
Casual members Vs User with Membership



When comparing each user, the number of paid members is greater than casual users.

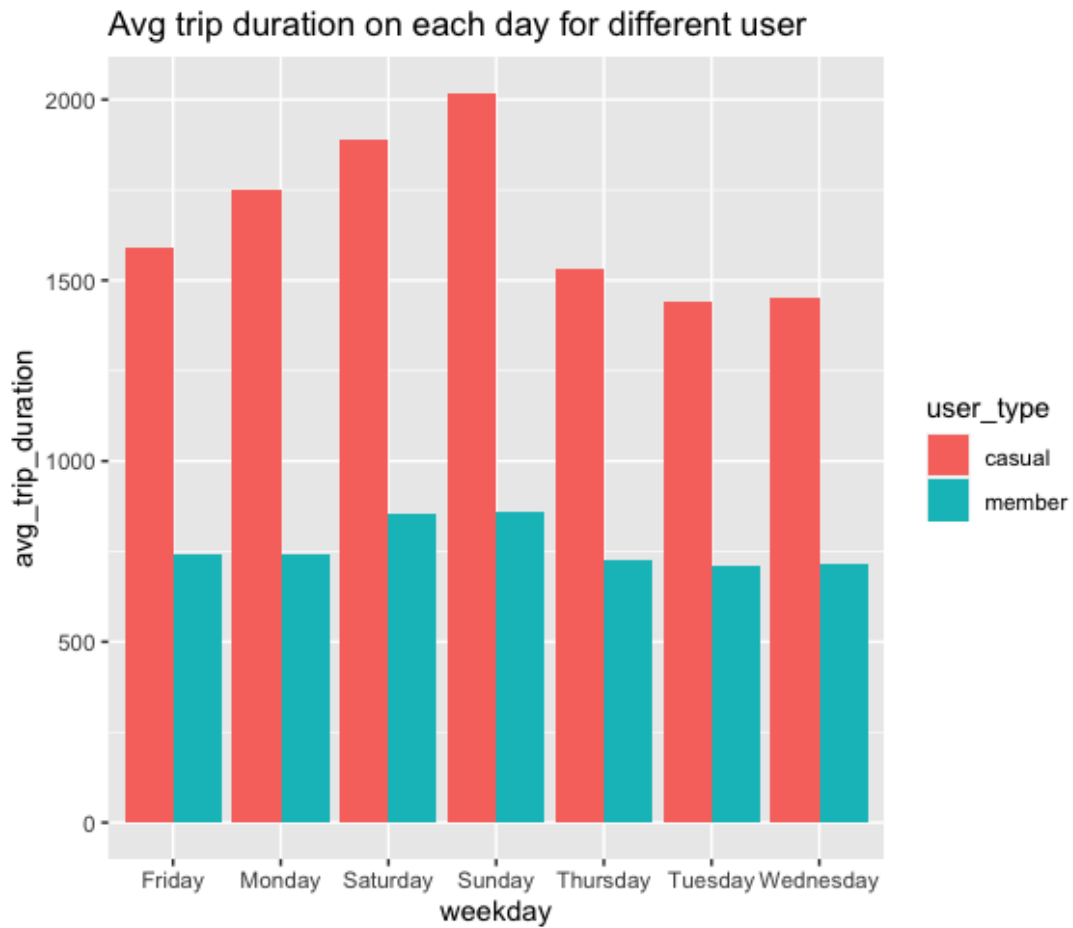
- Comparing bike types-classic, Docked, Electric bikes

Comparing each bike types



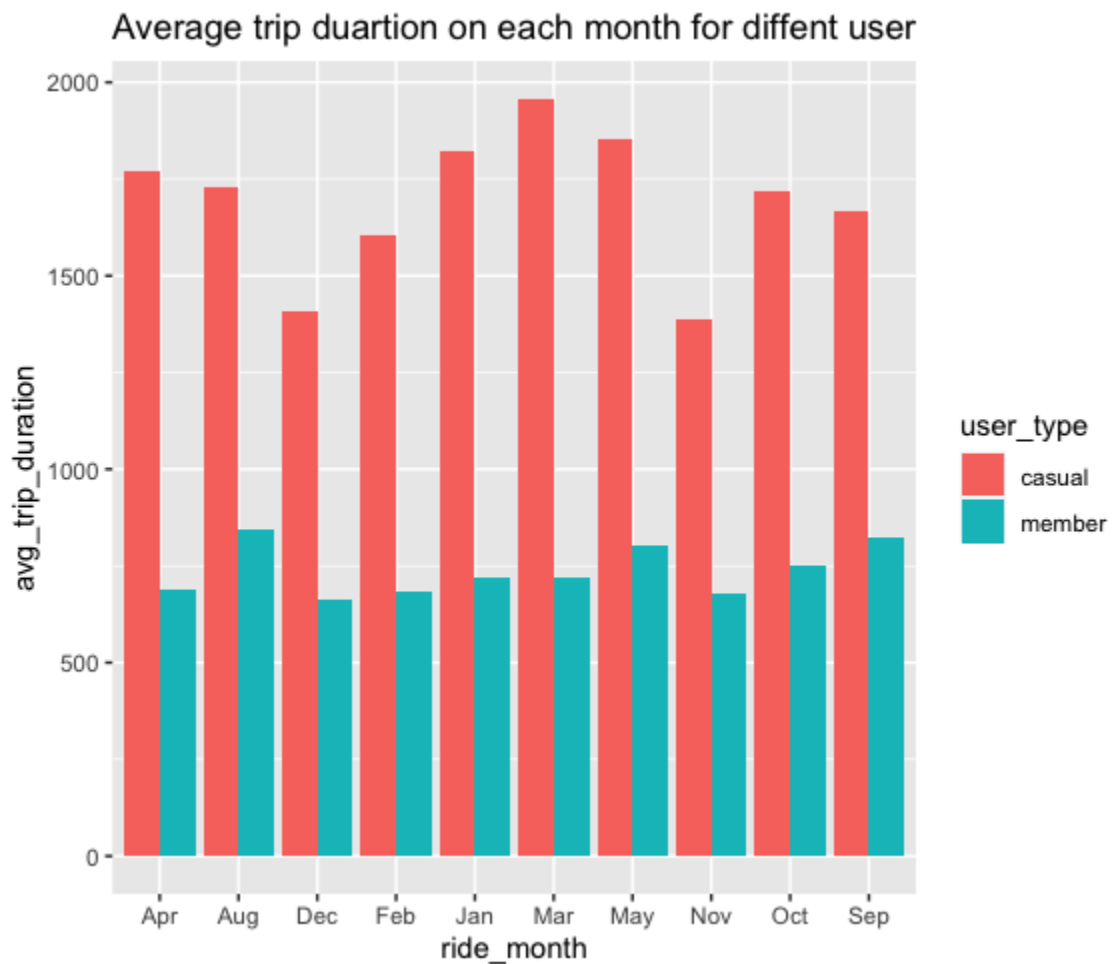
The most commonly used bike type is the classic. Promotion for docked bike types has to be worked in order to increase docked bike users.

- Analyzing trip duration for each day for different user types.



Trip duration is longer for casual users when compared to paid members.  
Trip duration is high on weekends for both user types.

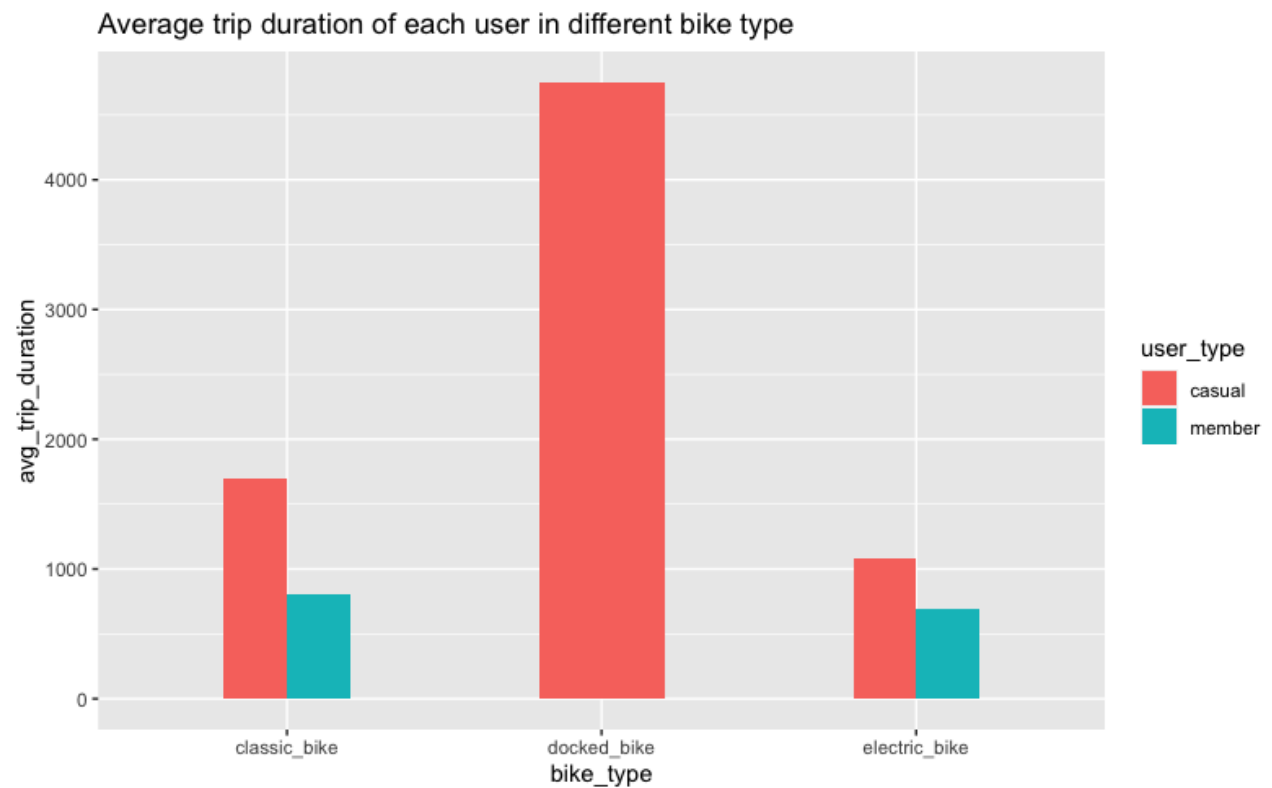
- Average trip duration on each month for different user



Trips are high in numbers in the months of January to May and are fewer in the months of November and December. As per the weather report, November and December are extremely cold in Chicago and that may be the reason for less trip duration in those months.

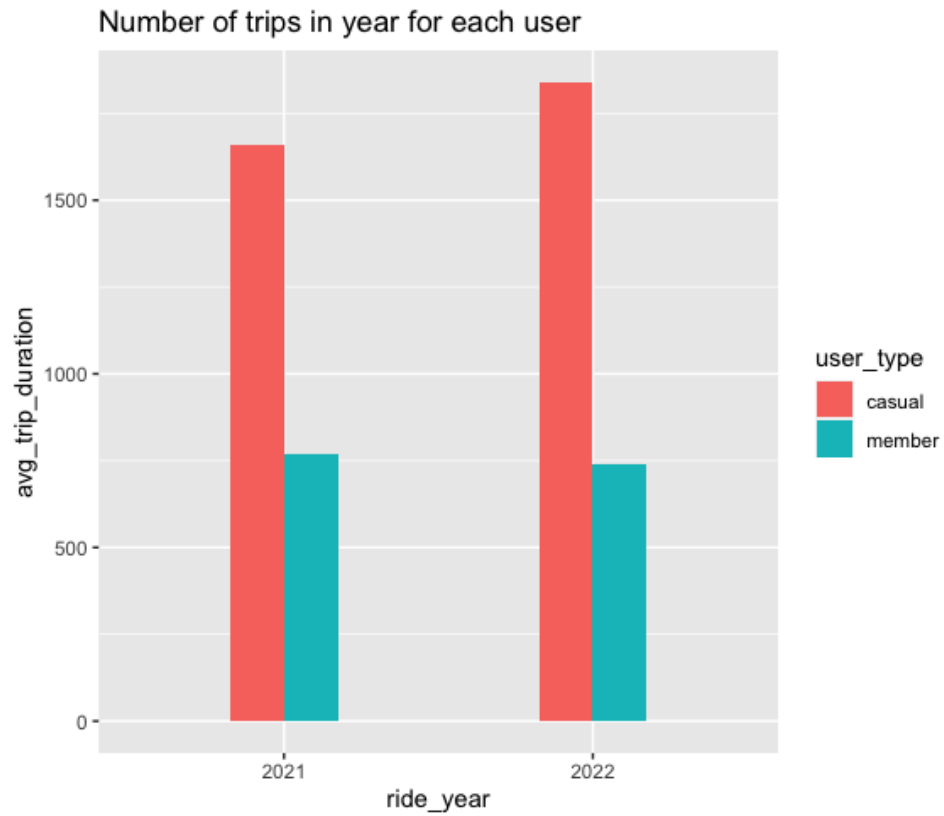
- Calculated average trip duration of each bike type based on user type





Each type of bike is mostly used by casual users. Users with membership highly prefer classic bike types.

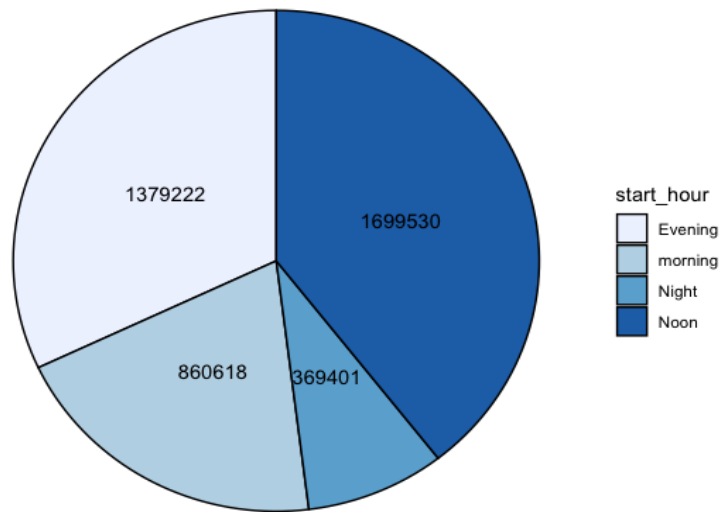
- Average trip in each year for different users



For casual and paid members, the year 2022 has the longest trip duration when compared to the year 2021.

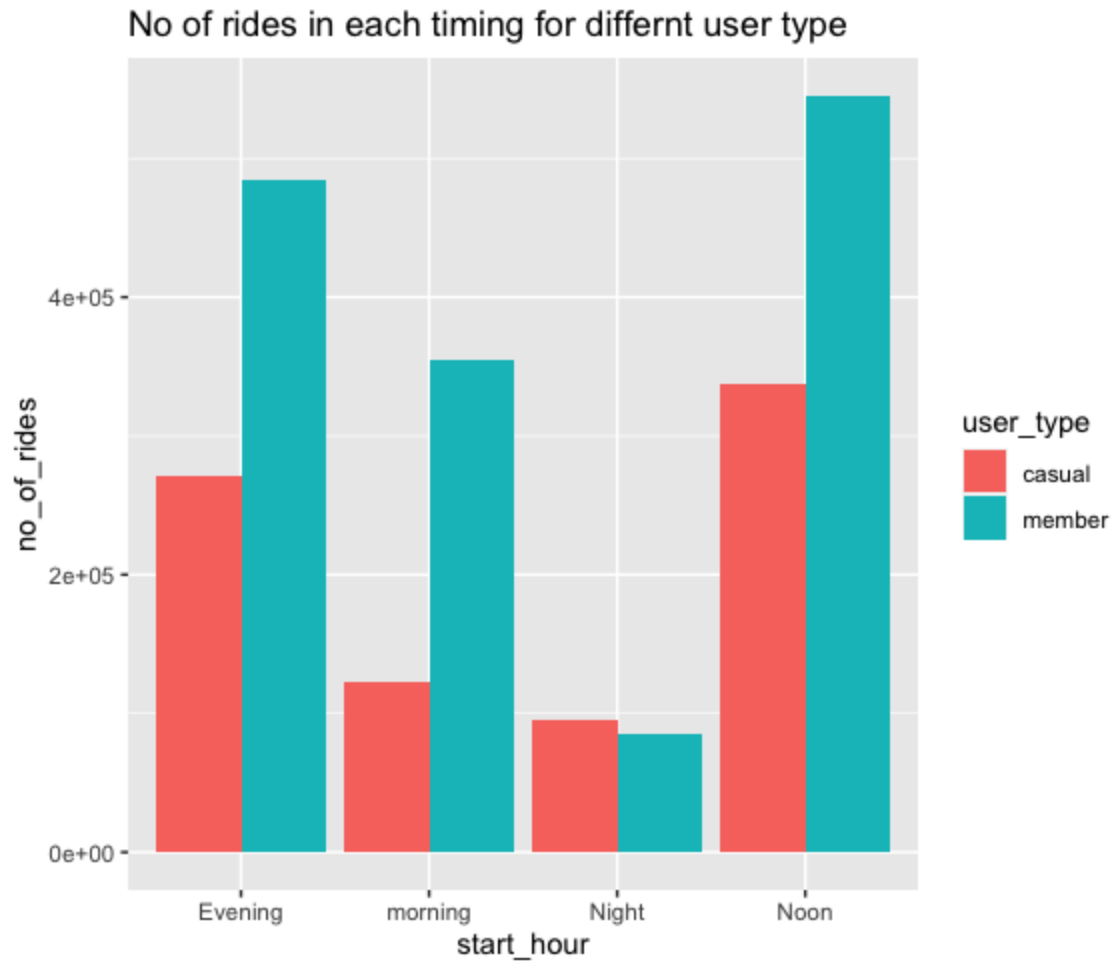
- Number of rides for each timing of a day.

Rides in each time of a day



The majority of rides occur between noon and evening, with the fewest occurring at night.

- User type and total trips in each time a day



Both casual and member users use bikes at noon and in the evening.

## Act Phase:

After the analysis, I've reached the following **insights**:

- Paid members are more than casual users in numbers, but when it comes to trip duration, casual users utilize the bikes more than paid members.
- The most used bike types are classic bikes, and the least used are docked bikes, which are only used by casual users.

- When comparing the years, the number of casual users increased while the paid members decreased from 2021 to 2022.
- High trip duration for casual users is from Jan to May, and for paid members is in the months of August, whereas less usage of bikes is in the months of November and December.
- Trip duration is high on weekends for both casual and paid users.
- More trips are made during the noon hour of the day and fewer at night.

## **Conclusion/ Recommendation:**

Top four recommendations from analysis are as follows:

Giving attractive offers for paid members will help casual users convert into paid members, and all the offers should be advertised in each bike station. Giving discounts on docked and electric bikes increases the usage of both and has to make sure that each electric bike is charged and well maintained.

Users can be provided discounts if they exceed a particular kilometer. Cyclistic can introduce a 'One Day Trip Plan' where paid members can use unlimited rides in a day with a fixed amount for each bike type, and the plan expires at midnight.

Cyclistic can increase the number of bike stations in places like office hubs, parks, and residential areas, which makes transportation easier and can increase the profit on weekdays.

Advertising for the 'One Day Trip Plan,' discounts, and special offers must be placed in strategic locations where the majority of people congregate. It would be more attractive if it was advertised in metro and bus stations.

## **Acknowledgement**

I'd like to thank Google and Coursera for the Data Analytics Certification course, where I've learned about what data is; phases of data analysis; data cleaning; data visualization; and getting insights from data. Adding on, I'd like to thank all the mentors who took me through each phase of analysis. This was a great journey for a beginner like me to explore data analysis.