

Cyclic Bikes Analysis

2024-05-09

Objective

The objective is to analyze annual member and casual bike riders bike data to identify trends.

Setting up My Environment

I installed and loaded the tidyverse package, in order to use the read_csv function to import the csv data.

```
install.packages("tidyverse")  
library(tidyverse)
```

Data Importation and Processing

Importation

The bike trip data was downloaded as a CSV document and then imported in to R studio, using the read.csv function from the tidyverse package, and then assigned each dataset to a variable name.

```
tripdata_2022_01 <- read_csv("202201-divvy-tripdata.csv")  
tripdata_2022_02 <- read_csv("202202-divvy-tripdata.csv")  
tripdata_2022_03 <- read_csv("202203-divvy-tripdata.csv")
```

Join the Tables

I joined all three tables together using the **full_join** function to join all rows from all three tables. The function recognizes the common column, the bike_id field in all three tables and joins all tables by the bike_id column.

```
join1<-full_join(tripdata_2022_01,tripdata_2022_02)  
tripdata_2022q1 <- full_join(join1, tripdata_2022_03)
```

Mutate

The mutate function was used to add three columns; **ride_length** to calculate the duration of the ride, **ride_date** to separate the start date from the time in the started_at column, and **day_of_week** to record the day of the week that the ride started

```
tripdata_2022q1_new <- mutate(tripdata_2022q1, ride_length=difftime(ended_at, started_at, units = "mins"), ride_date=tripdata_2022q1$started_at<-as_date(tripdata_2022q1$started_at))  
  
tripdata_2022q1_new <- mutate(tripdata_2022q1_new, day_of_week=yday(tripdata_2022q1_new$ride_date, label = TRUE))
```

Filtering and Sorting

Firstly, I created new dataframes, **num_of_rides** to calculate the total number of rides for each unique rider using the **n_distinct** to count the number of unique ride_id's and **avg_ride_length** to calculate average ride length using the **mean** function to , and grouped them by their membership status, ride type and day of the week.

```
num_of_rides<-tripdata_2022q1_new %>%
  select(ride_id, member_casual, rideable_type, day_of_week) %>%
  group_by(member_casual, rideable_type, day_of_week) %>%
  summarise(no_rides=n_distinct(ride_id))

avg_ride_length<-tripdata_2022q1_new %>%
  select(member_casual, rideable_type, ride_length, day_of_week) %>%
  group_by(day_of_week, member_casual, rideable_type) %>%
  summarise(avg_time=mean(ride_length))
```

Used the **head** function to get a quick summary of the new dataframes

```
head(num_of_rides)
```

```
## # A tibble: 6 × 4
## # Groups:   member_casual, rideable_type [1]
##   member_casual rideable_type day_of_week no_rides
##   <chr>         <chr>         <ord>         <int>
## 1 casual        classic_bike Sun           9368
## 2 casual        classic_bike Mon            8262
## 3 casual        classic_bike Tue            5236
## 4 casual        classic_bike Wed            7371
## 5 casual        classic_bike Thu            5982
## 6 casual        classic_bike Fri            4343
```

```
head(avg_ride_length)
```

```
## # A tibble: 6 × 4
## # Groups:   day_of_week, member_casual [3]
##   day_of_week member_casual rideable_type avg_time
##   <ord>         <chr>         <chr>         <drtn>
## 1 Sun          casual        classic_bike  33.98631 mins
## 2 Sun          casual        docked_bike  124.62373 mins
## 3 Sun          casual        electric_bike 18.84908 mins
## 4 Sun          member        classic_bike  14.15745 mins
## 5 Sun          member        electric_bike 11.85812 mins
## 6 Mon          casual        classic_bike  32.74451 mins
```

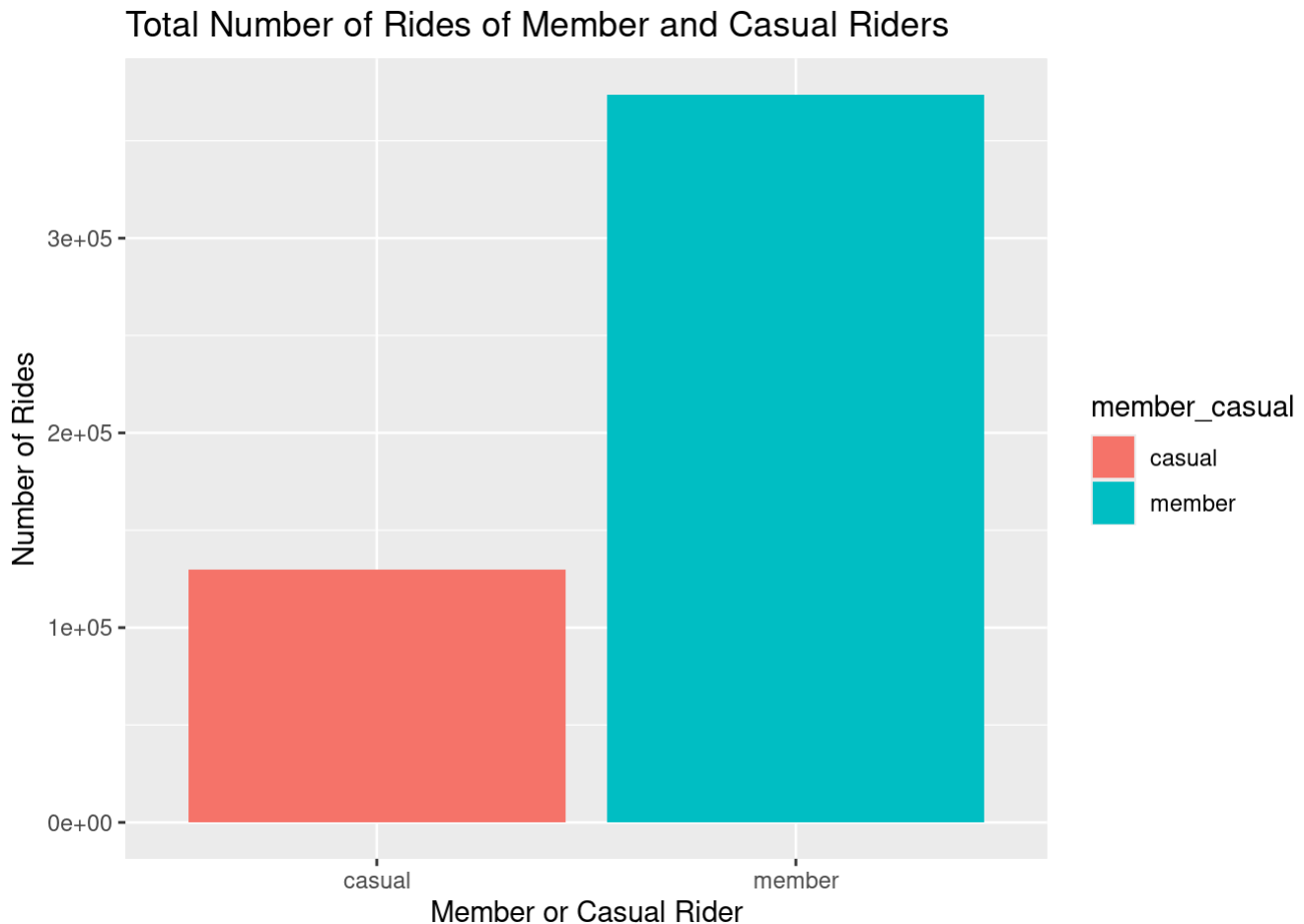
Analysis

Analysis was conducted to observe and discover trends in the datasets. With the use of visuals from the **ggplot** function, I created visuals to display the Total Number of Rides, Total Number of Rides on Each Bike Type, Number of Rides Per Day of Week, Average Ride Length, Average Ride Length on Each Bike, Average Ride Length Per Day of Week for Each Bike on column charts.

Total Number of Rides

Referencing the **num_of_rides**, An analysis was conducted to compare the total number of rides made by annual members versus casual riders. This chart shows that the total number of rides by annual members was significantly greater than the total number of rides by casual riders as shown below;

```
ggplot(data=num_of_rides)+
  geom_col(mapping = aes(x=member_casual, y=no_rides, fill=member_casual))+
  labs(title = "Total Number of Rides of Member and Casual Riders", x="Member or Casual Rider", y="Number of Rides")
```



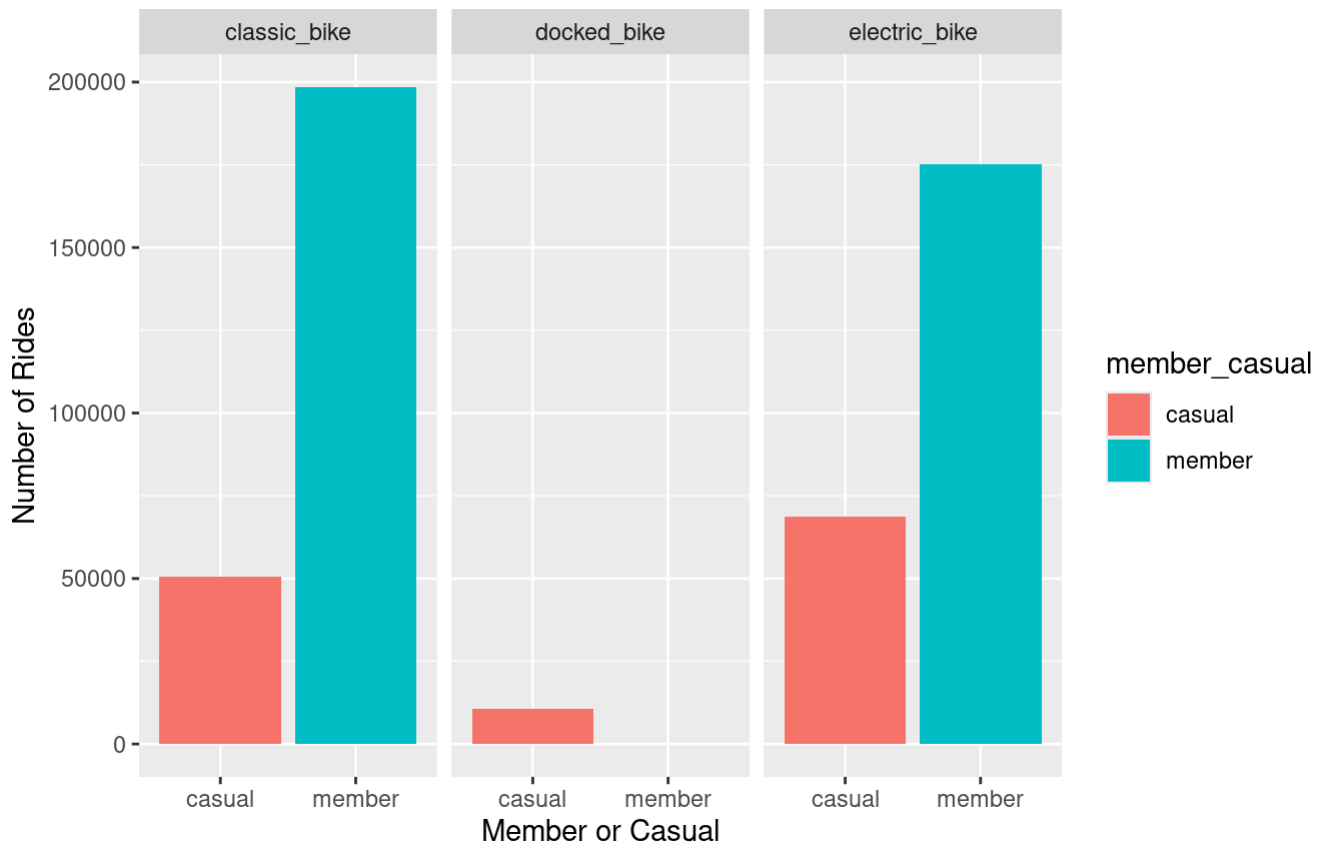
Total Number of Rides on Each Bike Type

Based on further analysis, it was found that annual members take more trips on classic bikes, followed by electric bikes, and no trips were recorded on docked bikes. Moreover, the analysis showed that the electric bikes are used the most by casual riders, followed by the classic bikes, and docked bikes are used the least.

```
ggplot(data=num_of_rides)+
  geom_col(mapping = aes(x=member_casual, y=no_rides, fill=member_casual))+
  facet_wrap(~rideable_type)+
  labs(title = "Number of Rides on Each Bike Type", subtitle = "The total number of rides on each bike type by member and casual riders", x="Member or Casual", y="Number of Rides")
```

Number of Rides on Each Bike Type

The total number of rides on each bike type by member and casual riders



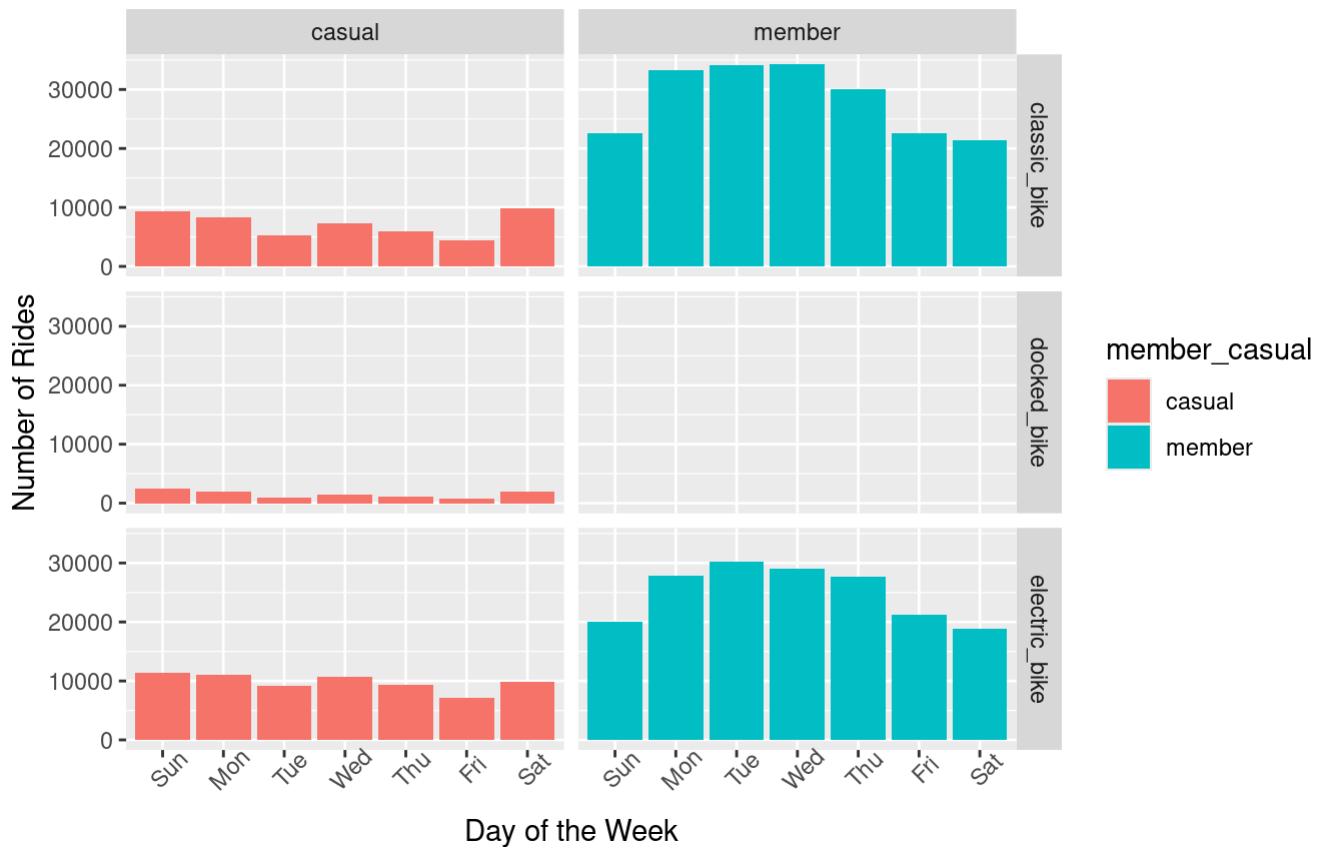
Number of Rides Per Day of Week

For annual members, the average number of rides on classic and electric bikes is higher during the weekdays than on the weekends, whereas for casual riders on electric bikes, the average number of rides remains steady during the week, but decreases on Fridays. Classic bikes seem to be more active during the weekend and see a decline during the week, while docked bikes continue to operate steadily.

```
ggplot(data = num_of_rides)+
  geom_col(mapping = aes(x=day_of_week, y=no_rides, fill=member_casual))+
  facet_grid(rideable_type~member_casual)+
  labs(title = "Number of Rides on Each Bike Per Day of Week", subtitle = "The total
number of rides on each bike per day of week by member and casual riders", x="Day of
the Week", y="Number of Rides")+
  theme(axis.text.x = element_text(angle = 45))
```

Number of Rides on Each Bike Per Day of Week

The total number of rides on each bike per day of week by member and casual riders

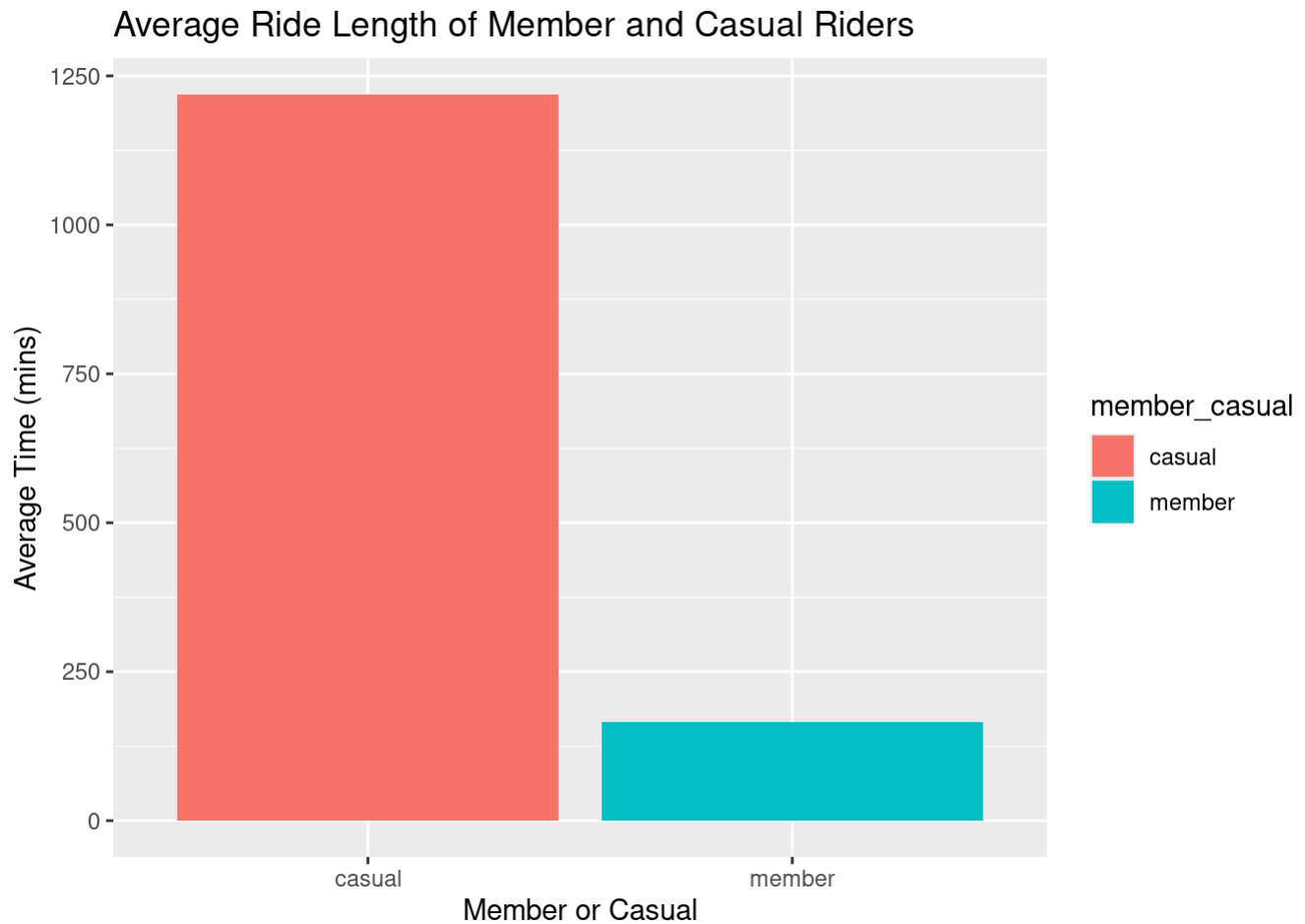


Average Ride Length

Referencing the **avg_ride_length** dataframe, the average duration of rides between annual members and casual riders was compared. It was discovered that the average duration of rides by casual riders was significantly longer than the average duration of rides by annual members as shown below;

```
ggplot(data = avg_ride_length)+
  geom_col(mapping = aes(x=member_casual, y=avg_time, fill=member_casual))+
  labs(title = "Average Ride Length of Member and Casual Riders", x="Member or Casual", y="Average Time (mins)")
```

```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```



Average Ride Length on Each Bike

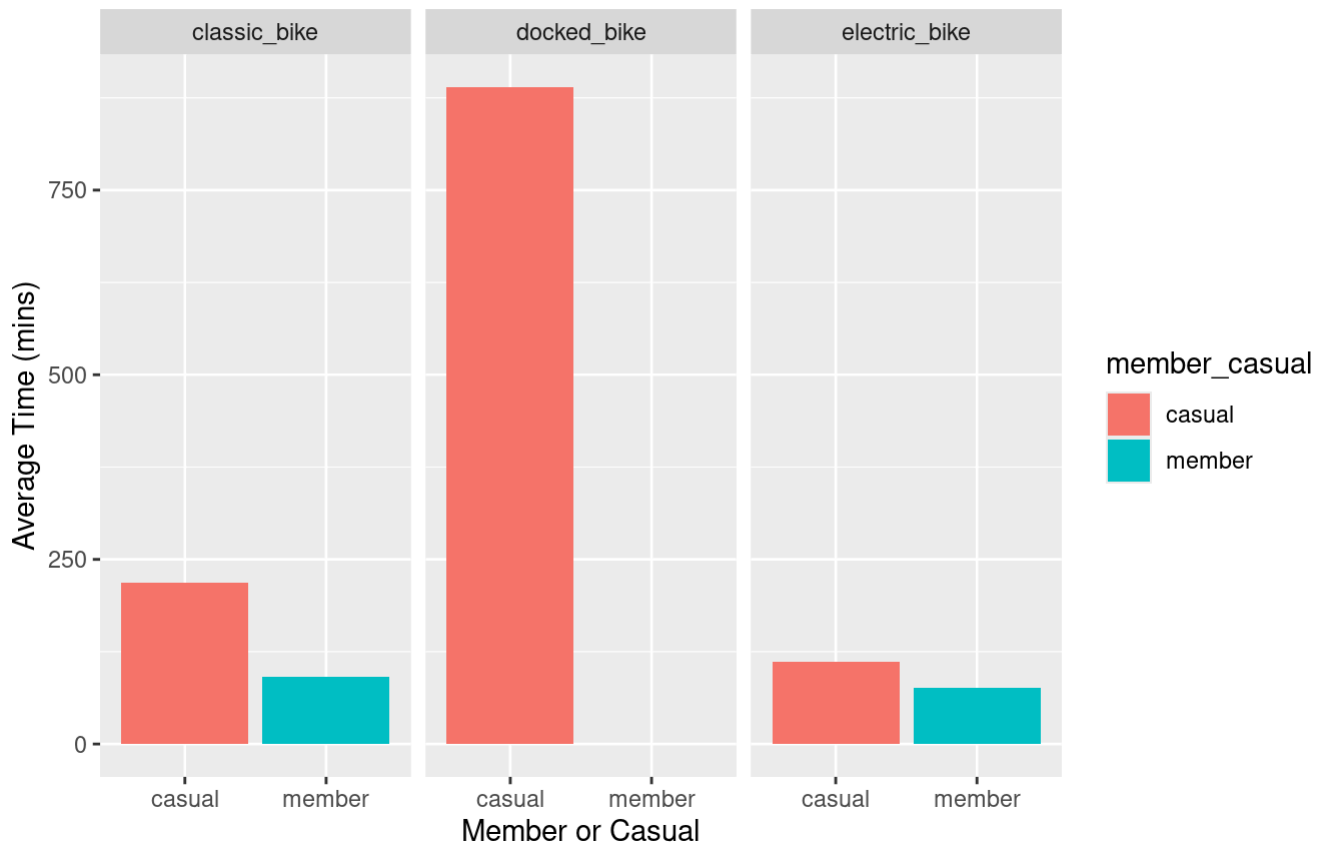
According to the analysis, casual riders on docked bikes had a significantly longer ride duration than those on classic bikes, followed by electric bikes with the shortest ride duration. Members also took longer trips on classic bikes than on electric bikes on average, but there was no significant difference between them.

```
ggplot(data = avg_ride_length)+
  geom_col(mapping = aes(x=member_casual, y=avg_time, fill=member_casual))+
  facet_wrap(~rideable_type)+
  labs(title = "Average Ride Length on Each Bike", subtitle = "The average ride length on each bike by member and casual riders", x="Member or Casual", y="Average Time (mins)")
```

```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```

Average Ride Length on Each Bike

The average ride length on each bike by member and casual riders



Average Ride Length Per Day of Week for Each Bike

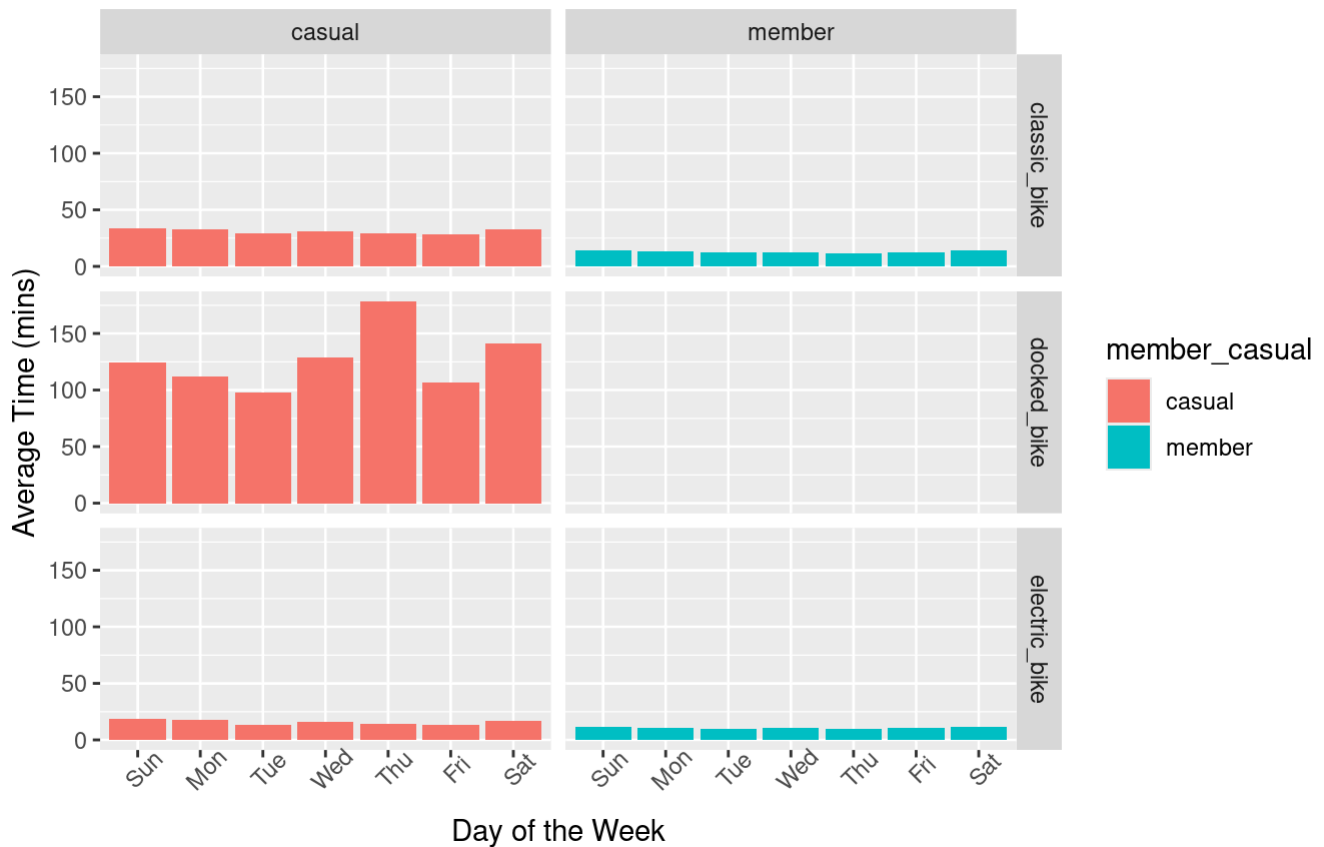
Analysis showed that the average ride duration of casual riders on classic and electric bikes were relatively the same throughout the week with little fluctuation while on docked bikes, Thursday was higher than the rest of the days which might be influenced by factors unknown. The average ride duration of annual members on classic and electric bikes were also relatively the same throughout the week with little fluctuation.

```
ggplot(data = avg_ride_length)+
  geom_col(mapping = aes(x=day_of_week, y=avg_time, fill=member_casual))+
  facet_grid(rideable_type~member_casual)+
  labs(title = "Average Ride Length on Each Bike Per Day of Week", subtitle = "The average ride length on each bike per day of week by member and casual riders", x="Day of the Week", y="Average Time (mins)")+
  theme(axis.text.x = element_text(angle = 45))
```

```
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```

Average Ride Length on Each Bike Per Day of Week

The average ride length on each bike per day of week by member and casual riders



Conclusion

- Annual members ride more than casual riders
- Annual members ride more during the week, while casual riders ride more during the weekend
- Annual members use classic bikes most often, whereas casual users use electric bikes most often
- Docked bikes are used only by casual riders
- Casual riders take longer trips than annual members
- Casual riders use docked bikes for significantly longer trips