

capstone project trips-2024

Raul

2025-09-25

Cyclist bike-share Analysis

Case Study for Google Data Analytics Professional Certificate. Data use for this case study obtained from the fictional company Cyclist bike-share. The I will follow the analisis process that consist of 7 Steps: Ask, Prepare, Process, Analyze, Share and Act

ASK

Business Task: *Design a marketing strategies aimed at converting casual riders into members by answering the following question. How do members and casual riders use Cyclist bike differently?*

PREPARE

Data obtained from [DIVV-TripData License](#)

- Tools use
 - Exploring Data
 - RStudio
 - Data Cleaning and Transforms Data
 - RStudio
 - Data visualizacion
 - RStudio

PROCESS

1. Data Exploration

- The first step of my work is to check if tables are consistent with each other
- This is achieved by looking at the columns and the data type each column contains

So first let's see if all the columns are equal

- **January**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **February**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **March**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **April**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **May**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **June**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **July**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- **August**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

- September**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		
- October**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		
- November**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		
- December**

[1]	"ride_id"	"rideable_type"	"started_at"
[4]	"ended_at"	"start_station_name"	"start_station_id"
[7]	"end_station_name"	"end_station_id"	"start_lat"
[10]	"start_lng"	"end_lat"	"end_lng"
[13]	"member_casual"		

Now let's see that all the data is consistent

- January**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			
- February**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **March**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **April**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **May**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **June**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **July**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **August**

ride_id	rideable_type	started_at	ended_at
"character"	"character"	"character"	"character"
start_station_name	start_station_id	end_station_name	end_station_id
"character"	"character"	"character"	"character"
start_lat	start_lng	end_lat	end_lng
"numeric"	"numeric"	"numeric"	"numeric"
member_casual			
"character"			

- **September**

```

ride_id      rideable_type      started_at      ended_at
"character"  "character"  "character"    "character"
start_station_name start_station_id end_station_name end_station_id
"character"  "character"  "character"    "character"
start_lat    start_lng    end_lat        end_lng
"numeric"    "numeric"    "numeric"      "numeric"
member_casual
"character"

```

- **October**

```

ride_id      rideable_type      started_at      ended_at
"character"  "character"  "character"    "character"
start_station_name start_station_id end_station_name end_station_id
"character"  "character"  "character"    "character"
start_lat    start_lng    end_lat        end_lng
"numeric"    "numeric"    "numeric"      "numeric"
member_casual
"character"

```

- **November**

```

ride_id      rideable_type      started_at      ended_at
"character"  "character"  "character"    "character"
start_station_name start_station_id end_station_name end_station_id
"character"  "character"  "character"    "character"
start_lat    start_lng    end_lat        end_lng
"numeric"    "numeric"    "numeric"      "numeric"
member_casual
"character"

```

- **December**

```

ride_id      rideable_type      started_at      ended_at
"character"  "character"  "character"    "character"
start_station_name start_station_id end_station_name end_station_id
"character"  "character"  "character"    "character"
start_lat    start_lng    end_lat        end_lng
"numeric"    "numeric"    "numeric"      "numeric"
member_casual
"character"

```

The data is consistent, so we can continue to the next step

2. Data Combination

The data combined of January 2024 to Diciember 2024, is join by *bind_rows* in RStudio.

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	member_casual
1	electric_bike	2024-01-12 15:30:27	2024-01-12 15:37:59	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90327	-87.63474	41.88918	-87.63851	member
2	electric_bike	2024-01-08 15:45:46	2024-01-08 15:52:59	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90294	-87.63444	41.88918	-87.63851	member
3	FA9ACE78061F17F7	2024-01-27 12:27:19	2024-01-27 12:35:19	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90295	-87.63447	41.88918	-87.63851	member
4	0A0D9E15EE5B0171	2024-01-29 16:26:17	2024-01-29 16:56:06	Wells St & Randolph St	TA1305000030	Larrabee St & Webster Ave	13193	41.88430	-87.63396	41.92182	-87.64414	member
5	33FFC9805E3EFF9A	2024-01-31 05:43:23	2024-01-31 06:09:35	Lincoln Ave & Waveland Ave	13253	Kingsbury St & Kinzie St	KA1503000043	41.94880	-87.67528	41.88918	-87.63851	member
6	C96080812CD285C5	2024-01-07 11:21:24	2024-01-07 11:30:03	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90322	-87.63432	41.88918	-87.63851	member
7	0EA7CB313D4F456A	2024-01-05 14:44:12	2024-01-05 14:53:06	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90322	-87.63432	41.88918	-87.63851	member
8	EE11F3A3839CF8D8	2024-01-04 18:19:53	2024-01-04 18:28:04	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90337	-87.63486	41.88918	-87.63851	member
9	63E83D68E3279F15	2024-01-01 14:46:53	2024-01-01 14:57:02	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90322	-87.63432	41.88918	-87.63851	member
10	8005682869122D93	2024-01-03 19:31:08	2024-01-03 19:40:05	Clark St & Ida B Wells Dr	TA1305000009	Kingsbury St & Kinzie St	KA1503000043	41.87603	-87.63087	41.88918	-87.63851	member
11	22885668AED0490	2024-01-03 07:39:20	2024-01-03 07:47:12	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90303	-87.63461	41.88918	-87.63851	member
12	133CD03CA430172	2024-01-03 17:03:11	2024-01-03 17:13:15	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90322	-87.63432	41.88918	-87.63851	member
13	32D578F9285802D5	2024-01-10 17:04:09	2024-01-10 17:11:40	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90315	-87.63458	41.88918	-87.63851	member
14	B11085685C38D698	2024-01-12 12:35:14	2024-01-12 12:43:34	Clark St & Ida B Wells Dr	TA1305000009	Kingsbury St & Kinzie St	KA1503000043	41.87565	-87.63093	41.88918	-87.63851	member
15	B66087108FA0938	2024-01-07 08:00:19	2024-01-07 08:06:46	Clark St & Ida B Wells Dr	TA1305000009	Kingsbury St & Kinzie St	KA1503000043	41.87557	-87.63078	41.88918	-87.63851	member
16	EA733AC59F174164	2024-01-24 08:28:31	2024-01-24 08:32:13	Clark St & Randolph St	TA1305000030	Kingsbury St & Kinzie St	KA1503000043	41.88417	-87.63209	41.88918	-87.63851	member
17	BAC65083CF6E160E	2024-01-13 11:18:54	2024-01-13 11:25:16	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90303	-87.63462	41.88918	-87.63851	member
18	C6C4B54F874526D	2024-01-24 18:52:19	2024-01-24 19:04:16	Sheffield Ave & Kingsbury St	13154	Aberdeen St & Randolph St	18062	41.91062	-87.65314	41.88411	-87.65426	member
19	A78D0CFFD9F1FC38	2024-01-27 13:12:24	2024-01-27 13:19:35	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	41.90322	-87.63432	41.88918	-87.63851	member
20	E247F348F1F9D880	2024-01-05 13:15:24	2024-01-05 13:21:41	Clark St & Randolph St	TA1305000030	Aberdeen St & Randolph St	18062	41.88530	-87.63225	41.88411	-87.65426	member

Showing 1 to 20 of 5,860,568 entries, 13 total columns

The result of the combined data is 5,860,568 rows and 13 columns

3. Data Cleaning and Transforming

- Transform the **started_at** and **ended_at** columns from **char** to **POSIXct** to perform the following calculation
- Calculate the duration of the rides

```
# Transform data type (char to POSIXct)
all_trips_24$started_at <- ymd_hms(all_trips_24$started_at)
all_trips_24$ended_at <- ymd_hms(all_trips_24$ended_at)

# Calculate the ride length
all_trips_24$ride_length <- as.numeric(difftime(all_trips_24$ended_at,
all_trips_24$started_at, units="secs"))
all_trips_24 %>%
  select(started_at, ended_at, ride_length) %>%
  head()

##           started_at           ended_at ride_length
## 1 2024-01-12 15:30:27 2024-01-12 15:37:59         452
## 2 2024-01-08 15:45:46 2024-01-08 15:52:59         433
## 3 2024-01-27 12:27:19 2024-01-27 12:35:19         480
## 4 2024-01-29 16:26:17 2024-01-29 16:56:06        1789
## 5 2024-01-31 05:43:23 2024-01-31 06:09:35        1572
## 6 2024-01-07 11:21:24 2024-01-07 11:30:03         519
```

- We create new columns from the **started_at** column, where we obtain the year, month, day, hour, and day of the week.

```
##   year month day    hour day_week
## 1 2024     1  12 15:30:27  Friday
## 2 2024     1   8 15:45:46   Monday
## 3 2024     1  27 12:27:19  Saturday
## 4 2024     1  29 16:26:17   Monday
## 5 2024     1  31 05:43:23 Wednesday
## 6 2024     1   7 11:21:24   Sunday
```

Before the next step in the process, I have to clean the data:

- Remove the columns that are not needed
 - start_lat
 - start_lng
 - end_lat
 - end_lng

```
all_trips_24 <- all_trips_24 %>%
  select(-c(start_lat, start_lng, end_lat, end_lng))
```

And finally, we create another dataframe with all the clean data

- The ride_length can not be negative so we have to exclude all rows that have negative values or 0 values
- The new and clean data frame is 5,859,845 row and 15 columns, which mean that 723 row were removed

ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	member_casual	ride_length	year	month	day	hour	day_week	
1	C1D650626CB099A	electric_bike	2024-01-12 15:30:27	2024-01-12 15:37:59	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	452	2024	1	12	15:30:27	Friday
2	EECD3880B258FCB0	electric_bike	2024-01-08 15:45:46	2024-01-08 15:55:29	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	433	2024	1	8	15:45:46	Monday
3	F4A9CE78061F17F7	electric_bike	2024-01-27 12:27:19	2024-01-27 12:35:19	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	480	2024	1	27	12:27:19	Saturday
4	0A0D9E15EE50B171	classic_bike	2024-01-29 16:26:17	2024-01-29 16:56:06	Wells St & Randolph St	TA1305000030	Larrabee St & Webster Ave	13193	member	1789	2024	1	29	16:26:17	Monday
5	33FFC9805E3EFF9A	classic_bike	2024-01-31 05:43:23	2024-01-31 06:09:35	Lincoln Ave & Waveland Ave	13253	Kingsbury St & Kinzie St	KA1503000043	member	1572	2024	1	31	05:43:23	Wednesday
6	C96080812CD285C5	classic_bike	2024-01-07 11:21:24	2024-01-07 11:30:03	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	519	2024	1	7	11:21:24	Sunday
7	0EA7CB313D4F456A	classic_bike	2024-01-05 14:44:12	2024-01-05 14:53:06	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	534	2024	1	5	14:44:12	Friday
8	EE11F3A3B39CF8D8	electric_bike	2024-01-04 18:19:53	2024-01-04 18:28:04	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	491	2024	1	4	18:19:53	Thursday
9	63E83DE8E3279F15	classic_bike	2024-01-01 14:46:53	2024-01-01 14:57:02	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	609	2024	1	1	14:46:53	Monday
10	800568269122D93	electric_bike	2024-01-03 19:31:08	2024-01-03 19:40:05	Clark St & Ida B Wells Dr	TA1305000009	Kingsbury St & Kinzie St	KA1503000043	member	537	2024	1	3	19:31:08	Wednesday
11	228565685AEDD490	electric_bike	2024-01-03 07:39:20	2024-01-03 07:47:12	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	472	2024	1	3	07:39:20	Wednesday
12	133CD03CA430172	classic_bike	2024-01-03 17:03:11	2024-01-03 17:13:15	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	604	2024	1	3	17:03:11	Wednesday
13	32D57B92858025D	electric_bike	2024-01-10 17:04:09	2024-01-10 17:11:40	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	451	2024	1	10	17:04:09	Wednesday
14	B11085685C38D698	electric_bike	2024-01-12 12:35:14	2024-01-12 12:43:34	Clark St & Ida B Wells Dr	TA1305000009	Kingsbury St & Kinzie St	KA1503000043	member	500	2024	1	12	12:35:14	Friday
15	B660871085FA0938	electric_bike	2024-01-07 08:00:19	2024-01-07 08:06:46	Clark St & Ida B Wells Dr	TA1305000009	Kingsbury St & Kinzie St	KA1503000043	member	387	2024	1	7	08:00:19	Sunday
16	E7A33AC59F174164	electric_bike	2024-01-24 08:28:31	2024-01-24 08:32:13	Clark St & Randolph St	TA1305000030	Kingsbury St & Kinzie St	KA1503000043	member	222	2024	1	24	08:28:31	Wednesday
17	BAC65083CFE6160E	electric_bike	2024-01-13 11:18:54	2024-01-13 11:25:16	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	382	2024	1	13	11:18:54	Saturday
18	C6CC4B54F74526D	electric_bike	2024-01-24 18:52:19	2024-01-24 19:04:16	Sheffield Ave & Kingsbury St	13154	Aberdeen St & Randolph St	18062	member	717	2024	1	24	18:52:19	Wednesday
19	A78DDCFEDF01FC88	classic_bike	2024-01-27 13:12:24	2024-01-27 13:19:35	Wells St & Elm St	KA1504000135	Kingsbury St & Kinzie St	KA1503000043	member	431	2024	1	27	13:12:24	Saturday
20	E247F34BF19D0880	electric_bike	2024-01-05 13:15:24	2024-01-05 13:21:41	Clark St & Randolph St	TA1305000030	Aberdeen St & Randolph St	18062	member	377	2024	1	5	13:15:24	Friday

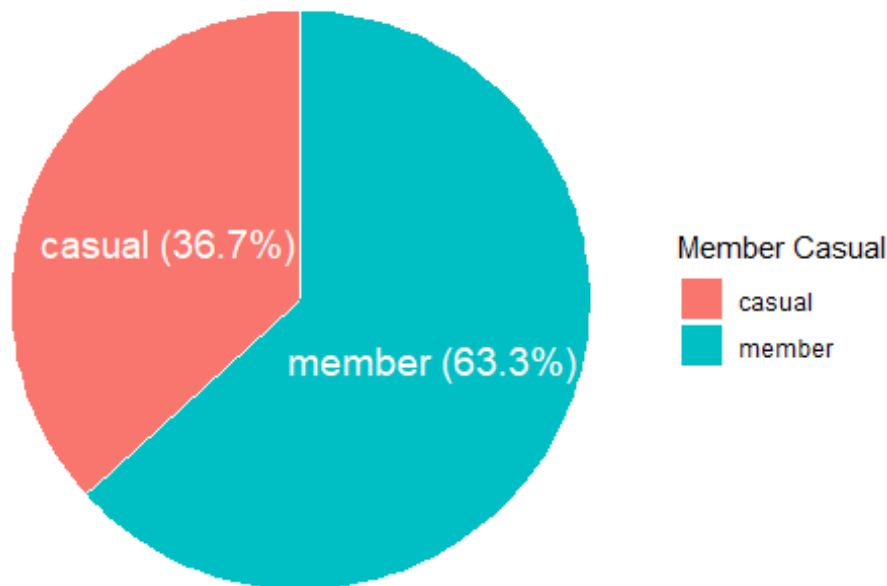
Showing 1 to 20 of 5,859,845 entries, 15 total columns

ANALYZE

Question to analyze: How do members and casual riders use Cyclist bike differently?

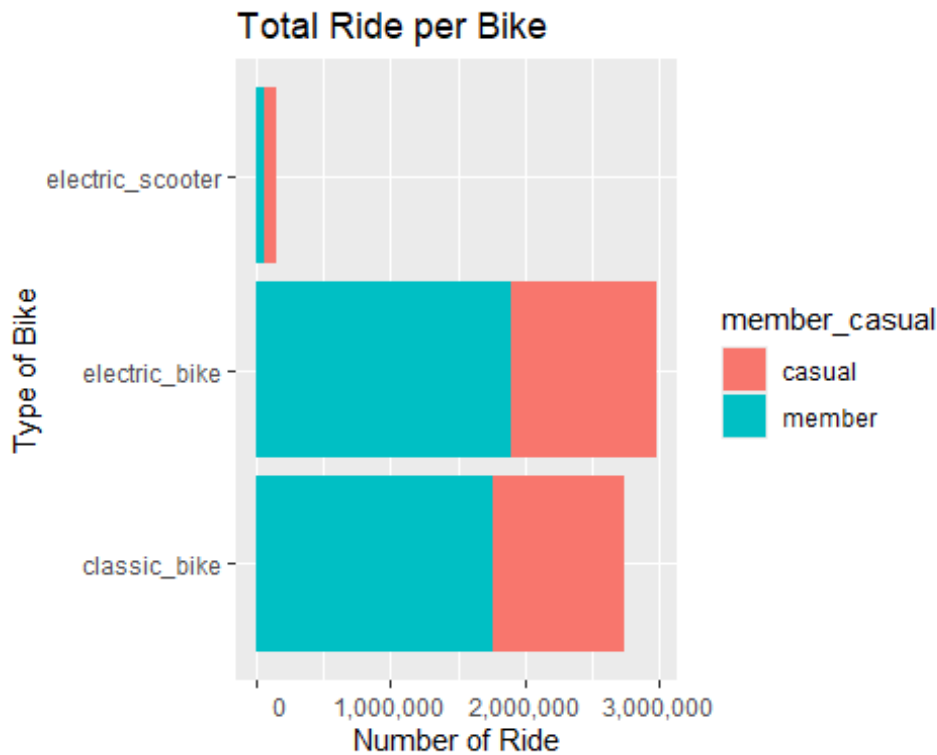
To answer this question first let's see the percentage of casual and member used

Percentage of ride by type of user



- Membership users have a higher use of bicycles compared to casual users
- Membership users use bicycles 63.3% more than casual users

In the following graph we can see which type are the most used bicycles



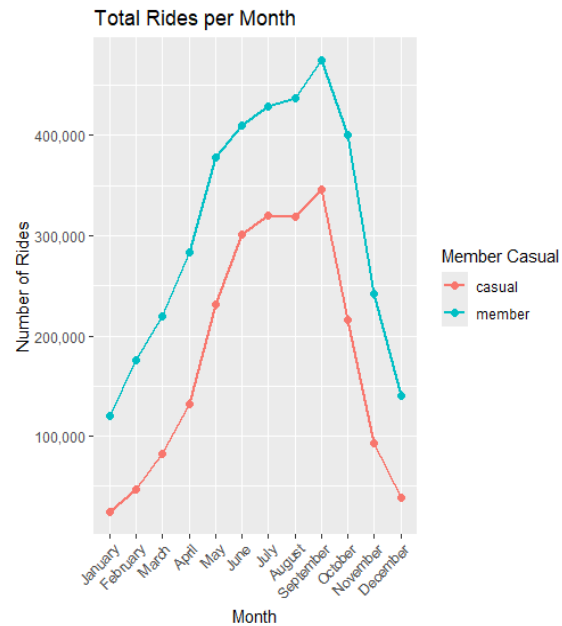
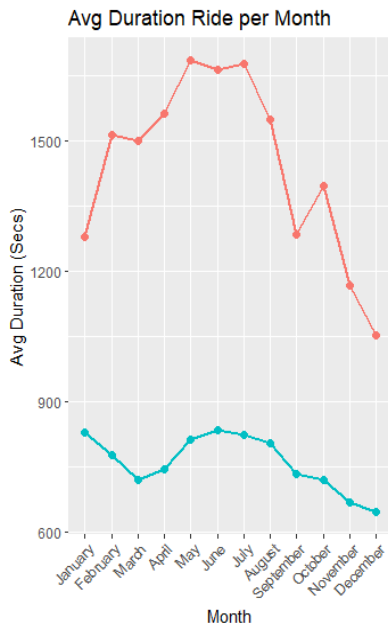
- We can see that all Cyclist users prefer bicycles over scooters
- The percentage of members who rent electric bikes is slightly higher than those who rent classic bikes
- The percentage of casual users who rent electric bikes is much higher than those who rent classic bikes

Now I will calculate the average duration of trips per month made by users with membership and casual users

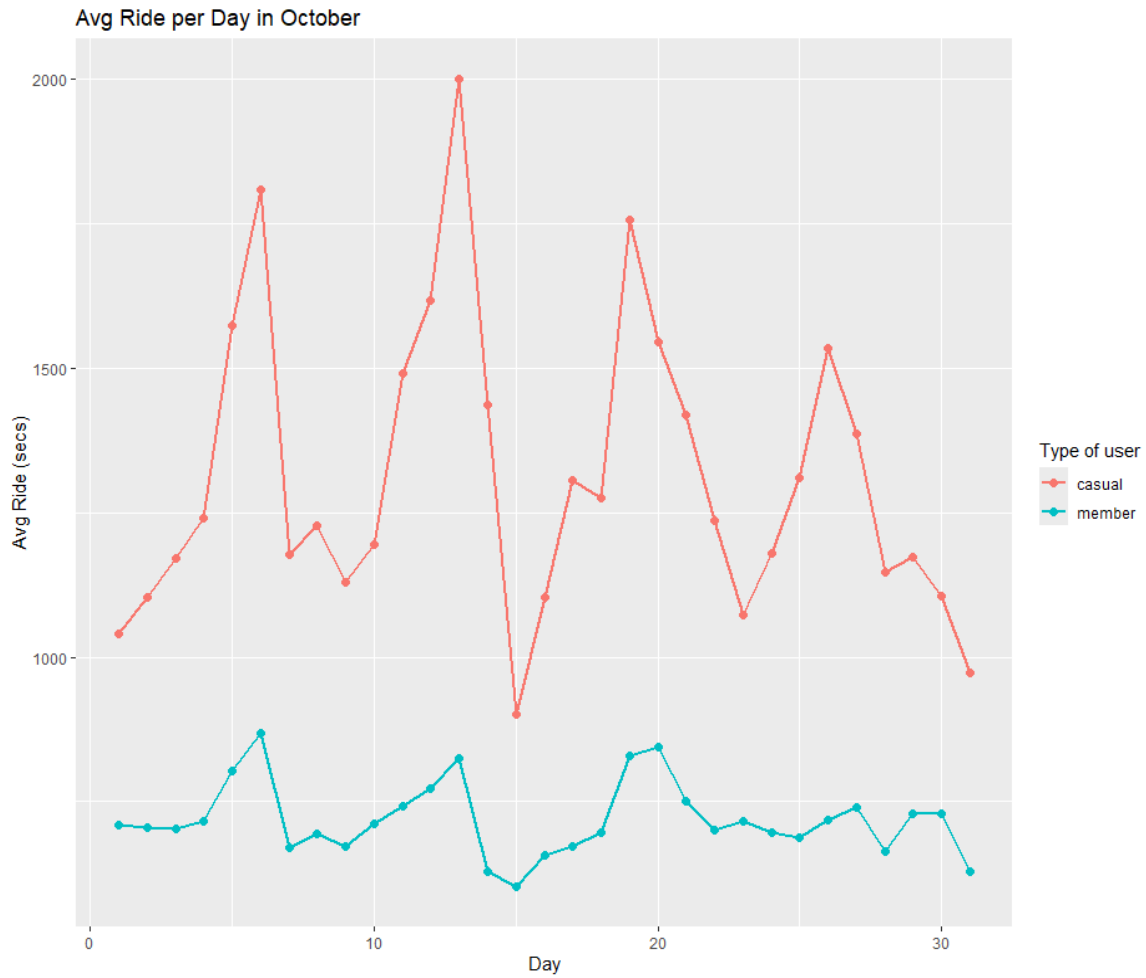
```
avg_duration_ride <- all_trips_24_v2 %>%
  group_by(member_casual, month) %>%
  summarise(avg_duration = mean(ride_length), .groups = "drop")
```

Now let's compare the number of rides per month and the average duration of the rides between members and casual users


```
# Agrupamos los datos por mes y se calcula el numero total de rides
month_rides <- all_trips_24_v2 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_ride = n(), .groups = "drop") %>%
  arrange(member_casual, month)
```



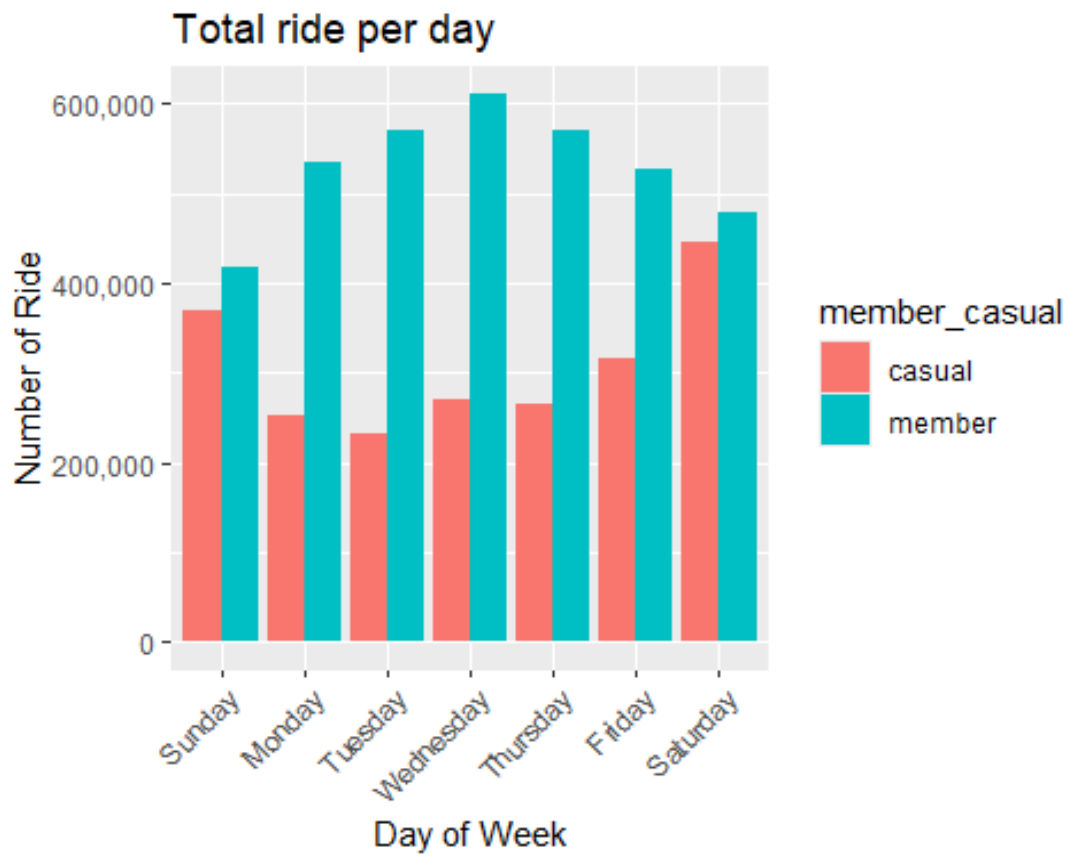
- It can be observed that users with memberships make more trips per month
- Casual users make longer trips
- We can also observe that there is a decrease in the number of trips in the month of September - October
- It can be observed that in October there was an increase in the ride time of casual users, this may be due to the holidays, Let's see if there is a correlation



- As we can see in the graph, there is no correlation between the holidays and the time spent using the bikes

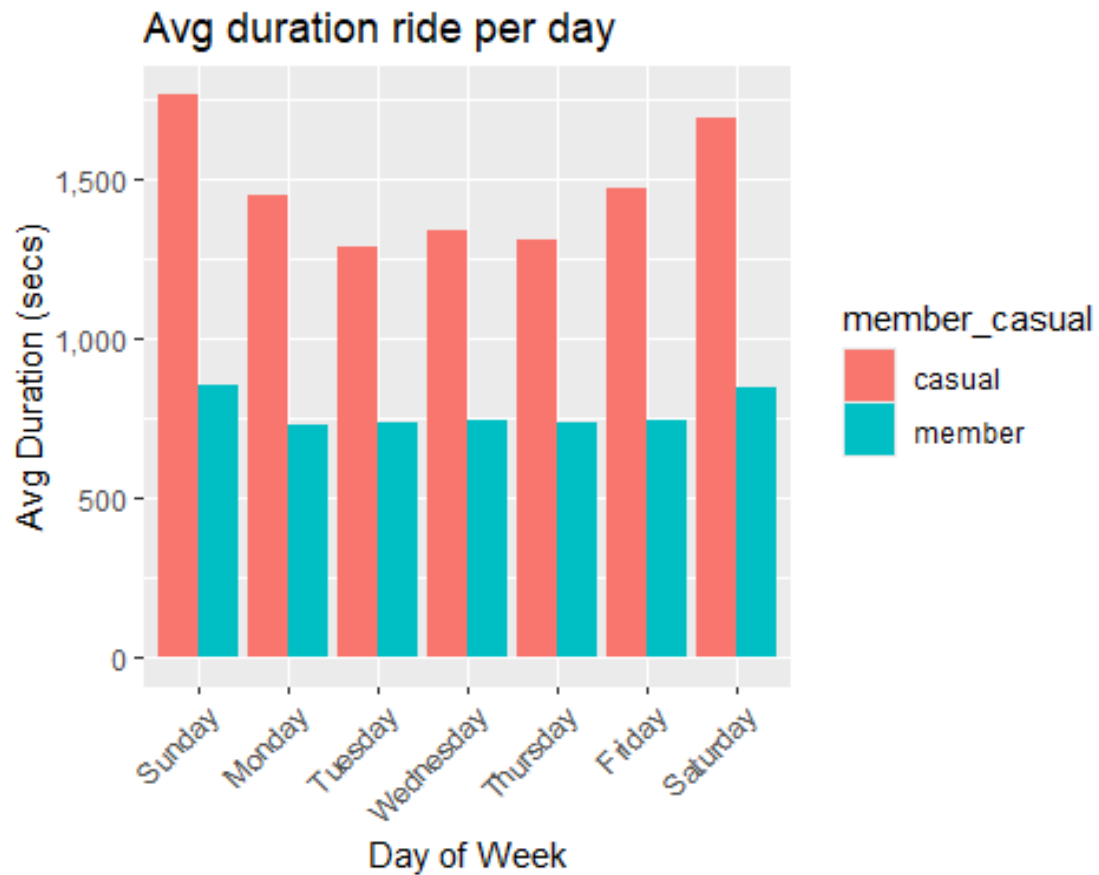
Now let's compare the number of rides per day and the average duration of the rides between members and casual user

Number of Rides per day



- We can see that member make more rides per day that casual
- We can see that Wednesdays are the days where members take the most rides
- Also, we can see that Saturdays are the days where casuals take the most rides

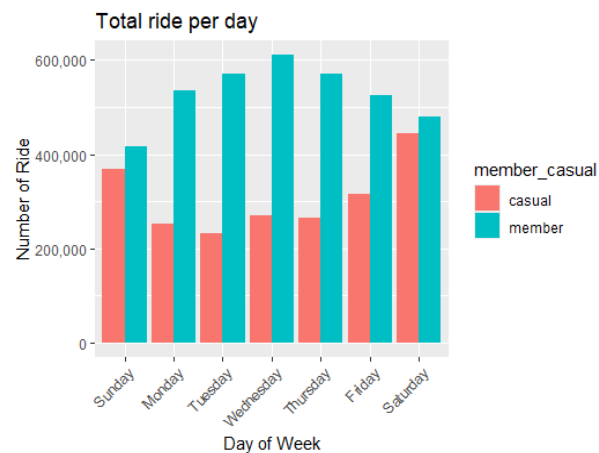
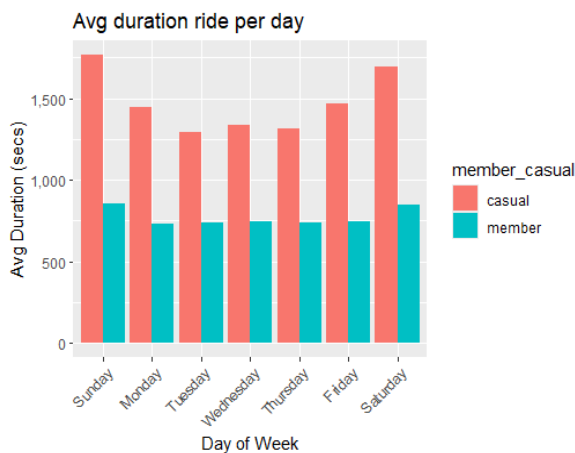
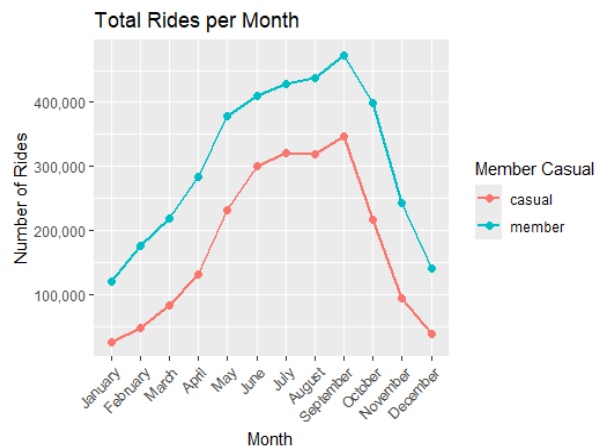
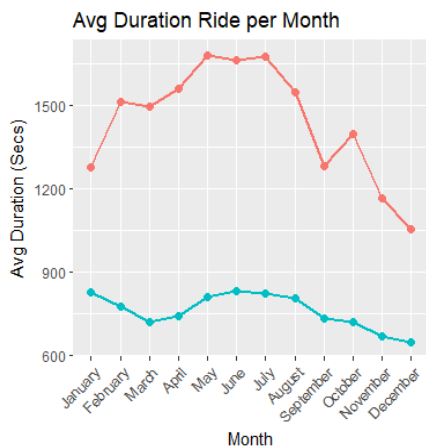
Avg Duration of Ride per day



- In the graph, we can see that casual users take much longer rides than member users
- It can also be seen that the days with the longest average trips are Sundays, followed by Saturdays

Share

Dashboard



How do members and casual riders use Cyclist bike differently?

- Based on the analysis, we can see that members use bicycles more frequently during the week, while casual users use them on weekends, with a higher percentage on Saturdays
- This may be because casual users prefer to use bicycles as a means of tourism or touring, while members use them as a mode of transportation
- It can also be observed that there is a decrease in use starting in September by both types of users, this may be because it is the end of summer

Act

Recommendations

Based on the analysis done the marketing campaign to convert casual users to members users, my suggestions would be the following:

- **Create custom memberships**
 - Custom memberships can be created to target a different group of users, for example, a monthly or seasonal membership, since there is a large group of users who use the service during the summer and may be interested in obtaining a membership during that time only
- **Give a discount for adding friends**
 - As can be seen in the graphs, casual users use the Cyclist service on Saturdays, just like member users. It's very likely that in a large group of rides that take place on that day will include both member and casual users, so it would be a good strategy to offer discounts to member users for adding their casual user friends to the member users
- **Seasonal Campaigns**
 - Create seasonal campaigns, where prices can be lowered during periods such as winter, when membership and casual users use the Cyclist service less frequently. Launch promotions for weekends, holidays, or longer trips for members that encourage cycling
- **Social Media**
 - Based on the previous points, you can create marketing campaigns on social media that tell stories about how casual users and members use the Cyclist service. These stories can include how the Cyclist service has helped them get around the city or how it has served as a distraction from a stressful week at work