

## **CYCLISTIC ANALYSIS REPORT**

### **COMPANY:**

Cyclistic, a bike-share company in Chicago.

### **PROBLEM STATEMENT:**

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.

### **AUDIENCE AND CHARACTERS:**

- Cyclistic Executive Team
- Lily Moreno: The director of marketing
- Cyclistic marketing analytics team

### **DETAILS AVAILABLE:**

Cyclistic currently has three pricing plans,

- 1- Single-Ride Pass
- 2- Full Day Pass
- 3- Annual Membership

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

Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers.

Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

### **BUSINESS USE CASE:**

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

### **GOALS AND METRICS:**

Design marketing strategies aimed at converting casual riders into annual members. To do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics.

## DELIVERABLES:

A report with the following:

1. A clear statement of the business task.
2. A description of all data sources used.
3. Documentation of any cleaning or manipulation of data.
4. A summary of analysis.
5. Supporting visualizations and key findings.
6. Top three recommendations based on analysis.

---







## BUSINESS TASK:

- Improve the future success of the company by increasing the number of people purchasing annual memberships.
- Converting casual riders into annual members through digital marketing.
- To identify the differences between annual members and casual rider's usage of cyclicistic bikes.
- How to make casual riders buy annual membership by providing extensive benefits.







---

## DATA SOURCES USED:

- Cyclicistic Bike usage of last six months i.e. 2 quarters datasets based on the availability is taken as data input for analysing. Each CSV file represents a trip data for each month from June to November 2023.

Name	Date modified	Type	Size
 202306-divvy-tripdata	28-12-2023 12:18	Microsoft Excel ...	1,40,974...
 202307-divvy-tripdata	28-12-2023 12:19	Microsoft Excel ...	1,50,255...
 202308-divvy-tripdata	28-12-2023 12:19	Microsoft Excel ...	1,51,472...
 202309-divvy-tripdata	28-12-2023 12:20	Microsoft Excel ...	1,31,016...
 202310-divvy-tripdata	28-12-2023 12:19	Microsoft Excel ...	1,05,363...
 202311-divvy-tripdata	28-12-2023 12:18	Microsoft Excel ...	71,319 KB

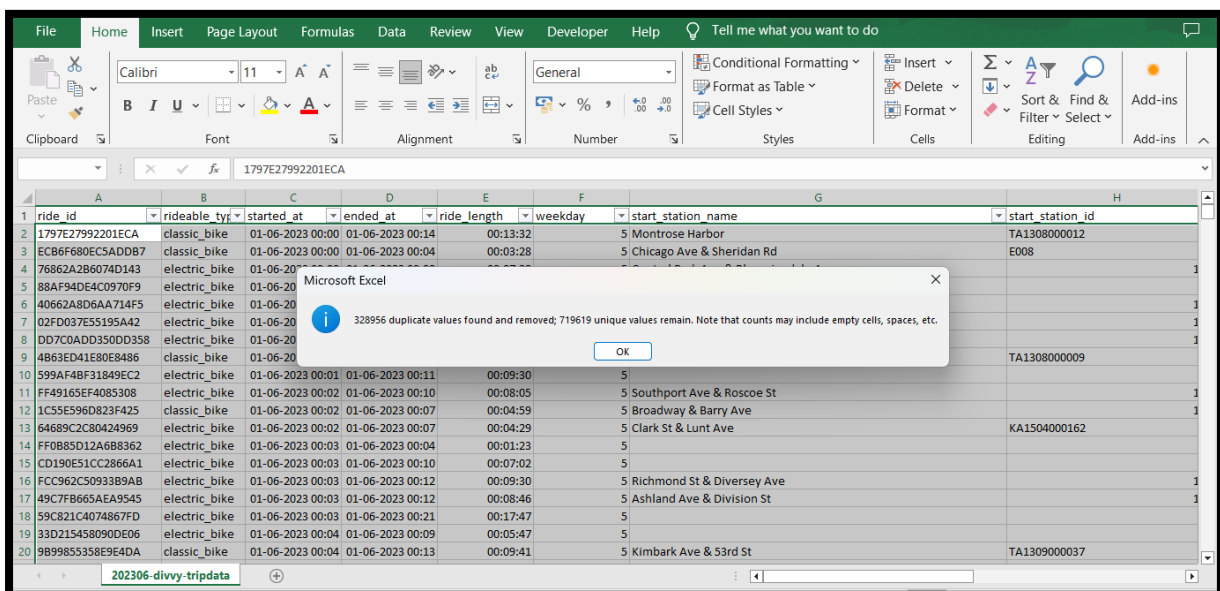
- Renamed all the 6 CSV files according to their month names (June to November).

Name	Date modified	Type	Size
 06-2023	28-12-2023 15:21	Microsoft Excel ...	70,645 KB
 07-2023	28-12-2023 15:23	Microsoft Excel ...	75,449 KB
 08-2023	28-12-2023 15:23	Microsoft Excel ...	75,667 KB
 09-2023	28-12-2023 15:22	Microsoft Excel ...	65,453 KB
 10-2023	28-12-2023 15:24	Microsoft Excel ...	52,343 KB
 11-2023	28-12-2023 15:22	Microsoft Excel ...	35,709 KB

- The data is well organized and no issues can be raised in terms of bias or integrity as this is a first-party data collected directly from inside the organization.
- The credibility comes under the radar which is corrected in the upcoming steps.
- The data gathered is **Reliable, Original, Comprehensive, Current and Cited**.
- The data gathered can be used to separate and analyse the behaviour of causal rider and annual member's ride.

## PROCESSING DATA FOR ANALYSIS:

- Removed duplicate rows from the data.



ride_id	rideable_type	started_at	ended_at	ride_length	weekday	start_station_name	start_station_id
1797E27992201ECA	classic_bike	01-06-2023 00:00	01-06-2023 00:14	00:13:32	5	Montrose Harbor	TA1308000012
ECB6F680EC5ADD87	classic_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
76862A2B6074D143	electric_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
88AF94DE4C0970F9	electric_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
40662A8D6AA714F5	electric_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
02FD037E55195A42	electric_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
DD7C0ADD350DD358	electric_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
4B63ED41E80E8486	classic_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28	5	Chicago Ave & Sheridan Rd	E008
599AF4BF31849EC2	electric_bike	01-06-2023 00:01	01-06-2023 00:11	00:09:30	5	Southport Ave & Roscoe St	TA1308000009
FF49165EF4085308	electric_bike	01-06-2023 00:02	01-06-2023 00:10	00:08:05	5	Southport Ave & Roscoe St	TA1308000009
1C55E596D823F425	classic_bike	01-06-2023 00:02	01-06-2023 00:07	00:04:59	5	Broadway & Barry Ave	KA1504000162
64689C2C80424969	electric_bike	01-06-2023 00:02	01-06-2023 00:07	00:04:29	5	Clark St & Lunt Ave	KA1504000162
FF0885D12A6B8362	electric_bike	01-06-2023 00:03	01-06-2023 00:04	00:01:23	5	Clark St & Lunt Ave	KA1504000162
CD190E51C2B86A1	electric_bike	01-06-2023 00:03	01-06-2023 00:10	00:07:02	5	Clark St & Lunt Ave	KA1504000162
FCC962C50933B9A8	electric_bike	01-06-2023 00:03	01-06-2023 00:12	00:09:30	5	Richmond St & Diversey Ave	TA1309000037
49C7F8665AEA9545	electric_bike	01-06-2023 00:03	01-06-2023 00:12	00:08:46	5	Ashland Ave & Division St	TA1309000037
59C821C4074867FD	electric_bike	01-06-2023 00:03	01-06-2023 00:21	00:17:47	5	Ashland Ave & Division St	TA1309000037
33D215458090DE06	electric_bike	01-06-2023 00:04	01-06-2023 00:09	00:05:47	5	Kimbark Ave & 53rd St	TA1309000037
9B9985358E9E4DA	classic_bike	01-06-2023 00:04	01-06-2023 00:13	00:09:41	5	Kimbark Ave & 53rd St	TA1309000037

- Created two different columns “ride\_length” indicating the time duration of each ride and “Weekday” day of the week in which the ride is happening.
- Deleted columns start\_station\_id, end\_station\_id, start\_lat, start\_lng, end\_lat, end\_lng which might not be used during the analysis phase.
- Sorted the data by starting date of each month by “started\_at” column values.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	ride_id	rideable_type	started_at	ended_at	ride_length	weekday	member_casual								
1	1797E27992201ECA	classic_bike	01-06-2023 00:00	01-06-2023 00:14	00:13:32		5 member								
2	ECB6F680EC5ADD8B7	classic_bike	01-06-2023 00:00	01-06-2023 00:04	00:03:28		5 member								
3	76862A2B6074D0143	electric_bike	01-06-2023 00:00	01-06-2023 00:08	00:07:28		5 casual								
4	88AF94DE4C0970F9	electric_bike	01-06-2023 00:00	01-06-2023 00:11	00:10:19		5 member								
5	40662A8D6AA714F5	electric_bike	01-06-2023 00:00	01-06-2023 00:08	00:07:22		5 casual								
6	02FD037E55195A42	electric_bike	01-06-2023 00:01	01-06-2023 00:02	00:01:53		5 member								
7	DD7C0ADD350DD358	electric_bike	01-06-2023 00:01	01-06-2023 00:08	00:07:06		5 member								
8	4B63ED41E80E8486	classic_bike	01-06-2023 00:01	01-06-2023 00:11	00:09:40		5 casual								
9	599AF48F31849EC2	electric_bike	01-06-2023 00:01	01-06-2023 00:11	00:09:30		5 member								
10	FF49165EF4085308	electric_bike	01-06-2023 00:02	01-06-2023 00:10	00:08:05		5 casual								
11	1C55E596D823F425	classic_bike	01-06-2023 00:02	01-06-2023 00:07	00:04:59		5 casual								
12	64689C2C80424969	electric_bike	01-06-2023 00:02	01-06-2023 00:07	00:04:29		5 casual								
13	FF0B85D12A6B8362	electric_bike	01-06-2023 00:03	01-06-2023 00:04	00:01:23		5 casual								
14	CD190E51CC2866A1	electric_bike	01-06-2023 00:03	01-06-2023 00:10	00:07:02		5 member								
15	FCC962C50933B9AB	electric_bike	01-06-2023 00:03	01-06-2023 00:12	00:09:30		5 member								
16	49C7F8665AEA9545	electric_bike	01-06-2023 00:03	01-06-2023 00:12	00:08:46		5 casual								
17	59C821C4074867FD	electric_bike	01-06-2023 00:03	01-06-2023 00:21	00:17:47		5 member								
18	33D215458090DE06	electric_bike	01-06-2023 00:04	01-06-2023 00:09	00:05:47		5 member								
19	9B99855358E9E4DA	classic_bike	01-06-2023 00:04	01-06-2023 00:13	00:09:41		5 member								

- Deleted ride\_length with negative values.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	ride_id	rideable_type	started_at	ended_at	ride_length	weekday	member_casual								
719607	53910A01EFA24C57	electric_bike	29-06-2023 19:20	29-06-2023 19:20	00:00:00		5 casual								
719608	0553FC62006EFF5E	electric_bike	29-06-2023 23:51	29-06-2023 23:51	00:00:00		5 member								
719609	E15B674214400250	electric_bike	30-06-2023 12:49	30-06-2023 12:49	00:00:00		6 casual								
719610	E2D9AAA7AF8CBD58	electric_bike	30-06-2023 14:12	30-06-2023 14:12	00:00:00		6 member								
719611	C69A757F13070A38	electric_bike	30-06-2023 14:46	30-06-2023 14:46	00:00:00		6 casual								
719612	0F2B4F735EF92042	electric_bike	30-06-2023 17:51	30-06-2023 17:51	00:00:00		6 member								
719613	FAC4E90497237BD1	classic_bike	01-06-2023 16:47	01-06-2023 16:47	#####		5 member								
719614	199D68CEC1DACF3	electric_bike	10-06-2023 15:49	10-06-2023 15:49	#####		7 casual								
719615	0F2C8AC039F63D8F	electric_bike	08-06-2023 16:31	08-06-2023 16:31	#####		5 member								
719616	3FC95E908B945FE2	electric_bike	04-06-2023 18:14	04-06-2023 18:14	#####		1 casual								
719617	318E29DFACCD7C24	electric_bike	02-06-2023 23:26	02-06-2023 23:10	#####		6 member								
719618	D12C3767DC204369	electric_bike	02-06-2023 19:29	02-06-2023 18:40	#####		6 casual								
719619	8B6E5BA70093AAB7	electric_bike	02-06-2023 19:29	02-06-2023 18:28	#####		6 casual								
719620															
719621															
719622															
719623															
719624															
719625															
719626															

Repeated the same process in all 5-remaining data.

Dataset Name	Total Duplicates	Negative values in Ride Length
06 June	3,28,956	6
07 July	0	30
08 August	0	60
09 September	0	50
10 October	0	36
11 November	0	64

The June data is the only dataset where duplicate values were found which needs to be verified again.

The above data are removed as a process of removing values with errors and unwanted data.

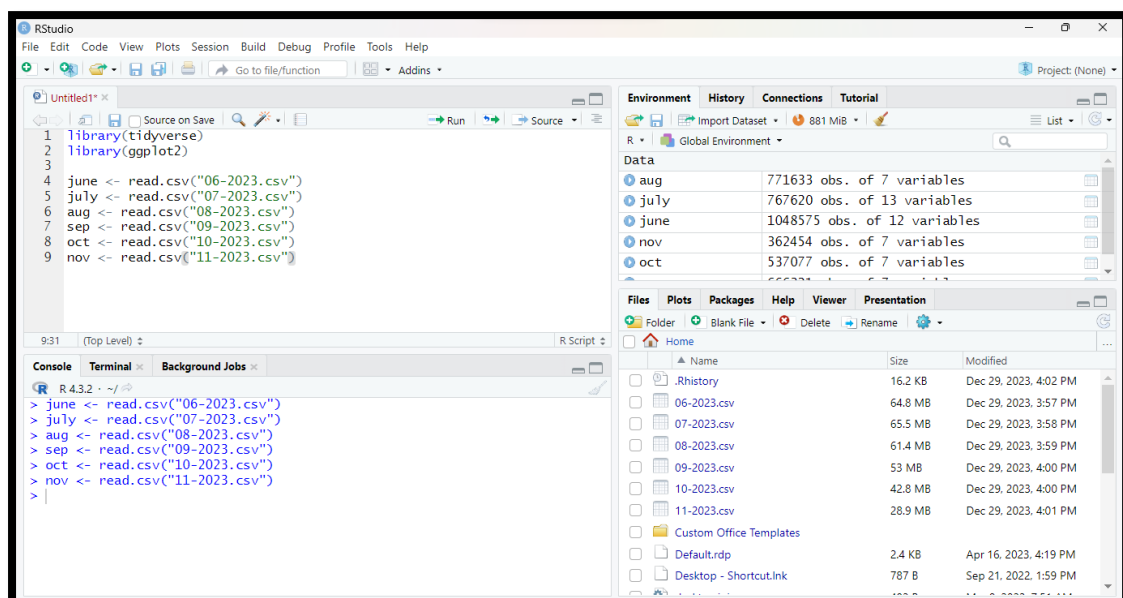
**ROWS BEFORE CLEANING - 41,15,368**

**ROWS AFTER CLEANING – 31,58,395**

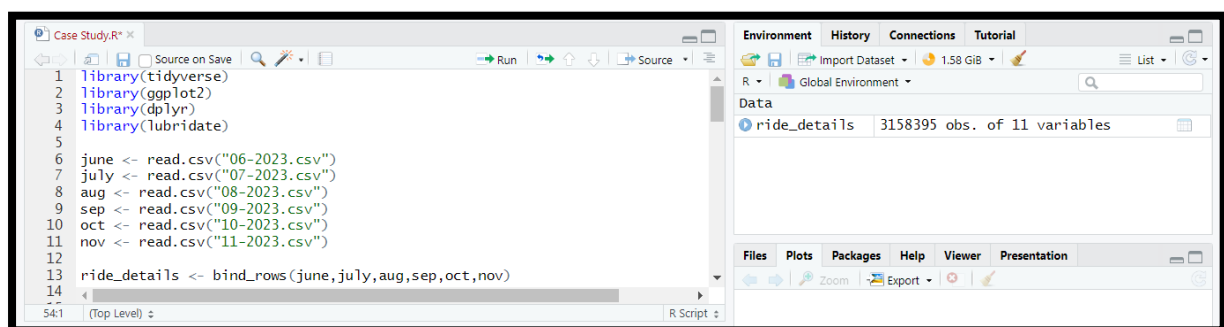
---

## ANALYSING DATA FOR INSIGHTS

- With around 7,00,000+ data in each data it is preferred to use R Studio to analyse as Excel cannot be efficient with more than 10,00,000 of data.



- All data has been imported into R Studio and stored in respective data frames.

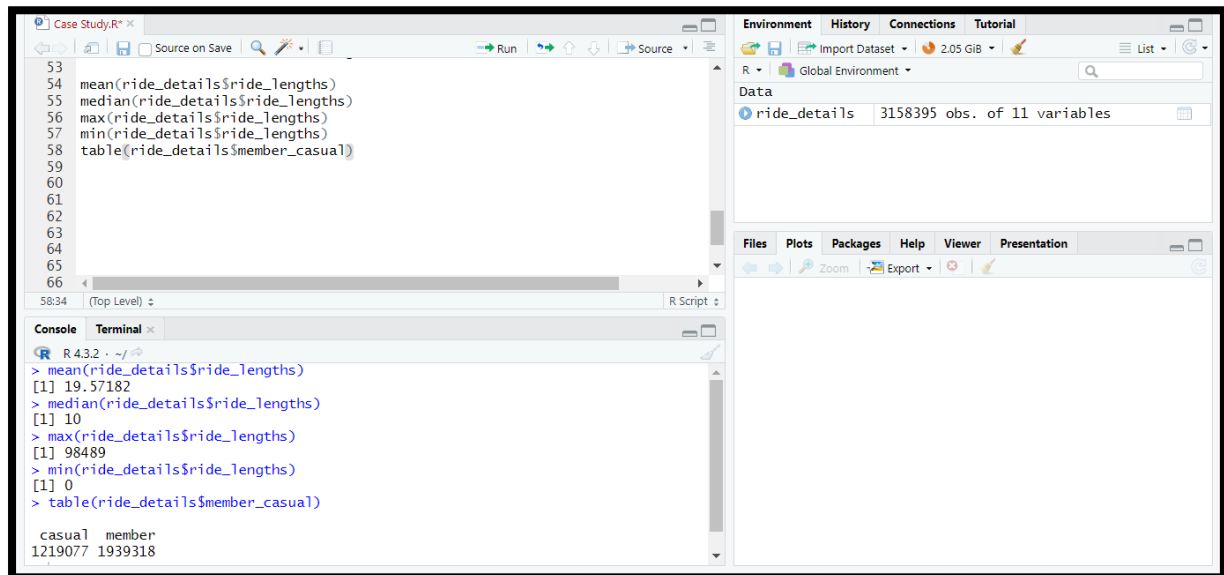


- All 6 data frames are combined into a single data frame named “ride\_details” and total number of rows is 31,58,395.

## INSIGHTS GATHERED:

Total Rides by Casual Riders: 12,19,077

Total Rides by Annual Members: 19,39,318



The screenshot shows the RStudio interface. The script editor contains the following R code:

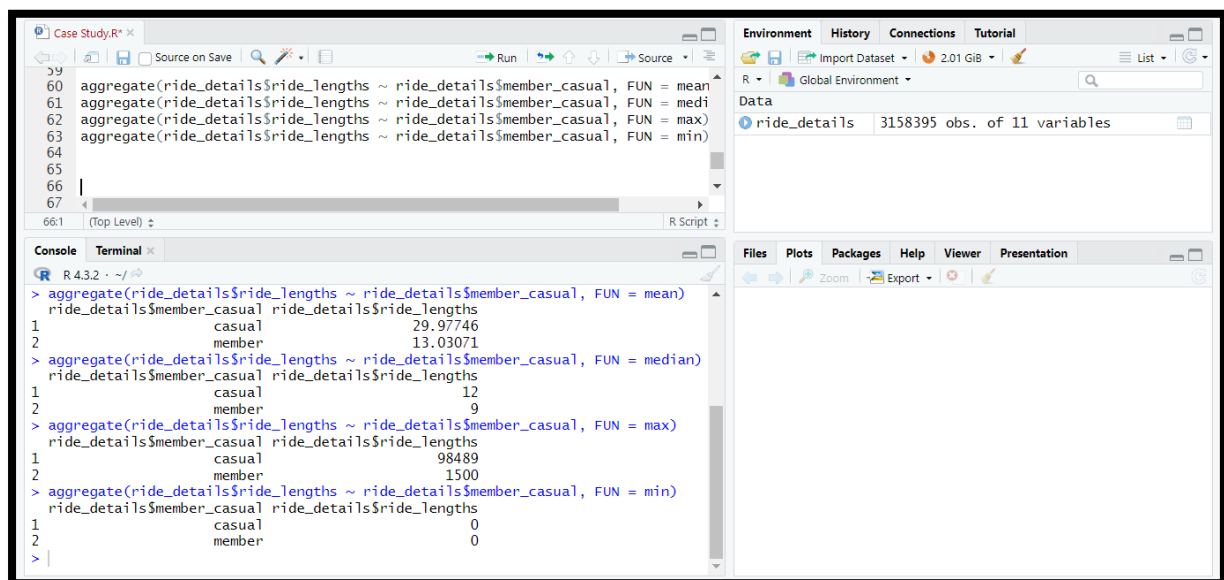
```
53  
54 mean(ride_details$ride_lengths)  
55 median(ride_details$ride_lengths)  
56 max(ride_details$ride_lengths)  
57 min(ride_details$ride_lengths)  
58 table(ride_details$member_casual)  
59  
60  
61  
62  
63  
64  
65  
66
```

The console shows the output of these commands:

```
R 4.3.2 ~ /  
> mean(ride_details$ride_lengths)  
[1] 19.57182  
> median(ride_details$ride_lengths)  
[1] 10  
> max(ride_details$ride_lengths)  
[1] 98489  
> min(ride_details$ride_lengths)  
[1] 0  
> table(ride_details$member_casual)  
  
casual member  
1219077 1939318
```

The Environment pane on the right shows the 'ride\_details' dataset with 3158395 observations and 11 variables.

The maximum ride duration is 98,489 minutes which is about 1,641.483 hours which is about 68.3 days done by a Casual Rider which rises an alert that the bikes are being properly returned to a n end station.



The screenshot shows the RStudio interface. The script editor contains the following R code:

```
59  
60 aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = mean)  
61 aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = median)  
62 aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = max)  
63 aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = min)  
64  
65  
66  
67
```

The console shows the output of these aggregate commands:

```
R 4.3.2 ~ /  
> aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = mean)  
ride_details$member_casual ride_details$ride_lengths  
1 casual 29.97746  
2 member 13.03071  
> aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = median)  
ride_details$member_casual ride_details$ride_lengths  
1 casual 12  
2 member 9  
> aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = max)  
ride_details$member_casual ride_details$ride_lengths  
1 casual 98489  
2 member 1500  
> aggregate(ride_details$ride_lengths ~ ride_details$member_casual, FUN = min)  
ride_details$member_casual ride_details$ride_lengths  
1 casual 0  
2 member 0  
>
```

The Environment pane on the right shows the 'ride\_details' dataset with 3158395 observations and 11 variables.

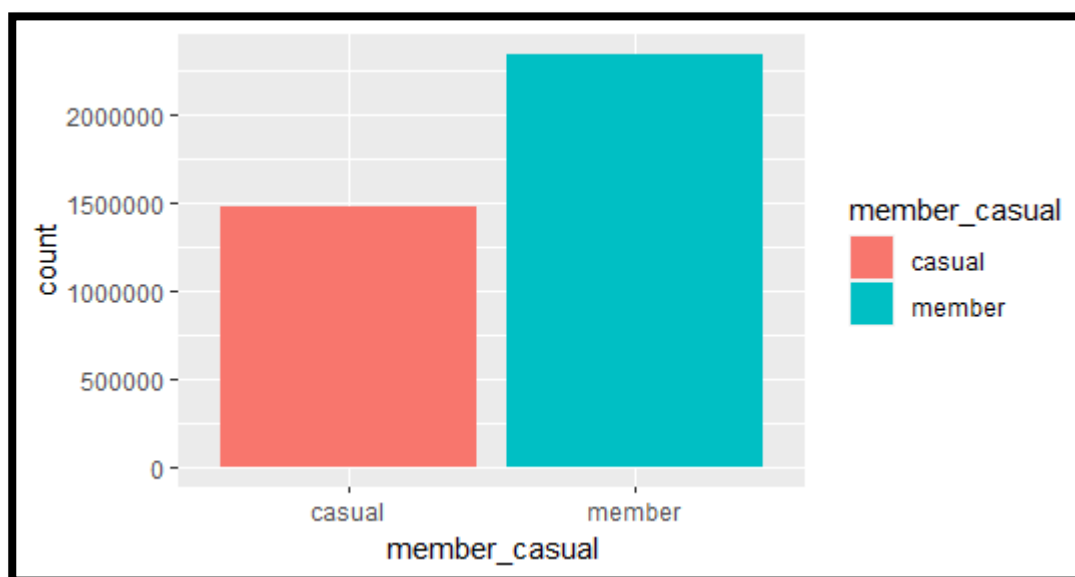
The maximum ride duration by an Annual Member is 1500 minutes which is around 25 hours

```

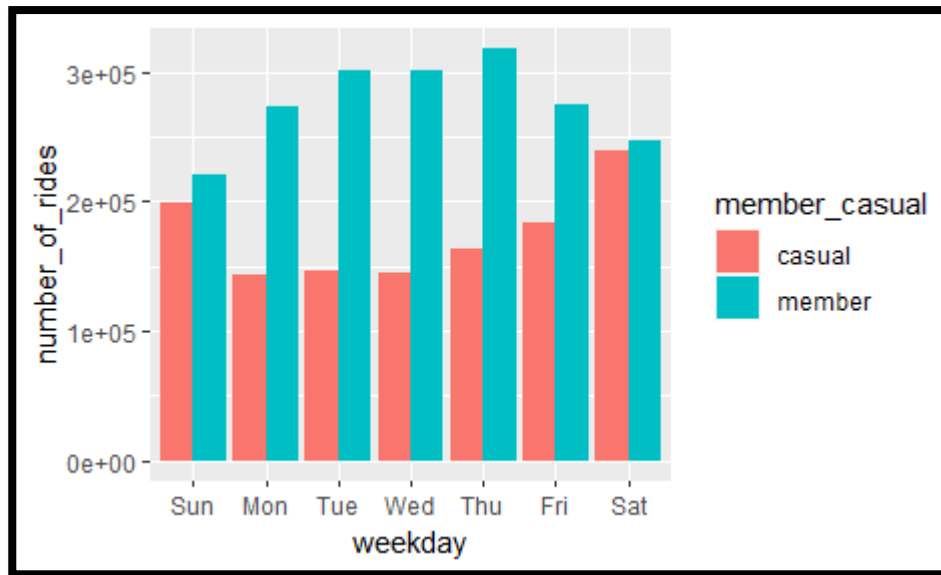
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Case Study.R x ride_details x
68 ride_details%>%
69 mutate(weekday = wday(starts_at, label = TRUE)) %>%
70 group_by(member_casual, weekday) %>%
71 summarise(number_of_rides = n(), average_duration =
72 mean(ride_lengths)) %>%
73 arrange(member_casual, weekday) |
74
75
76
73:35 (Top Level) R Script
Console Terminal
R 4.3.2 ~ /
member_casual weekday number_of_rides average_duration
<chr> <ord> <int> <dbl>
1 casual Sun 199199 34.3
2 casual Mon 142696 29.1
3 casual Tue 145841 26.8
4 casual Wed 144373 26.3
5 casual Thu 164150 26.4
6 casual Fri 183656 29.7
7 casual Sat 239162 33.8
8 member Sun 220852 14.5
9 member Mon 274171 12.4
10 member Tue 301908 12.5
11 member Wed 301546 12.5
12 member Thu 318375 12.5
13 member Fri 275165 13.0
14 member Sat 247301 14.5
>

```

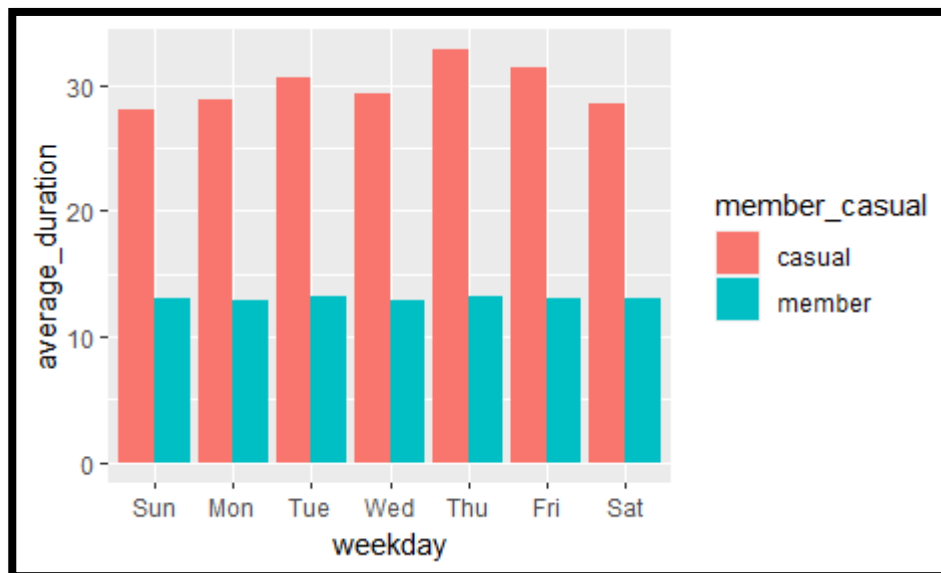
The number of rides by casual riders increase during weekends and decreases for the annual riders during weekends.



Total Rides vs Member Types



Number of Rides vs Member Types by Weekdays



Average Ride Duration vs Member Types by Weekdays

- Based on the data we can clearly see that Annual Members are consistent throughout the week in terms of travel duration and number of rides see's a slight dip at weekends.
- Whereas for casual riders there is a spike in weekends in terms of no. of rides, which is where we can make an improvement to convert casual riders into annual members.

### RECOMMENDATIONS:

- We can add a weekend annual pass which might attract casual riders.
- An annual pass which can be bought for minimum price and can be recharged whenever required as the credits gets low, which can make the casual riders think of the new pass.
- This might not be the best of other ideas where we can reduce the pass available on Weekends and increase the Weekend annual passes which might customers buy it.