

Case Study: How Does a Bike-Share Navigate Speedy Success?

Alexandre Santana

2022-08-21

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.

Characters and teams

- **Cyclistic:** A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.
- **Lily Moreno:** The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.
- **Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic's mission and business goals — as well as how you, as a junior data analyst, can help Cyclistic achieve them.
- **Cyclistic executive team:** The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

About the company

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. Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day

passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members. Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, 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. Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order 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. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

1) Ask

Three questions will guide the future marketing program:

- 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?

2) Prepare

Downloaded 12 months of Cyclistic trip data from August 2021 until July 2022. (Note: The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this license).

3) Process

Data processing, analysis and visualizations will be performed using RStudio/R programming language.

- Setting up environment

Remarks: Setting up R environment by loading the 'tidyverse', 'lubridate', 'skimr' and 'scales' packages.

```
library(tidyverse)
library(lubridate)
library(skimr)
library(scales)
```

- Load datasets

```
tripdata_202108 <- read.csv("202108-divvy-tripdata.csv")
tripdata_202109 <- read.csv("202109-divvy-tripdata.csv")
```

```
tripdata_202110 <- read.csv("202110-divvy-tripdata.csv")
tripdata_202111 <- read.csv("202111-divvy-tripdata.csv")
tripdata_202112 <- read.csv("202112-divvy-tripdata.csv")
tripdata_202201 <- read.csv("202201-divvy-tripdata.csv")
tripdata_202202 <- read.csv("202202-divvy-tripdata.csv")
tripdata_202203 <- read.csv("202203-divvy-tripdata.csv")
tripdata_202204 <- read.csv("202204-divvy-tripdata.csv")
tripdata_202205 <- read.csv("202205-divvy-tripdata.csv")
tripdata_202206 <- read.csv("202206-divvy-tripdata.csv")
tripdata_202207 <- read.csv("202207-divvy-tripdata.csv")
```

- Observe raw data

```
str(tripdata_202108)
```

```
## 'data.frame': 804352 obs. of 13 variables:
## $ ride_id : chr "99103BB87CC6C1BB" "EAFCCCFB0A3FC5A1"
## $ rideable_type : chr "electric_bike" "electric_bike"
## $ started_at : chr "2021-08-10 17:15:49" "2021-08-10 17:23:14"
## $ ended_at : chr "2021-08-10 17:22:44" "2021-08-10 17:39:24"
## $ start_station_name: chr "" "" "" "" ...
## $ start_station_id : chr "" "" "" "" ...
## $ end_station_name : chr "" "" "" "" ...
## $ end_station_id : chr "" "" "" "" ...
## $ start_lat : num 41.8 41.8 42 42 41.8 ...
## $ start_lng : num -87.7 -87.7 -87.7 -87.7 -87.6 ...
## $ end_lat : num 41.8 41.8 42 42 41.8 ...
## $ end_lng : num -87.7 -87.6 -87.7 -87.7 -87.6 ...
## $ member_casual : chr "member" "member" "member" "member" ...
```

```
str(tripdata_202109)
```

```
## 'data.frame': 756147 obs. of 13 variables:
## $ ride_id : chr "9DC7B962304CBFD8" "F930E2C6872D6B32"
## $ rideable_type : chr "electric_bike" "electric_bike"
## $ started_at : chr "2021-09-28 16:07:10" "2021-09-28 14:24:51"
## $ ended_at : chr "2021-09-28 16:09:54" "2021-09-28 14:40:05"
## $ start_station_name: chr "" "" "" "" ...
## $ start_station_id : chr "" "" "" "" ...
## $ end_station_name : chr "" "" "" "" ...
## $ end_station_id : chr "" "" "" "" ...
## $ start_lat : num 41.9 41.9 41.8 41.8 41.9 ...
## $ start_lng : num -87.7 -87.6 -87.7 -87.7 -87.7 ...
```

Google Data Analytics Capstone Project

```
## $ end_lat      : num  41.9 42 41.8 41.8 41.9 ...
## $ end_lng      : num  -87.7 -87.7 -87.7 -87.7 -87.7 ...
## $ member_casual : chr   "casual" "casual" "casual" "casual" ...

str(tripdata_202110)

## 'data.frame':   631226 obs. of  13 variables:
## $ ride_id      : chr   "620BC6107255BF4C" "4471C70731AB2E45"
##                "26CA69D43D15EE14" "362947F0437E1514" ...
## $ rideable_type : chr   "electric_bike" "electric_bike"
##                "electric_bike" "electric_bike" ...
## $ started_at    : chr   "2021-10-22 12:46:42" "2021-10-21 09:12:37"
##                "2021-10-16 16:28:39" "2021-10-16 16:17:48" ...
## $ ended_at      : chr   "2021-10-22 12:49:50" "2021-10-21 09:14:14"
##                "2021-10-16 16:36:26" "2021-10-16 16:19:03" ...
## $ start_station_name: chr   "Kingsbury St & Kinzie St" "" "" "" ...
## $ start_station_id : chr   "KA1503000043" "" "" "" ...
## $ end_station_name : chr   "" "" "" "" ...
## $ end_station_id   : chr   "" "" "" "" ...
## $ start_lat        : num   41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng         : num  -87.6 -87.7 -87.7 -87.7 -87.7 ...
## $ end_lat           : num   41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng           : num  -87.6 -87.7 -87.7 -87.7 -87.7 ...
## $ member_casual     : chr   "member" "member" "member" "member" ...

str(tripdata_202111)

## 'data.frame':   359978 obs. of  13 variables:
## $ ride_id      : chr   "7C00A93E10556E47" "90854840DFD508BA"
##                "0A7D10CDD144061C" "2F3BE33085BCFF02" ...
## $ rideable_type : chr   "electric_bike" "electric_bike"
##                "electric_bike" "electric_bike" ...
## $ started_at    : chr   "2021-11-27 13:27:38" "2021-11-27 13:38:25"
##                "2021-11-26 22:03:34" "2021-11-27 09:56:49" ...
## $ ended_at      : chr   "2021-11-27 13:46:38" "2021-11-27 13:56:10"
##                "2021-11-26 22:05:56" "2021-11-27 10:01:50" ...
## $ start_station_name: chr   "" "" "" "" ...
## $ start_station_id : chr   "" "" "" "" ...
## $ end_station_name : chr   "" "" "" "" ...
## $ end_station_id   : chr   "" "" "" "" ...
## $ start_lat        : num   41.9 42 42 41.9 41.9 ...
## $ start_lng         : num  -87.7 -87.7 -87.7 -87.8 -87.6 ...
## $ end_lat           : num   42 41.9 42 41.9 41.9 ...
## $ end_lng           : num  -87.7 -87.7 -87.7 -87.8 -87.6 ...
## $ member_casual     : chr   "casual" "casual" "casual" "casual" ...

str(tripdata_202112)

## 'data.frame':   247540 obs. of  13 variables:
## $ ride_id      : chr   "46F8167220E4431F" "73A77762838B32FD"
##                "4CF42452054F59C5" "3278BA87BF698339" ...
```

Google Data Analytics Capstone Project

```
## $ rideable_type      : chr "electric_bike" "electric_bike"
"electric_bike" "classic_bike" ...
## $ started_at        : chr "2021-12-07 15:06:07" "2021-12-11 03:43:29"
"2021-12-15 23:10:28" "2021-12-26 16:16:10" ...
## $ ended_at          : chr "2021-12-07 15:13:42" "2021-12-11 04:10:23"
"2021-12-15 23:23:14" "2021-12-26 16:30:53" ...
## $ start_station_name: chr "Laflin St & Cullerton St" "LaSalle Dr & Huron
St" "Halsted St & North Branch St" "Halsted St & North Branch St" ...
## $ start_station_id  : chr "13307" "KP1705001026" "KA1504000117"
"KA1504000117" ...
## $ end_station_name  : chr "Morgan St & Polk St" "Clarendon Ave & Leland
Ave" "Broadway & Barry Ave" "LaSalle Dr & Huron St" ...
## $ end_station_id    : chr "TA1307000130" "TA1307000119" "13137"
"KP1705001026" ...
## $ start_lat         : num 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng         : num -87.7 -87.6 -87.6 -87.6 -87.7 ...
## $ end_lat           : num 41.9 42 41.9 41.9 41.9 ...
## $ end_lng           : num -87.7 -87.7 -87.6 -87.6 -87.6 ...
## $ member_casual     : chr "member" "casual" "member" "member" ...

str(tripdata_202201)

## 'data.frame': 103770 obs. of 13 variables:
## $ ride_id           : chr "C2F7DD78E82EC875" "A6CF8980A652D272"
"BD0F91DFF741C66D" "CBB80ED419105406" ...
## $ rideable_type     : chr "electric_bike" "electric_bike" "classic_bike"
"classic_bike" ...
## $ started_at        : chr "2022-01-13 11:59:47" "2022-01-10 08:41:56"
"2022-01-25 04:53:40" "2022-01-04 00:18:04" ...
## $ ended_at          : chr "2022-01-13 12:02:44" "2022-01-10 08:46:17"
"2022-01-25 04:58:01" "2022-01-04 00:33:00" ...
## $ start_station_name: chr "Glenwood Ave & Touhy Ave" "Glenwood Ave &
Touhy Ave" "Sheffield Ave & Fullerton Ave" "Clark St & Bryn Mawr Ave" ...
## $ start_station_id  : chr "525" "525" "TA1306000016" "KA1504000151" ...
## $ end_station_name  : chr "Clark St & Touhy Ave" "Clark St & Touhy Ave"
"Greenview Ave & Fullerton Ave" "Paulina St & Montrose Ave" ...
## $ end_station_id    : chr "RP-007" "RP-007" "TA1307000001"
"TA1309000021" ...
## $ start_lat         : num 42 42 41.9 42 41.9 ...
## $ start_lng         : num -87.7 -87.7 -87.7 -87.7 -87.6 ...
## $ end_lat           : num 42 42 41.9 42 41.9 ...
## $ end_lng           : num -87.7 -87.7 -87.7 -87.7 -87.6 ...
## $ member_casual     : chr "casual" "casual" "member" "casual" ...

str(tripdata_202202)

## 'data.frame': 115609 obs. of 13 variables:
## $ ride_id           : chr "E1E065E7ED285C02" "1602DCDC5B30FFE3"
"BE7DD2AF4B55C4AF" "A1789BDF844412BE" ...
## $ rideable_type     : chr "classic_bike" "classic_bike" "classic_bike"
"classic_bike" ...
```

```
## $ started_at      : chr  "2022-02-19 18:08:41" "2022-02-20 17:41:30"
"2022-02-25 18:55:56" "2022-02-14 11:57:03" ...
## $ ended_at        : chr  "2022-02-19 18:23:56" "2022-02-20 17:45:56"
"2022-02-25 19:09:34" "2022-02-14 12:04:00" ...
## $ start_station_name: chr  "State St & Randolph St" "Halsted St &
Wrightwood Ave" "State St & Randolph St" "Southport Ave & Waveland Ave" ...
## $ start_station_id  : chr  "TA1305000029" "TA1309000061" "TA1305000029"
"13235" ...
## $ end_station_name  : chr  "Clark St & Lincoln Ave" "Southport Ave &
Wrightwood Ave" "Canal St & Adams St" "Broadway & Sheridan Rd" ...
## $ end_station_id    : chr  "13179" "TA1307000113" "13011" "13323" ...
## $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng         : num  -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ end_lat          : num  41.9 41.9 41.9 42 41.9 ...
## $ end_lng          : num  -87.6 -87.7 -87.6 -87.6 -87.6 ...
## $ member_casual     : chr  "member" "member" "member" "member" ...

str(tripdata_202203)

## 'data.frame': 284042 obs. of 13 variables:
## $ ride_id           : chr  "47EC0A7F82E65D52" "8494861979B0F477"
"EFE527AF80B66109" "9F446FD9DEE3F389" ...
## $ rideable_type      : chr  "classic_bike" "electric_bike" "classic_bike"
"classic_bike" ...
## $ started_at        : chr  "2022-03-21 13:45:01" "2022-03-16 09:37:16"
"2022-03-23 19:52:02" "2022-03-01 19:12:26" ...
## $ ended_at          : chr  "2022-03-21 13:51:18" "2022-03-16 09:43:34"
"2022-03-23 19:54:48" "2022-03-01 19:22:14" ...
## $ start_station_name: chr  "Wabash Ave & Wacker Pl" "Michigan Ave & Oak
St" "Broadway & Berwyn Ave" "Wabash Ave & Wacker Pl" ...
## $ start_station_id  : chr  "TA1307000131" "13042" "13109" "TA1307000131"
...
## $ end_station_name  : chr  "Kingsbury St & Kinzie St" "Orleans St &
Chestnut St (NEXT Apts)" "Broadway & Ridge Ave" "Franklin St & Jackson Blvd"
...
## $ end_station_id    : chr  "KA1503000043" "620" "15578" "TA1305000025"
...
## $ start_lat         : num  41.9 41.9 42 41.9 41.9 ...
## $ start_lng         : num  -87.6 -87.6 -87.7 -87.6 -87.6 ...
## $ end_lat          : num  41.9 41.9 42 41.9 41.9 ...
## $ end_lng          : num  -87.6 -87.6 -87.7 -87.6 -87.7 ...
## $ member_casual     : chr  "member" "member" "member" "member" ...

str(tripdata_202204)

## 'data.frame': 371249 obs. of 13 variables:
## $ ride_id           : chr  "3564070EEFD12711" "0B820C7FCF22F489"
"89EEEE32293F07FF" "84D4751AEB31888D" ...
## $ rideable_type      : chr  "electric_bike" "classic_bike" "classic_bike"
"classic_bike" ...
## $ started_at        : chr  "2022-04-06 17:42:48" "2022-04-24 19:23:07"
```

```

"2022-04-20 19:29:08" "2022-04-22 21:14:06" ...
## $ ended_at      : chr  "2022-04-06 17:54:36" "2022-04-24 19:43:17"
"2022-04-20 19:35:16" "2022-04-22 21:23:29" ...
## $ start_station_name: chr  "Paulina St & Howard St" "Wentworth Ave &
Cermak Rd" "Halsted St & Polk St" "Wentworth Ave & Cermak Rd" ...
## $ start_station_id  : chr  "515" "13075" "TA1307000121" "13075" ...
## $ end_station_name  : chr  "University Library (NU)" "Green St & Madison
St" "Green St & Madison St" "Delano Ct & Roosevelt Rd" ...
## $ end_station_id    : chr  "605" "TA1307000120" "TA1307000120"
"KA1706005007" ...
## $ start_lat        : num  42 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num  -87.7 -87.6 -87.6 -87.6 -87.6 ...
## $ end_lat          : num  42.1 41.9 41.9 41.9 41.9 ...
## $ end_lng          : num  -87.7 -87.6 -87.6 -87.6 -87.6 ...
## $ member_casual    : chr  "member" "member" "member" "casual" ...

str(tripdata_202205)

## 'data.frame':   634858 obs. of  13 variables:
## $ ride_id        : chr  "EC2DE40644C6B0F4" "1C31AD03897EE385"
"1542FBEC830415CF" "6FF59852924528F8" ...
## $ rideable_type   : chr  "classic_bike" "classic_bike" "classic_bike"
"classic_bike" ...
## $ started_at      : chr  "2022-05-23 23:06:58" "2022-05-11 08:53:28"
"2022-05-26 18:36:28" "2022-05-10 07:30:07" ...
## $ ended_at        : chr  "2022-05-23 23:40:19" "2022-05-11 09:31:22"
"2022-05-26 18:58:18" "2022-05-10 07:38:49" ...
## $ start_station_name: chr  "Wabash Ave & Grand Ave" "DuSable Lake Shore
Dr & Monroe St" "Clinton St & Madison St" "Clinton St & Madison St" ...
## $ start_station_id  : chr  "TA1307000117" "13300" "TA1305000032"
"TA1305000032" ...
## $ end_station_name  : chr  "Halsted St & Roscoe St" "Field Blvd & South
Water St" "Wood St & Milwaukee Ave" "Clark St & Randolph St" ...
## $ end_station_id    : chr  "TA1309000025" "15534" "13221" "TA1305000030"
...
## $ start_lat        : num  41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng        : num  -87.6 -87.6 -87.6 -87.6 -87.6 ...
## $ end_lat          : num  41.9 41.9 41.9 41.9 41.9 ...
## $ end_lng          : num  -87.6 -87.6 -87.7 -87.6 -87.7 ...
## $ member_casual    : chr  "member" "member" "member" "member" ...

str(tripdata_202206)

## 'data.frame':   769204 obs. of  13 variables:
## $ ride_id        : chr  "600CFD130D0FD2A4" "F5E6B5C1682C6464"
"B6EB6D27BAD771D2" "C9C320375DE1D5C6" ...
## $ rideable_type   : chr  "electric_bike" "electric_bike"
"electric_bike" "electric_bike" ...
## $ started_at      : chr  "2022-06-30 17:27:53" "2022-06-30 18:39:52"
"2022-06-30 11:49:25" "2022-06-30 11:15:25" ...
## $ ended_at        : chr  "2022-06-30 17:35:15" "2022-06-30 18:47:28"

```

```
"2022-06-30 12:02:54" "2022-06-30 11:19:43" ...
## $ start_station_name: chr "" "" "" "" ...
## $ start_station_id : chr "" "" "" "" ...
## $ end_station_name : chr "" "" "" "" ...
## $ end_station_id : chr "" "" "" "" ...
## $ start_lat : num 41.9 41.9 41.9 41.8 41.9 ...
## $ start_lng : num -87.6 -87.6 -87.7 -87.7 -87.6 ...
## $ end_lat : num 41.9 41.9 41.9 41.8 41.9 ...
## $ end_lng : num -87.6 -87.6 -87.6 -87.7 -87.6 ...
## $ member_casual : chr "casual" "casual" "casual" "casual" ...

str(tripdata_202207)

## 'data.frame': 823488 obs. of 13 variables:
## $ ride_id : chr "954144C2F67B1932" "292E027607D218B6"
## $ rideable_type : chr "classic_bike" "classic_bike" "classic_bike"
## $ started_at : chr "2022-07-05 08:12:47" "2022-07-26 12:53:38"
## $ ended_at : chr "2022-07-05 08:24:32" "2022-07-26 12:55:31"
## $ start_station_name: chr "Ashland Ave & Blackhawk St" "Buckingham
## $ start_station_id : chr "13224" "15541" "15541" "15541" ...
## $ end_station_name : chr "Kingsbury St & Kinzie St" "Michigan Ave & 8th
## $ end_station_id : chr "KA1503000043" "623" "623" "TA1307000164" ...
## $ start_lat : num 41.9 41.9 41.9 41.9 41.9 ...
## $ start_lng : num -87.7 -87.6 -87.6 -87.6 -87.6 ...
## $ end_lat : num 41.9 41.9 41.9 41.8 41.9 ...
## $ end_lng : num -87.6 -87.6 -87.6 -87.6 -87.7 ...
## $ member_casual : chr "member" "casual" "casual" "casual" ...
```

Remarks: Columns and data types are consistent.

Data Cleaning & Transformation

- Combine all data files into a single data frame

```
all_tripdata <-
bind_rows(tripdata_202108,tripdata_202109,tripdata_202110,tripdata_202111,tri
pdata_202112,

tripdata_202201,tripdata_202202,tripdata_202203,tripdata_202204,
tripdata_202205,tripdata_202206,tripdata_202207)
```

- Analyse combined data frame

```
skim_without_charts(all_tripdata)
```


Google Data Analytics Capstone Project

Data summary

Name	all_tripdata
Number of rows	5901463
Number of columns	13

Column type frequency:

character	9
numeric	4

Group variables	None
-----------------	------

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ride_id	0	1	16	16	0	5901463	0
rideable_type	0	1	11	13	0	3	0
started_at	0	1	19	19	0	4923992	0
ended_at	0	1	19	19	0	4928247	0
start_station_name	0	1	0	64	860786	1382	0
start_station_id	0	1	0	44	860784	1227	0
end_station_name	0	1	0	64	919896	1397	0
end_station_id	0	1	0	44	919896	1237	0
member_casual	0	1	6	6	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
start_lat	0	1	41.90	0.05	41.64	41.88	41.90	41.93	45.64
start_lng	0	1	-87.65	0.03	-87.84	-87.66	-87.64	-87.63	-73.80

skim_variabl e	n_missin g	complete_rat e	mea n	sd	p0	p25	p50	p75	p100
end_lat	5590	1	41.9 0	0.0 5	41.3 9	41.8 8	41.9 0	41.9 3	42.3 7
end_lng	5590	1	- 87.6 5	0.0 3	- 88.9 7	- 87.6 6	- 87.6 4	- 87.6 3	- 87.5 0

- Creating a new column for the trip start hour

```
all_tripdata <- all_tripdata %>%
  mutate(start_hour = hour(started_at))
```

- Add column for month (ride start)

```
all_tripdata$month <- format(as.Date(all_tripdata$started_at), '%y_%m')
```

- Add column for day of the week

```
all_tripdata <- all_tripdata %>%
  mutate(day_of_week = wday(started_at, label=TRUE, abbr=FALSE))
```

- Modify columns as date type

```
all_tripdata <- all_tripdata %>%
  mutate(started_at = ymd_hms(started_at), ended_at = ymd_hms(ended_at))
```

- Add column for ride length

```
all_tripdata <- all_tripdata %>%
  mutate(ride_length = difftime(ended_at, started_at, units="mins"))
```

- Check ride_length < 0

```
nrow(subset(all_tripdata, ride_length < 0))
```

```
## [1] 149
```

- Remove ride_length < 0

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

- Rename columns

```
all_tripdata <- all_tripdata %>%
  rename(customer_type = member_casual)
```

- Remove unused columns

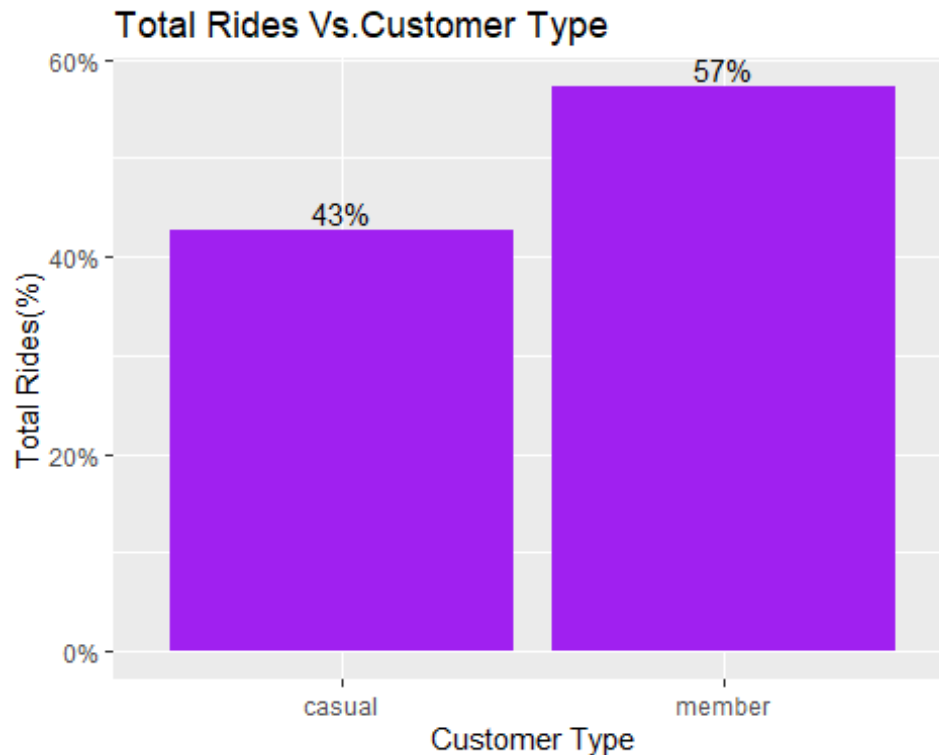
```
all_tripdata <- all_tripdata %>%
  select(-c(ride_id, start_station_name, start_station_id, end_station_name,
            end_station_id, start_lat, start_lng, end_lat, end_lng))
```

4 & 5) Analyze and Share

Data Exploration through Visualization

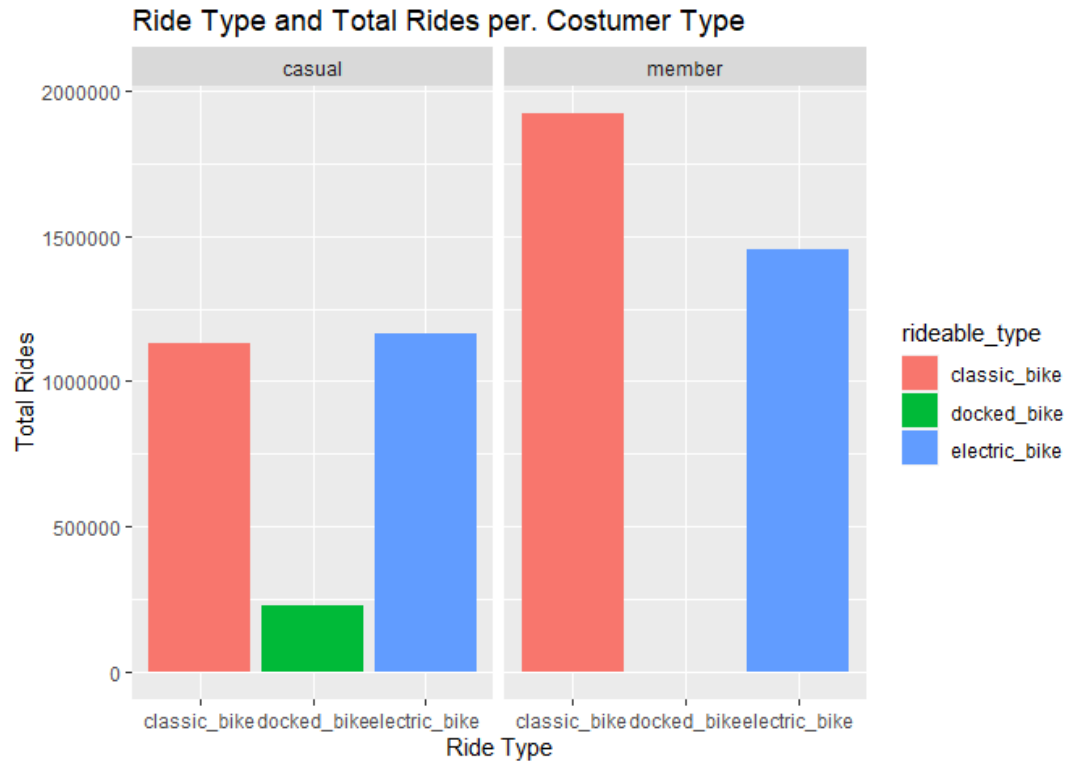
- Ride Type Vs. Costumers

```
ggplot(all_tripdata, aes(customer_type)) +
  geom_bar(aes(y = (..count..)/sum(..count..)), fill="purple")+
  scale_y_continuous(label=scales::percent)+
  geom_text(aes(label = scales::percent((..count..)/sum(..count..)),
               y = ((..count..)/sum(..count..)), stat="count",
               vjust = -.25)+
  labs(title = "Total Rides Vs.Customer Type ",
       x = "Customer Type",
       y = "Total Rides(%)")
```



Annual memberships make 14% more use of the service compared to casual customers.

```
all_tripdata %>%
  ggplot() + geom_bar(aes(x=rideable_type, fill=rideable_type)) +
  facet_wrap(~customer_type) + labs(x="Ride Type", y="Total Rides")+
  labs(title = "Ride Type and Total Rides per. Costumer Type")
```



Annual memberships customers prefer classic bikes, but also use electric bikes often not using docked bikes. Casual customers choose classic and electric bikes in equal proportions, and also use docked bikes but less frequently.

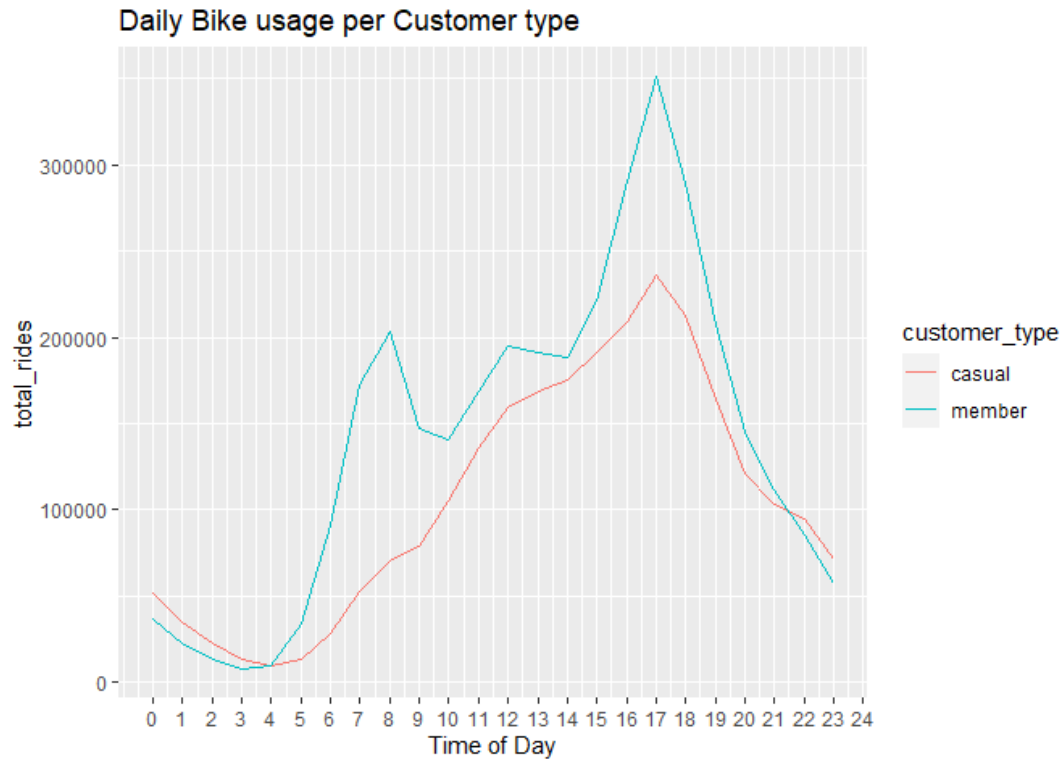
```
all_tripdata %>%
  ggplot(aes(x=customer_type,y=ride_length,color=rideable_type)) +
  geom_boxplot() + labs(x="Costumer Type",y="Ride length (min.)")+
  labs(title ="Ride length (min.) per Costumer Type")+
  scale_y_continuous(name="Ride lenth (min.)", limits=c(0, 100))
```



Casual customers are those who use the service for longer periods of time, especially using docked bikes.

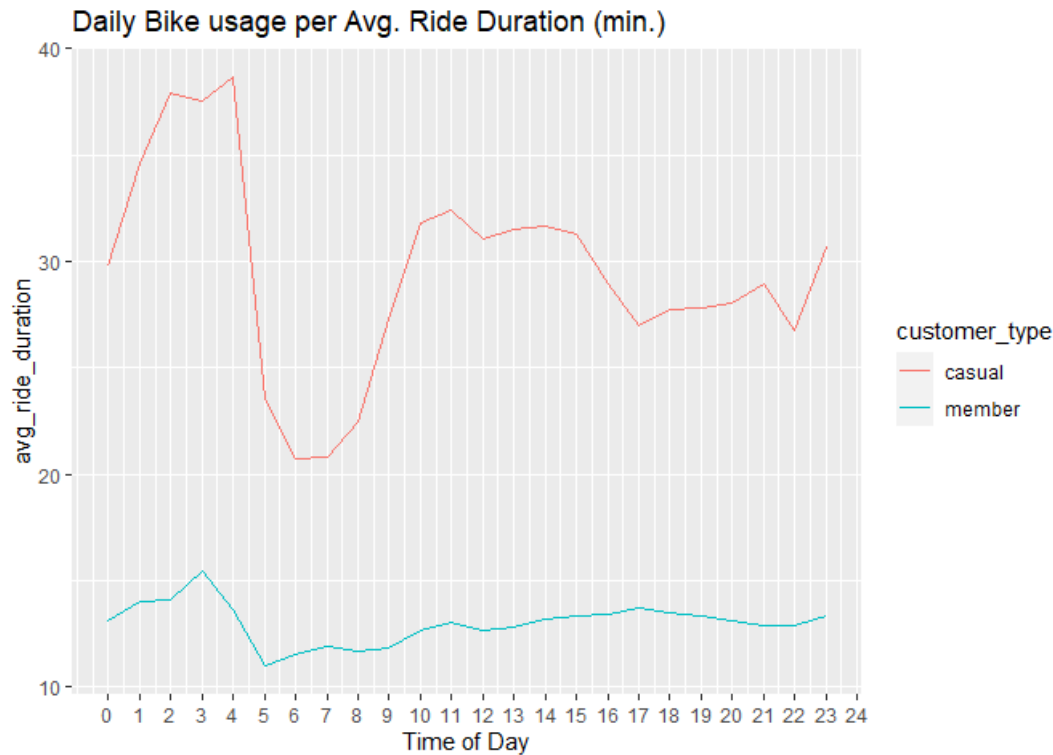
- Daily bike usage

```
all_tripdata %>%
  group_by(start_hour, customer_type)%>%
  summarise(total_rides = n()) %>%
  ggplot(aes(x=start_hour, y= total_rides, group = customer_type, color =
customer_type))+
  geom_line() + scale_x_continuous(breaks =
c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24))+
  labs(title = "Daily Bike usage per Customer type", x = "Time of Day")+
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```



It is in the period between 3 pm and 7 pm that the greatest number of uses of both customer types occur.

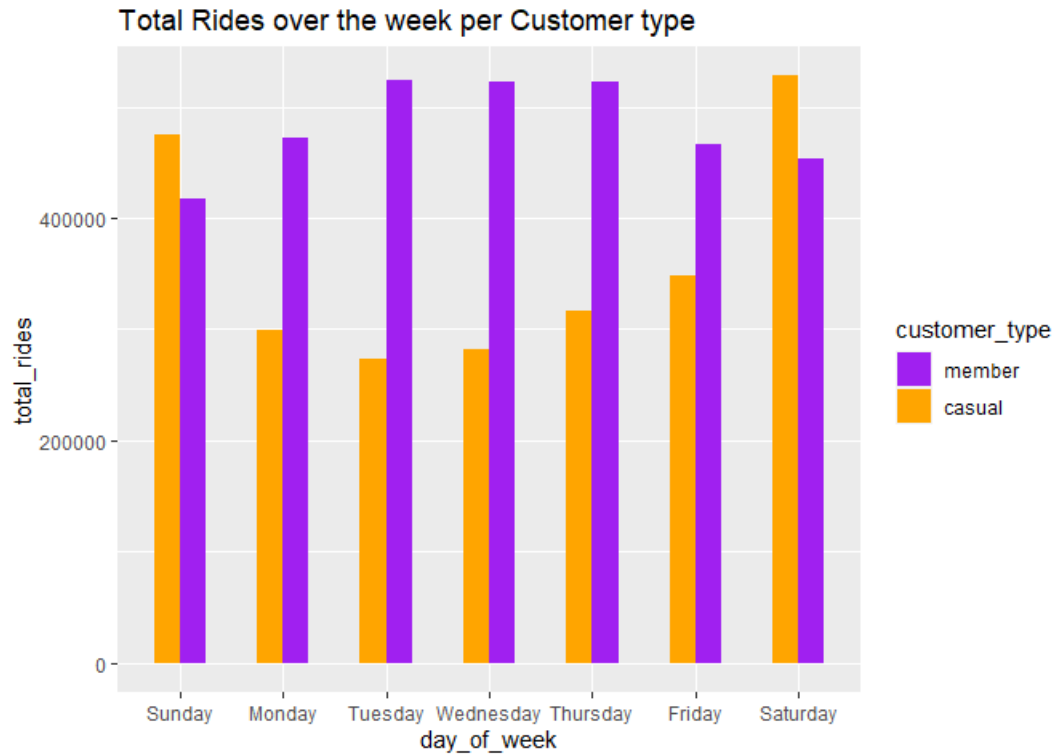
```
all_tripdata %>%
  group_by(start_hour, customer_type)%>%
  summarise(avg_ride_duration = mean(ride_length)) %>%
  ggplot(aes(x=start_hour, y= avg_ride_duration, group = customer_type, color
= customer_type))+
  geom_line() + scale_x_continuous(breaks =
c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24))+
  labs(title = "Daily Bike usage per Avg. Ride Duration (min.)", x = "Time of
Day")+
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```



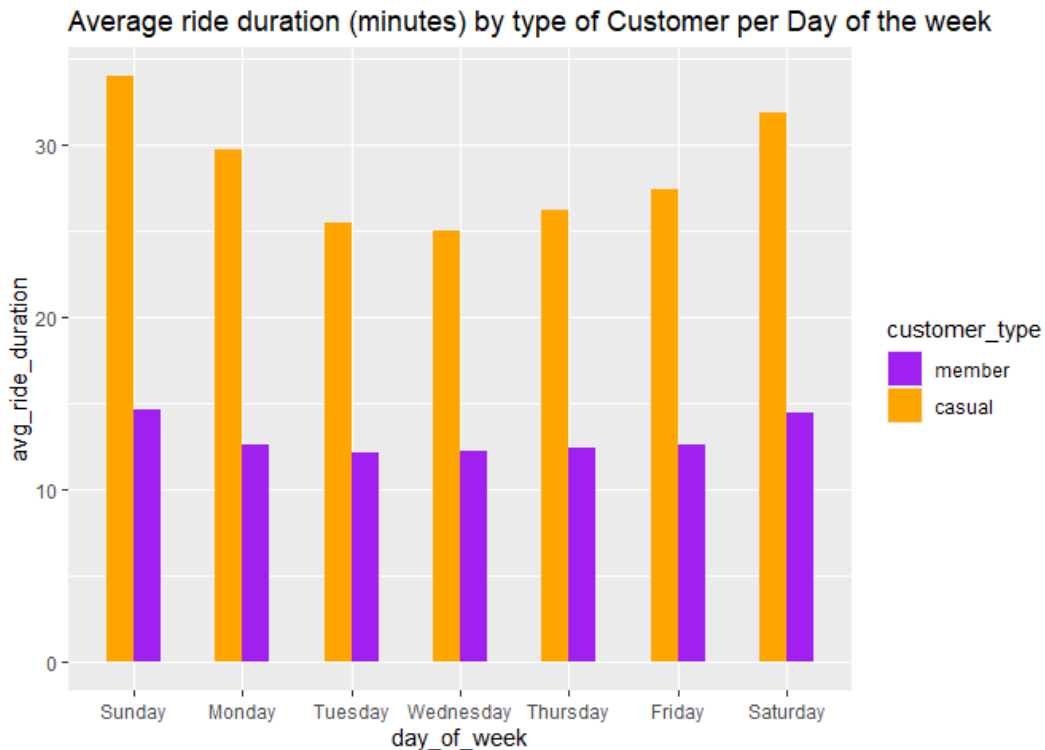
With consistently longer rides on average, casual customers ride the bikes the longest between the 2 am and 4 am period and for less time between the 5 am and 8 am period. The afternoon and early evening period have a constant average duration of use.

- Weekly bike usage

```
all_tripdata %>%
  group_by(customer_type, day_of_week) %>%
  summarise(total_rides = n()) %>%
  arrange(customer_type, day_of_week) %>%
  ggplot(aes(x = day_of_week, y = total_rides, fill = customer_type)) +
  geom_col(width=0.5, position = position_dodge(width=0.5)) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  labs(title = "Total Rides over the week per Customer type") +
  scale_fill_manual("customer_type", values = c("member" = 'purple', 'casual'
= 'orange'))
```



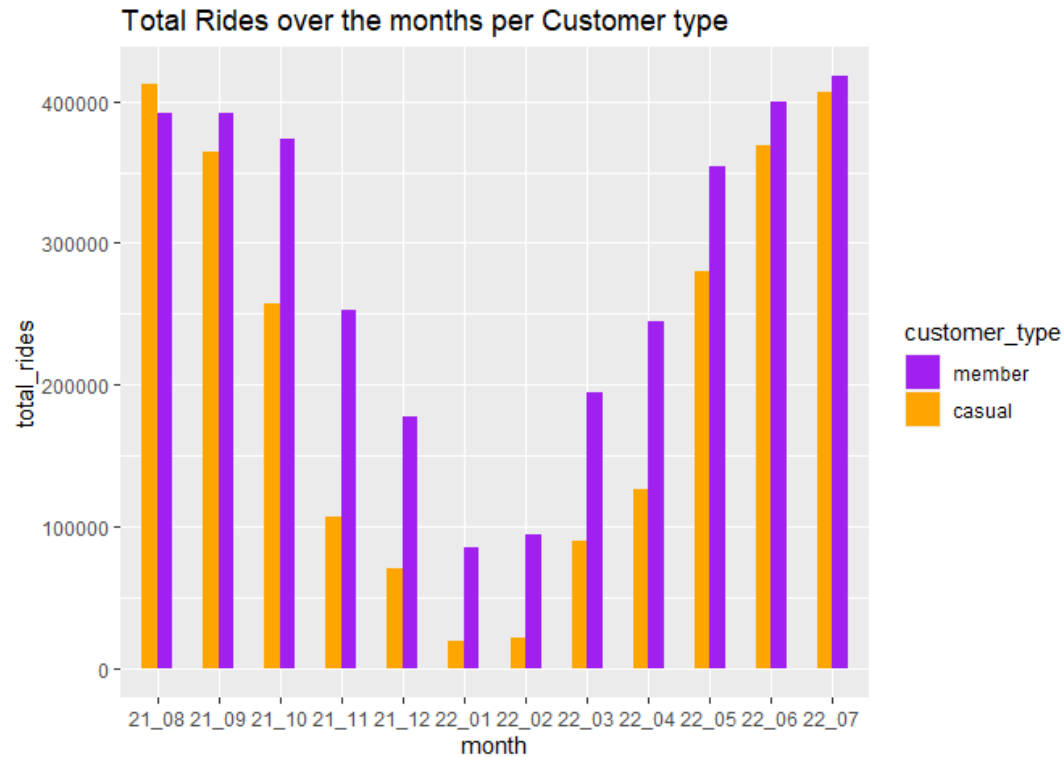
```
all_tripdata %>%
  group_by(customer_type, day_of_week) %>%
  summarise(avg_ride_duration = mean(ride_length)) %>%
  ggplot(aes(x=day_of_week, y= avg_ride_duration, fill= customer_type)) +
  geom_col(width = 0.5, position = position_dodge(width = 0.5))+
  labs(title= "Average ride duration (minutes) by type of Customer per Day of
the week") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))+
  scale_fill_manual("customer_type", values = c("member" = 'purple', 'casual'
= 'orange'))
```

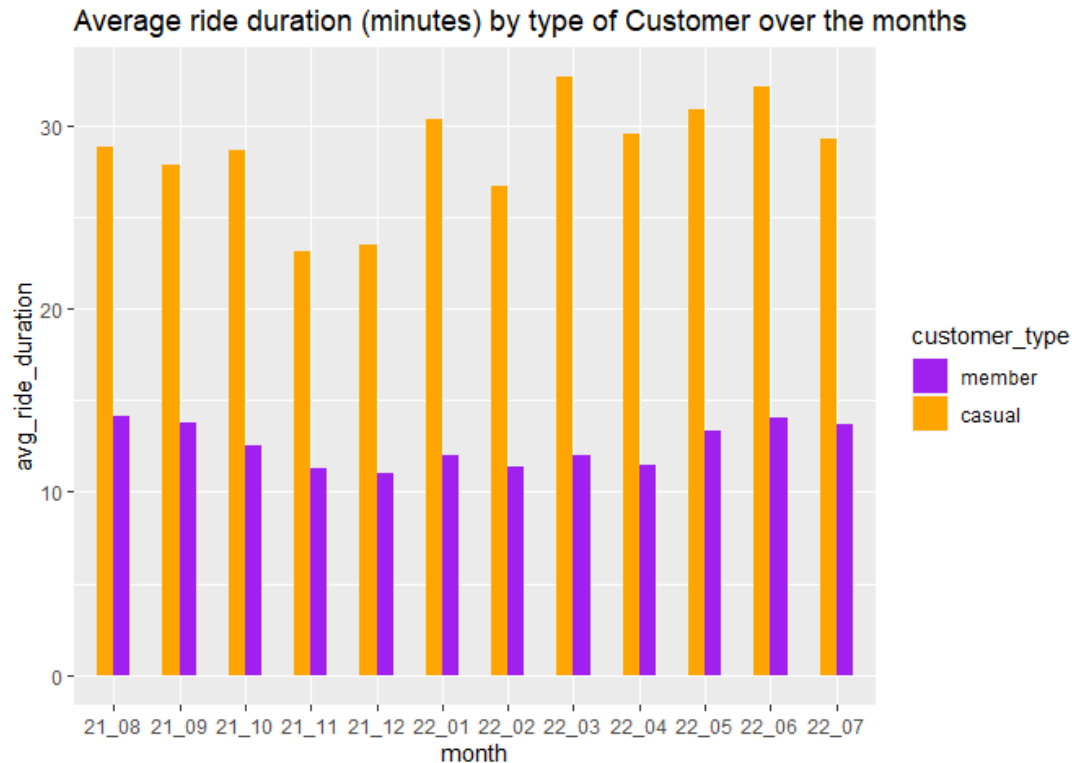
Customers with annual memberships use the service consistently throughout the week. Casual customers use the service more often on weekends and during longer periods of time.

- Monthly bike usage

```
all_tripdata %>%
  group_by(customer_type, month) %>%
  summarise(total_rides = n()) %>%
  arrange(customer_type, month) %>%
  ggplot(aes(x = month, y = total_rides, fill = customer_type)) +
  geom_col(width=0.5, position = position_dodge(width=0.5)) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  labs(title = "Total Rides over the months per Customer type")+
  scale_fill_manual("customer_type", values = c("member" = 'purple', 'casual'
= 'orange'))
```



```
all_tripdata %>%
  group_by(customer_type, month) %>%
  summarise(avg_ride_duration = mean(ride_length)) %>%
  ggplot(aes(x = month, y = avg_ride_duration, fill = customer_type)) +
  geom_col(width = 0.5, position = position_dodge(width = 0.5)) +
  labs(title = "Average ride duration (minutes) by type of Customer over the
months") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE)) +
  scale_fill_manual("customer_type", values = c("member" = 'purple', 'casual'
= 'orange'))
```



In the summer and spring period, customers use the services more frequently. However, the average duration of use is not substantially different throughout the year.

6) Act

Conclusions

- Casual customers use the service less frequently but for longer periods.
- Casual customers use the service most often on weekends.
- There is a substantial reduction in the volume of users during the autumn and winter months (probably due to weather conditions).

Recommendations

- Promotion of lower prices, progressive over time of use, applied to annual members in order to encourage use of the service for longer periods of time.
- Special membership package applied to casual weekend customers with the aim of converting the largest number of casual customers into fixed customers.
- Lower prices for the entire range of products during the autumn and winter months, in order to encourage the use of the service in the period of lower affluence.

Thanks for reading.