Case 1 - Cyclistic: Member and casual consumers

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

Epale! In this text you will see my step by step in **Case 1** of Google Data Analytics Capstore from Coursera. I will use the **R programming language** for data processing and in **Public Tableau** I will perform a deeper analysis together with visualizations from the **database** provided by in the course.

Scenery

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.

About the company

In Cyclistic, our purpose is to generate general awareness and appeal to broad segments of consumers. We specialize in sharing, geolocated and locked bikes in a network of 692 stations throughout Chicago. Our goal is to continue growing and attracting consumers in one of our three modalities: single-ride passes, full-day passes, and annual memberships.

Steps

In this text I describe the steps: Ask, Prepare, Process, Analyze, Share, Act.

I only used the last month, because the visualizations did not run well in Public Tableau, which was where I did them. Likewise, the Prepare and Process steps do not change almost anything for the all data or only one month.

The last month corresponds to **June 2023**.

Ask

The objective that they have set in the company is to design marketing strategies aimed at converting casual riders (single-ride passes and full-day passes) into annual members, since these members are much more profitable than occasional ones. However, for this you must understand:

- How do annual members and casual riders use Cyclistic bikes differently?

- Why would casual riders buy Cyclistic annual memberships?

- How can Cyclistic use digital media to influence casual riders to become members?

The business tasks

In this case, the first task is to find the difference between usage for annual members and casual passengers.

And the stakeholders in it are both the owners of the company and the entire marketing department. Since it depends on this question to be able to move on to the next task, which would be to get casual riders to buy an annual membership in Cyclistic.

Finally, the last question would be answered by the marketing team, as well as some possible solutions will be presented.

Prepare

For this case I used only the **last month of historical travel data**, the reason is explained at the beginning. Here you can see all the available data.

(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 license).

Type of data

The data is in a CSV file (comma-separated values).

It's organized

The data is organized in rows and columns (data-frame).

Sort and filter the data

It is organized by date. Contains 13 columns and 719618 rows.

Final of the Process: `534719 rows and 20 columns`

Determine the credibility of the data

To determine the credibility of the data in the course they use a process called "**ROCCC**".

The **ROCCC** is an acronym, which refers to **R**eliable, **O**riginal, **C**omprehensive, **C**urrent, **C**ited.

The data is **R**eliable because it is under the license of Motivate International Inc., it is **O**riginal since it is captured by them, it is **C**omprehensive because it covers the entire study that we want to carry out, **C**urrent because it's from the month of June 2023 and is **C**ited from a source specific.

Prepare the data

Having done all this we will now move on to data preparation, which was done in R through the following steps. If it is not necessary, they will not be explained.

Installing and loading required packages

```
install.packages("tidyverse")

install.packages("lubridate")

install.packages("janitor")

install.packages("dplyr")

install.packages("ggplot2")
```

```
library(tidyverse)

library(lubridate)

library(janitor)

library(dplyr)

library(ggplot2)
```

Importing data to data-frame in R

```
setwd("C:/Users/alfre/Documents/R/DATOS")

data_ciclystic <- read.csv("202306-divvy-tripdata.csv")
```

Process

Since we only use one month, the data-frame is already created and we name it:
`data_ciclystic`

The head of data-frame

```
head(data_ciclystic)

dim(dataciclystic)
```

Console:

```
  ride_id rideable_type        started_at          ended_at start_station_name
1 6F1682AC40EB6F71 electric_bike 2023-06-05 13:34:12 2023-06-05 14:31:56
2 622A1686D64948EB electric_bike 2023-06-05 01:30:22 2023-06-05 01:33:06
3 3C88859D926253B4 electric_bike 2023-06-20 18:15:49 2023-06-20 18:32:05
4 EAD8A5E0259DEC88 electric_bike 2023-06-19 14:56:00 2023-06-19 15:00:35
5 5A36F21930D6A55C electric_bike 2023-06-19 15:03:34 2023-06-19 15:07:16
6 CF682EA7D0F961DB electric_bike 2023-06-09 21:30:25 2023-06-09 21:49:52

  start_station_id end_station_name end_station_id start_lat start_lng end_lat end_lng member_casual
1                                                      41.91   -87.69   41.91  -87.70        member
2                                                      41.94   -87.65   41.94  -87.65        member
3                                                      41.95   -87.68   41.92  -87.63        member
4                                                      41.99   -87.65   41.98  -87.66        member
5                                                      41.98   -87.66   41.99  -87.65        member
6                                                      41.99   -87.68   41.94  -87.65        member
[1] 719618      13
```

Seeing this, I understand that the columns "start_station_name", "start_station_id", "end_station_name", "end_station_id" have a problems with data.

Removing the empty columns and rows using janitor

```
data_ciclystic <- janitor::remove_empty(dat = data_ciclystic,which = c("cols"))

data_ciclystic <- janitor::remove_empty(dat = data_ciclystic,which = c("rows"))

dim(data_ciclystic)
```

Console:
```
[1] 719618      13
```

The janitor not found good, because is the same mount of data. Then we will use **na.omit()**:

```
data_ciclystic <- na.omit(data_ciclystic)

dim(data_ciclystic)
```

Console:
```
[1] 718729      13
```

It's a change!, but when I see de head(), again.

Console:
```
        ride_id rideable_type        started_at          ended_at start_station_name
1 6F1682AC40EB6F71 electric_bike 2023-06-05 13:34:12 2023-06-05 14:31:56
2 622A1686D64948EB electric_bike 2023-06-05 01:30:22 2023-06-05 01:33:06
3 3C88859D926253B4 electric_bike 2023-06-20 18:15:49 2023-06-20 18:32:05
4 EAD8A5E0259DEC88 electric_bike 2023-06-19 14:56:00 2023-06-19 15:00:35
5 5A36F21930D6A55C electric_bike 2023-06-19 15:03:34 2023-06-19 15:07:16
6 CF682EA7D0F961DB electric_bike 2023-06-09 21:30:25 2023-06-09 21:49:52
```

```
   start_station_id end_station_name end_station_id start_lat start_lng end_lat end_lng member_casual
1                                                      41.91    -87.69   41.91  -87.70        member
2                                                      41.94    -87.65   41.94  -87.65        member
3                                                      41.95    -87.68   41.92  -87.63        member
4                                                      41.99    -87.65   41.98  -87.66        member
5                                                      41.98    -87.66   41.99  -87.65        member
6                                                      41.99    -87.68   41.94  -87.65        member
```

This dataframe have dirty data again, for last stance I use **filter.all()** with exclude the NA values:

```
data_ciclystic_clean <- data_ciclystic %>%

  filter_all(all_vars(!is.na(.) & . != ""))

head(data_ciclystic_clean)

dim(data_ciclystic_clean)
```

```
Console:
       ride_id rideable_type          started_at            ended_at           start_station_name
1 055E6783FCD761FD electric_bike 2023-06-06 20:58:13 2023-06-06 21:04:00 California Ave & Milwaukee Ave
2 65BC2A1FC12CFFA4  classic_bike 2023-06-28 18:21:26 2023-06-28 18:56:06      Cottage Grove Ave & 51st St
3 4D149B57532FE9C9  classic_bike 2023-06-30 09:53:16 2023-06-30 09:53:38          Western Ave & Roscoe St
4 BAE48AFCA5C36414  classic_bike 2023-06-21 13:23:57 2023-06-21 13:24:16      Cottage Grove Ave & 51st St
5 E3491C5273700783  classic_bike 2023-06-21 13:24:46 2023-06-21 13:40:47      Cottage Grove Ave & 51st St
6 4AE5E4B5B30DBFEF  classic_bike 2023-06-02 23:39:41 2023-06-02 23:40:26 California Ave & Milwaukee Ave
  start_station_id            end_station_name end_station_id start_lat start_lng  end_lat   end_lng
1            13084      California Ave & Division St          13256  41.92262 -87.69711 41.90303 -87.69747
2     TA1309000067      Cottage Grove Ave & 51st St   TA1309000067  41.80304 -87.60662 41.80304 -87.60662
3            15634          Western Ave & Roscoe St          15634  41.94303 -87.68729 41.94303 -87.68729
4     TA1309000067      Cottage Grove Ave & 51st St   TA1309000067  41.80304 -87.60662 41.80304 -87.60662
5     TA1309000067      Cottage Grove Ave & 51st St   TA1309000067  41.80304 -87.60662 41.80304 -87.60662
6            13084 California Ave & Milwaukee Ave          13084  41.92269 -87.69715 41.92269 -87.69715
  member_casual
1        member
2        member
3        member
4        member
5        member
6        member
[1]  534758      13
```

Yep, that is a change. The head look good, I have 200K less rows. **Rename!**

```
data_ciclystic <- data_ciclystic_clean
```

Converting started_at and ended_at columns to date data type (POSIXct) using lubridate

```
data_ciclystic$started_at <- lubridate::as_datetime(data_ciclystic$started_at)

data_ciclystic_clean$ended_at <- lubridate::as_datetime(data_ciclystic$ended_at)
```

Getting the start time and end time components from date-time

The start and end time (hour) of the trip

```
data_ciclystic$start_hr <- lubridate::hour(data_ciclystic$started_at)

data_ciclystic$end_hr <- lubridate::hour(data_ciclystic$ended_at)
```

Getting the start date and end date

The start and end date of the trip

```
data_ciclystic$start_date <- lubridate::date(data_ciclystic$started_at)
```

```
data_ciclystic$end_date <- lubridate::date(data_ciclystic$ended_at)
```

Get the duration of the trip and then change the column format to numeric, the time is in minutes

```
data_ciclystic$length_min <- difftime(data_ciclystic$ended_at,data_ciclystic$sta
rted_at,units = c("mins"))

data_ciclystic$length_min <- as.numeric(data_ciclystic$length_min)´

summary(data_ciclystic)
```

When viewing the summary in the `Console`:

```
   length_min
Min.    :   -0.067
1st Qu.:    6.250
Median :   10.933
Mean    :   17.519
3rd Qu.:   19.350
Max.    :11152.267
```

The data has even negative trips, which is impossible, there can be no trips less than zero minutes or zero.

Eliminating trips with duration of zero or less than zero (0)

```
data_ciclystic_clean <- data_ciclystic %>%

  filter(length_min > 0)

summary(data_ciclystic_clean)
```

When viewing the new summary in the `Console`:

```
   length_min
Min.    :    0.017
1st Qu.:    6.250
Median :   10.933
Mean    :   17.520
3rd Qu.:   19.350
Max.    :11152.267
```

Rename!

```
data_ciclystic <- data_ciclystic_clean
```

Assign the days of the week to start_date and end_date in other columns called start_day and end_day

```
data_ciclystic$start_day <- wday(data_ciclystic$start_date, label = TRUE, abbr =
FALSE)
```

```
data_ciclystic$end_day <- wday(data_ciclystic$end_date, label = TRUE, abbr = FAL
SE)

dim(data_ciclystic)

View(data_ciclystic)
```

```
Console:
[1] 534719     20
View:
```


Export as CSV!

```
write.csv(data_ciclystic, file = "C:…/Documents/R/DATA/data_ciclystic_202306.csv
", row.names = FALSE)
```

Ussing summary()

Ussing

```
summary(data_ciclystic)
```

We can see the min, max, mean and median.

```
   length_min
Min.    :     0.017
1st Qu.:     6.250
Median :    10.933
Mean   :    17.520
3rd Qu.:    19.350
Max.   :11152.267
```

Analyze

For the analysis we will use Tableau. Loading the CSV file that we have extracted from RStudio.

Measures

Count of trips and percentage of riders:

TOTAL TRIPS

534.719

RIDERS

CASUAL	MEMBER
219.778	314.941
41,10%	58,90%

- The total count of trips is: 534719.
- The count of riders:
  - Casual: 219778
  - Member: 314941
- The count in percentage of riders:
  - Casual: 41,10%
  - Members: 58,90%

The total number of trips made by members is greater than the number of casual trips, by more than **15% (95163)**.

Count of trips and percentage of bike:

Type of bike

Classic	Docked	Electric
313.042	14.576	207.101
58,54%	2,73%	38,73%

14.576 ▬▬▬▬▬▬▬ 313.042

Trips with:

- Classic bike: 3133042 (58,54%)
- Docked bike: 14576 (2,73%)
- Electric bike: 207101 (38,73%)

The most used type of bicycle is the classic (58,54%), followed by the electric (38,73%) and finally the docked (2,73%).

Count of trips and percentage on days of week:

Days of the week

Mon.	64.943 12,15%
Tue.	57.908 10,83%
Wed.	71.115 13,30%
Thu.	87.250 16,32%
Fri.	95.935 17,94%
Sat.	95.260 17,81%
Sun.	62.308 11,65%
Total general	534.719 100,00%

57.908 ▮▮▮▮▮▮ 95.935

- Sunday:
  - 62308 trips
  - 11,65%
- Monday:
  - 64943 trips
  - 12,15%
- Tuesday:
  - 57908 trips
  - 10,83%
- Wednesday:
  - 71115 trips
  - 13,30%
- Thursday:
  - 87250 trips
  - 16,32%
- Friday:
  - 95935 trips
  - 17,94%
- Saturday:
  - 95260 trips
  - 17,81%

The number of trips per day of the week are in the following order:

Friday > Saturday > Thursday > Wednesday > Monday > Sunday > Tuesday.

Being **Friday**, **Saturday** and **Thursday** the days with the most trips, in that order.

Count of trips and percentage on hours of the day

The hours are in numbers from zero to twenty-three (0-23), first you have the count of trips and next the percentage.

Hours of the day

Hour	Count	Percentage
00	6.800	1,27%
01	3.768	0,70%
02	1.986	0,37%
03	1.057	0,20%
04	1.150	0,22%
05	4.337	0,81%
06	13.215	2,47%
07	23.183	4,34%
08	29.110	5,44%
09	21.408	4,00%
10	21.793	4,08%
11	26.223	4,90%
12	29.824	5,58%
13	30.005	5,61%
14	31.684	5,93%
15	36.461	6,82%
16	47.946	8,97%
17	55.099	10,30%
18	45.355	8,48%
19	34.060	6,37%
20	24.848	4,65%
21	19.173	3,59%
22	15.119	2,83%
23	11.115	2,08%

By far the time of day with the most trips is at 17:00 (5:00 p.m.) and the hours around this are those with the most trips. It behaves in an increasing way, with the exception of 8-9, where there is a small decrease.

AVG of the trips on days of the week and total for riders

AVG of time trip on day of week

	Casual	Member
Mon.	23,44	12,54
Tue.	20,07	11,74
Wed.	20,82	12,25
Thu.	21,23	12,28
Fri.	24,04	12,84
Sat.	27,53	15,12
Sun.	26,75	14,32
Total ..	24,08	12,95

The travel time for riders is almost double for casuals than for members. Both have higher numbers on the weekends, but in the members, they remain more equal.

Median of the trips on days of the week and total for riders

Median of time trip on day of week

	Casual	Member
Mon.	12,983	8,983
Tue.	11,367	8,783
Wed.	11,925	9,083
Thu.	12,200	9,167
Fri.	13,767	9,383
Sat.	16,500	11,033
Sun.	15,600	10,133
Total ..	13,800	9,433

The behavior is very similar to the previous AVG chart, but now the numbers are lower. This refers to most of the trips being short, which concludes that the AVG seems a bit biased because there are extremely long trips.

5 stations with more trips for riders

5 stations with more trips

	Casual	Member
Streeter Dr & Grand Ave	**6.243**	2.327
DuSable Lake Shore Dr & North Blvd	3.354	2.389
DuSable Lake Shore Dr & Monroe St	3.990	1.455
Michigan Ave & Oak St	3.472	1.970
Theater on the Lake	2.437	1.933

Questions

What type of bicycle does each type of cyclist use?

With these numbers the question arose, what would be the bicycle that the type of cyclist uses the most? The type of bike for riders (and the total general).

Type of bike for riders

	Classic	Docked	Electric
Casual	116.193	14.576	89.009
	52,87%	6,63%	40,50%
Member	196.849		118.092
	62,50%		37,50%
Total general	313.042	14.576	207.101
	58,54%	2,73%	38,73%

CASUAL:

- Classic: 52,87%
- Docked: 6,63%
- Electric: 40,50%

MEMBER:

- Classic: 62,50%
- Electric: 37,50%
- Docked: 0,00%

The riders, both casual and members, use classic bikes more than electric (Classic > Electric), but members don't use docked bikes.

Total general represent: Count of trips and percentage of bike.

What day of the week does each type of rider make the most trips?

Days of week for riders

Week	Casual	Member	Total general
Mon.	23.759 10,81%	41.184 13,08%	64.943 12,15%
Tue.	18.217 8,29%	39.691 12,60%	57.908 10,83%
Wed.	24.644 11,21%	**46.471** **14,76%**	71.115 13,30%
Thu.	31.055 14,13%	**56.195** **17,84%**	87.250 16,32%
Fri.	40.456 18,41%	**55.479** **17,62%**	95.935 17,94%
Sat.	**50.750** **23,09%**	44.510 14,13%	95.260 17,81%
Sun.	30.897 14,06%	31.411 9,97%	62.308 11,65%

Casual: Saturday > Friday > Thursday > Sunday > Wednesday > Monday > Tuesday.

Members: Thursday > Friday > Wednesday > Saturday > Monday > Tuesday > Sunday.

That is to say:

For the casual, the day with the most trips is Saturday, followed by Friday.

For members the day with the most trips is Thursday, followed by Friday.

This, when adding both types of riders, remains Friday as the best day, since it is one of the best for both.

We can also highlight that the days for members remain more equitable in the number of trips. While the casual ones have very few on some days, Tuesday for example, and too many on others, like Saturday.

Trends and relationships

- There are more member riders than casuals.

- Bike type most used is classic and the members don't use Docked bike.

- The day with more trips is Friday for riders, but Thursday for members and Saturday for casuals.

- At 17:00 (5:00p.m.) the hour with the most trips of the day is achieved. Where the data keeps growing from 3:00 to 17:00, less than 8-9 there is a small decrease.

- Casuals riders take longer rides than members.

- The day with the longest trips is Saturday and followed by Sunday.

Days of week vs Type of bike

Riders	Week	Classic	Docked	Electric	Total general
Casual	Mon.	11.785 3,76%	1.581 10,85%	10.393 5,02%	23.759 4,44%
	Tue.	8.671 2,77%	1.058 7,26%	8.488 4,10%	18.217 3,41%
	Wed.	12.104 3,87%	1.436 9,85%	11.104 5,36%	24.644 4,61%
	Thu.	15.488 4,95%	1.919 13,17%	13.648 6,59%	31.055 5,81%
	Fri.	21.456 6,85%	2.720 18,66%	16.280 7,86%	40.456 7,57%
	Sat.	29.810 9,52%	3.503 24,03%	17.437 8,42%	50.750 9,49%
	Sun.	16.879 5,39%	2.359 16,18%	11.659 5,63%	30.897 5,78%
Member	Mon.	25.674 8,20%		15.510 7,49%	41.184 7,70%
	Tue.	24.049 7,68%		15.642 7,55%	39.691 7,42%
	Wed.	27.955 8,93%		18.516 8,94%	46.471 8,69%
	Thu.	34.654 11,07%		21.541 10,40%	56.195 10,51%
	Fri.	34.795 11,12%		20.684 9,99%	55.479 10,38%
	Sat.	29.634 9,47%		14.876 7,18%	44.510 8,32%
	Sun.	20.088 6,42%		11.323 5,47%	31.411 5,87%
Total general		313.042 100,00%	14.576 100,00%	207.101 100,00%	534.719 100,00%

Share

Here's how riders use bikes differently:

- Members mostly use the bikes in their daily life, possibly to go to work or similar.

- Casuals ones use it for recreation, they often take walks on Friday, Saturday and Sunday.

You can see the graphs on the next dashboard on Tableau: Link!

Type of graphs

Cake: Type of riders

RIDERS %

41,10%

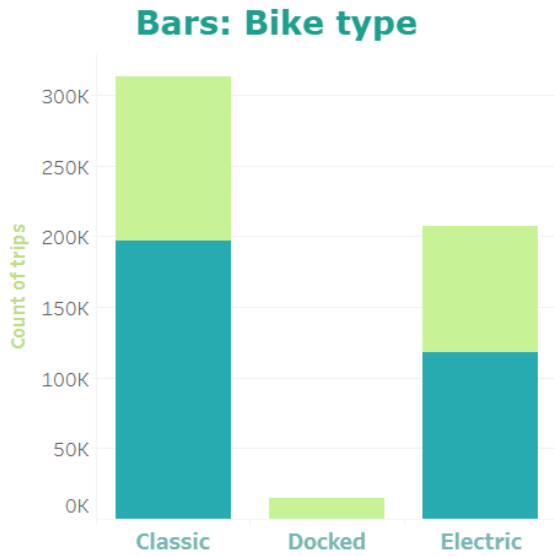58,90%

☐ Casual   ☐ Member

Bars: AVG of trips

Bars: AVG of trip

Bars: Type of bike

Bars: Bike type

Stacked bars & line: Week trips vs AVG of trips

Week trips: Count vs AVG (min)

- AVG Time (min), Casual
- AVG Time (min), Member
- Count trips, Casual
- Count trips, Member
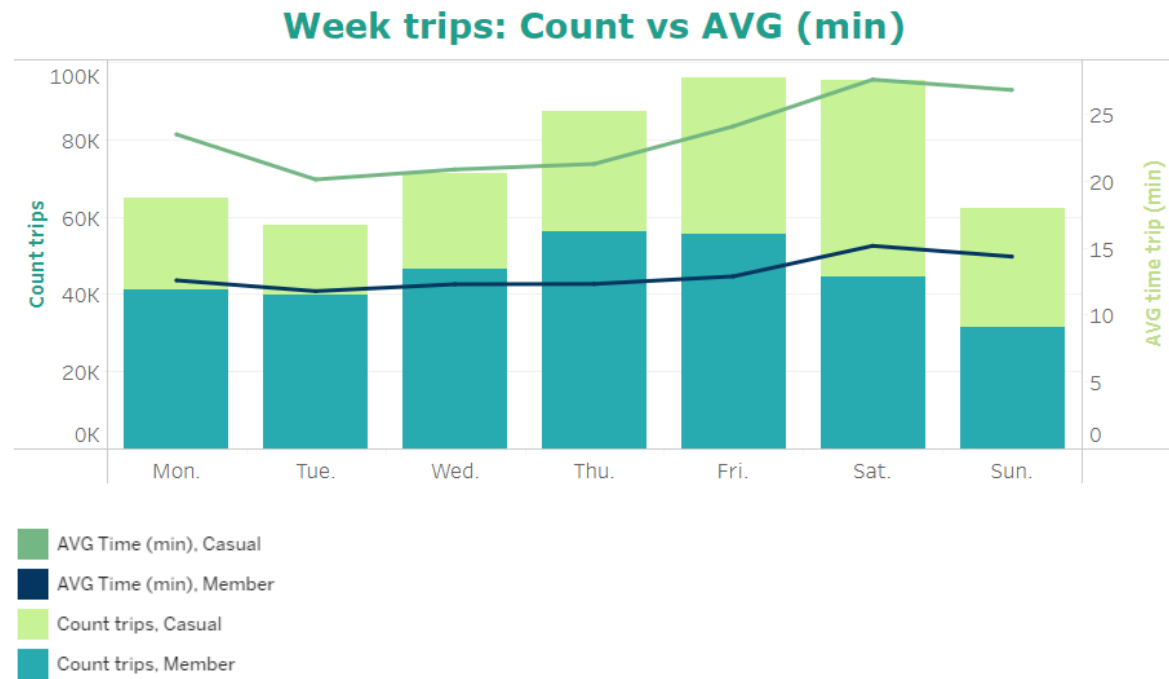
Stacked bars & line: Day trips vs AVG of trips

Day trips: Count vs AVG (min)

- AVG Time (min), Casual
- AVG Time (min), Member
- Count trips, Casual
- Count trips, Member

Final dashboard

TOTAL TRIPS
534.719

RIDERS %
41,10%
58,90%
Casual   Member

RIDERS
Casual   Member
219.778   314.941

Day trips: Count vs AVG (min)

Bars: Bike type
Classic   Docked   Electric

Bars: AVG of trip
Casual   Member

Week trips: Count vs AVG (min)
Mon. Tue. Wed. Thu. Fri. Sat. Sun.
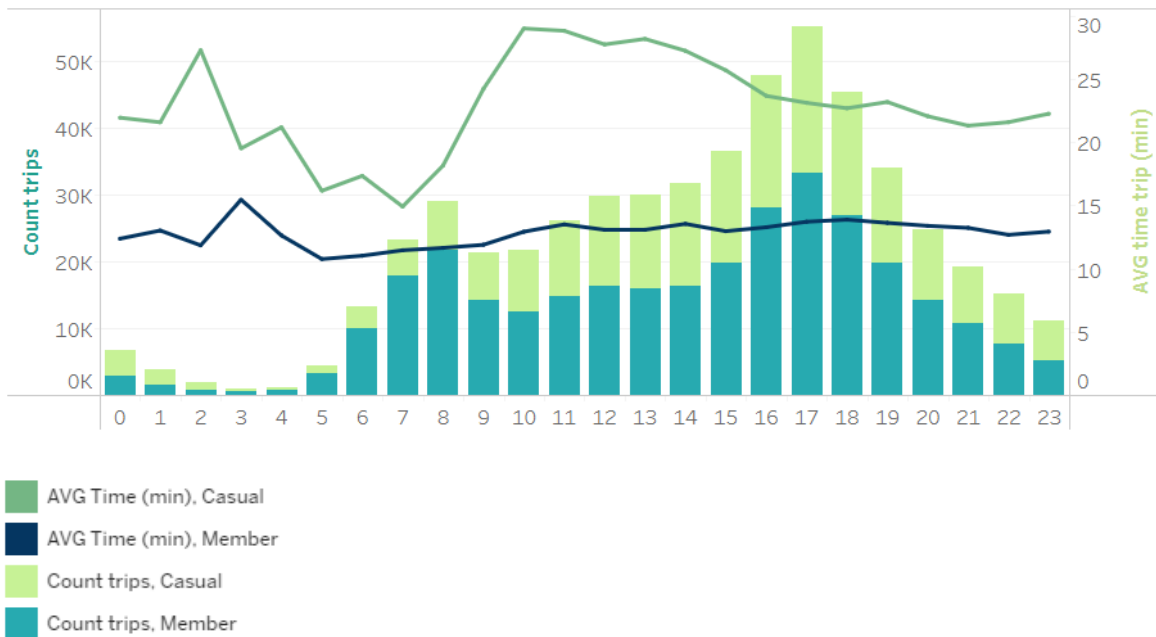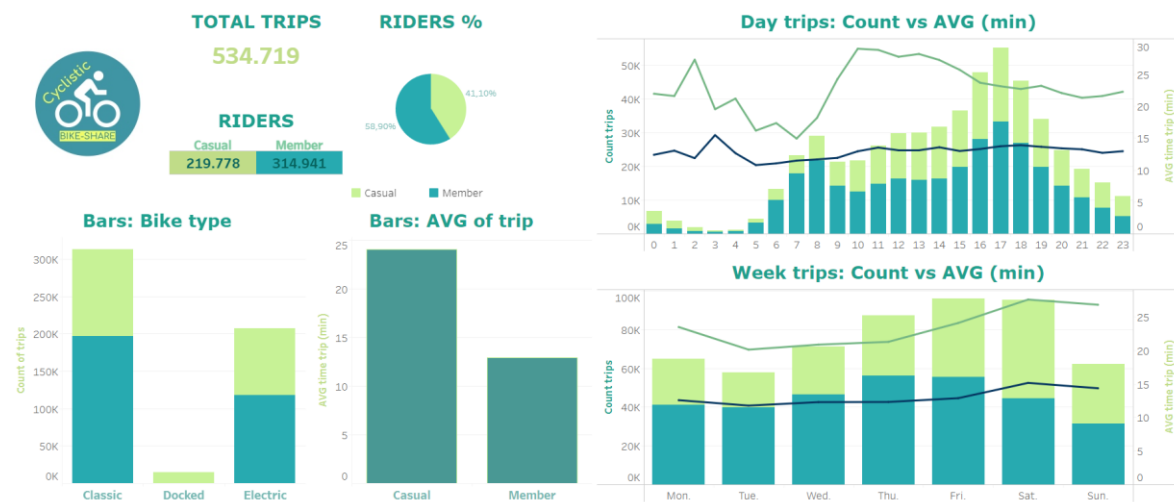
Act

We already concluded what is the difference between the use of bicycles for riders. But some possible recommendations that occur to me in order for casual riders to acquire memberships is:

- Offer a type of membership only for weekends (including Fridays), since it would be more adapted to your needs.

- Offer some kind of discount on the membership for that casual rider who is recurring or who has an accumulated amount of time using our bikes.

This could be done through social media, social networks, emailing or carry out a small campaign in the 5 busiest stations by casual riders.