Divvy Bikes 2019 Data Report

David Teixeira

8/9/2021

DIVVY BIKES

A BRIEF LOOK AT THE COMPANY

Divvy bikes is a bicycle sharing system in Chicago. A vision that Chicago Mayor Richard M. Daley brought from Pairs - France in 2007.

On June 28, 2013, Divvy launched with 750 bikes at 75 stations in an area from the Loop north to Berwyn Ave, west to Kedzie Ave, and south to 59th St.

As of September 2019, Divvy had 594 stations in Chicago.

Visit Divvy Bikes Chicago

GOALS

WHAT I'M TRYIN TO ACHIEVE

This presentation serves as a portfolio case for the Google Data Analytics - Professional Certificate.

Design marketing strategies aimed at converting casual riders into annual members.

For that I was tasked with answering the following.

How do annual members and casual riders use Cyclist bikes differently?

My goal is to answer the question trough the data provided and come up with ideas on to converting riders.

THE DATA

ABOUT THE DATA USED

The data available for this project is available at https://divvy-tripdata.s3.amazonaws.com/index.html.

The chosen year was 2019.

Close to 4 million (3818004) observations in 12 columns.

Names of the columns:

trip_id, start_time, end_time, bikeid, tripduration, from_station_id, from_station_name, to_station_id, to_station_name, usertype, gender, birthyear

INDEX

WHAT ARE WE GONNA SEE

We will be looking at general population before we move on to look at behavior of the subscribers and then the casual.

We will look at their differences based on numbers and make assumptions on how and why they use the service.

This will answer the main question.

> How are they different?

And in the end I will offer some actions in order to solve the secondary problem.

> How to turn casual riders in to subscribers.

In the appendix of this presentation you will be able to look at my code.

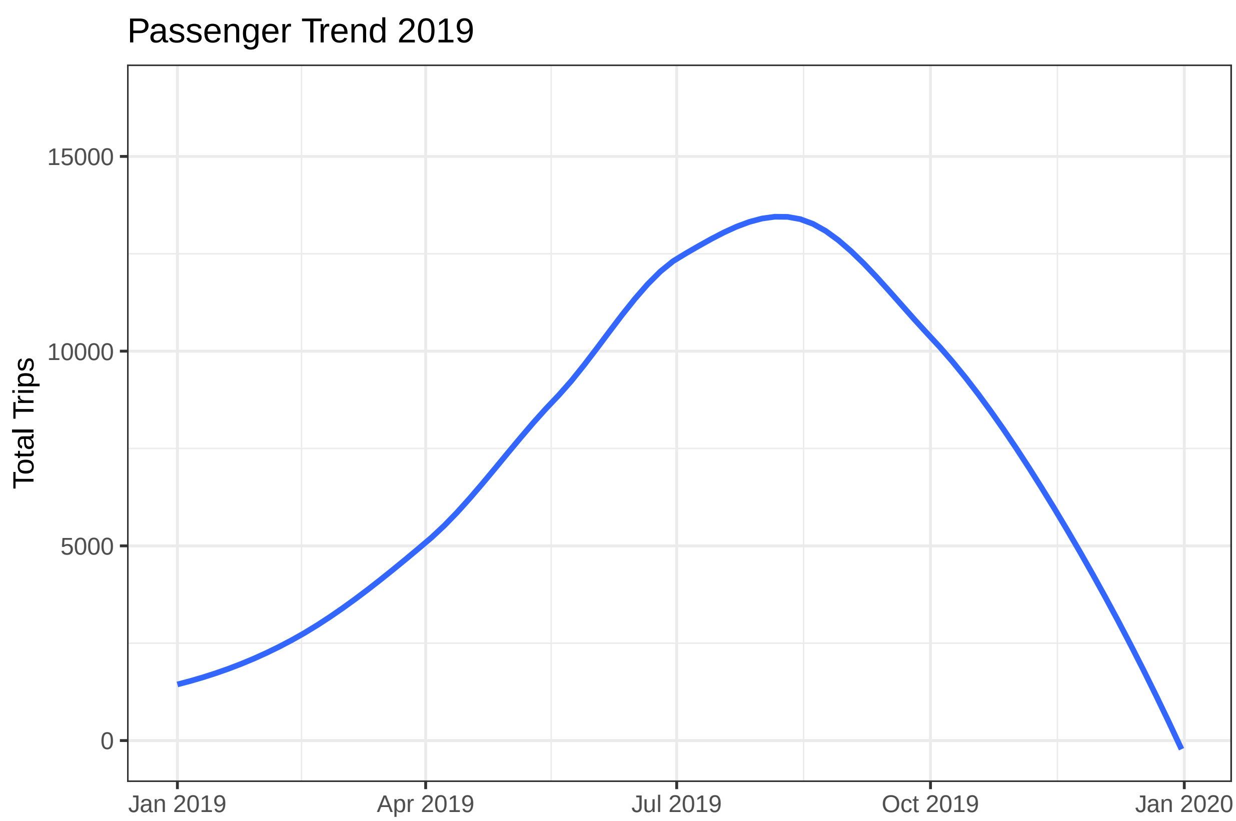
The raw code in R and the Markdown Code that made this report.

This means that everything that is represented here will be reproducible.

DIVVY CLIENTS

Let's look at general population

- In 2019 there where **3818004** individual rides all over Chicago.
- There's a noticeable trend towards the summer months with a peak around the month of August.



Total Rides per Month showing a peak in August

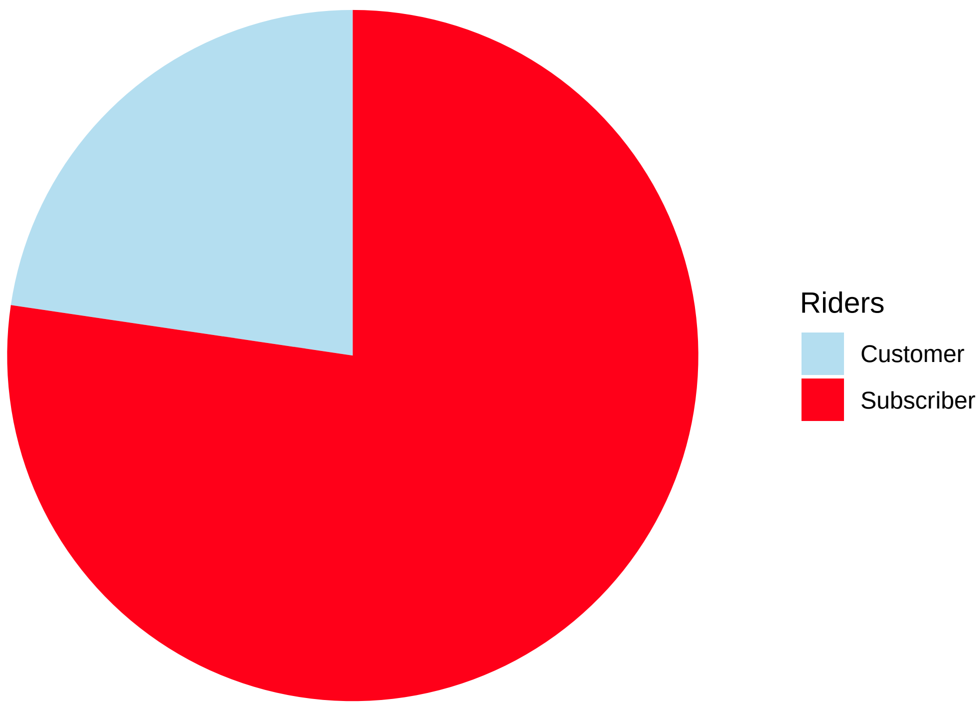
- The mean trip in 2019 was **16.55 minutes** but the most usual time was actually **7 minutes.**
- The top 10 Starting and End Stations where:

```
##        Top 10 Starting Stations Rides
## 1        Streeter Dr & Grand Ave 48589
## 2            Canal St & Adams St 43298
## 3  Clinton St & Washington Blvd 37071
```

```
## 4     Lake Shore Dr & Monroe St 35818
## 5       Clinton St & Madison St 35048
## 6  Michigan Ave & Washington St 28948
## 7    Lake Shore Dr & North Blvd 28688
## 8     Columbus Dr & Randolph St 27995
## 9           Theater on the Lake 26643
## 10      Franklin St & Monroe St 26284
```

```
##          Top 10 End Stations Rides
## 1       Streeter Dr & Grand Ave 56944
## 2           Canal St & Adams St 39560
## 3  Clinton St & Washington Blvd 38804
## 4    Lake Shore Dr & North Blvd 34131
## 5       Clinton St & Madison St 32566
## 6  Michigan Ave & Washington St 30151
## 7           Theater on the Lake 28625
## 8     Lake Shore Dr & Monroe St 27510
## 9         Michigan Ave & Oak St 26792
## 10              Millennium Park 25774
```

- Total Subscribers vs Casual Riders

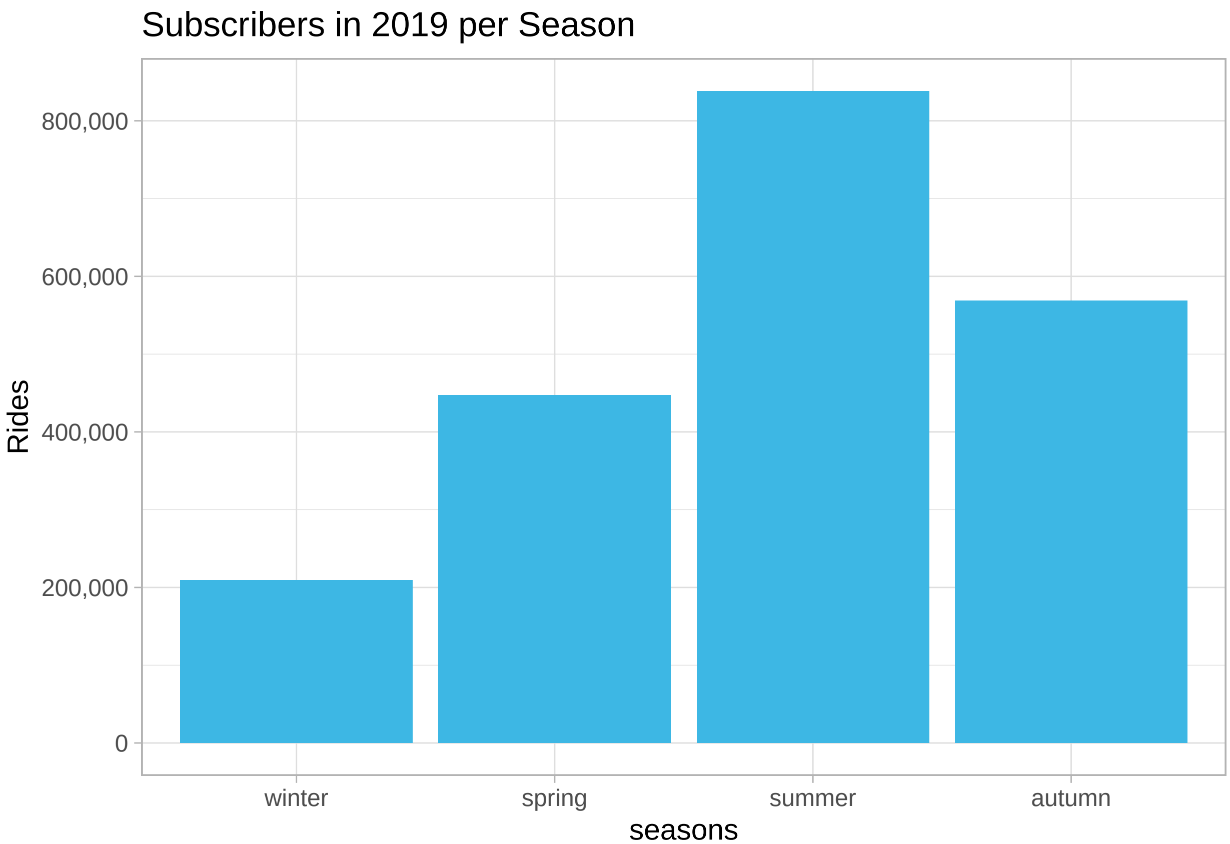



```
## $title
## [1] "Total Riders distribution"
##
## attr(,"class")
## [1] "labels"
```

More than 3 times Subscribers versus Customers

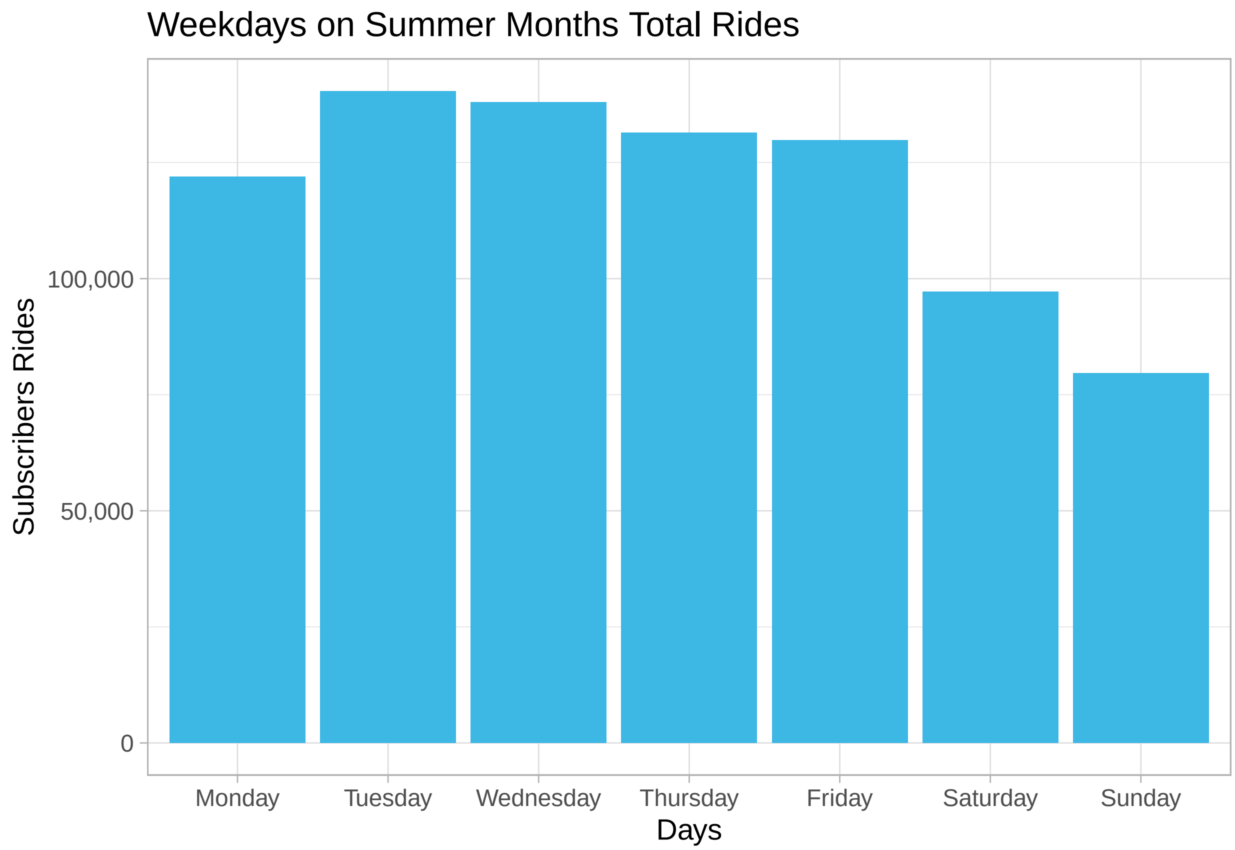
DIVVY SUBSCRIBERS

Let's take a look at Subscribers subset

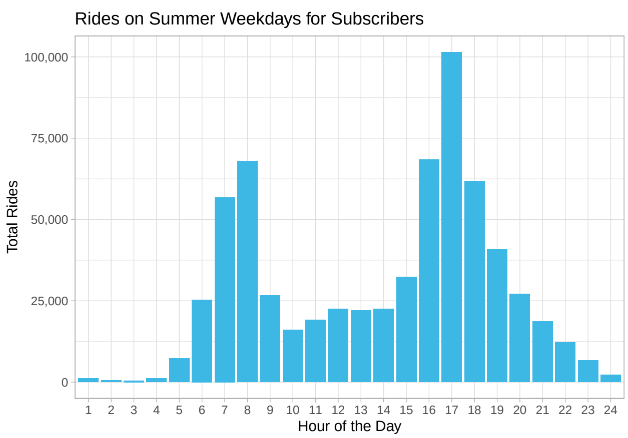


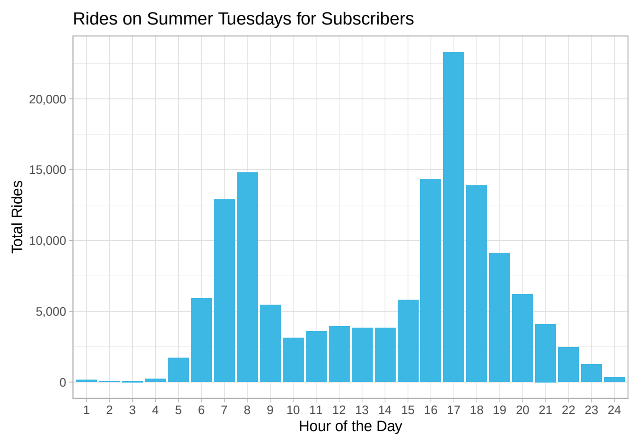
- As we can see from the representation above, the original trend towards the summer maintains.
- The mean trip in 2019 was **15.06 minutes** and the most usual time was **7 minutes.**

Let's look at the summer months, as they are the top season



It is somewhat visible a downward trend towards the weekend. Being weekend the least popular days.

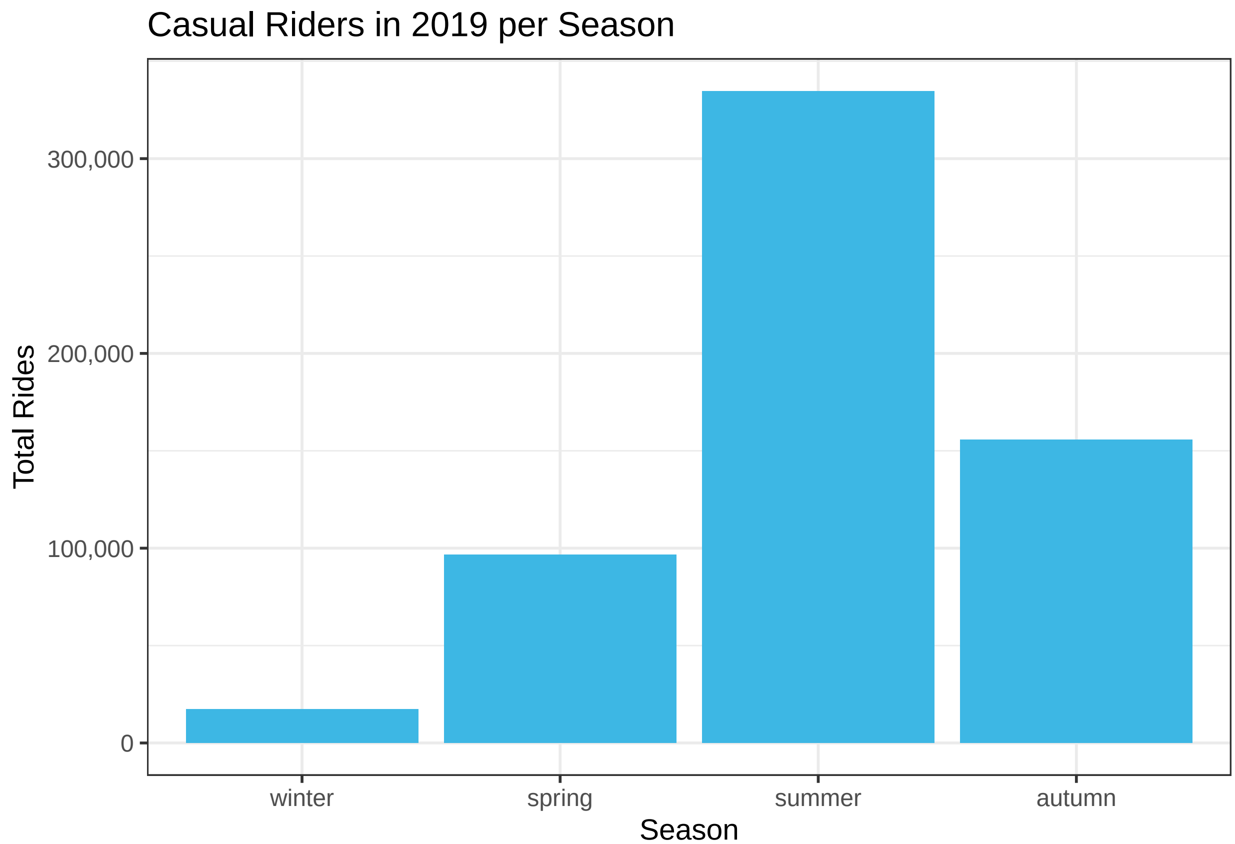




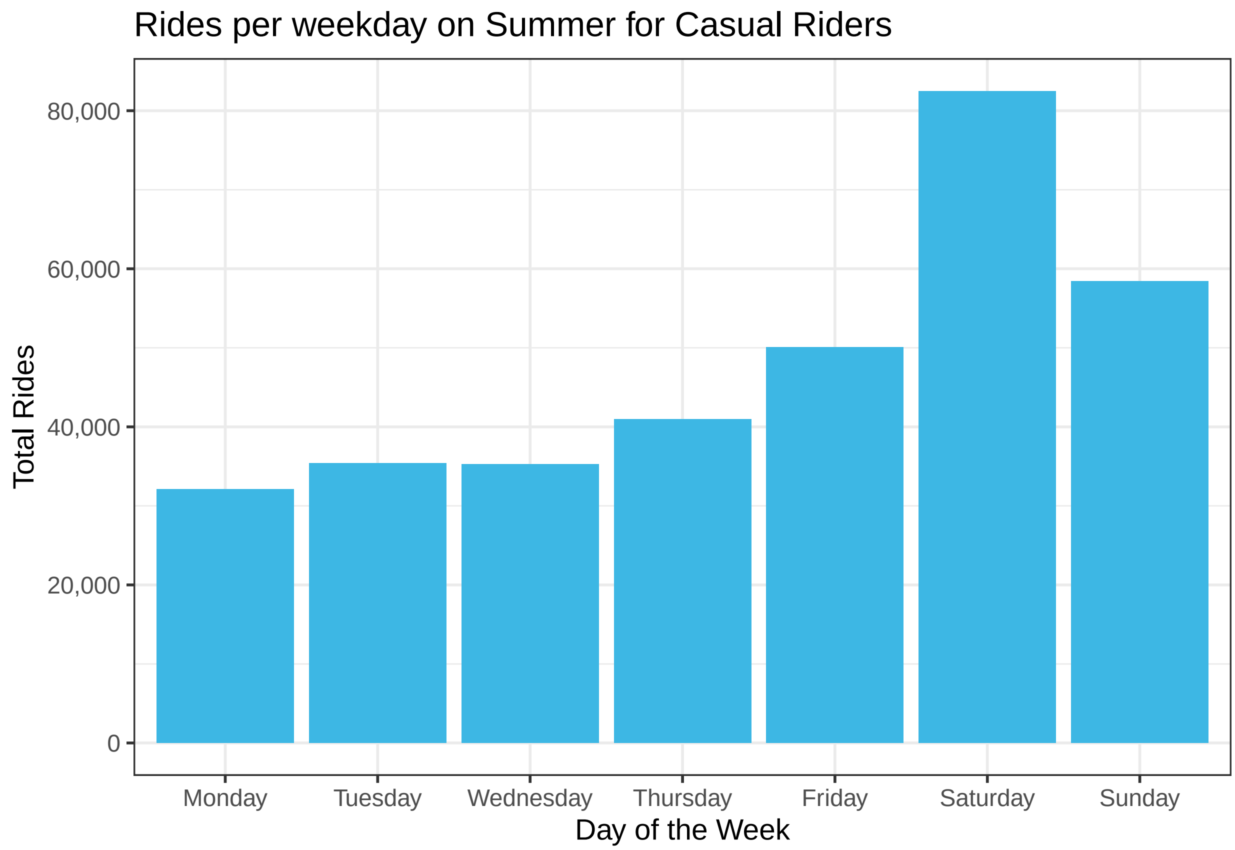
By looking at the hour usage on weekdays and then on Tuesdays we can see that the rush hour, morning and afternoon is the most used time of the day for subscribers.

DIVVY NON SUBSCRIBERS

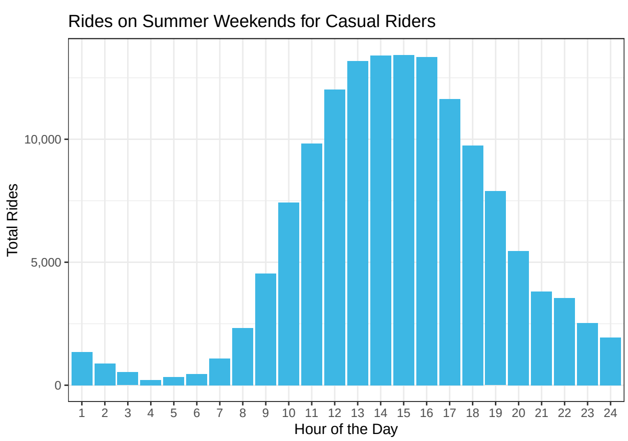
Let's move on to casual riders usage of Divvy Bikes

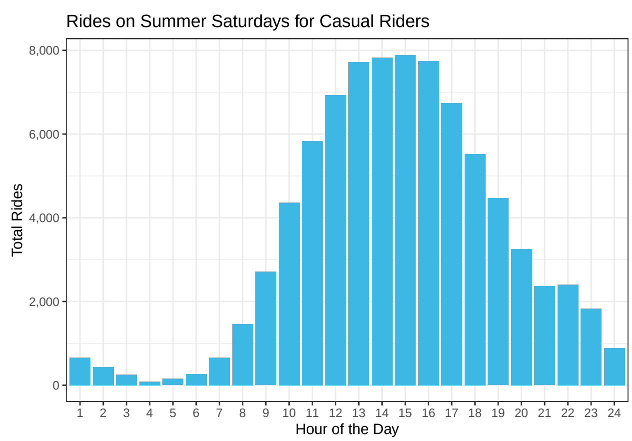


Here again we can see with no surprise that the best time of the year is the Summer months



Here we start to see something different, casual riders seem to prefer the weekends and not the weekdays.

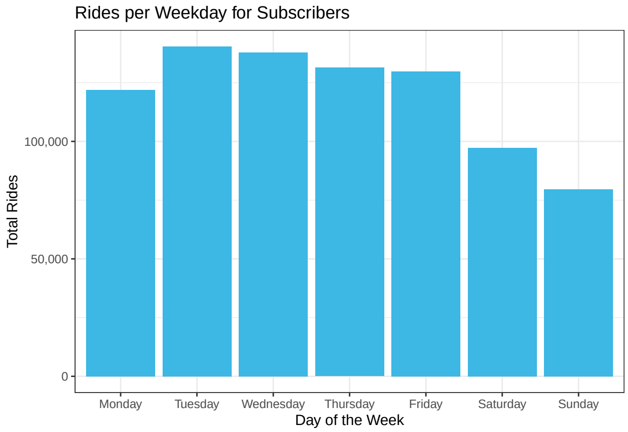


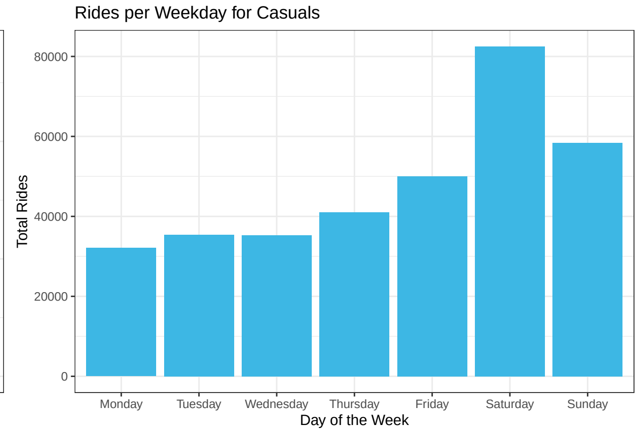


Here we can see that they also prefer different hours of the day when it comes to ride. Casuals do share the late rush hour but they ride all afternoon in mass.

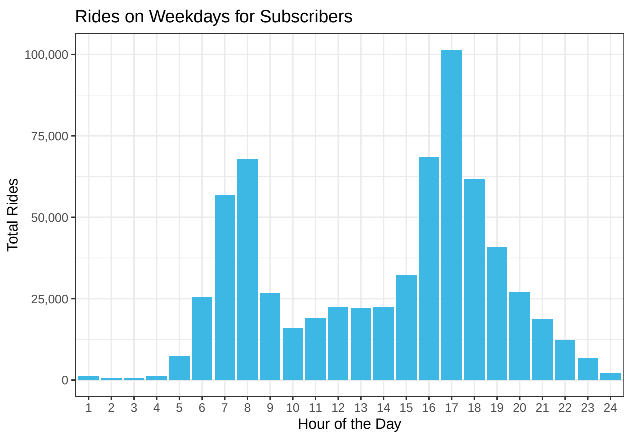
DIVVY COMPARISON

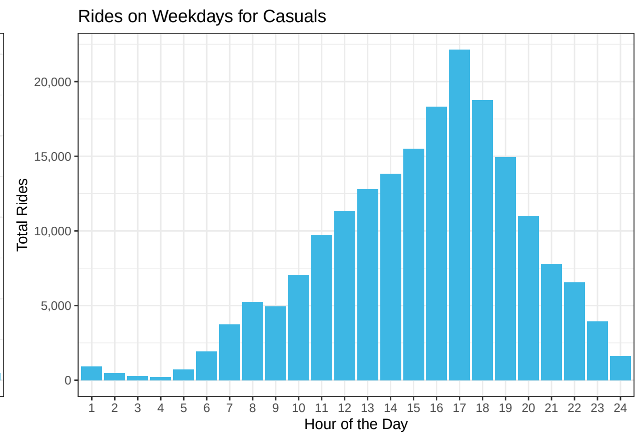
HOW ARE SUBSCRIBERS DIFFERENT FROM CASUALS?



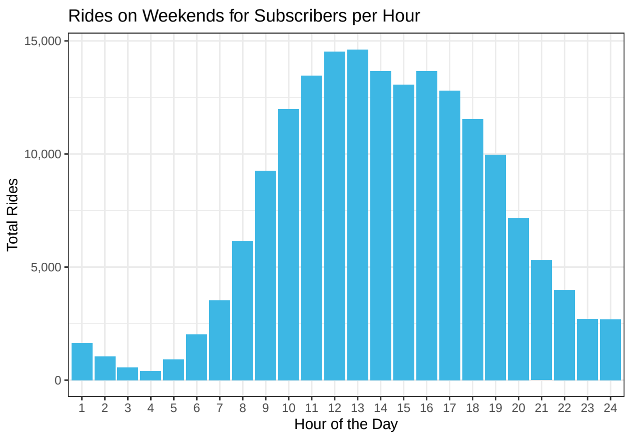


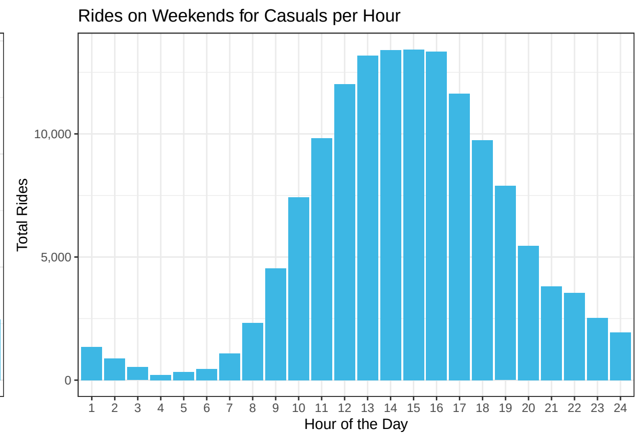
This is our first finding: Subscribers prefer the weekdays, Casuals prefer the weekends.





This is our second difference between them. Here we can observe two distinct peaks for Subscribers and only major for Casual Riders.





But if we look at weekends only we can say that Subscriber do act like Casual riders

Subs_Fav_Sattion	Total_Subs	Cust_Fav_Station	Total_Cust
Canal St & Adams St	40344	Streeter Dr & Grand Ave	35776
Clinton St & Washington Blvd	34827	Lake Shore Dr & Monroe St	26621
Clinton St & Madison St	31895	Shedd Aquarium	15765
Columbus Dr & Randolph St	23568	Lake Shore Dr & North Blvd	14707

Subs_Fav_Sattion	Total_Subs	Cust_Fav_Station	Total_Cust
Franklin St & Monroe St	23562	Millennium Park	13122
Daley Center Plaza	23045	Michigan Ave & Oak St	12976
Michigan Ave & Washington St	21463	Theater on the Lake	11188
Canal St & Madison St	20537	Adler Planetarium	8754
Kingsbury St & Kinzie St	17834	Dusable Harbor	8470
Orleans St & Merchandise Mart Plaza	16424	Michigan Ave & Washington St	7485

In terms of stations we can see as well that they do have their unique choices.

DIVVY COMPARISON

HOW ARE SUBSCRIBERS DIFFERENT FROM CASUALS?

So we can definitely conclude that based on the data available they do act different but only when it comes to weekdays vs weekends, and the time of the day chosen for riding. Which would imply that they probably have different reasons to use the bicycle.

Those are their main differences. And off course as we have seen above their choice of stations do diverge as well.

CONCLUSIONS AND SUGESTIONS

HOW CAN WE TURN CASUALS INTO SUBSCRIBERS?

This was the second problem/question to answer. In order to be completely honest, the data do show some aspects of the type of riders, but one can only assume without asking.

My recommendation would be go out and survey the clients and find out their reasons.

Nonetheless we can say (theoretical) that since there are a peak towards the summer, one could offer a simple summer pass, from June to August.

Would they quit after that, maybe not.

If we look at the data that is available her we can see a that the summer trend do extend towards the Autumn.

That's why after the Special Summer Pass one would be offered a full year subscription for the next 9 months with a special discount on what they already had payed for the Special Summer Pass.

In this logic, one would first go for a "test drive" and then would have a supposed 25% discount on deciding to stay on because Autumn is still a nice ride in Chicago.

Other steps could be taken into further study's if more data is made available.

If it's true (in data), that one person would be more healthy just because of riding a bicycle 15 min on average per day.

One could say to employers in the area, support and offer Divvy Subscriptions to your employees and they will be less absent from work due to the fact that their health would improve and they would save money.

Campaigning people on the fact that by subscribing to divvy and riding they would be healthier, help the planet, get better summers and winter and save on their own health bills. A win win situation.

BEFORE YOU GO

A NOTE ON THE PROJECT

Just like in the first preliminary presentation I would like to finish with some notes on the project and a thank you.

This time I would like to address my sincere Thank you to "Professor (Instructor) Charles Lanfear" from University of Washington.

His lectures made available by him trough is YouTube Chanel as changed and helped my code by so much.

You will be able to notice exactly that on the appendix.

What is my take on all of this. - I need to practice and study more when in comes to Programming in R. - My Statistics background is not enough, even if the core should be left to the Data Scientist a Data Analyst needs to know more than the basic, it's a requirement.

Where will I go from here: - Well more projects for sure. - Statistics is another one.

APPENDIX

From here on it get's boring for some, interesting for others

I would appreciate that if you have any suggestion related to the project or code to address me at davesticteixeira@gmail.com.

The Code in R as I've written from the beginning, be advised that despite of finding new and better ways of writing I've chosen not to change anything that had been already written.

```
# Adding Library's

library(tidyverse)
library(skimr)
library(hydroTSM)
library(lubridate)

# Add Data - Using read.csv

Q1 <- read_csv("data/Divvy_Trips_2019_Q1.csv", col_names = TRUE)
Q2 <- read_csv("data/Divvy_Trips_2019_Q2.csv", col_names = TRUE)
Q3 <- read_csv("data/Divvy_Trips_2019_Q3.csv", col_names = TRUE)
Q4 <- read_csv("data/Divvy_Trips_2019_Q4.csv", col_names = TRUE)

# Correcting columns names from Q2 to match with others

col_names_list <-
  as.list(names(Q1)) # First create a list with the original col names

colnames(Q2) <-
  c(col_names_list) # Use the list to correct the col names in Q2

# Adding all data sets together in one using bind_rows

df <- bind_rows(Q1, Q2, Q3, Q4)

# Counting rows of Q data and df to make sure it all matches

sum(count(Q1), count(Q2), count(Q3), count(Q4))

count(df)

# It seems all data is accountable

remove(Q1,Q2,Q3,Q4,col_names_list) # Remove Data from Global Environment
```

```r
# Let's make a copy of the original data

bike_trip <- df

# Save the bike_trip df in order to be easier to load data if restarted

# Don't Run
# write_csv(bike_trip, "bike_trip.csv")

# Don't Run
# bike_trip <- read.csv("bike_trip.csv") # Save time by loading the df in case R is restarted

# Get the trip duration in minutes

bike_trip$trip <- difftime(bike_trip$end_time, bike_trip$start_time)

bike_trip$trip <- round(bike_trip$trip) # Round to minutes

bike_trip$trip <- as.numeric(bike_trip$trip)

# Remove Original Trip_Duration column

bike_trip <- subset(bike_trip, select = -c(tripduration))

# Remove negative ride times

bike_trip <-
  bike_trip %>%
  filter(trip >= 0)

# Remove the lower and upper quartile based on IQR

bike_trip <-
  bike_trip %>%
  filter(trip <= (IQR(trip)*1.5) + (quantile(trip, 0.75))) %>%
  filter(trip >= quantile(trip,0.25))

max(bike_trip$trip)
min(bike_trip$trip)

# Make new columns Weekday / Month / Hour

bike_trip <-
  bike_trip %>%
  mutate(
    month = month(start_time),
    weekday = weekdays(start_time),
    hour = hour(start_time)
  )

bike_trip$weekday <-
  factor(bike_trip$weekday,
         levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))
```

```r
# Reformat Hour 0:23 to 1:24

bike_trip$hour[bike_trip$hour == "0"] <- "24"

bike_trip$hour <-
  factor(bike_trip$hour,
         levels = c("1","2","3","4","5","6","7","8",
                    "9","10","11","12","13","14","15","16",
                    "17","18","19","20","21","22","23","24"))

# Adding Season with a different package

bike_trip$season <- time2season(bike_trip$start_time,
                               out.fmt = "seasons",
                               type = "default")

unique(bike_trip$season) # confirmation / Autumn is misspelling

bike_trip$season[bike_trip$season == "autumm"] <- "autumn" # Correction of the misspelling

bike_trip$season <-
  factor(bike_trip$season,
         levels = c("winter","spring","summer","autumn"))

# Add Date column

bike_trip$date <- as.Date(bike_trip$start_time)

# Stations

# Top 10 Stations Gen Pop Start and End

top_10_stations_popgen_start <-
    as.data.frame(table(
      bike_trip$from_station_name))

top_10_stations_popgen_start <-
        top_10_stations_popgen_start %>%
          rename( "Top 10 Starting Stations" = Var1,
                  "Rides" = Freq)

top_10_stations_popgen_end <-
  as.data.frame(table(
    bike_trip$to_station_name))

top_10_stations_popgen_end <-
  top_10_stations_popgen_end %>%
  rename( "Top 10 End Stations" = Var1,
          "Rides" = Freq)

# This filters Customer

top_10_stat_cust <-
```

```
  bike_trip %>%
  filter(str_detect
         (usertype, "Customer")) # Next table with stations use

top_10_stat_cust <-
  as.data.frame(table
                (top_10_stat_cust$from_station_name)) # Next table with stations use

top_10_stat_cust <-
  head(arrange
      (top_10_stat_cust, desc(Freq)), n = 10) # And finally the top 10 Per Customer

top_10_stat_cust <-
  top_10_stat_cust %>%
  rename( "Top 10 Starting Stations" = Var1, "Total Rides" = Freq)



# Time to start messing with the plots

# Annual Trend for total number of trips in 2019

day_tri_cum <- (as.data.frame(table(bike_trip$date)))

day_tri_cum$Var1 <- as.Date(day_tri_cum$Var1)

day_tri_cum %>%
  ggplot(aes(x = Var1, y = Freq))+
  geom_point(alpha = 0)+
  geom_smooth(se = FALSE, method = "loess", formula = y~x)+
  labs(title = "Passenger Trend 2019", x = "", y = "Total Trips")+
  theme_bw()

# Avg and Mode Trip_Time for 2019

#Add mode function

getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

mean(bike_trip$trip)
getmode(bike_trip$trip)

# Subscribers Vs Customer

subs_vs_cust <-
  as.data.frame(table(bike_trip$usertype)) %>%
  rename(Riders = Var1)


ggplot(subs_vs_cust, aes(x="", y=Freq, fill=Riders), color = "white") +
```

```r
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  theme_void()+
  scale_fill_manual(values=c("#b4def0", "#ff0019"))


# Subscribers behavior

#DF filtered for subscribers

bike_trip_subs <-
  filter(bike_trip, usertype == "Subscriber")

bike_trip_subs$season <-
  factor(bike_trip_subs$season,
  levels = c("winter","spring","summer","autumn"))

bike_trip_subs$weekday <-
  factor(bike_trip_subs$weekday,
  levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))

bike_trip_subs$hour <-
  factor(bike_trip_subs$hour,
  levels = c("1","2","4","5","6","7","8",
             "9","10","11","12","13","14","15","16",
             "17","18","19","20","21","22","23","24"))

# Create DF for season and plot

bike_trip_subs_season <-
  as.data.frame(table(bike_trip_subs$season))%>%
  rename(seasons = Var1, Rides = Freq)

bike_trip_subs_season %>%
  ggplot(aes(x = seasons, y = Rides))+
  geom_col(fill = "#3db7e4")+
  theme_light()

# Subsetting for the summer months only and looking at the weekdays

bike_trip_subs_summer_weekdays <-
  bike_trip_subs %>% filter(season == "summer")

bike_trip_subs_summer_weekdays <-
  as.data.frame(table(bike_trip_subs_summer_weekdays$weekday)) %>%
  rename(Days = Var1, Rides = Freq)

bike_trip_subs_summer_weekdays %>%
  ggplot(aes(x = Days, y = Rides))+
  geom_col(fill = "#3db7e4")+
  theme_light()

# Rides on Summer Months by hour on weekdays.
```

```
# This is by far the best approach so far.

  # Filtler out the desired conditions.

  bike_trip %>%
  filter(usertype == "Subscriber",
         season == "summer",
         !weekday %in% c("Saturday", "Sunday")) %>%

  # And then Plot

  ggplot(aes(x = hour))+
  geom_bar(fill = "#3db7e4")+
  theme_light()

  # Repeat with Tuesday only

  bike_trip %>%
  filter(usertype == "Subscriber",
         season == "summer",
         weekday == "Tuesday") %>%

  # And then Plot

  ggplot(aes(x = hour))+
  geom_bar(fill = "#3db7e4")+
  labs(title = "Rides on Tuesdays Summer Months for Subscribers",
       x = "Hour of the Day",
       y = "Total Rides")+
  theme_bw()

  # Use the same logic to get the mean trip.

  bike_trip %>%
  filter(usertype == "Subscriber",
         season == "summer",
         weekday == "Tuesday") %>%
  summarise(mean = mean(trip))

# -- A Note before moving forward --
# I should probably go back and change all the code that is behind
# the logic (mine) for not doing it, is it only to show and see how far did I came
# from the beginning of this project. And for me that is important.

# Looking at Casual Riders

  # By Season

  bike_trip %>%
    filter(usertype == "Customer") %>%

    ggplot(aes(x = season))+
    geom_bar(fill = "#3db7e4")+
```

```r
  labs(title = "Rides on Tuesdays Summer Months for Subscribers",
       x = "Hour of the Day",
       y = "Total Rides")+
  theme_bw()

# By weekday in Summer

bike_trip %>%
  filter(usertype == "Customer",
         season == "summer") %>%

  ggplot(aes(x = weekday))+
  geom_bar(fill = "#3db7e4")+
  labs(title = "Rides on Tuesdays Summer Months for Subscribers",
       x = "Hour of the Day",
       y = "Total Rides")+
  theme_bw()

# By hour on weekends and on Saturday Summer season

# Weekends

bike_trip %>%
  filter(usertype == "Customer",
         season == "summer",
         weekday %in% c("Saturday", "Sunday")) %>%

  ggplot(aes(x = hour))+
  geom_bar(fill = "#3db7e4")+
  labs(title = "Rides on Tuesdays Summer Months for Subscribers",
       x = "Hour of the Day",
       y = "Total Rides")+
  theme_bw()

  # Saturday

bike_trip %>%
  filter(usertype == "Customer",
         season == "summer",
         weekday == "Saturday") %>%

  ggplot(aes(x = hour))+
  geom_bar(fill = "#3db7e4")+
  labs(title = "Rides on Tuesdays Summer Months for Subscribers",
       x = "Hour of the Day",
       y = "Total Rides")+
  theme_bw()

# Weekdays

bike_trip %>%
  filter(usertype == "Customer",
         season == "summer",
```

```
          !weekday %in% c("Saturday", "Sunday")) %>%

    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Tuesdays Summer Months for Subscribers",
         x = "Hour of the Day",
         y = "Total Rides")+
    theme_bw()

  #END
```

And here is the RMarkdown with all the visible code used to generate him.

```
# Library's to be loaded in order to make the document function.

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

```
#Function to get the mode in R

function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

options(scipen = 999999)
```

DIVVY BIKES

A BRIEF LOOK AT THE COMPANY

Divvy bikes is a bicycle sharing system in Chicago. A vision that Chicago Mayor Richard M. Daley brought from Pairs - France in 2007.

On June 28, 2013, Divvy launched with 750 bikes at 75 stations in an area from the Loop north to Berwyn Ave, west to Kedzie Ave, and south to 59th St.

As of September 2019, Divvy had 594 stations in Chicago.

Visit Divvy Bikes Chicago

GOALS

WHAT I'M TRYIN TO ACHIEVE

This presentation serves as a portfolio case for the Google Data Analytics - Professional Certificate.

Design marketing strategies aimed at converting casual riders into annual members.

For that I was tasked with answering the following.

How do annual members and casual riders use Cyclist bikes differently?

My goal is to answer the question trough the data provided and come up with ideas on to converting riders.

THE DATA

ABOUT THE DATA USED

```
colnames_list <- colnames(df)
number_obs_df <- nrow(df)
number_of_columns_df <- length(df)
```

The data available for this project is available at https://divvy-tripdata.s3.amazonaws.com/index.html.

The chosen year was 2019.

Close to 4 million (3818004) observations in 12 columns.

Names of the columns:

trip_id, start_time, end_time, bikeid, tripduration, from_station_id, from_station_name, to_station_id, to_station_name, usertype, gender, birthyear

INDEX

WHAT ARE WE GONNA SEE

We will be looking at general population before we move on to look at behavior of the subscribers and then the casual.

We will look at their differences based on numbers and make assumptions on how and why they use the service.

This will answer the main question.

> How are they different?

And in the end I will offer some actions in order to solve the secondary problem.

> How to turn casual riders in to subscribers.

In the appendix of this presentation you will be able to look at my code.

The raw code in R and the Markdown Code that made this report.

This means that everything that is represented here will be reproducible.

DIVVY CLIENTS

Let's look at general population

- In 2019 there where **3818004** individual rides all over Chicago.
- There's a noticeable trend towards the summer months with a peak around the month of August.

```
day_tri_cum %>%
  ggplot(aes(x = Var1, y = Freq))+
  geom_point(alpha = 0)+
  geom_smooth(se = FALSE, method = "loess", formula = y~x)+
  labs(title = "Passenger Trend 2019", x = "", y = "Total Trips")+
  theme_bw()
```

Total Rides per Month showing a peak in August

```
mean_trip <- round(mean(bike_trip$trip),2)
mode_trip <- getmode(bike_trip$trip)
```

- The mean trip in 2019 was **15.06 minutes** but the most usual time was actually **7 minutes.**

- The top 10 Starting and End Stations where:

```r
head(arrange(top_10_stations_popgen_start, desc(Rides)), n = 10)

head(arrange(top_10_stations_popgen_end, desc(Rides)), n = 10)
```

- Total Subscribers vs Casual Riders

```r
ggplot(subs_vs_cust, aes(x="", y=Freq, fill=Riders), color = "white") +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)+
  theme_void()+
  scale_fill_manual(values=c("#b4def0", "#ff0019"))
  labs(title = "Total Riders distribution")
```

More than 3 times Subscribers versus Customers

DIVVY SUBSCRIBERS

Let's take a look at Subscribers subset

```r
bike_trip_subs_season %>%
  ggplot(aes(x = seasons, y = Rides))+
  geom_col(fill = "#3db7e4")+
  labs(title = "Subscribers in 2019 per Season")+
  scale_y_continuous(labels = comma)+
  theme_light()
```

- As we can see from the representation above, the original trend towards the summer maintains.

```r
mean_trip <- round(mean(bike_trip_subs$trip),2)
mode_trip <- getmode(bike_trip_subs$trip)
```

- The mean trip in 2019 was **15.06 minutes** and the most usual time was **7 minutes.**

Let's look at the summer months, as they are the top season

```r
bike_trip_subs_summer_weekdays <-
  bike_trip_subs %>% filter(season == "summer")

bike_trip_subs_summer_weekdays <-
  as.data.frame(table(bike_trip_subs_summer_weekdays$weekday)) %>%
  rename(Days = Var1, Rides = Freq)

bike_trip_subs_summer_weekdays %>%
  ggplot(aes(x = Days, y = Rides))+
  geom_col(fill = "#3db7e4")+
  scale_y_continuous(labels = comma)+
  labs(title = "Weekdays on Summer Months Total Rides", y = "Subscribers Rides")+
  theme_light()
```

It is somewhat visible a downward trend towards the weekend. Being weekend the least popular days.

```r
bike_trip %>%
  filter(usertype == "Subscriber",
         season == "summer",
         !weekday %in% c("Saturday", "Sunday")) %>%
  ggplot(aes(x = hour))+
  geom_bar(fill = "#3db7e4")+
```

```
  labs(title = "Rides on Summer Weekdays for Subscribers",
          x = "Hour of the Day",
          y = "Total Rides")+
  scale_y_continuous(labels = comma)+
  theme_light()

bike_trip %>%
  filter(usertype == "Subscriber",
         season == "summer",
         weekday == "Tuesday") %>%
  ggplot(aes(x = hour))+
  geom_bar(fill = "#3db7e4")+
  labs(title = "Rides on Summer Tuesdays for Subscribers",
          x = "Hour of the Day",
          y = "Total Rides")+
  scale_y_continuous(labels = comma)+
  theme_light()
```

By looking at the hour usage on weekdays and then on Tuesdays we can see that the rush hour, morning and afternoon is the most used time of the day for subscribers.

DIVVY NON SUBSCRIBERS

Let's move on to casual riders usage of Divvy Bikes

```
bike_trip %>%
    filter(usertype == "Customer") %>%
    ggplot(aes(x = season))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Casual Riders in 2019 per Season",
        x = "Season",
        y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()
```

Here again we can see with no surprise that the best time of the year is the Summer months

```
bike_trip %>%
    filter(usertype == "Customer",
           season == "summer") %>%

    ggplot(aes(x = weekday))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides per weekday on Summer for Casual Riders",
        x = "Day of the Week",
        y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()
```

Here we start to see something different, casual riders seem to prefer the weekends and not the weekdays.

```
bike_trip %>%
    filter(usertype == "Customer",
           season == "summer",
           weekday %in% c("Saturday", "Sunday")) %>%
```

```
    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Summer Weekends for Casual Riders",
         x = "Hour of the Day",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()

bike_trip %>%
    filter(usertype == "Customer",
           season == "summer",
           weekday == "Saturday") %>%

    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Summer Saturdays for Casual Riders",
         x = "Hour of the Day",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()
```

Here we can see that they also prefer different hours of the day when it comes to ride. Casuals do share the late rush hour but they ride all afternoon in mass.

DIVVY COMPARISON

HOW ARE SUBSCRIBERS DIFFERENT FROM CASUALS?

```
bike_trip %>%
    filter(usertype == "Subscriber",
           season == "summer") %>%
    ggplot(aes(x = weekday))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides per Weekday for Subscribers",
         x = "Day of the Week",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()

bike_trip %>%
    filter(usertype == "Customer",
           season == "summer") %>%

    ggplot(aes(x = weekday))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides per Weekday for Casuals",
         x = "Day of the Week",
         y = "Total Rides")+
    theme_bw()
```

This is our first finding: Subscribers prefer the weekdays, Casuals prefer the weekends.

```
bike_trip %>%
    filter(usertype == "Subscriber",
           season == "summer",
```

```
           !weekday %in% c("Saturday", "Sunday")) %>%
    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Weekdays for Subscribers",
         x = "Hour of the Day",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()

bike_trip %>%
    filter(usertype == "Customer",
           season == "summer",
           !weekday %in% c("Saturday", "Sunday")) %>%
    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Weekdays for Casuals",
         x = "Hour of the Day",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()
```

This is our second difference between them. Here we can observe two distinct peaks for Subscribers and only major for Casual Riders.

```
bike_trip %>%
    filter(usertype == "Subscriber",
           season == "summer",
           weekday %in% c("Saturday", "Sunday")) %>%
    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Weekends for Subscribers per Hour",
         x = "Hour of the Day",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()

bike_trip %>%
    filter(usertype == "Customer",
           season == "summer",
           weekday %in% c("Saturday", "Sunday")) %>%
    ggplot(aes(x = hour))+
    geom_bar(fill = "#3db7e4")+
    labs(title = "Rides on Weekends for Casuals per Hour",
         x = "Hour of the Day",
         y = "Total Rides")+
    scale_y_continuous(labels = comma)+
    theme_bw()
```

But if we look at weekends only we can say that Subscriber do act like Casual riders

```
x <- bike_trip %>%
  filter(usertype == "Subscriber") %>%
  select(from_station_name) %>%
  count(from_station_name) %>%
  arrange(desc(n)) %>%
```

```
  head(10)

y <- bike_trip %>%
  filter(usertype == "Customer") %>%
  select(from_station_name) %>%
  count(from_station_name) %>%
  arrange(desc(n)) %>%
  head(10)

xy <-as.data.frame(c(x,y))

colnames(xy) <- c("Subs_Fav_Sattion", "Total_Subs", "Cust_Fav_Station", "Total_Cust")

knitr::kable(xy)
```

In terms of stations we can see as well that they do have their unique choices.

DIVVY COMPARISON

HOW ARE SUBSCRIBERS DIFFERENT FROM CASUALS?

So we can definitely conclude that based on the data available they do act different but only when it comes to weekdays vs weekends, and the time of the day chosen for riding. Which would imply that they probably have different reasons to use the bicycle.

Those are their main differences. And off course as we have seen above their choice of stations do diverge as well.

CONCLUSIONS AND SUGESTIONS

HOW CAN WE TURN CASUALS INTO SUBSCRIBERS?

This was the second problem/question to answer. In order to be completely honest, the data do show some aspects of the type of riders, but one can only assume without asking.

My recommendation would be go out and survey the clients and find out their reasons.

Nonetheless we can say (theoretical) that since there are a peak towards the summer, one could offer a simple summer pass, from June to August.

Would they quit after that, maybe not.

If we look at the data that is available her we can see a that the summer trend do extend towards the Autumn.

That's why after the Special Summer Pass one would be offered a full year subscription for the next 9 months with a special discount on what they already had payed for the Special Summer Pass.

In this logic, one would first go for a "test drive" and then would have a supposed 25% discount on deciding to stay on because Autumn is still a nice ride in Chicago.

Other steps could be taken into further study's if more data is made available.

If it's true (in data), that one person would be more healthy just because of riding a bicycle 15 min on average per day.

One could say to employers in the area, support and offer Divvy Subscriptions to your employees and they will be less absent from work due to the fact that their health would improve and they would save money.

Campaigning people on the fact that by subscribing to divvy and riding they would be healthier, help the planet, get better summers and winter and save on their own health bills. A win win situation.

BEFORE YOU GO

A NOTE ON THE PROJECT

Just like in the first preliminary presentation I would like to finish with some notes on the project and a thank you.

This time I would like to address my sincere Thank you to "Professor (Instructor) Charles Lanfear" from University of Washington.

His lectures made available by him trough is YouTube Chanel as changed and helped my code by so much.

You will be able to notice exactly that on the appendix.

What is my take on all of this. - I need to practice and study more when in comes to Programming in R. - My Statistics background is not enough, even if the core should be left to the Data Scientist a Data Analyst needs to know more than the basic, it's a requirement.

Where will I go from here: - Well more projects for sure. - Statistics is another one.

APPENDIX

From here on it get's boring for some, interesting for others

I would appreciate that if you have any suggestion related to the project or code to address me at davesticteixeira@gmail.com.

The Code in R as I've written from the beginning, be advised that despite of finding new and better ways of writing I've chosen not to change anything that had been already written.