cycling\_company

Mohamed Salah

4/28/2021

# ASK phase

##Business task In 2016, Cyclistic (fictional company) launched a successful bike-share offering, and finance team notice an opportunity of revenue growing by increasing annual memberships, thus based on that the marketing manager (primary stakeholder) notice that converting the casual riders to an annual membership program will achieve this goal because casual riders already aware of Cyclistic program and they have a needs of mobility by using company products. Therefore, one of questions need to be answered is How do annual members and casual riders use Cyclistic bikes differently? And to answer that I will analyzing the Cyclistic historical bike trip data to identify trends.

## Preparing Rstudio

(following results under R version 4.0.3 (2020-10-10) – “Bunny-Wunnies Freak Out”Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit)) *> getwd() [1] “D:/userdata/malshark/My Documents”* Loding below packages > library(tidyverse) – Attaching packages ——————————————————————————– tidyverse 1.3.0 – v ggplot2 3.3.3 v purrr 0.3.4 v tibble 3.1.0 v dplyr 1.0.4 v tidyr 1.1.2 v stringr 1.4.0 v readr 1.4.0 v forcats 0.5.1 – Conflicts ———————————————————————————– tidyverse\_conflicts() – x dplyr::filter() masks stats::filter() x dplyr::lag() masks stats::lag()

#Prepare phase ##Need to check data under ROCCC process: \* Reliable: I notice that data is complete and trusted and reflects all customers using company products. \* Original: Data sets already inside company (DIVVY) \* Comprehensive: Data have all important information need to answer the question task. \* Current: Data source information represented all data belong to 2015 which are need it for 2016 marketing plan. \* Cited: data already checked and examined. ## Need to check licensing, privacy, security, and accessibility \* Under license from this [link](https://www.divvybikes.com/data-license-agreement) \* Term of use in this [link](https://www.divvybikes.com/terms-of-use) \* Privacy in this [link](https://www.lyft.com/privacy). In addition, all customers with annual memberships data have been hidden such as sex and birthday. \*I confirm integrity based on About the Owner section: Divvy is a program of the Chicago Department of Transportation (CDOT), which owns the city’s bikes, stations and vehicles. Initial funding for the program came from federal grants for projects that promote economic recovery, reduce traffic congestion and improve air quality, as well as additional funds from the City’s Tax Increment Financing program. In 2016, Divvy expanded to the neighboring suburb of Evanston with a grant from the State of Illinois. ## A description of all data sources used: For analyzing purpose I will use the following data from [link](https://divvy-tripdata.s3.amazonaws.com/index.html) Before begin I notice I have Q1.csv file represent month (1,2,3) and .csv file per month so I decide it is better for me to work under one .csv. per quarter Therefore, I will collect months (3,4,5) under Q2 and months (6,7,8) under Q3 I face a problem with month 11 so I will keep months (10,11,12) separates. Finally, below details of every file

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.0.5

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.1.1 v dplyr 1.0.5  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

Divvy\_Trips\_2020\_Q1 <- read.csv(“D:/userdata/malshark/My Documents/Divvy/Divvy\_Trips\_2020\_Q1.csv”) head(Divvy\_Trips\_2020\_Q1) X ride\_id started\_at ended\_at start\_station\_id end\_station\_id member\_casual 1 1 EACB19130B0CDA4A 2020-01-21 20:06:59 2020-01-21 20:14:30 239 326 member 2 2 8FED874C809DC021 2020-01-30 14:22:39 2020-01-30 14:26:22 234 318 member 3 3 789F3C21E472CA96 2020-01-09 19:29:26 2020-01-09 19:32:17 296 117 member 4 4 C9A388DAC6ABF313 2020-01-06 16:17:07 2020-01-06 16:25:56 51 24 member 5 5 943BC3CBECCFD662 2020-01-30 08:37:16 2020-01-30 08:42:48 66 212 member 6 6 6D9C8A6938165C11 2020-01-10 12:33:05 2020-01-10 12:37:54 212 96 member

* I will join 3 .csv files (months 4,5,6) together to have one .csv file for 2020-Q2

tbl2 <- sapply(files, read\_csv, simplify=FALSE) %>% bind\_rows(.id = “id”) write.csv(tbl2, ‘Divvy\_Trips\_2020\_Q2’)

\*Once I finish write.csv I Import the new Dataset from Rstudio Environment. head (Divvy\_Trips\_2020\_Q2) ## New file after combine (202004-divvy-tripdata,202005-divvy-tripdata,202006-divvy-tripdata) X ride\_id started\_at ended\_at start\_station\_id end\_station\_id member\_casual 1 1 A847FADBBC638E45 2020-04-26 17:45:00 2020-04-26 18:12:00 86 152 member 2 2 5405B80E996FF60D 2020-04-17 17:08:00 2020-04-17 17:17:00 503 499 member 3 3 5DD24A79A4E006F4 2020-04-01 17:54:00 2020-04-01 18:08:00 142 255 member 4 4 2A59BBDF5CDBA725 2020-04-07 12:50:00 2020-04-07 13:02:00 216 657 member 5 5 27AD306C119C6158 2020-04-18 10:22:00 2020-04-18 11:15:00 125 323 casual 6 6 356216E875132F61 2020-04-30 17:55:00 2020-04-30 18:01:00 173 35 member

* I will join 3 .csv files (months 7,8,9) together to have one .csv file for 2020-Q3

files <- list.files(path = “D:/userdata/malshark/My Documents/Divvy”, pattern = "\*.csv", full.names = T) write.csv(tbl2,‘Divvy\_Trips\_2020\_Q3.csv’)

\*Once I finish write.csv I Import the new Dataset from Rstudio Environment.

glimpse(Divvy\_Trips\_2020\_Q3) ## New file after combine (202007-divvy-tripdata, 202008-divvy-tripdata,202009-divvy-tripdata)## Rows: 11,528,415 Columns: 18 $ V1 NA, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, ~ $ V2 “id”, “D:/userdata/malshark/My Documents/Divvy/Divvy\_trips\_2020.csv”, “D:/userdata/malshark/My Documents/Di~ $ V3 ”X1“,”1“,”2“,”3“,”4“,”5“,”6“,”7“,”8“,”9“,”10“,”11“,”12“,”13“,”14“,”15“,”16“,”17“,”18“,”1~ $ V4 “X”, “2”, “3”, “4”, “5”, “6”, “7”, “8”, “9”, “10”, “11”, “12”, “13”, “14”, “15”, “16”, “17”, “18”, “19”, “2~ $ V5 ”trip\_id“,”EACB19130B0CDA4A“,”8FED874C809DC021“,”789F3C21E472CA96“,”C9A388DAC6ABF313“,”943BC3CBECCFD66~ $ V6 “start\_time”, “2020-01-21 20:06:59”, “2020-01-30 14:22:39”, “2020-01-09 19:29:26”, “2020-01-06 16:17:07”, “~ $ V7 ”end\_time“,”2020-01-21 20:14:30“,”2020-01-30 14:26:22“,”2020-01-09 19:32:17“,”2020-01-06 16:25:56“,”20~ $ V8 “from\_station\_id”, “239”, “234”, “296”, “51”, “66”, “212”, “96”, “96”, “212”, “38”, “117”, “181”, “91”, “91~ $ V9 ”to\_station\_id“,”326“,”318“,”117“,”24“,”212“,”96“,”212“,”212“,”96“,”100“,”632“,”91“,”181“,”18~ $ V10 “user\_type”, “member”, “member”, “member”, “member”, “member”, “member”, “member”, “member”, “member”, “mem~ $ V11 ”ride\_length“,”451“,”223“,”171“,”529“,”332“,”289“,”289“,”297“,”295“,”203“,”196“,”380“,”529“,”~ $ V12 “day\_of\_week”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~ $ V13 “ride\_id”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~ $ V14 “started\_at”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~ $ V15 “ended\_at”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~ $ V16 “start\_station\_id”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~ $ V17 “end\_station\_id”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~ $ V18 “member\_casual”, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~

* I will join 2 .csv files (months 10,11) together to have one .csv file for 2020-M10M11 because I notice some problems in M12 which is column (sart station id) in not and whenever I try to combine with other files (10,11) it give me an error and since all files contains the same column in mode except this file I decide to ignor and drop this file.

files <- list.files(path = “D:/userdata/malshark/My Documents/Divvy”, pattern = "\*.csv“, full.names = T) tbl3 <- sapply(files, read\_csv, simplify=FALSE) %>% bind\_rows(.id =”id") write.csv(tbl3,‘Divvy\_Trips\_2020\_M10M11.csv’)

\*Once I finish write.csv I Import the new Dataset from Rstudio Environment.

glimpse(Divvy\_Trips\_2020\_M10M11) ## New file after combine (202010-divvy-tripdata, 202011-divvy-tripdata)## Rows: 648,369 Columns: 15 $ X 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 2~ $ id “D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv”, “D:/userdata/malshark/M~ $ ride\_id ”ACB6B40CF5B9044C“,”DF450C72FD109C01“,”B6396B54A15AC0DF“,”44A4AEE261B9E854“,”10B7DD76A6A~ $ rideable\_type “electric\_bike”, “electric\_bike”, “electric\_bike”, “electric\_bike”, “electric\_bike”, “electr~ $ started\_at ”2020-10-31 19:39:43“,”2020-10-31 23:50:08“,”2020-10-31 23:00:01“,”2020-10-31 22:16:43“, ~ $ ended\_at ”2020-10-31 19:57:12“,”2020-11-01 00:04:16“,”2020-10-31 23:08:22“,”2020-10-31 22:19:35“, ~ $ start\_station\_name ”Lakeview Ave & Fullerton Pkwy“,”Southport Ave & Waveland Ave“,”Stony Island Ave & 67th St~ $ start\_station\_id 313, 227, 102, 165, 190, 359, 313, 125, NA, 174, 114, 349, 49, 344, 701, 29, NA, 216, 676, 6~ $ end\_station\_name “Rush St & Hubbard St”, “Kedzie Ave & Milwaukee Ave”, “University Ave & 57th St”, “Broadway ~ $ end\_station\_id 125, 260, 423, 256, 185, 53, 125, 313, 199, 635, 303, 114, 344, 49, NA, 182, 69, 69, 676, 67~ $ start\_lat 41.92610, 41.94817, 41.77346, 41.95085, 41.92886, 41.90353, 41.92584, 41.89031, 41.90000, 41~ $ start\_lng -87.63898, -87.66391, -87.58537, -87.65924, -87.66396, -87.64335, -87.63904, -87.62626, -87.~ $ end\_lat 41.89035, 41.92953, 41.79145, 41.95281, 41.91778, 41.89440, 41.89047, 41.92574, 41.89122, 41~ $ end\_lng -87.62607, -87.70782, -87.60005, -87.65010, -87.69143, -87.63431, -87.62590, -87.63891, -87.~ $ member\_casual ”casual“,”casual“,”casual“,”casual“,”casual“,”casual“,”casual“,”casual“,”casual",

head(Divvy\_Trips\_2020\_M10M11) X id ride\_id rideable\_type started\_at 1 1 D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv ACB6B40CF5B9044C electric\_bike 2020-10-31 19:39:43 2 2 D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv DF450C72FD109C01 electric\_bike 2020-10-31 23:50:08 3 3 D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv B6396B54A15AC0DF electric\_bike 2020-10-31 23:00:01 4 4 D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv 44A4AEE261B9E854 electric\_bike 2020-10-31 22:16:43 5 5 D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv 10B7DD76A6A2EB95 electric\_bike 2020-10-31 19:38:19 6 6 D:/userdata/malshark/My Documents/Divvy/202010-divvy-tripdata.csv DA6C3759660133DA electric\_bike 2020-10-29 17:38:04 ended\_at start\_station\_name start\_station\_id end\_station\_name end\_station\_id 1 2020-10-31 19:57:12 Lakeview Ave & Fullerton Pkwy 313 Rush St & Hubbard St 125 2 2020-11-01 00:04:16 Southport Ave & Waveland Ave 227 Kedzie Ave & Milwaukee Ave 260 3 2020-10-31 23:08:22 Stony Island Ave & 67th St 102 University Ave & 57th St 423 4 2020-10-31 22:19:35 Clark St & Grace St 165 Broadway & Sheridan Rd 256 5 2020-10-31 19:54:32 Southport Ave & Wrightwood Ave 190 Stave St & Armitage Ave 185 6 2020-10-29 17:45:43 Larrabee St & Division St 359 Wells St & Huron St 53 start\_lat start\_lng end\_lat end\_lng member\_casual 1 41.92610 -87.63898 41.89035 -87.62607 casual 2 41.94817 -87.66391 41.92953 -87.70782 casual 3 41.77346 -87.58537 41.79145 -87.60005 casual 4 41.95085 -87.65924 41.95281 -87.65010 casual 5 41.92886 -87.66396 41.91778 -87.69143 casual 6 41.90353 -87.64335 41.89440 -87.63431 casual

# Process phase

## Cleaning procedure

* Remove some columns not related to my analysis.

Divvy\_Trips\_Q1 <- Divvy\_Trips\_2020\_Q1 [,c(1,3,4,6,8,13)] ## Create new df with only need it columns ##

* Transform from (chr) to (dttm) on both columns(started\_at, ended\_at) in all files.

install.packages(“lubridate”) library (lubridate) suppressPackageStartupMessages({library(lubridate)}) ## to avoid NA while transform from (chr) to (dttm) ##

Divvy\_Trips\_Q1 $started\_at <- ymd\_hms(Divvy\_Trips\_Q1 $started\_at) ## Transform to dttm class ##

Divvy\_Trips\_Q1 $ended\_at <- ymd\_hms(Divvy\_Trips\_Q1 $ended\_at) ## Transform to dttm class ##

glimpse(Divvy\_Trips\_Q1) ## To confirm transformation ##

write.csv(Divvy\_Trips\_Q1, ‘Divvy\_trips\_q1.csv’) ## Save new df ##

Divvy\_Trips\_Q2 <- Divvy\_Trips\_2020\_Q2 [,c(2,3,4,5,7,9,14)] ## Remove some columns and create new df ##

Divvy\_trips\_q2 $started\_at <- ymd\_hms(Divvy\_trips\_q2 $started\_at) ## transform from (chr) to (dttm) ##

Divvy\_trips\_q2 $ended\_at <- ymd\_hms(Divvy\_trips\_q2 $ended\_at) ## transform from (chr) to (dttm) ##

glimpse (Divvy\_trips\_q2) ## confirm transformation ##

Divvy\_Trips\_Q3 <- Divvy\_Trips\_2020\_Q3 [,c(3,5,6,8,10,15)] ## Create new df with need it columns ##

Divvy\_Trips\_Q3 $started\_at <- ymd\_hms(Divvy\_Trips\_Q3 $started\_at)

Divvy\_Trips\_Q3 $ended\_at <- ymd\_hms(Divvy\_Trips\_Q3 $ended\_at)

glimpse(Divvy\_Trips\_Q3) ## Confirm transformation ##

write.csv(Divvy\_Trips\_Q3,‘Divvy\_Trips\_q3.csv’) ## save new df ##

* Transform the last file from df source (202012.divvy.tripdata)

Divvy\_Trips\_M12 <- 202012.divvy.tripdata [,c(1,3,4,6,8,13)] ## This file will join Divvy\_Trips\_10M11##

glimpse (Divvy\_Trips\_M12)

Divvy\_Trips\_M12 $started\_at <- ymd\_hms(Divvy\_Trips\_M12 $started\_at)

Divvy\_Trips\_M12 $ended\_at <- ymd\_hms(Divvy\_Trips\_M12 $ended\_at)

glimpse(Divvy\_Trips\_M12)

write.csv(Divvy\_Trips\_M12, ‘Divvy\_trips\_M12.csv’) ## Save file after transformation complete ##

\*Transform and create Q4

Divvy\_Trips\_M10M11 <- Divvy\_Trips\_2020\_M10M11 [,c(3,5,6,8,10,15)] ## New df ##

Divvy\_Trips\_M10M11 $started\_at <- ymd\_hms(Divvy\_Trips\_M10M11 $started\_at) ## Transform from (chr) to (dttm) ##

Divvy\_Trips\_M10M11 $ended\_at <- ymd\_hms(Divvy\_Trips\_M10M11 $ended\_at) ## Transform from (chr) to (dttm) ##

glimpse(Divvy\_Trips\_M10M11)

write.csv(Divvy\_Trips\_M10M11,‘Divvy\_trips\_M10M11.csv’) ## save new df ##

* I will combine files (Divvy\_trips\_M10M11.csv, Divvy\_trips\_M12) to be one file (Divvy\_Trips\_Q4)
* I notice that Divvy\_trips\_M12 have two columns (start\_station\_id & end\_station\_id) in (chr) which contradicts with other columns on other files which are in (int) mode so I convert both columns.

Divvy\_Trips\_M12 $start\_station\_id <- as.integer (Divvy\_Trips\_M12 $start\_station\_id) ## Transform this column to be same as other columns##

Divvy\_Trips\_M12 $end\_station\_id <- as.integer (Divvy\_Trips\_M12 $end\_station\_id)  
  
glimpse(Divvy\_Trips\_M12) ## confirm that both columns in the right class (int).##  
  
files <- list.files(path = "D:/userdata/malshark/My Documents/Divvy", pattern = "\*.csv", full.names = T) ## Join Divvy\_Trips\_M10M11 and Divvy\_Trips\_M12 in oe file ##tbl5 <- sapply(files, read\_csv, simplify=FALSE) %>% bind\_rows(.id = "id") ## create new df ##  
  
write.csv(tbl5,'Divvy\_trips\_q4.csv') ## Save the new df ##  
  
glimpse(Divvy\_trips\_q4) ## confirm new df and keep only columns need to analysis ##  
  
\* I will Join all file under one file (Divvy\_Trips\_2020)  
  
files <- list.files(path = "D:/userdata/malshark/My Documents/Divvy", pattern = "\*.csv", full.names = T)  
  
tbl <- sapply(files, read\_csv, simplify=FALSE) %>% bind\_rows(.id = "id")  
  
write.csv(table, 'Divvy\_trips\_2020.csv')  
  
Divvy\_Trips\_2020 $ ride\_length <- difftime(Divvy\_Trips\_2020 $ end\_time,Divvy\_Trips\_2020 $ start\_time) # Add a "ride\_length" calculation to all\_trips (in seconds)  
  
head(Divvy\_trips\_2020) #See the first 6 rows of data frame.  
  
tail(Divvy\_trips\_2020)  
  
str(Divvy\_trips\_2020) #See list of columns and data types (numeric, character, etc)  
  
summary(Divvy\_trips\_2020) #Statistical summary of data. Mainly for numerics.  
  
Divvy\_Trips\_2020\_v4 <- Divvy\_trips\_2020[(Divvy\_trips\_2020$ride\_length<0),] ## to show a table for all ride\_length < 0  
  
  
```r  
Divvy\_Trips\_2020\_v4 <- Divvy\_trips\_2020[!(Divvy\_trips\_2020$ride\_length<0),]

## Error in eval(expr, envir, enclos): object 'Divvy\_trips\_2020' not found

summary(Divvy\_Trips\_2020\_v4$ride\_length) Min. 1st Qu. Median Mean 3rd Qu. Max. 0 480 840 1696 1560 9387024

median(Divvy\_Trips\_2020\_v4$ride\_length) #midpoint number in the ascending array of ride lengths [1] 840

max(Divvy\_Trips\_2020\_v4$ride\_length) #longest ride [1] 9387024

min(Divvy\_Trips\_2020\_v4$ride\_length) #shortest ride [1] 0

Divvy\_Trips\_2020\_v4start\_time), “%A”)## To add new column

aggregate(Divvy\_Trips\_2020\_v4user\_type + Divvy\_Trips\_2020\_v4$day\_of\_week, FUN = mean)

Divvy\_Trips\_2020\_v4day\_of\_week Divvy\_Trips\_2020\_v4$ride\_length 1 casual Friday 2823.3506 2 member Friday 932.9811 3 casual Monday 2910.8165 4 member Monday 884.0107 5 casual Saturday 2969.5668 6 member Saturday 1074.6389 7 casual Sunday 3401.7184 8 member Sunday 1079.7364 9 casual Thursday 2917.0429 10 member Thursday 886.7974 11 casual Tuesday 2711.0042 12 member Tuesday 872.4097 13 casual Wednesday 2645.1722 14 member Wednesday 888.7004

Divvy\_Trips\_2020\_v4day\_of\_week, levels=c(“Sunday”, “Monday”, “Tuesday”, “Wednesday”, “Thursday”, “Friday”, “Saturday”)) # Notice that the days of the week are out of order.

aggregate(Divvy\_Trips\_2020\_v4user\_type + Divvy\_Trips\_2020\_v4$day\_of\_week, FUN = mean)

Divvy\_Trips\_2020\_v4day\_of\_week Divvy\_Trips\_2020\_v4$ride\_length 1 casual Sunday 3401.7184 2 member Sunday 1079.7364 3 casual Monday 2910.8165 4 member Monday 884.0107 5 casual Tuesday 2711.0042 6 member Tuesday 872.4097 7 casual Wednesday 2645.1722 8 member Wednesday 888.7004 9 casual Thursday 2917.0429 10 member Thursday 886.7974 11 casual Friday 2823.3506 12 member Friday 932.9811 13 casual Saturday 2969.5668 14 member Saturday 1074.6389

* Analyze ridership data by type and weekday

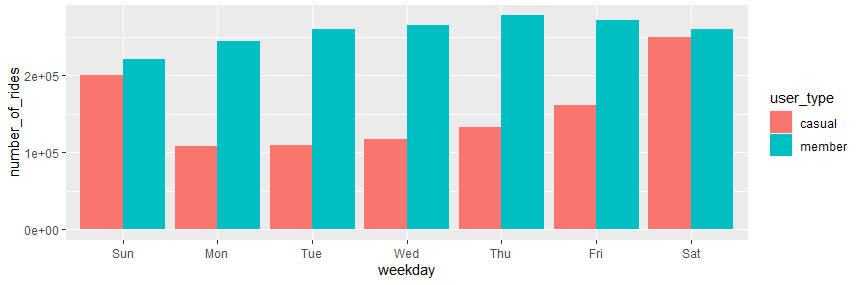
Divvy\_Trips\_2020\_v4 %>% mutate(weekday = wday(start\_time, label = TRUE)) %>% #creates weekday field using wday() group\_by(user\_type, weekday) %>% #groups by usertype and weekday summarise(number\_of\_rides = n() #calculates the number of rides and average duration ,average\_duration = mean(ride\_length)) %>% # calculates the average duration arrange(user\_type, weekday) # sorts

# A tibble: 14 x 4

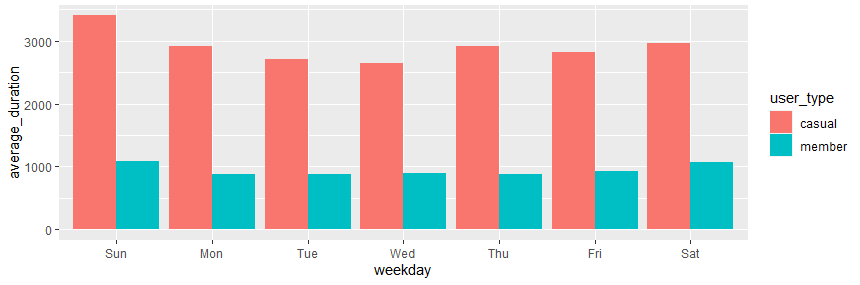
# Groups: user\_type [2]

user\_type weekday number\_of\_rides average\_duration 1 casual Sun 200472 3402. 2 casual Mon 108640 2911. 3 casual Tue 109578 2711. 4 casual Wed 117080 2645. 5 casual Thu 132863 2917. 6 casual Fri 161545 2823. 7 casual Sat 249364 2970. 8 member Sun 221344 1080. 9 member Mon 244361 884. 10 member Tue 260300 872. 11 member Wed 265641 889. 12 member Thu 277435 887. 13 member Fri 272147 933. 14 member Sat 259978 1075.

* Let’s visualize the number of rides by rider type Divvy\_Trips\_2020\_v4 %>% mutate(weekday = wday(start\_time, label = TRUE)) %>% group\_by(user\_type, weekday) %>% summarise(number\_of\_rides = n() ,average\_duration = mean(ride\_length)) %>% arrange(user\_type, weekday) %>% ggplot(aes(x = weekday, y = number\_of\_rides, fill = user\_type)) + geom\_col(position = “dodge”)



* Let’s create a visualization for average duration Divvy\_Trips\_2020\_v4 %>% mutate(weekday = wday(start\_time, label = TRUE)) %>% group\_by(user\_type, weekday) %>% summarise(number\_of\_rides = n(),average\_duration = mean(ride\_length)) %>% arrange(user\_type, weekday) %>% ggplot(aes(x = weekday, y = average\_duration, fill = user\_type)) +geom\_col(position = “dodge”)



counts <- aggregate(Divvy\_Trips\_2020\_v4 user\_type + Divvy\_Trips\_2020\_v4$day\_of\_week, FUN = mean)

counts

Divvy\_Trips\_2020\_v4day\_of\_week Divvy\_Trips\_2020\_v4$ride\_length 1 casual Sunday 3401.7184 2 member Sunday 1079.7364 3 casual Monday 2910.8165 4 member Monday 884.0107 5 casual Tuesday 2711.0042 6 member Tuesday 872.4097 7 casual Wednesday 2645.1722 8 member Wednesday 888.7004 9 casual Thursday 2917.0429 10 member Thursday 886.7974 11 casual Friday 2823.3506 12 member Friday 932.9811 13 casual Saturday 2969.5668 14 member Saturday 1074.6389

write.csv(Divvy\_Trips\_2020\_v4, ‘Divvy\_Trips\_2020\_final.csv’) ## Last save of file with all changes.