

Case study: How does a bike-share navigate speedy success?

Yasin MASLAK

2024-04-27

Scenario

You are a junior data analyst working on 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 nsights, 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 bebacked 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 the bikes 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 solid opportunity 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 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.

Question of Analysis

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?

**Now, I finish analyze and I share each step with Cyclistic marketing analytics team and Lily Moreno because of importance of project and I newly became part of this project I would like to share code if they would like to change my codes**

**Kütüphanelerin yüklenmesi/ Load Packages**

install.packages(“tidyverse”)

library(readr)

library(tidyverse)

library(dplyr)

library(lubridate)

library(skimr) l

ibrary(janitor)

library(ggplot2)

install.packages(“tidyverse”)

**Verisetlerini Okuma /Read Datasets**

# File name or full path of the

td2023\_1 <- read\_csv(“202301-divvy-tripdata.csv”)

td2023\_2 <- read\_csv(“202302-divvy-tripdata.csv”)

td2023\_3 <- read\_csv(“202303-divvy-tripdata.csv”)

td2023\_4 <- read\_csv(“202304-divvy-tripdata.csv”)

td2023\_5 <- read\_csv(“202305-divvy-tripdata.csv”)

td2023\_6 <- read\_csv(“202306-divvy-tripdata.csv”)

td2023\_7 <- read\_csv(“202307-divvy-tripdata.csv”)

td2023\_8 <- read\_csv(“202308-divvy-tripdata.csv”)

td2023\_9 <- read\_csv(“202309-divvy-tripdata.csv”)

td2023\_10 <- read\_csv(“202310-divvy-tripdata.csv”)

td2023\_11 <- read\_csv(“202311-divvy-tripdata.csv”)

td2023\_12 <- read\_csv(“202312-divvy-tripdata.csv”)

Understanding Dataset

The skim without charts function gives us a pretty comprehensive summary of a dataset.

Output of Calnames : “ride\_id” “rideable\_type” “started\_at” “ended\_at” “start\_station\_name” “start\_station\_id” “end\_station\_name”, “end\_station\_id” “start\_lat” “start\_lng” “end\_lat” “end\_lng” “member\_casual”

skim\_without\_charts(td2023\_1)

summary(td2023\_12)

glimpse(td2023\_10)

colnames(td2023\_8)

str(td2023\_5)

colnames(td2023\_7)

**After check all datasets and be sure all of same columns we can bind all of them**

data\_2023 <- rbind(td2023\_1,td2023\_2,td2023\_3,td2023\_4,td2023\_5, td2023\_6, td2023\_7, td2023\_8 ,td2023\_9,td2023\_10,td2023\_11,td2023\_12)

str(data\_2023) dim(data\_2023)

Rename of columnnames

(data\_2023 <- rename(data\_2023,

bikeid = ride\_id ,

biketype =rideable\_type,

start\_time= started\_at, end\_time = ended\_at,

from\_station\_name= start\_station\_name ,

from\_station\_id =start\_station\_id ,

to\_station\_name = end\_station\_name ,

to\_station\_id =end\_station\_id, usertype = member\_casual ))

head(data\_2023)

#It is always good to backup raw data before data cleaning. #Write as a csv file

write\_csv(data\_2023, “total\_data\_2023.csv”)

head(data\_2023,12) #First 12 row we can see

Data Cleaning

Preprocessing of the data is important before analysis, so null values have to be checked and removed.

str(data\_2023)

View(data\_2023)

glimpse(data\_2023)

dim(data\_2023)

sum(is.null(data\_2023)) #be sure that get rid of null datas.

#In this step we generate new colums that dates split to see under part of date when we analyze

data\_2023date<-as.Date(data\_2023start\_time)

data\_2023day<-format(as.Date(data\_2023date),“%d”)

data\_2023month<-format(as.Date(data\_2023date),“%m”)

data\_2023year<-format(as.Date(data\_2023date),“%Y”) #to get the year as separate column data\_2023day\_o f\_w eek<-format(as.Date(data\_2023date),“%A”)

colnames(data\_2023)

head(data\_2023)

View(data\_2023)

write\_csv(data\_2023, “extra\_columns\_total\_data\_2023.csv”)

After adding extra date columns, we save again. because we may want to continue the analysis with different tools. Remove lat, long, birthyear, and gender fields as this data was dropped beginning in 2020

data\_2023 <- data\_2023 %>% select(-c(start\_lat,start\_lng, end\_lat,end\_lng, from\_station\_id, from\_station\_name, to\_station\_name, to\_station\_id ))

colnames(data\_2023)

STEP 4: Clean up and prepare for data analysis

colnames(data\_2023) #List of column names

nrow(data\_2023) #How many rows are in data frame?

dim(data\_2023) #Dimensions of the data frame? head(data\_2023) #See the first 6 rows of data frame.

tail(div\_2019\_2)

str(data\_2023) #See list of columns and data types (numeric, character,etc)

summary(data\_2023) #Statistical summary of data. Mainly for numerics

Before Analyze Check Data Structure

to save original datas we analyze copy to datas a new file

trip\_datas <- data\_2023 # Our Dataset copy

colnames(trip\_datas)

table(trip\_datasusertype) #Tableofusertypetable(trip\_datasbiketype)

Table of biketype

bikeofusertype <- table(trip\_datasusertype,trip\_d atasbiketype ) #which user choose which bike

?write.table() #Check funck values

write.table(bikeofusertype, file = “Usertype\_ofbik\_chose.csv”, sep = “,”, quote = FALSE, row.names = TRUE ) #For other Analyzes save to tables

View(trip\_datas)

colSums(is.na(data\_2023))# calculate the trip time \*

?difftime()

data\_2023tripduration<-difftime(data\_2023end\_time, data\_2023$start\_time, units= “mins”)

summary(data\_2023$tripduration)

head(data\_2023)

unique(data\_2023$usertype ) # To know the unique values from ‘usertype’

unique(data\_2023$usertype ) # To know the unique values from ‘biketype’

Convert “tripduration” from Factor to numeric so we can run calculations on the data

is.factor(data\_2023tripduration)

\_2023tripduration <-as.numeric(as.character(data\_2023$tripduration))

is.factor(data\_2023tripduration)is.numeric(data\_2023tripduration)

Removed rows which had negative tripduration but I dont want to so under 10 minutes trips effect analyze

data\_2023 <- data\_2023 %>% filter(tripduration > 10)

|  |  |  |  |
| --- | --- | --- | --- |
| Output of Bike of user type table | | | |
|  | classic\_bike | docked\_bike | electric\_bike |
| casual | 876805 | 78287 | 1103529 |
| member | 1819026 | 0 | 1840961 |

*min\_trip\_duration <- aggregate (trip\_datas$tripduration tripdatas$usertype,FUN=min)*

*average\_trip\_duration <- aggregate (trip\_datas$tripduration tripdatas$usertype,FUN=mean)*

*median\_trip\_duration aggregate (trip\_datas$tripduration tripdatas$usertype,FUN=median)*

*max\_trip\_duration <- aggregate (trip\_datas$tripduration tripdatas$usertype,FUN=max)*

*#We prepare new df for analyze*

*# I take one more copy this data to use analyze*

*trip\_duration\_datas <- rbind(min\_trip\_duration, average\_trip\_duration,median\_trip\_duration,max\_trip\_duration )*

*write.table(trip\_duration\_datas, file = “trip\_duration\_datas.csv”, sep = “,”, quote = FALSE, row.names = TRUE )*

*summary(trip\_datas$tripduration)*

trip\_datasday\_of\_week,

levels=c(“Sunday”, “Monday”, “Tuesday”, “Wednesday”, “Thursday”, “Friday”, “Saturday”))

**STEP 5: CONDUCT DESCRIPTIVE ANALYSIS**

**Compare members and casual users**

**Sort days of the week # Notice that the days of the week are out of order. Let’s fix that.**

**#Total number of rides for the year 2023**

#Total number of rides for the year 2023

ggplot(data = trip\_datas)+

geom\_bar(mapping = aes(x=usertype, fill=usertype),show.legend = FALSE,width = 0.8)+ labs(y=“total”,title = “Total number of rides for the year 2023”)

ggsave(“Total number of rides for the year 2023.png”)

pie\_labels <- paste0(round(100\*type\_of\_users\_summarycount),2),“%”)

pie(type\_of\_users\_summary$count, labels = as.character(pie\_labels), lty=2, col = c(“lightblue”, “pink”), main = “% of ride by customer type”)

metin, ekran görüntüsü, dikdörtgen, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, diyagram, daire, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Comparing by each day of week for member vs casual**

datasof\_usertype\_days <- trip\_datas %>%

group\_by(usertype, day\_of\_week) %>%

summarise( average\_ride\_duration = mean(tripduration),

max = max(tripduration), min = min(tripduration)) %>%

arrange(usertype, day\_of\_week)

View(datasof\_usertype\_days)

write.table(datasof\_usertype\_days, file = “datasof\_usertype\_day.csv”, sep = “,”, quote = FALSE, row.names = TRUE )

ggplot(datasof\_usertype\_days, aes(x = day\_of\_week, y = average\_ride\_duration, fill = usertype, colour = usertype)) + geom\_bar(position = “dodge”, stat=“identity”)

datasof\_usertype\_days <- trip\_datas %>%

group\_by(usertype, day\_of\_week) %>%

summarise( average\_ride\_duration = mean(tripduration), max = max(tripduration), min = min(tripduration)) %>%

arrange(usertype, day\_of\_week)

View(datasof\_usertype\_days)

metin, ekran görüntüsü, yazı tipi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Total rides taken per month for each type of customers**

Select the required columns & reshape the data

Select the required columns & reshape the data

monthly\_ride\_count <- trip\_datas %>% group\_by(month,usertype) %>% summarise(count\_of\_ride =n())

Plot the line chart

ggplot(monthly\_ride\_count,aes(x=month,y=count\_of\_ride,group=usertype,xlim(0,400000)))+ geom\_point(aes(color=usertype),size=1.5)+

geom\_line(aes(color=usertype),size=1)+

labs(x=“Month”,y=“Count\_of\_rides”,title = “Total No. of rides per month”)

metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**3) Total rides taken each day in a week for each type of customers**

colnames(trip\_datas)

day\_wise\_ride\_count <- trip\_datas %>%

group\_by(day\_of\_week,usertype) %>%

summarise(count\_of\_ride =n()) %>%

pivot\_wider(names\_from = usertype,values\_from = count\_of\_ride) %>% mutate(total\_rides = casual+member) %>%

mutate(casual\_percentage= (casual/total\_rides)*100) %>% mutate(member\_percentage= (member/total\_rides)*100)

print(day\_wise\_ride\_count)

day\_wise\_ride\_count\_2 <- trip\_datas %>%

group\_by(day\_of\_week,usertype) %>%

summarise(count\_of\_ride =n())

ggplot(data = day\_wise\_ride\_count\_2)+ geom\_col(mapping=aes(x=day\_of\_week,y=count\_of\_ride,fill=usertype))+ labs(x=“day\_of\_week”,y=“Count\_of\_rides”,title = “Day wise total ride count”)

metin, ekran görüntüsü, diyagram, paralel içeren bir resim

Açıklama otomatik olarak oluşturuldu

**4) Types of bikes per type of customers**

bike\_type\_count <- trip\_datas %>%

group\_by(biketype) %>%

summarise(ride\_count=n()) %>%

mutate(ride\_count\_percentage = round(100\*ride\_count/sum(ride\_count),1))

print(bike\_type\_count)

daire, diyagram, grafik, renklilik içeren bir resim

Açıklama otomatik olarak oluşturuldu

bike\_type\_count\_2 <- trip\_datas %>%

group\_by(biketype,usertype) %>%

summarise(ride\_count=n()) %>%

pivot\_wider(names\_from = usertype,values\_from = ride\_count) %>% mutate(total\_ride\_count = sum(casual,member,na.rm=TRUE)) %>% mutate(casual\_percentage = round(100*(casual/total\_ride\_count),2)) %>% mutate(member\_percentage = round(100*(member/total\_ride\_count),2)) print(bike\_type\_count\_2)

|  |
| --- |
| biketype casual member total\_ride\_count casual\_percentage member\_percentage  *<chr>* *<int>* *<int>* *<int>* *<dbl>* *<dbl>*  1 classic\_bike 581313 809497 1390810 41.8 58.2  2 docked\_bike 68838 NA 68838 100 NA  3 electric\_bike 537362 728775 1266137 42.4 57.6 |
|  |
| |  | | --- | |  | |

bike\_type\_count\_3 <- trip\_datas %>%

group\_by(biketype,usertype) %>%

summarise(ride\_count=n())

ggplot(data = bike\_type\_count\_3)+ geom\_col(mapping=aes(x=biketype,y=ride\_count,fill=usertype))+ labs(x=“biketype”,y=“total\_rides”,title = “Total number of rides/rideable type”)

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

Share & Act

• I will skip the following steps Share and Act because, the analysis is a personal project.

• I will go directly to findings and recommendations.



**Findings and Recommendations**

**How do annual members and casual riders use Cyclistic bikes differently?\_\_**

\* For the year 2023, which is our study time, we have annual members are (%56 ,casual riders %43,75 rides

\* When we go down to a finer level, at a month, we observe the following between the casual & annual members:

+Casual user much more use than members.

+Daily usege are not different than each other.

+Members and Casual users have depict same customer behaviors

\* At july Casual and Members have equal number of ride other months members ride number rise than casuals but differences is not huge.

\* Docked bikes are only used by casual riders, it represents \_\_2,5%\_\_ of rides (electric 51% and classic 46%).

\* 34.3% of rides by classic bikes are casual riders and 65.7% are annual members.

\* 43.4% of rides by electric bikes are casual riders and 56.6% are annual members

**.**

**2) Why would casual riders buy Cyclistic annual memberships?\_\_**

\* We previously saw that casual riders ride more on weekends, if they have to ride at the same pace during the weekdays, it may motivate them to become annual members.

\* If their preference for docked bikes shift to classic or electric bikes, Cyclistic can hope of having an increase in annual members.

\* To understand more about the customer's choice of becoming annual members or casual riders and their ride time, the following information can help to do a finer analysis:

+ The reason behind each rides, example: home, work, leisure..

+ We have not have the cost details for rider type and bike type. But Members and Casuals nearlest same rate. We can say there is little difference of price. Some users not believe annual membership give advantages or they dont know. Especially social media marketing focus on mention about these advantages. Cyclistic make much more agreement or discount annual members to convince others.

**3) How can Cyclistic use digital media to influence casual riders to become members?**

Through influencer marketing, advertising and environmental awareness campaign on social media and TV,Cyclistic can work on the following:

\* The advantages of using more electric bikes (environmentally friendly) than docked bikes (from our sample data, we do not have annual members for docked bikes only for electric & classic bikes).

\* Encourage casual riders to ride throughout the week as they do during the weekend.

\* The advantages in becoming an annual member, example: it can be less costly when we compare the average price (in a year, month, week, day) for each ride as an annual member as compared to a casual rider.