Cyclistic Case Study Report

Jesslyn

12/13/2021

# Cyclistic Case Study

Cyclistic merupakan sebauh perusahaan fiktif yang menawaerkan peminjaman sepeda. Terdapat tiga *pricing plan* yang ditawarkan oleh Cyclistic: *single-ride passes*, *full-day passes*, dan *annual memberships*. Pelanggan yang membeli *single-ride* atau *full-day passes* disebut sebagai *casual riders*. Pelanggan yang membeli *annual membership* disebut *Cyclistic members*.

Seorang finance analyst dari Cyclistic menyimpulkan bahwa *annual members* lebih menguntungkan daripada *casual riders* walaupun fleksibilitas *pricing plan* membantu Cyclistic menarik banyak pelanggan. Oleh karena itu, jumlah *annual members* ingin dimaksimumkan.

Dibanding menarik pelanggan baru untuk menjadi *annual members*, finance analyst tersebut percaya bahwa terdapat kesempatan yang bagus untuk mengubah *casual riders* menjadi *annual members* . Oleh karena itu, tujuan dari case study ini adalah mencari cara agar dapat menarik *casual riders* ,emkado *annual members*.

## Ask

Terdapat tiga pertanyaan yang perlu dijawab dalam analisis ini agar dapat menemukan cara untuk menerik *casual riders* menjadi *annual members*:  
1. Apa perilaku yang membedakan *casual riders* dengan *annual members*?  
2. Menagapa *casual riders* mau membeli *annual membership*?  
3. Bagaimana cara Cyclistic menggunakan media digital agar dapat mempengaruhi *casual riders* untuk menjadi *annual members*?

## Prepare

[Dataset](https://divvy-tripdata.s3.amazonaws.com/index.html) yang digunakan dalam case study ini berasal dari Motivate International Inc. dengan [lisensi](https://www.divvybikes.com/data-license-agreement).

## Process

Pada bagian ini, dilakukan *data cleaning*. Berikut ini adalah tahap-tahap *data cleaning* dengan emggunakan R:

1. Mengimport setiap dataset ke RStudio;

data\_q1 <- read.csv("Divvy\_Trips\_2020\_Q1.csv")  
data\_04 <- read.csv("202004-divvy-tripdata.csv")  
data\_05 <- read.csv("202005-divvy-tripdata.csv")  
data\_06 <- read.csv("202006-divvy-tripdata.csv")  
data\_07 <- read.csv("202007-divvy-tripdata.csv")  
data\_08 <- read.csv("202008-divvy-tripdata.csv")  
data\_09 <- read.csv("202009-divvy-tripdata.csv")  
data\_10 <- read.csv("202010-divvy-tripdata.csv")  
data\_11 <- read.csv("202011-divvy-tripdata.csv")  
data\_12 <- read.csv("202012-divvy-tripdata.csv")

1. Menggabungkan setiap dataset ke dalam satu variabel;

trip\_data <- rbind(data\_q1, data\_04, data\_05, data\_06, data\_07, data\_08, data\_09, data\_10, data\_11, data\_12)

1. Menghilangkan setiap baris yang yang mengandung NA dan memasukkannya ke dalam variabel baru;

no\_null\_trip\_data <- na.omit(trip\_data)

1. Mengambil setiap baris dengan ended\_at > started\_at;

no\_null\_trip\_data <- subset(no\_null\_trip\_data, ended\_at > started\_at)

1. Menambah tiga kolom baru pada no\_null\_trip data, yaitu ride\_length (jarak antara ended\_at dan started\_at), day\_of\_week (hari pada started\_at), dan month (bulan pada started\_at).

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

no\_null\_trip\_data$ride\_length <- difftime(ymd\_hms(no\_null\_trip\_data$ended\_at),ymd\_hms(no\_null\_trip\_data$started\_at), units = "secs")  
no\_null\_trip\_data$day\_of\_week <- wday(as.Date(no\_null\_trip\_data$started\_at))  
no\_null\_trip\_data$month <- month(as.Date(no\_null\_trip\_data$started\_at))

## Analyze

Pada bagian ini, dilakukan analisis data. Berikut ini adalah tahap-tahao analisis dengan menggunakan R:

1. Memanggil dplyr dan lubridate yang akan digunakan sepanjang analisis ini;

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(lubridate)

1. Membuat fungsi modus yang akan digunakan dalam analisis ini:

mode = function(x) {  
 data.frame(sort(-table(x)))[1,1]  
}

1. Mencari hari apa yang paling banyak terdapat peminjaman sepeda dari *casual riders* dan *annual members*. Hal ini dilakukan dengan mencari modus. Keterangan untuk day\_of\_week: 1 adalah hari Minggu;

mode\_day\_of\_week <- no\_null\_trip\_data %>%   
 group\_by(member\_casual) %>%   
 summarise(mode = mode(day\_of\_week))  
print(mode\_day\_of\_week)

## # A tibble: 2 x 2  
## member\_casual mode   
## <chr> <fct>  
## 1 casual 7   
## 2 member 4

1. Mencari rata-rata jumlah pemakai sepeda per hari;

number\_of\_riders <- no\_null\_trip\_data %>%   
 group\_by(day\_of\_week) %>%   
 count(member\_casual)  
unique\_date <- unique(as.Date(no\_null\_trip\_data$started\_at))  
number\_of\_days <- data.frame(table(wday(unique\_date)))  
casual <- filter(number\_of\_riders, member\_casual=="casual")  
member<- filter(number\_of\_riders, member\_casual=="member")  
casual$mean\_riders <- casual$n / number\_of\_days$Freq  
member$mean\_riders <- member$n / number\_of\_days$Freq  
mean\_riders\_per\_day <- rbind(casual, member)  
print(mean\_riders\_per\_day)

## # A tibble: 14 x 4  
## # Groups: day\_of\_week [7]  
## day\_of\_week member\_casual n mean\_riders  
## <dbl> <chr> <int> <dbl>  
## 1 1 casual 245914 4729.  
## 2 2 casual 134652 2640.  
## 3 3 casual 130651 2562.  
## 4 4 casual 144813 2732.  
## 5 5 casual 154371 2913.  
## 6 6 casual 191934 3691.  
## 7 7 casual 300367 5776.  
## 8 1 member 260148 5003.  
## 9 2 member 281589 5521.  
## 10 3 member 304417 5969.  
## 11 4 member 318188 6004.  
## 12 5 member 316749 5976.  
## 13 6 member 313158 6022.  
## 14 7 member 298968 5749.

1. Mencari tipe sepeda apa yang paling banyak dipinjam oleh *casual riders* dan *annual members*. Hal ini juga dilakukan dengan mencari modus.

mode\_rideable\_type <- no\_null\_trip\_data %>%   
 group\_by(member\_casual) %>%   
 summarise(mode = mode(rideable\_type))  
print(mode\_rideable\_type)

## # A tibble: 2 x 2  
## member\_casual mode   
## <chr> <fct>   
## 1 casual docked\_bike  
## 2 member docked\_bike

1. Mencari tipe sepeda apa yang paling banyak dipinjam oleh *casual riders* dan *annual members* untuk setiap harinya. Hal ini untuk melihat apakah ada trend. Keterangan untuk day\_of\_week: 1 adalah hari Minggu;

mode\_rideable\_type\_day\_of\_week <- no\_null\_trip\_data %>%   
 group\_by(member\_casual, day\_of\_week) %>%   
 summarise(mode = mode(rideable\_type))

## `summarise()` has grouped output by 'member\_casual'. You can override using the `.groups` argument.

print(mode\_rideable\_type\_day\_of\_week)

## # A tibble: 14 x 3  
## # Groups: member\_casual [2]  
## member\_casual day\_of\_week mode   
## <chr> <dbl> <fct>   
## 1 casual 1 docked\_bike  
## 2 casual 2 docked\_bike  
## 3 casual 3 docked\_bike  
## 4 casual 4 docked\_bike  
## 5 casual 5 docked\_bike  
## 6 casual 6 docked\_bike  
## 7 casual 7 docked\_bike  
## 8 member 1 docked\_bike  
## 9 member 2 docked\_bike  
## 10 member 3 docked\_bike  
## 11 member 4 docked\_bike  
## 12 member 5 docked\_bike  
## 13 member 6 docked\_bike  
## 14 member 7 docked\_bike

1. Mencari rata-rata jumlah pemakai per hari berdasarkan tipe sepeda;

number\_of\_riders\_rideable\_type <- no\_null\_trip\_data %>%   
 group\_by(rideable\_type) %>%   
 count(member\_casual)  
mean\_riders\_rideable\_type <- number\_of\_riders\_rideable\_type  
mean\_riders\_rideable\_type$mean\_riders <- number\_of\_riders\_rideable\_type$n / 364  
print(mean\_riders\_rideable\_type)

## # A tibble: 6 x 4  
## # Groups: rideable\_type [3]  
## rideable\_type member\_casual n mean\_riders  
## <chr> <chr> <int> <dbl>  
## 1 classic\_bike casual 11288 31.0  
## 2 classic\_bike member 59248 163.   
## 3 docked\_bike casual 1140595 3134.   
## 4 docked\_bike member 1810763 4975.   
## 5 electric\_bike casual 150819 414.   
## 6 electric\_bike member 223206 613.

1. Mencari mean, min, dan max dari ride\_length dari *casual riders* dan *annual members*;

mean\_min\_max<- no\_null\_trip\_data %>%   
 group\_by(member\_casual) %>%   
 summarise(mean = mean(ride\_length),min = min(ride\_length), max =max(ride\_length))  
  
mean\_min\_max$mean <- seconds\_to\_period(mean\_min\_max$mean)  
mean\_min\_max$min <- seconds\_to\_period(mean\_min\_max$min)  
mean\_min\_max$max <- seconds\_to\_period(mean\_min\_max$max)  
print(mean\_min\_max)

## # A tibble: 2 x 4  
## member\_casual mean min max   
## <chr> <Period> <Period> <Period>   
## 1 casual 47M 49.7377097755252S 1S 108d 15H 30M 24S  
## 2 member 15M 37.2204520601538S 1S 65d 3H 13M 31S

1. Mencari mean, min, dan max dari ride\_length dari *casual riders* dan *annual members* untuk setiap harinya. Hal ini untuk melihat apakah ada trend. Keterangan untuk day\_of\_week: 1 adalah hari Minggu;

mean\_min\_max\_day\_of\_week<- no\_null\_trip\_data %>%   
 group\_by(member\_casual, day\_of\_week) %>%   
 summarise(mean = mean(ride\_length),min = min(ride\_length), max =max(ride\_length))

## `summarise()` has grouped output by 'member\_casual'. You can override using the `.groups` argument.

mean\_min\_max\_day\_of\_week$mean <- seconds\_to\_period(mean\_min\_max\_day\_of\_week$mean)  
mean\_min\_max\_day\_of\_week$min <- seconds\_to\_period(mean\_min\_max\_day\_of\_week$min)  
mean\_min\_max\_day\_of\_week$max <- seconds\_to\_period(mean\_min\_max\_day\_of\_week$max)  
print(mean\_min\_max\_day\_of\_week)

## # A tibble: 14 x 5  
## # Groups: member\_casual [2]  
## member\_casual day\_of\_week mean min max   
## <chr> <dbl> <Period> <Period> <Period>   
## 1 casual 1 54M 37.5833868750874S 1S 99d 22H 56M 45S   
## 2 casual 2 46M 45.9344829634911S 1S 32d 1H 24M 36S   
## 3 casual 3 43M 1.11767992590967S 1S 48d 6H 24M 51S   
## 4 casual 4 43M 1.63424554425364S 1S 51d 21H 3M 21S   
## 5 casual 5 46M 52.9658290741136S 1S 108d 15H 30M 24S  
## 6 casual 6 45M 48.2446361770185S 1S 81d 11H 23M 25S   
## 7 casual 7 48M 55.6852883306087S 1S 45d 14H 21M 42S   
## 8 member 1 18M 6.96443178498384S 1S 40d 18H 40M 2S   
## 9 member 2 14M 51.1490860793568S 1S 65d 3H 13M 31S   
## 10 member 3 14M 24.0778898681742S 1S 10d 11H 45M 4S   
## 11 member 4 14M 37.4394634618526S 1S 28d 15H 51M 0S   
## 12 member 5 14M 38.4655452740184S 1S 25d 2H 36M 7S   
## 13 member 6 15M 21.6452078503503S 1S 29d 19H 56M 14S   
## 14 member 7 17M 46.9771346766208S 1S 55d 0H 17M 31S

1. Mencari jumlah pemakai untuk setiap bulan.

number\_of\_riders\_per\_month <- no\_null\_trip\_data %>%   
 group\_by(month) %>%   
 count(member\_casual)  
print(number\_of\_riders\_per\_month)

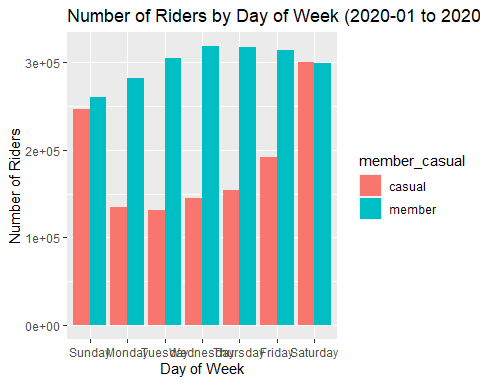
## # A tibble: 24 x 3  
## # Groups: month [12]  
## month member\_casual n  
## <dbl> <chr> <int>  
## 1 1 casual 7785  
## 2 1 member 136099  
## 3 2 casual 12860  
## 4 2 member 126715  
## 5 3 casual 27625  
## 6 3 member 115593  
## 7 4 casual 23566  
## 8 4 member 61054  
## 9 5 casual 86693  
## 10 5 member 113079  
## # ... with 14 more rows

## Share

Pada bagian ini, dilakukasi visualisasi data. Berikut ini adalah beberapa plot beserta penjelasannya:

1. Jumlah pemakai berdasarkan hari

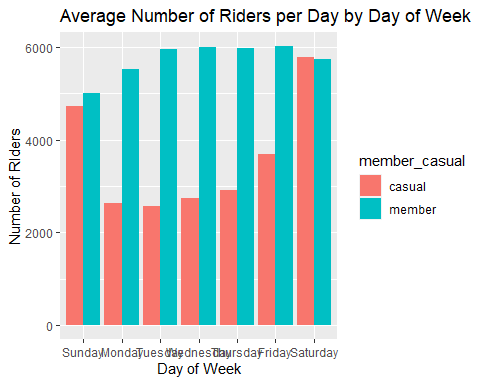
library(ggplot2)  
  
ggplot(data = number\_of\_riders, aes(x = factor(day\_of\_week), y = n, fill = member\_casual)) + geom\_bar(position ="dodge", stat="identity") +  
 scale\_x\_discrete(name = "Day of Week" ,labels = c("1"="Sunday", "2"="Monday", "3"="Tuesday", "4"="Wednesday", "5"="Thursday", "6"="Friday", "7"="Saturday")) +  
 ylab("Number of Riders") +   
 labs(title = "Number of Riders by Day of Week (2020-01 to 2020-12)")



Plot di atas memberikan beberapa informasi berikut:

1. Jumlah *casual riders* terbanyak pada hari *weekends* dalam peiode Januari 2020 hingga Desember 2020;
2. Jumlah *casual riders* jauh lebih rendah pada hari *weekdays* dalam periode Januari 2020 hingga Desember 2020;
3. Jumlah *annual members* terbanyak pada hari *weekdays* dalam periode Januari 2020 hingga Desember 2020;
4. Jumlah *annual members* pada hari *weekends* lebih rendah dibanding pada hari *weekdays*, tetapi perbedaannya tidak begitu jauh.
5. Rata-rata jumlah pemakai berdasarkan hari

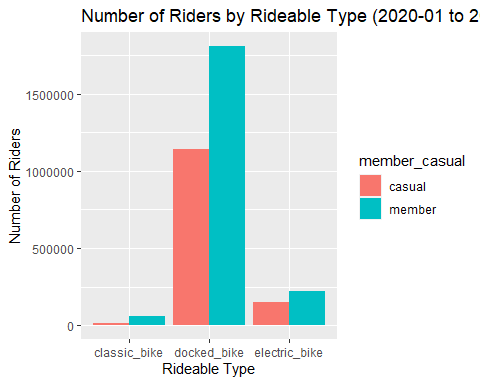
ggplot(data = mean\_riders\_per\_day, aes(x = factor(day\_of\_week), y = mean\_riders, fill = member\_casual)) + geom\_bar(position ="dodge", stat="identity") +  
 scale\_x\_discrete(name = "Day of Week" ,labels = c("1"="Sunday", "2"="Monday", "3"="Tuesday", "4"="Wednesday", "5"="Thursday", "6"="Friday", "7"="Saturday")) +  
 ylab("Number of RIders") +   
 labs(title = "Average Number of Riders per Day by Day of Week")



Plot di atas memberikan beberapa informasi berikut:

1. Rata-rata jumlah *casual riders* terbesar berada pada hari *weekends*;
2. Rata-rata jumlah *casual riders* jauh lebih rendah pada hari *weekdays* ;
3. Rata-rata jumlah *annual members* terbesar berada pada hari *weekdays*;
4. Rata-rata jumlah *annual members* pada hari *weekends* lebih rendah dibanding pada hari *weekdays*, tetapi perbedaannya tidak begitu jauh.
5. Jumlah pemakai berdasarkan tipe sepeda

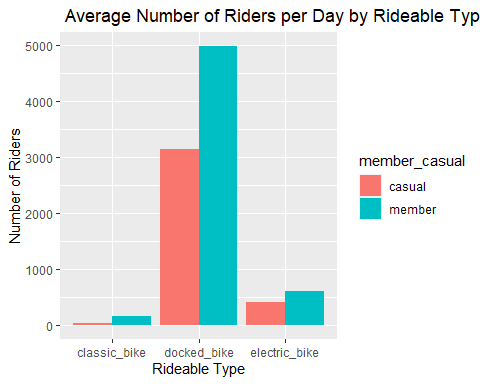
ggplot(data = number\_of\_riders\_rideable\_type, aes(x = rideable\_type, y = n, fill = member\_casual)) +   
 geom\_bar(position ="dodge", stat="identity") +  
 labs(title = "Number of Riders by Rideable Type (2020-01 to 2020-12)") +   
 xlab("Rideable Type") +   
 ylab("Number of Riders")



Plot di atas memberikan beberapa informasi berikut:

1. *Casual riders* paling banyak menggunakan *docked bike* dalam peiode Januari 2020 hingga Desember 2020;
2. *Annual members* paling banyak menggunakan *dockded bike*dalam peiode Januari 2020 hingga Desember 2020;
3. *Docked bike* merupakan tipe sepeda yang paling populer dalam peiode Januari 2020 hingga Desember 2020.
4. Rata-rata jumlah pemakai berdasarkan tipe sepeda

ggplot(data = mean\_riders\_rideable\_type, aes(x = rideable\_type, y = mean\_riders, fill = member\_casual)) +   
 geom\_bar(position ="dodge", stat="identity") +  
 labs(title = " Average Number of Riders per Day by Rideable Type") +   
 xlab("Rideable Type") +   
 ylab("Number of Riders")

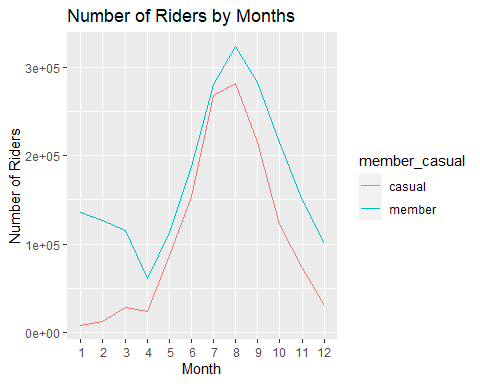


Plot di atas memberikan beberapa informasi berikut:

1. *casual riders* paling banyak menggunakan *docked bike*;
2. *Annual members* paling banyak menggunakan *dockded bike*;
3. *Docked bike* merupakan tipe sepeda yang paling populer.
4. Jumlah pemakai berdasarkan bulan

ggplot(data = number\_of\_riders\_per\_month, aes(x = as.numeric(month), y = n, color = member\_casual)) + geom\_line() + labs(title = "Number of Riders by Months") +  
 scale\_x\_discrete(name = "Month", limits = c(1,2,3,4,5,6,7,8,9,10,11,12)) +   
 ylab("Number of Riders")

## Warning: Continuous limits supplied to discrete scale.  
## Did you mean `limits = factor(...)` or `scale\_\*\_continuous()`?



Plot di atas memberikan beberapa informasi berikut:

1. Terjadi lonjakan jumlah *casual riders* pada liburan musim panas;
2. Terjadi lonjakan jumlah *annual members* pada libruan musim panas;
3. Selama 12 bulan, jumlah *annual memberss* selalu lebih banyak daripada jumlah *casual riders*.

## Act

Pada tahap ini, diberikan rekomendasi-rekomendasi berdasarkan hasil analisis. Untuk dapat memberikan rekomendasi, pertanyaan-pertanyaan pada tahap **Ask** perlu dijawab. Berikut ini adalah jawaban berdasarkan hasil analisis:

1. *Casual riders* cenderung melakukan peminjaman sepeda pada hari libur, sedangkan *annual members* cenderung melakukan peminjaman sepeda pada hari kerja. Selain itu, *casual riders* cenderung untuk bersepeda dalam waktu yang lama dibandingkan dengan *annual members*. Kemungkinan besar hal ini terjadi karena *casual riders* menggunakan sepeda untuk rekereasi, sedangkan *annual members* menggunakan sepeda untuk kegiatan sehari-hari.
2. Alasan yang mungkin untuk *casual\_riders* menjadi *annual members* adalah *annual membership* sesuai dengan kebutuhan mereka terkait penggunaan sepeda. Jika *casual riders* menggunakan sepeda untuk sehari-hari, maka *casual riders* sangat mungkin untuk membeli *annual membership*.
3. Cyclistic dapat menggunakan media digital untuk melakukan promosi.

Berdasarkan jawaban di atas, saya memberikan beberapa rekomendasi berikut:

1. Membuat *pricing plans* baru yang menargetkan pelanggan yang menggunakan sepeda di hari *weekends*. Hal itu bisa *annual membership* khusus untuk penggunaaan di hari *weekends* dan/atau *weekends passes*;
2. Membuat sistem *rewards* berdasarkan penggunaan sepeda sehingga dapat menarik pelanggan untuk tetap melakukan peminjaman sepeda. Hal ini untuk mendorong *casual riders* untuk membeli *annual membership*;
3. Melihat *casual riders* yang cenderung melakukan peminjaman sepeda pada hari libur, Cyclistic dapat mencoba untuk melakukan promosi melaui media digital pada musim liburan atau hari *weekends*