

# Case\_Study\_Google\_data\_analitik

Muhammad Jafar Shodiq

2022-10-27

## Bagaimana bisnis berbagi sepeda mencapai kesuksesan?

### Skenario

Anda adalah seorang analis data junior yang bekerja di tim analis pemasaran di Cyclistic, sebuah perusahaan berbagi sepeda di Chicago. Direktur pemasaran percaya bahwa kesuksesan perusahaan di masa depan tergantung pada upaya untuk memaksimalkan jumlah keanggotaan tahunan (Member). Oleh karena itu, tim Anda ingin memahami bagaimana pengendara biasa (Casual) dan Member menggunakan sepeda Cyclistic secara berbeda. Dari wawasan ini, tim Anda akan merancang strategi pemasaran baru untuk mengubah pengendara Casual menjadi pengendara Member. Tetapi pertama-tama, para eksekutif Cyclistic harus menyetujui rekomendasi Anda, sehingga rekomendasi tersebut harus didukung dengan wawasan data yang menarik dan visualisasi data profesional.

### Bertanya

#### Identifikasi tugas bisnis

Tim analitik pemasaran Cyclistic menyimpulkan bahwa pengendara Member lebih menguntungkan dari pada pengendara biasa, sehingga Lily Moreno sebagai Direktur Pemasaran memutuskan untuk mengubah pengendara biasa (Casual) menjadi pengendara tahunan (Member).

Analisis dari Tim analitik pemasaran Cyclistic akan membantu membuat strategi agar para pengguna casual beralih menjadi member.

#### Pemangku kepentingan utama

Pemangku kepentingan utama: Lily Moreno (Direktur pemasaran), Tim analitik pemasaran Cyclistic, dan Tim eksekutif Cyclistic.

### Mebersiapkan

#### Sumber data

Data yang digunakan adalah data perjalanan historis Cyclistic selama satu tahun yang disediakan oleh [Motivate International Inc.](#) Dataset tersebut adalah data publik yang dapat digunakan untuk mengeksplorasi bagaimana jenis pelanggan yang berbeda menggunakan sepeda Cyclistic.

#### Identifikasi pengaturan data

Data tersimpan dalam sebuah repositori. Penamaan file menyertakan tanggal, versi, dan tidak ada penamaan file yang berulang

## Kualitas data

Kualitas data dengan pendekatan ROCCC

1. Reliable

Data memiliki informasi yang akurat dan tidak bias.

2. Original

Data yang digunakan disediakan langsung oleh Motivate International Inc.

3. Comprehensive

Dataset mempunyai informasi yang dibutuhkan untuk menemukan solusi.

4. Current

Data yang digunakan mulai dari Oktober 2021 sampai September 2022.

5. Cite

Data yang digunakan adalah data public yang disediakan oleh Motivate International Inc.

## Proses

### Import library

```
library(tidyverse)

## — Attaching packages ————— tidyverse 1.
3.2 —
## ✓ ggplot2 3.3.6      ✓ purrr  0.3.5
## ✓ tibble  3.1.8      ✓ dplyr  1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.3      ✓ forcats 0.5.2
## — Conflicts ————— tidyverse_conflict
s() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()

library(ggplot2)
library(lubridate)

##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:base':
```

```
##
##      date, intersect, setdiff, union

library(scales)

##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##      discard
##
## The following object is masked from 'package:readr':
##
##      col_factor
```

## Import dataset

```
t2021_10 <- read.csv("dataset/202110-divvy-tripdata.csv")
t2021_11 <- read.csv("dataset/202111-divvy-tripdata.csv")
t2021_12 <- read.csv("dataset/202112-divvy-tripdata.csv")
t2022_01 <- read.csv("dataset/202201-divvy-tripdata.csv")
t2022_02 <- read.csv("dataset/202202-divvy-tripdata.csv")
t2022_03 <- read.csv("dataset/202203-divvy-tripdata.csv")
t2022_04 <- read.csv("dataset/202204-divvy-tripdata.csv")
t2022_05 <- read.csv("dataset/202205-divvy-tripdata.csv")
t2022_06 <- read.csv("dataset/202206-divvy-tripdata.csv")
t2022_07 <- read.csv("dataset/202207-divvy-tripdata.csv")
t2022_08 <- read.csv("dataset/202208-divvy-tripdata.csv")
t2022_09 <- read.csv("dataset/202209-divvy-publictripdata.csv")
```

## Menyatukan dataset

```
all_data <- bind_rows(t2021_10,t2021_11,t2021_12,t2022_01,t2022_02,t2022_03,t
2022_04,t2022_05,t2022_06,t2022_07,t2022_08,t2022_09)
```

## Menghapus kolom yang tidak diperlukan

```
all_data <- all_data %>%
  select(-c(start_lat, start_lng, end_lat, end_lng))
```

## Menganalisis

### Melihat 6 baris pertama

```
head(all_data)

##           ride_id rideable_type      started_at      ended_at
## 1 620BC6107255BF4C electric_bike 2021-10-22 12:46:42 2021-10-22 12:49:50
## 2 4471C70731AB2E45 electric_bike 2021-10-21 09:12:37 2021-10-21 09:14:14
## 3 26CA69D43D15EE14 electric_bike 2021-10-16 16:28:39 2021-10-16 16:36:26
## 4 362947F0437E1514 electric_bike 2021-10-16 16:17:48 2021-10-16 16:19:03
```

```
## 5 BB731DE2F2EC51C5 electric_bike 2021-10-20 23:17:54 2021-10-20 23:26:10
## 6 7176307BBC097313 electric_bike 2021-10-21 16:57:37 2021-10-21 17:11:58
##      start_station_name start_station_id end_station_name end_station_id
## 1 Kingsbury St & Kinzie St      KA1503000043
## 2
## 3
## 4
## 5
## 6
##      member_casual
## 1      member
## 2      member
## 3      member
## 4      member
## 5      member
## 6      member
```

### Melihat kategori pada kolom member\_casual

```
table(all_data$member_casual)

##
## casual member
## 2401286 3426949
```

kolom member\_casual hanya ada dua kategori yaitu casual dan member

### Menambahkan kolom tanggal, bulan, hari, dan tahun

```
all_data$date <- as.Date(all_data$started_at)
all_data$month <- format(as.Date(all_data$date), "%m")
all_data$day <- format(as.Date(all_data$date), "%d")
all_data$year <- format(as.Date(all_data$date), "%Y")
all_data$day_of_week <- format(as.Date(all_data$date), "%A")
```

### Membuat kolom ride\_length yang isinya panjang perjalanan dalam detik

```
all_data$ride_length <- as.numeric(difftime(all_data$ended_at, all_data$started_at))
```

### Mengecek apakah ada data negatif pada kolom ride\_length

```
nrow(all_data[all_data$ride_length < 0,])

## [1] 108
```

### Membuat tabel baru yang didalam kolom ride\_length tidak ada negatif

```
all_data_v2 <- all_data[!(all_data$ride_length < 0),]
```

### Analisis deskriptif kolom ride\_length

```
summary(all_data_v2$ride_length)/60
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##      0.00     5.93    10.48    19.60    18.85  40705.02
```

### Perbandingan pengguna Member dan pengendara Casual

```
aggregate(all_data_v2$ride_length ~ all_data_v2$member_casual, FUN = mean)
```

```
##  all_data_v2$member_casual all_data_v2$ride_length
## 1                        casual                1761.6450
## 2                        member                 766.1175
```

```
aggregate(all_data_v2$ride_length ~ all_data_v2$member_casual, FUN = median)
```

```
##  all_data_v2$member_casual all_data_v2$ride_length
## 1                        casual                   807
## 2                        member                   533
```

```
aggregate(all_data_v2$ride_length ~ all_data_v2$member_casual, FUN = max)
```

```
##  all_data_v2$member_casual all_data_v2$ride_length
## 1                        casual                2442301
## 2                        member                 93594
```

```
aggregate(all_data_v2$ride_length ~ all_data_v2$member_casual, FUN = min)
```

```
##  all_data_v2$member_casual all_data_v2$ride_length
## 1                        casual                   0
## 2                        member                   0
```

Dengan perhitungan diatas pengendara Casual mempunyai panjang perjalanan lebih lama dari pada pengendara Member.

### Urutkan berdasarkan hari

```
all_data_v2$day_of_week <- ordered(all_data_v2$day_of_week, levels=c("Sunday",
, "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
```

### Melihat rata-rata ride\_length perhari untuk pengguna Casual vs pengendara Member

```
aggregate(all_data_v2$ride_length/60 ~ all_data_v2$member_casual + all_data_v2$day_of_week, FUN = mean)
```

```
##  all_data_v2$member_casual all_data_v2$day_of_week all_data_v2$ride_length/60
## 1                        casual                Sunday                34.36456
## 2                        member                Sunday                14.21449
## 3                        casual                Monday                29.72834
## 4                        member                Monday                12.
```

32761			
## 5	casual	Tuesday	25.
80913			
## 6	member	Tuesday	12.
16304			
## 7	casual	Wednesday	25.
03412			
## 8	member	Wednesday	12.
12280			
## 9	casual	Thursday	25.
67967			
## 10	member	Thursday	12.
29406			
## 11	casual	Friday	28.
01101			
## 12	member	Friday	12.
52842			
## 13	casual	Saturday	32.
70886			
## 14	member	Saturday	14.
26368			

### Analisis pengendara berdasarkan tipe dan hari

```
all_data %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length/60)) %>%
  arrange(member_casual, weekday)
```

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

```
## # A tibble: 14 × 4
## # Groups:   member_casual [2]
##   member_casual weekday number_of_rides average_duration
##   <chr>          <ord>          <int>          <dbl>
## 1 casual        Sun             405040          34.4
## 2 casual        Mon             279789          29.7
## 3 casual        Tue             275787          25.8
## 4 casual        Wed             281660          25.0
## 5 casual        Thu             306694          25.7
## 6 casual        Fri             352510          28.0
## 7 casual        Sat             499806          32.7
## 8 member        Sun             393624          14.2
## 9 member        Mon             473060          12.3
## 10 member       Tue             541527          12.2
## 11 member       Wed             538490          12.1
## 12 member       Thu             530552          12.3
```

## 13 member	Fri	491467	12.5
## 14 member	Sat	458229	14.3

## Analisis

- Pengendara Casual mempunyai perjalanan terlama pada hari minggu diikuti hari sabtu.
- pengendara Member mempunyai perjalanan terlama pada hari sabtu diikuti hari minggu.
- Jumlah Pengendara Casual selalu bertambah mulai dari hari selasa dan puncaknya pada hari sabtu, setelah itu menurun pada hari minggu dan senin
- Jumlah pengendara Member berangsur menurun mulai dari hari selasa sampai hari minggu dan mengalami kenaikan pada hari senin.

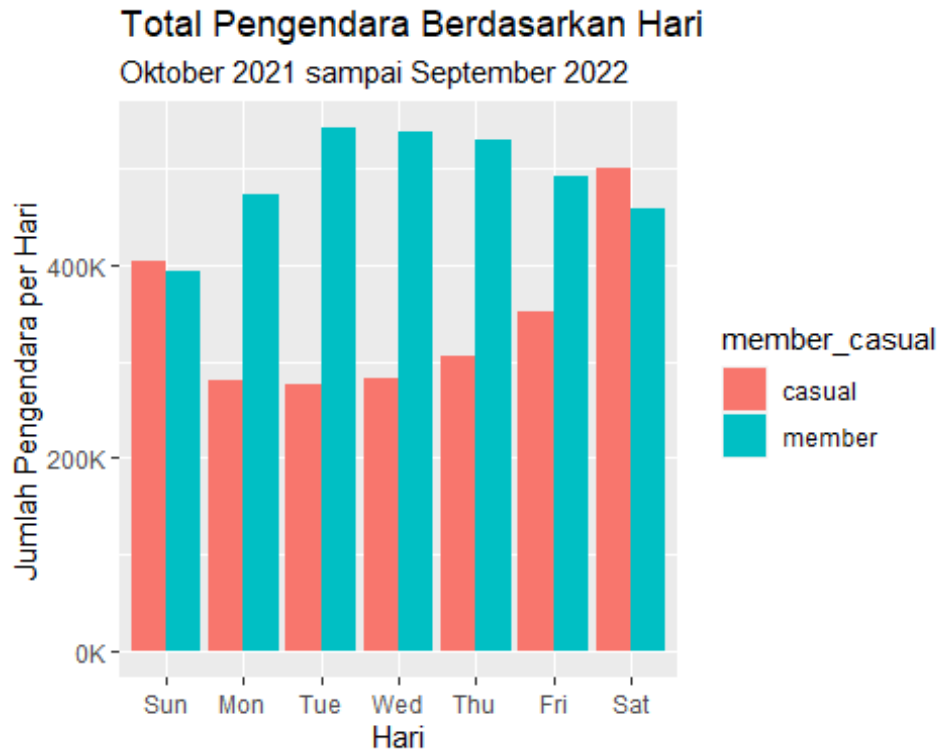
## Membagikan

### visualisasi jumlah pengendara per hari berdasarkan tipe pengendara

```
jumlah_pengendara <- all_data_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>%
  arrange(member_casual, weekday) %>%
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(x = "Hari", y = "Jumlah Pengendara per Hari",
       title = "Total Pengendara Berdasarkan Hari",
       subtitle = "Oktober 2021 sampai September 2022")

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

jumlah_pengendara + scale_y_continuous(labels = label_number(suffix = "K", scale = 1e-3))
```



Rata-rata pengendara Member lebih banyak memakai sepeda pada hari kerja, sedangkan Casual menggunakan sepeda pada hari akhir pekan.

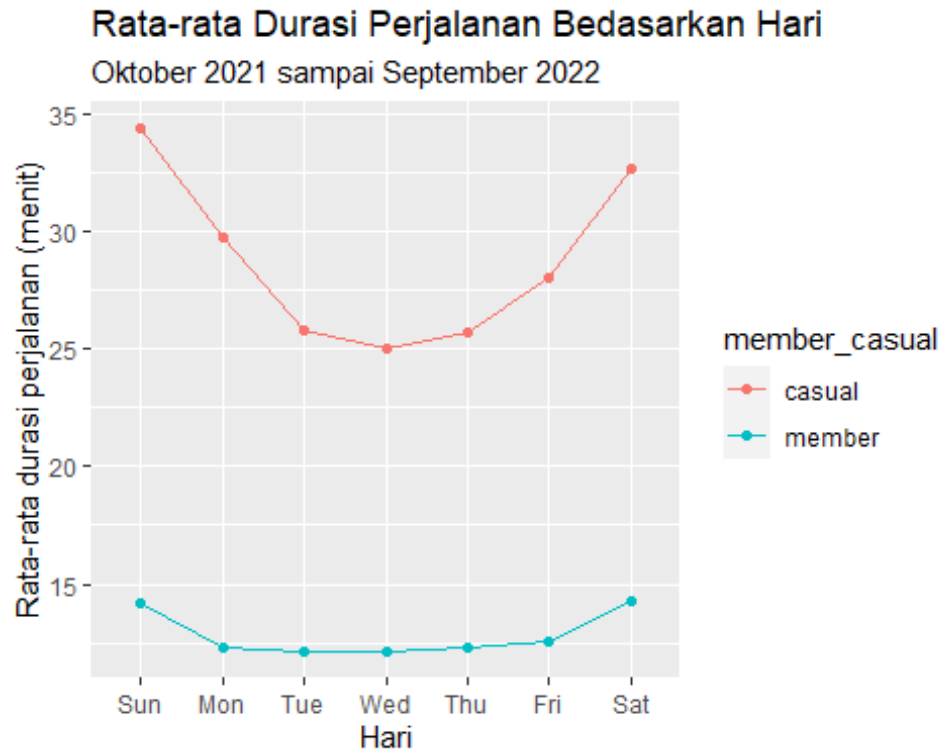
### visualisasi rata-rata durasi perjalanan berdasarkan hari dan tipe pengendara

```
durasi_perjalanan <- all_data_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>%
  group_by(member_casual, weekday) %>%
  summarise(number_of_rides = n(), average_duration = mean(ride_length)/60) %>%
  arrange(member_casual, weekday) %>%
  ggplot(aes(x = weekday, y = average_duration, group = member_casual)) +
  geom_line(aes(color = member_casual)) +
  geom_point(aes(color = member_casual)) +
  labs(x = "Hari", y = "Rata-rata durasi perjalanan (menit)",
       title = "Rata-rata Durasi Perjalanan Berdasarkan Hari",
       subtitle = "Oktober 2021 sampai September 2022")

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

durasi_perjalanan
```





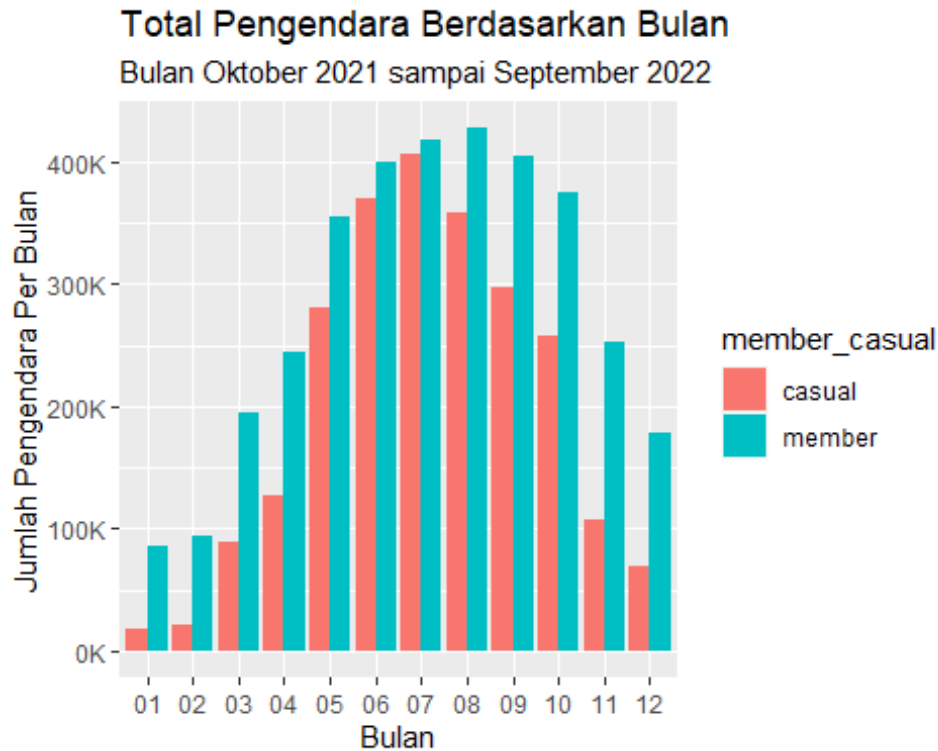
Rata-rata durasi perjalanan pengendara biasa yang lebih tinggi dari pada pengendara Member.

### visualisasi jumlah pengendara per bulan berdasarkan tipe pengendara

```
jumlah_pengendara_bulan <- all_data_v2 %>%
  group_by(member_casual, month) %>%
  summarise(number_of_rides = n()) %>%
  arrange(member_casual, month) %>%
  ggplot(aes(x = month, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(x = "Bulan", y = "Jumlah Pengendara Per Bulan",
       title = "Total Pengendara Berdasarkan Bulan",
       subtitle = "Bulan Oktober 2021 sampai September 2022")

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

jumlah_pengendara_bulan + scale_y_continuous(labels = label_number(suffix = "K", scale = 1e-3))
```



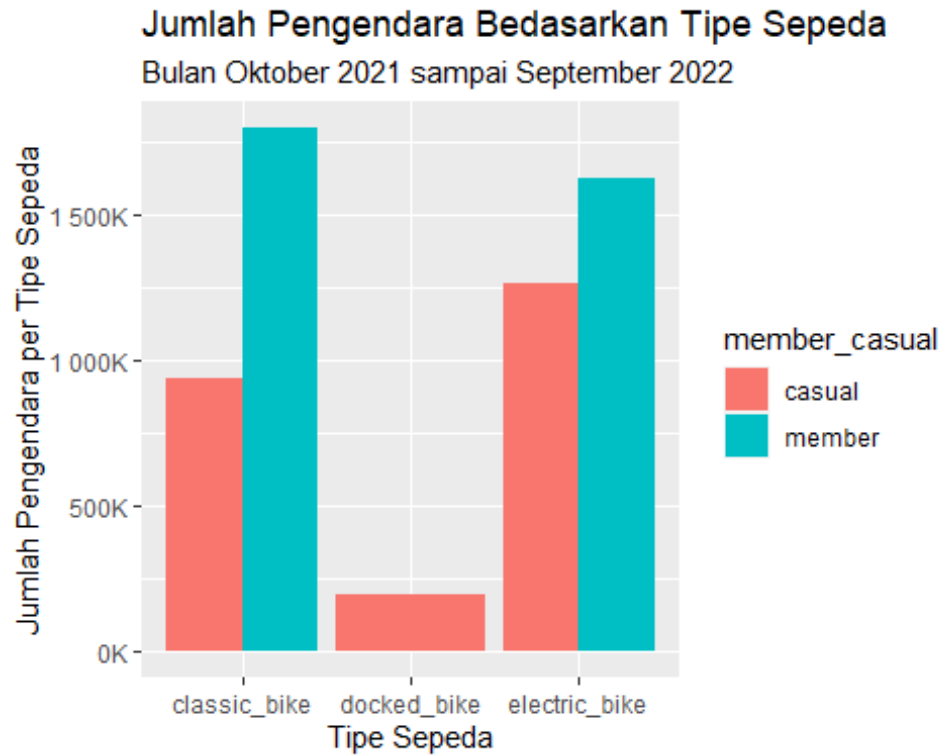
Para pesepeda kebanyakan berkendara pada bulan Mei sampai Oktober dan pada saat itu Kota Chicago tidak mengalami musim dingin.

### Visualisasi Jumlah tipe sepeda yang dipakai berdasarkan pengendara biasa dan pengendara member

```
tipe_sepeda <- all_data_v2 %>%
  group_by(member_casual, rideable_type) %>%
  summarise(number_of_rides = n()) %>%
  arrange(member_casual, rideable_type) %>%
  ggplot(aes(x = rideable_type, y = number_of_rides, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(x = "Tipe Sepeda", y = "Jumlah Pengendara per Tipe Sepeda",
       title = "Jumlah Pengendara Berdasarkan Tipe Sepeda",
       subtitle = "Bulan Oktober 2021 sampai September 2022")

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

tipe_sepeda + scale_y_continuous(labels = label_number(suffix = "K", scale = 1e-3))
```



Sepeda dengan tipe `docked_bike` mempunyai jumlah pengendara paling sedikit dibanding tipe lainnya. Tipe sepeda `docked_bike` juga hanya digunakan oleh pengendara Casual.

#### Bertindak

#### Berikut beberapa rekomendasi berdasarkan analisis

1. Membuat program Member bulanan untuk menarik pengendara biasa yang bersepeda sepanjang tahun.
2. Melakukan promosi pada musim semi sampai musim panas, dimana pada saat itu orang-orang banyak menggunakan sepeda.
3. Membuat kampanye yang menargetkan pengendara Casual untuk berangkat kerja dengan memakai sepeda.