study\_case\_cyclistic

Annisa Farah

2022-11-08

# STRATEGI PEMASARAN CYCLISTIC

# Mengubah Pengendara Biasa Menjadi Anggota Tahunan

## Overview

Report ini merupakan study case Google Data Analytics Professional Certificate tentang merancang strategi pemasaran baru untuk perusahaan Cyclistic

**Skenario**

Saya seorang analis data junior dan bekerja untuk sebuah perusahaan fiksi, Cyclistic. Direktur pemasaran ingin memaksimalkan jumlah keanggotaan tahunan karena dinilai lebih menguntungkan. Saya ditugaskan untuk merancang strategi pemasaran baru agar dapat memahami cara pelanggan menggunakan sepeda Cyclistic secara berbeda. Tujuannya untuk membuat kampanye yang mentargetkan member **casual** (pengendara biasa) bisa menjadi **member** (anggota tahunan) secara efisien.

## Step 1 - Ask

### 1.1. Latar belakang masalah

Program berbagi sepeda yang menampilkan lebih dari 5.800 sepeda dan 600 stasiun dok. Cyclistic membedakan diri dari perusahaan lain dengan menawarkan sepeda reclining, hand tricycles, dan sepeda kargo, menjadikan bike-share lebih inklusif bagi penyandang disabilitas dan pengendara yang tidak dapat menggunakan sepeda roda dua standar. Mayoritas pengendara memilih sepeda tradisional dan sekitar 8% pengendara menggunakan opsi bantuan. Pengguna sepeda lebih cenderung bersepeda untuk bersantai, tetapi sekitar 30% menggunakannya untuk bepergian ke tempat kerja setiap hari.

Sampai saat ini, strategi pemasaran Cyclistic mengandalkan upaya untuk membangun kesadaran umum dan menarik segmen konsumen yang luas. Salah satu pendekatan yang membantu mewujudkan hal ini adalah fleksibilitas harga: tiket sekali jalan, tiket sehari penuh, dan keanggotaan tahunan. Pelanggan yang membeli tiket sekali jalan atau tiket sehari penuh disebut sebagai pengendara **casual**. Pelanggan yang membeli keanggotaan tahunan adalah **member** Cyclistic.

Moreno yang meruupakan direktur pemasaran dan manajer saya, telah menetapkan tujuan yang jelas: *Merancang strategi pemasaran yang bertujuan untuk mengubah pengendara biasa menjadi anggota tahunan.* Namun, untuk melakukannya tim analis pemasaran perlu lebih memahami perbedaan antara anggota tahunan dan pengendara biasa, mengapa pengendara akan bersedia membeli keanggotaan, dan bagaimana media digital dapat memengaruhi taktik pemasaran mereka. Moreno dan timnya tertarik untuk menganalisis data perjalanan sepeda Cyclistic untuk mengidentifikasi tren.

### 1.2. Pertanyaan

1. Bagaimana pengendara member dan pengendara casual menggunakan sepeda secara berbeda?
2. Mengapa pengendara casual membeli keanggotaan tahunan Cyclistic?
3. Bagaimana Cyclistic menggunakan media digital untuk mempengaruhi pengendara casual untuk menjadi member?

## Step 2 - Prepare Data

1. Mendownload dataset [Cyclistic](https://divvy-tripdata.s3.amazonaws.com/index.html) sepanjang tahun 2021,
2. Melakukan ekstraksi file,
3. Pengecekan standarisasi data berdasarkan ROCCC (Reliable, Original, Comprehensive, Current, and Credible), dan

Tabel 1.1 Standarisasi Data Berdasarkan ROCC

| ROCCC | Deskripsi |
| --- | --- |
| *Reliable* | Validitas: Kolom *start\_station\_id* dan *end\_station\_id* memiliki format yang kurang konsisten  Kelengkapan: Terdapat nilai yang hilang (*null values*)  Duplikat: Semua nilai pada kolom *ride\_id* memiliki nilai yang unik |
| *Original* | Sumber Data dari pihak pertama |
| *Comprehensive* | Data cukup lengkap dan sebagian besar bebas dari kesalahan manusia |
| *Current* | Data terkini dan masih berkaitan dengan permasalahan |
| *Credible* | Data diperoleh dari pihak pertama secara langsung sehingga data memiliki kredibilitas yang cukup baik |

1. Import dataset ke dalam environment R.

## Step 3 - Process

### 3.1. Import library yang dibutuhkan

## Install required packages  
library(tidyverse) ## helps wrangle data

## Warning: package 'tidyverse' was built under R version 4.2.2

## ── 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\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(lubridate) ## helps wrangle date attributes

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

library(ggplot2) ## helps visualize data  
library(rmarkdown)

## Warning: package 'rmarkdown' was built under R version 4.2.2

library(skimr) ## helps summary dataframe structure and others atribute

## Warning: package 'skimr' was built under R version 4.2.2

library(geosphere) ## helps compute distances

## Warning: package 'geosphere' was built under R version 4.2.2

library(gridExtra) ## to arrange multiple grid-based plots on a page

## Warning: package 'gridExtra' was built under R version 4.2.2

##   
## Attaching package: 'gridExtra'  
##   
## The following object is masked from 'package:dplyr':  
##   
## combine

### 3.2. Import data set

## Import data set ##  
X202101\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202101-divvy-tripdata.csv")

## Rows: 96834 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202102\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202102-divvy-tripdata.csv")

## Rows: 49622 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202103\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202103-divvy-tripdata.csv")

## Rows: 228496 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202104\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202104-divvy-tripdata.csv")

## Rows: 337230 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202105\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202105-divvy-tripdata.csv")

## Rows: 531633 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202106\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202106-divvy-tripdata.csv")

## Rows: 729595 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202107\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202107-divvy-tripdata.csv")

## Rows: 822410 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202108\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202108-divvy-tripdata.csv")

## Rows: 804352 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202109\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202109-divvy-tripdata.csv")

## Rows: 756147 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202110\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202110-divvy-tripdata.csv")

## Rows: 631226 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202111\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202111-divvy-tripdata.csv")

## Rows: 359978 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

X202112\_divvy\_tripdata <- read\_csv("D:/Data Analitik/Course 8/Project\_course\_8/tripdata\_2021/202112-divvy-tripdata.csv")

## Rows: 247540 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ride\_id, rideable\_type, start\_station\_name, start\_station\_id, end\_...  
## dbl (4): start\_lat, start\_lng, end\_lat, end\_lng  
## dttm (2): started\_at, ended\_at  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

### 3.3. Memeriksa kelengkapan dataframe

Penting untuk melihat struktur dan ringkasan dari setiap data sebelum memprosesnya. Seluruh nama kolom dan type data harus konsisten. Kolom *ride\_id* semua nilainya harus unik. *rideable\_type* hanya ada 3 jenis yaitu: *electric\_bike, classic\_bike,* dan *docked\_bike*. Hanya ada 2 tipe *member\_casual* yaitu: *member* dan *casual*

## Resume all data  
skim\_without\_charts(X202101\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202101\_divvy\_tripdata |
| Number of rows | 96834 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 96834 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 8625 | 0.91 | 10 | 51 | 0 | 640 | 0 |
| start\_station\_id | 8625 | 0.91 | 3 | 35 | 0 | 638 | 0 |
| end\_station\_name | 10277 | 0.89 | 10 | 53 | 0 | 632 | 0 |
| end\_station\_id | 10277 | 0.89 | 3 | 35 | 0 | 629 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.05 | 41.64 | 41.88 | 41.90 | 41.93 | 42.06 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.78 | -87.66 | -87.64 | -87.63 | -87.53 |
| end\_lat | 103 | 1 | 41.90 | 0.05 | 41.64 | 41.88 | 41.90 | 41.93 | 42.07 |
| end\_lng | 103 | 1 | -87.65 | 0.03 | -87.81 | -87.66 | -87.64 | -87.63 | -87.51 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-01-01 00:02:05 | 2021-01-31 23:57:00 | 2021-01-15 06:05:04 | 93736 |
| ended\_at | 0 | 1 | 2021-01-01 00:08:39 | 2021-02-01 15:33:15 | 2021-01-15 06:19:58 | 93582 |

skim\_without\_charts(X202102\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202102\_divvy\_tripdata |
| Number of rows | 49622 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 49622 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 4046 | 0.92 | 10 | 51 | 0 | 582 | 0 |
| start\_station\_id | 4046 | 0.92 | 3 | 35 | 0 | 582 | 0 |
| end\_station\_name | 5358 | 0.89 | 10 | 53 | 0 | 584 | 0 |
| end\_station\_id | 5358 | 0.89 | 3 | 35 | 0 | 584 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.04 | 41.65 | 41.88 | 41.90 | 41.93 | 42.06 |
| start\_lng | 0 | 1 | -87.64 | 0.03 | -87.77 | -87.66 | -87.64 | -87.63 | -87.53 |
| end\_lat | 214 | 1 | 41.90 | 0.04 | 41.54 | 41.88 | 41.90 | 41.93 | 42.07 |
| end\_lng | 214 | 1 | -87.64 | 0.03 | -87.77 | -87.66 | -87.64 | -87.63 | -87.53 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-02-01 00:55:44 | 2021-02-28 23:59:41 | 2021-02-22 13:17:53 | 48139 |
| ended\_at | 0 | 1 | 2021-02-01 01:22:48 | 2021-03-05 15:11:45 | 2021-02-22 13:39:20 | 48035 |

skim\_without\_charts(X202103\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202103\_divvy\_tripdata |
| Number of rows | 228496 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 228496 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 14848 | 0.94 | 10 | 53 | 0 | 673 | 0 |
| start\_station\_id | 14848 | 0.94 | 3 | 35 | 0 | 673 | 0 |
| end\_station\_name | 16727 | 0.93 | 10 | 53 | 0 | 673 | 0 |
| end\_station\_id | 16727 | 0.93 | 3 | 35 | 0 | 673 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.04 | 41.65 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.64 | 0.03 | -87.78 | -87.66 | -87.64 | -87.63 | -87.53 |
| end\_lat | 167 | 1 | 41.90 | 0.04 | 41.64 | 41.88 | 41.90 | 41.93 | 42.08 |
| end\_lng | 167 | 1 | -87.65 | 0.03 | -88.07 | -87.66 | -87.64 | -87.63 | -87.53 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-03-01 00:01:09 | 2021-03-31 23:59:08 | 2021-03-19 17:37:20 | 209025 |
| ended\_at | 0 | 1 | 2021-03-01 00:06:28 | 2021-04-06 11:00:11 | 2021-03-19 17:55:05 | 208629 |

skim\_without\_charts(X202104\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202104\_divvy\_tripdata |
| Number of rows | 337230 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 337230 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 26056 | 0.92 | 10 | 53 | 0 | 681 | 0 |
| start\_station\_id | 26056 | 0.92 | 3 | 35 | 0 | 681 | 0 |
| end\_station\_name | 28174 | 0.92 | 10 | 53 | 0 | 681 | 0 |
| end\_station\_id | 28174 | 0.92 | 3 | 35 | 0 | 681 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.05 | 41.64 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.64 | 0.03 | -87.78 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 267 | 1 | 41.90 | 0.05 | 41.59 | 41.88 | 41.90 | 41.93 | 42.15 |
| end\_lng | 267 | 1 | -87.65 | 0.03 | -87.85 | -87.66 | -87.64 | -87.63 | -87.52 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-04-01 00:03:18 | 2021-04-30 23:59:53 | 2021-04-15 22:37:04 | 298722 |
| ended\_at | 0 | 1 | 2021-04-01 00:14:29 | 2021-05-05 22:14:39 | 2021-04-15 23:00:10 | 298625 |

skim\_without\_charts(X202105\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202105\_divvy\_tripdata |
| Number of rows | 531633 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 531633 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 53744 | 0.90 | 10 | 53 | 0 | 687 | 0 |
| start\_station\_id | 53744 | 0.90 | 3 | 35 | 0 | 686 | 0 |
| end\_station\_name | 58194 | 0.89 | 10 | 53 | 0 | 683 | 0 |
| end\_station\_id | 58194 | 0.89 | 3 | 35 | 0 | 682 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.05 | 41.65 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.64 | 0.03 | -87.78 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 452 | 1 | 41.90 | 0.05 | 41.56 | 41.88 | 41.90 | 41.93 | 42.09 |
| end\_lng | 452 | 1 | -87.64 | 0.03 | -87.85 | -87.66 | -87.64 | -87.63 | -87.52 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-05-01 00:00:11 | 2021-05-31 23:59:16 | 2021-05-19 07:44:31 | 447224 |
| ended\_at | 0 | 1 | 2021-05-01 00:03:26 | 2021-06-10 22:17:11 | 2021-05-19 07:59:43 | 447217 |

skim\_without\_charts(X202106\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202106\_divvy\_tripdata |
| Number of rows | 729595 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 729595 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 80093 | 0.89 | 10 | 53 | 0 | 689 | 0 |
| start\_station\_id | 80093 | 0.89 | 3 | 35 | 0 | 689 | 0 |
| end\_station\_name | 86387 | 0.88 | 10 | 53 | 0 | 690 | 0 |
| end\_station\_id | 86387 | 0.88 | 3 | 35 | 0 | 690 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.04 | 41.64 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.64 | 0.03 | -87.78 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 717 | 1 | 41.90 | 0.04 | 41.51 | 41.88 | 41.90 | 41.93 | 42.08 |
| end\_lng | 717 | 1 | -87.64 | 0.03 | -87.86 | -87.66 | -87.64 | -87.63 | -87.49 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-06-01 00:00:38 | 2021-06-30 23:59:59 | 2021-06-14 19:46:47 | 589805 |
| ended\_at | 0 | 1 | 2021-06-01 00:06:22 | 2021-07-13 22:51:35 | 2021-06-14 20:13:55 | 589069 |

skim\_without\_charts(X202107\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202107\_divvy\_tripdata |
| Number of rows | 822410 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 822410 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 87263 | 0.89 | 10 | 53 | 0 | 717 | 0 |
| start\_station\_id | 87262 | 0.89 | 3 | 36 | 0 | 710 | 0 |
| end\_station\_name | 93158 | 0.89 | 10 | 53 | 0 | 714 | 0 |
| end\_station\_id | 93158 | 0.89 | 3 | 36 | 0 | 707 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.04 | 41.65 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.84 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 731 | 1 | 41.90 | 0.04 | 41.63 | 41.88 | 41.90 | 41.93 | 42.15 |
| end\_lng | 731 | 1 | -87.65 | 0.03 | -87.85 | -87.66 | -87.64 | -87.63 | -87.49 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-07-01 00:00:22 | 2021-07-31 23:59:58 | 2021-07-17 13:58:37 | 659640 |
| ended\_at | 0 | 1 | 2021-07-01 00:04:51 | 2021-08-12 17:45:41 | 2021-07-17 14:28:04 | 658663 |

skim\_without\_charts(X202108\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202108\_divvy\_tripdata |
| Number of rows | 804352 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 804352 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 88458 | 0.89 | 3 | 53 | 0 | 727 | 0 |
| start\_station\_id | 88458 | 0.89 | 3 | 35 | 0 | 726 | 0 |
| end\_station\_name | 94115 | 0.88 | 10 | 53 | 0 | 727 | 0 |
| end\_station\_id | 94115 | 0.88 | 3 | 35 | 0 | 727 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.04 | 41.65 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.84 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 706 | 1 | 41.90 | 0.04 | 41.58 | 41.88 | 41.90 | 41.93 | 42.15 |
| end\_lng | 706 | 1 | -87.65 | 0.03 | -87.85 | -87.66 | -87.64 | -87.63 | -87.51 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-08-01 00:00:04 | 2021-08-31 23:59:35 | 2021-08-16 07:57:11 | 646516 |
| ended\_at | 0 | 1 | 2021-08-01 00:03:11 | 2021-09-01 17:37:35 | 2021-08-16 08:12:14 | 645299 |

skim\_without\_charts(X202109\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202109\_divvy\_tripdata |
| Number of rows | 756147 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 756147 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 93113 | 0.88 | 10 | 53 | 0 | 758 | 0 |
| start\_station\_id | 93111 | 0.88 | 3 | 35 | 0 | 758 | 0 |
| end\_station\_name | 99261 | 0.87 | 10 | 53 | 0 | 756 | 0 |
| end\_station\_id | 99261 | 0.87 | 3 | 35 | 0 | 756 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.05 | 41.65 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.84 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 595 | 1 | 41.90 | 0.05 | 41.57 | 41.88 | 41.90 | 41.93 | 42.17 |
| end\_lng | 595 | 1 | -87.65 | 0.03 | -87.87 | -87.66 | -87.64 | -87.63 | -87.50 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-09-01 00:00:06 | 2021-09-30 23:59:48 | 2021-09-15 16:43:37 | 611240 |
| ended\_at | 0 | 1 | 2021-09-01 00:00:41 | 2021-10-01 22:55:35 | 2021-09-15 17:01:16 | 610277 |

skim\_without\_charts(X202110\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202110\_divvy\_tripdata |
| Number of rows | 631226 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 631226 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 108210 | 0.83 | 10 | 53 | 0 | 793 | 0 |
| start\_station\_id | 108210 | 0.83 | 3 | 35 | 0 | 793 | 0 |
| end\_station\_name | 114834 | 0.82 | 10 | 53 | 0 | 790 | 0 |
| end\_station\_id | 114834 | 0.82 | 3 | 35 | 0 | 790 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.05 | 41.65 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.83 | -87.66 | -87.64 | -87.63 | -87.52 |
| end\_lat | 484 | 1 | 41.90 | 0.05 | 41.60 | 41.88 | 41.90 | 41.93 | 42.13 |
| end\_lng | 484 | 1 | -87.65 | 0.03 | -87.96 | -87.66 | -87.64 | -87.63 | -87.52 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-10-01 00:00:09 | 2021-10-31 23:59:49 | 2021-10-15 05:31:57 | 524629 |
| ended\_at | 0 | 1 | 2021-10-01 00:03:11 | 2021-11-03 21:45:48 | 2021-10-15 05:56:26 | 523397 |

skim\_without\_charts(X202111\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202111\_divvy\_tripdata |
| Number of rows | 359978 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 359978 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 75290 | 0.79 | 10 | 53 | 0 | 815 | 0 |
| start\_station\_id | 75290 | 0.79 | 3 | 35 | 0 | 815 | 0 |
| end\_station\_name | 79187 | 0.78 | 10 | 53 | 0 | 805 | 0 |
| end\_station\_id | 79187 | 0.78 | 3 | 35 | 0 | 805 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.89 | 0.05 | 41.65 | 41.88 | 41.89 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.84 | -87.66 | -87.64 | -87.63 | -87.53 |
| end\_lat | 191 | 1 | 41.89 | 0.05 | 41.39 | 41.88 | 41.89 | 41.93 | 42.12 |
| end\_lng | 191 | 1 | -87.65 | 0.03 | -88.97 | -87.66 | -87.64 | -87.63 | -87.53 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-11-01 00:00:14 | 2021-11-30 23:59:56 | 2021-11-12 08:32:12 | 320477 |
| ended\_at | 0 | 1 | 2021-11-01 00:04:06 | 2021-12-02 06:41:33 | 2021-11-12 08:46:55 | 320071 |

skim\_without\_charts(X202112\_divvy\_tripdata)

Data summary

|  |  |
| --- | --- |
| Name | X202112\_divvy\_tripdata |
| Number of rows | 247540 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 4 |
| POSIXct | 2 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ride\_id | 0 | 1.00 | 16 | 16 | 0 | 247540 | 0 |
| rideable\_type | 0 | 1.00 | 11 | 13 | 0 | 3 | 0 |
| start\_station\_name | 51063 | 0.79 | 10 | 53 | 0 | 818 | 0 |
| start\_station\_id | 51063 | 0.79 | 3 | 35 | 0 | 816 | 0 |
| end\_station\_name | 53498 | 0.78 | 10 | 53 | 0 | 800 | 0 |
| end\_station\_id | 53498 | 0.78 | 3 | 35 | 0 | 798 | 0 |
| member\_casual | 0 | 1.00 | 6 | 6 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| start\_lat | 0 | 1 | 41.90 | 0.05 | 41.64 | 41.88 | 41.90 | 41.93 | 42.07 |
| start\_lng | 0 | 1 | -87.65 | 0.03 | -87.84 | -87.67 | -87.64 | -87.63 | -87.52 |
| end\_lat | 144 | 1 | 41.90 | 0.05 | 41.48 | 41.88 | 41.90 | 41.93 | 42.07 |
| end\_lng | 144 | 1 | -87.65 | 0.03 | -87.85 | -87.67 | -87.64 | -87.63 | -87.52 |

**Variable type: POSIXct**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| started\_at | 0 | 1 | 2021-12-01 00:00:01 | 2021-12-31 23:59:48 | 2021-12-13 13:04:54 | 228845 |
| ended\_at | 0 | 1 | 2021-12-01 00:02:40 | 2022-01-03 17:32:18 | 2021-12-13 13:18:39 | 228657 |

## Memeriksa struktur dataframe  
str(X202101\_divvy\_tripdata)

## spec\_tbl\_df [96,834 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:96834] "E19E6F1B8D4C42ED" "DC88F20C2C55F27F" "EC45C94683FE3F27" "4FA453A75AE377DB" ...  
## $ rideable\_type : chr [1:96834] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:96834], format: "2021-01-23 16:14:19" "2021-01-27 18:43:08" ...  
## $ ended\_at : POSIXct[1:96834], format: "2021-01-23 16:24:44" "2021-01-27 18:47:12" ...  
## $ start\_station\_name: chr [1:96834] "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" ...  
## $ start\_station\_id : chr [1:96834] "17660" "17660" "17660" "17660" ...  
## $ end\_station\_name : chr [1:96834] NA NA NA NA ...  
## $ end\_station\_id : chr [1:96834] NA NA NA NA ...  
## $ start\_lat : num [1:96834] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:96834] -87.7 -87.7 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:96834] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:96834] -87.7 -87.7 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:96834] "member" "member" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202102\_divvy\_tripdata)

## spec\_tbl\_df [49,622 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:49622] "89E7AA6C29227EFF" "0FEFDE2603568365" "E6159D746B2DBB91" "B32D3199F1C2E75B" ...  
## $ rideable\_type : chr [1:49622] "classic\_bike" "classic\_bike" "electric\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:49622], format: "2021-02-12 16:14:56" "2021-02-14 17:52:38" ...  
## $ ended\_at : POSIXct[1:49622], format: "2021-02-12 16:21:43" "2021-02-14 18:12:09" ...  
## $ start\_station\_name: chr [1:49622] "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Clark St & Lake St" "Wood St & Chicago Ave" ...  
## $ start\_station\_id : chr [1:49622] "525" "525" "KA1503000012" "637" ...  
## $ end\_station\_name : chr [1:49622] "Sheridan Rd & Columbia Ave" "Bosworth Ave & Howard St" "State St & Randolph St" "Honore St & Division St" ...  
## $ end\_station\_id : chr [1:49622] "660" "16806" "TA1305000029" "TA1305000034" ...  
## $ start\_lat : num [1:49622] 42 42 41.9 41.9 41.8 ...  
## $ start\_lng : num [1:49622] -87.7 -87.7 -87.6 -87.7 -87.6 ...  
## $ end\_lat : num [1:49622] 42 42 41.9 41.9 41.8 ...  
## $ end\_lng : num [1:49622] -87.7 -87.7 -87.6 -87.7 -87.6 ...  
## $ member\_casual : chr [1:49622] "member" "casual" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202103\_divvy\_tripdata)

## spec\_tbl\_df [228,496 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:228496] "CFA86D4455AA1030" "30D9DC61227D1AF3" "846D87A15682A284" "994D05AA75A168F2" ...  
## $ rideable\_type : chr [1:228496] "classic\_bike" "classic\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:228496], format: "2021-03-16 08:32:30" "2021-03-28 01:26:28" ...  
## $ ended\_at : POSIXct[1:228496], format: "2021-03-16 08:36:34" "2021-03-28 01:36:55" ...  
## $ start\_station\_name: chr [1:228496] "Humboldt Blvd & Armitage Ave" "Humboldt Blvd & Armitage Ave" "Shields Ave & 28th Pl" "Winthrop Ave & Lawrence Ave" ...  
## $ start\_station\_id : chr [1:228496] "15651" "15651" "15443" "TA1308000021" ...  
## $ end\_station\_name : chr [1:228496] "Stave St & Armitage Ave" "Central Park Ave & Bloomingdale Ave" "Halsted St & 35th St" "Broadway & Sheridan Rd" ...  
## $ end\_station\_id : chr [1:228496] "13266" "18017" "TA1308000043" "13323" ...  
## $ start\_lat : num [1:228496] 41.9 41.9 41.8 42 42 ...  
## $ start\_lng : num [1:228496] -87.7 -87.7 -87.6 -87.7 -87.7 ...  
## $ end\_lat : num [1:228496] 41.9 41.9 41.8 42 42.1 ...  
## $ end\_lng : num [1:228496] -87.7 -87.7 -87.6 -87.6 -87.7 ...  
## $ member\_casual : chr [1:228496] "casual" "casual" "casual" "casual" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202104\_divvy\_tripdata)

## spec\_tbl\_df [337,230 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:337230] "6C992BD37A98A63F" "1E0145613A209000" "E498E15508A80BAD" "1887262AD101C604" ...  
## $ rideable\_type : chr [1:337230] "classic\_bike" "docked\_bike" "docked\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:337230], format: "2021-04-12 18:25:36" "2021-04-27 17:27:11" ...  
## $ ended\_at : POSIXct[1:337230], format: "2021-04-12 18:56:55" "2021-04-27 18:31:29" ...  
## $ start\_station\_name: chr [1:337230] "State St & Pearson St" "Dorchester Ave & 49th St" "Loomis Blvd & 84th St" "Honore St & Division St" ...  
## $ start\_station\_id : chr [1:337230] "TA1307000061" "KA1503000069" "20121" "TA1305000034" ...  
## $ end\_station\_name : chr [1:337230] "Southport Ave & Waveland Ave" "Dorchester Ave & 49th St" "Loomis Blvd & 84th St" "Southport Ave & Waveland Ave" ...  
## $ end\_station\_id : chr [1:337230] "13235" "KA1503000069" "20121" "13235" ...  
## $ start\_lat : num [1:337230] 41.9 41.8 41.7 41.9 41.7 ...  
## $ start\_lng : num [1:337230] -87.6 -87.6 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:337230] 41.9 41.8 41.7 41.9 41.7 ...  
## $ end\_lng : num [1:337230] -87.7 -87.6 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:337230] "member" "casual" "casual" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202105\_divvy\_tripdata)

## spec\_tbl\_df [531,633 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:531633] "C809ED75D6160B2A" "DD59FDCE0ACACAF3" "0AB83CB88C43EFC2" "7881AC6D39110C60" ...  
## $ rideable\_type : chr [1:531633] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:531633], format: "2021-05-30 11:58:15" "2021-05-30 11:29:14" ...  
## $ ended\_at : POSIXct[1:531633], format: "2021-05-30 12:10:39" "2021-05-30 12:14:09" ...  
## $ start\_station\_name: chr [1:531633] NA NA NA NA ...  
## $ start\_station\_id : chr [1:531633] NA NA NA NA ...  
## $ end\_station\_name : chr [1:531633] NA NA NA NA ...  
## $ end\_station\_id : chr [1:531633] NA NA NA NA ...  
## $ start\_lat : num [1:531633] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:531633] -87.6 -87.6 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:531633] 41.9 41.8 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:531633] -87.6 -87.6 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:531633] "casual" "casual" "casual" "casual" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202106\_divvy\_tripdata)

## spec\_tbl\_df [729,595 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:729595] "99FEC93BA843FB20" "06048DCFC8520CAF" "9598066F68045DF2" "B03C0FE48C412214" ...  
## $ rideable\_type : chr [1:729595] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:729595], format: "2021-06-13 14:31:28" "2021-06-04 11:18:02" ...  
## $ ended\_at : POSIXct[1:729595], format: "2021-06-13 14:34:11" "2021-06-04 11:24:19" ...  
## $ start\_station\_name: chr [1:729595] NA NA NA NA ...  
## $ start\_station\_id : chr [1:729595] NA NA NA NA ...  
## $ end\_station\_name : chr [1:729595] NA NA NA NA ...  
## $ end\_station\_id : chr [1:729595] NA NA NA NA ...  
## $ start\_lat : num [1:729595] 41.8 41.8 41.8 41.8 41.8 ...  
## $ start\_lng : num [1:729595] -87.6 -87.6 -87.6 -87.6 -87.6 ...  
## $ end\_lat : num [1:729595] 41.8 41.8 41.8 41.8 41.8 ...  
## $ end\_lng : num [1:729595] -87.6 -87.6 -87.6 -87.6 -87.6 ...  
## $ member\_casual : chr [1:729595] "member" "member" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202107\_divvy\_tripdata)

## spec\_tbl\_df [822,410 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:822410] "0A1B623926EF4E16" "B2D5583A5A5E76EE" "6F264597DDBF427A" "379B58EAB20E8AA5" ...  
## $ rideable\_type : chr [1:822410] "docked\_bike" "classic\_bike" "classic\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:822410], format: "2021-07-02 14:44:36" "2021-07-07 16:57:42" ...  
## $ ended\_at : POSIXct[1:822410], format: "2021-07-02 15:19:58" "2021-07-07 17:16:09" ...  
## $ start\_station\_name: chr [1:822410] "Michigan Ave & Washington St" "California Ave & Cortez St" "Wabash Ave & 16th St" "California Ave & Cortez St" ...  
## $ start\_station\_id : chr [1:822410] "13001" "17660" "SL-012" "17660" ...  
## $ end\_station\_name : chr [1:822410] "Halsted St & North Branch St" "Wood St & Hubbard St" "Rush St & Hubbard St" "Carpenter St & Huron St" ...  
## $ end\_station\_id : chr [1:822410] "KA1504000117" "13432" "KA1503000044" "13196" ...  
## $ start\_lat : num [1:822410] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:822410] -87.6 -87.7 -87.6 -87.7 -87.7 ...  
## $ end\_lat : num [1:822410] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:822410] -87.6 -87.7 -87.6 -87.7 -87.7 ...  
## $ member\_casual : chr [1:822410] "casual" "casual" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202108\_divvy\_tripdata)

## spec\_tbl\_df [804,352 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:804352] "99103BB87CC6C1BB" "EAFCCCFB0A3FC5A1" "9EF4F46C57AD234D" "5834D3208BFAF1DA" ...  
## $ rideable\_type : chr [1:804352] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:804352], format: "2021-08-10 17:15:49" "2021-08-10 17:23:14" ...  
## $ ended\_at : POSIXct[1:804352], format: "2021-08-10 17:22:44" "2021-08-10 17:39:24" ...  
## $ start\_station\_name: chr [1:804352] NA NA NA NA ...  
## $ start\_station\_id : chr [1:804352] NA NA NA NA ...  
## $ end\_station\_name : chr [1:804352] NA NA NA NA ...  
## $ end\_station\_id : chr [1:804352] NA NA NA NA ...  
## $ start\_lat : num [1:804352] 41.8 41.8 42 42 41.8 ...  
## $ start\_lng : num [1:804352] -87.7 -87.7 -87.7 -87.7 -87.6 ...  
## $ end\_lat : num [1:804352] 41.8 41.8 42 42 41.8 ...  
## $ end\_lng : num [1:804352] -87.7 -87.6 -87.7 -87.7 -87.6 ...  
## $ member\_casual : chr [1:804352] "member" "member" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202109\_divvy\_tripdata)

## spec\_tbl\_df [756,147 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:756147] "9DC7B962304CBFD8" "F930E2C6872D6B32" "6EF72137900BB910" "78D1DE133B3DBF55" ...  
## $ rideable\_type : chr [1:756147] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:756147], format: "2021-09-28 16:07:10" "2021-09-28 14:24:51" ...  
## $ ended\_at : POSIXct[1:756147], format: "2021-09-28 16:09:54" "2021-09-28 14:40:05" ...  
## $ start\_station\_name: chr [1:756147] NA NA NA NA ...  
## $ start\_station\_id : chr [1:756147] NA NA NA NA ...  
## $ end\_station\_name : chr [1:756147] NA NA NA NA ...  
## $ end\_station\_id : chr [1:756147] NA NA NA NA ...  
## $ start\_lat : num [1:756147] 41.9 41.9 41.8 41.8 41.9 ...  
## $ start\_lng : num [1:756147] -87.7 -87.6 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:756147] 41.9 42 41.8 41.8 41.9 ...  
## $ end\_lng : num [1:756147] -87.7 -87.7 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:756147] "casual" "casual" "casual" "casual" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202110\_divvy\_tripdata)

## spec\_tbl\_df [631,226 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:631226] "620BC6107255BF4C" "4471C70731AB2E45" "26CA69D43D15EE14" "362947F0437E1514" ...  
## $ rideable\_type : chr [1:631226] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:631226], format: "2021-10-22 12:46:42" "2021-10-21 09:12:37" ...  
## $ ended\_at : POSIXct[1:631226], format: "2021-10-22 12:49:50" "2021-10-21 09:14:14" ...  
## $ start\_station\_name: chr [1:631226] "Kingsbury St & Kinzie St" NA NA NA ...  
## $ start\_station\_id : chr [1:631226] "KA1503000043" NA NA NA ...  
## $ end\_station\_name : chr [1:631226] NA NA NA NA ...  
## $ end\_station\_id : chr [1:631226] NA NA NA NA ...  
## $ start\_lat : num [1:631226] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:631226] -87.6 -87.7 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:631226] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:631226] -87.6 -87.7 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:631226] "member" "member" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202111\_divvy\_tripdata)

## spec\_tbl\_df [359,978 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:359978] "7C00A93E10556E47" "90854840DFD508BA" "0A7D10CDD144061C" "2F3BE33085BCFF02" ...  
## $ rideable\_type : chr [1:359978] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:359978], format: "2021-11-27 13:27:38" "2021-11-27 13:38:25" ...  
## $ ended\_at : POSIXct[1:359978], format: "2021-11-27 13:46:38" "2021-11-27 13:56:10" ...  
## $ start\_station\_name: chr [1:359978] NA NA NA NA ...  
## $ start\_station\_id : chr [1:359978] NA NA NA NA ...  
## $ end\_station\_name : chr [1:359978] NA NA NA NA ...  
## $ end\_station\_id : chr [1:359978] NA NA NA NA ...  
## $ start\_lat : num [1:359978] 41.9 42 42 41.9 41.9 ...  
## $ start\_lng : num [1:359978] -87.7 -87.7 -87.7 -87.8 -87.6 ...  
## $ end\_lat : num [1:359978] 42 41.9 42 41.9 41.9 ...  
## $ end\_lng : num [1:359978] -87.7 -87.7 -87.7 -87.8 -87.6 ...  
## $ member\_casual : chr [1:359978] "casual" "casual" "casual" "casual" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

str(X202112\_divvy\_tripdata)

## spec\_tbl\_df [247,540 × 13] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:247540] "46F8167220E4431F" "73A77762838B32FD" "4CF42452054F59C5" "3278BA87BF698339" ...  
## $ rideable\_type : chr [1:247540] "electric\_bike" "electric\_bike" "electric\_bike" "classic\_bike" ...  
## $ started\_at : POSIXct[1:247540], format: "2021-12-07 15:06:07" "2021-12-11 03:43:29" ...  
## $ ended\_at : POSIXct[1:247540], format: "2021-12-07 15:13:42" "2021-12-11 04:10:23" ...  
## $ start\_station\_name: chr [1:247540] "Laflin St & Cullerton St" "LaSalle Dr & Huron St" "Halsted St & North Branch St" "Halsted St & North Branch St" ...  
## $ start\_station\_id : chr [1:247540] "13307" "KP1705001026" "KA1504000117" "KA1504000117" ...  
## $ end\_station\_name : chr [1:247540] "Morgan St & Polk St" "Clarendon Ave & Leland Ave" "Broadway & Barry Ave" "LaSalle Dr & Huron St" ...  
## $ end\_station\_id : chr [1:247540] "TA1307000130" "TA1307000119" "13137" "KP1705001026" ...  
## $ start\_lat : num [1:247540] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:247540] -87.7 -87.6 -87.6 -87.6 -87.7 ...  
## $ end\_lat : num [1:247540] 41.9 42 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:247540] -87.7 -87.7 -87.6 -87.6 -87.6 ...  
## $ member\_casual : chr [1:247540] "member" "casual" "member" "member" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ride\_id = col\_character(),  
## .. rideable\_type = col\_character(),  
## .. started\_at = col\_datetime(format = ""),  
## .. ended\_at = col\_datetime(format = ""),  
## .. start\_station\_name = col\_character(),  
## .. start\_station\_id = col\_character(),  
## .. end\_station\_name = col\_character(),  
## .. end\_station\_id = col\_character(),  
## .. start\_lat = col\_double(),  
## .. start\_lng = col\_double(),  
## .. end\_lat = col\_double(),  
## .. end\_lng = col\_double(),  
## .. member\_casual = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

### 3.4. Menggabungkan seluruh data menjadi satu big dataframe

tripdata <- rbind(X202101\_divvy\_tripdata, X202102\_divvy\_tripdata, X202103\_divvy\_tripdata, X202104\_divvy\_tripdata, X202105\_divvy\_tripdata, X202106\_divvy\_tripdata, X202107\_divvy\_tripdata, X202108\_divvy\_tripdata, X202109\_divvy\_tripdata, X202110\_divvy\_tripdata, X202111\_divvy\_tripdata, X202112\_divvy\_tripdata)

terdapat 13 variabel dan 5595063 baris pada big dataset tripdata.

### 3.5. Clean up data

## data ride\_id setiap baris harus unique value sehingga tidak boleh ada nilai yang sama  
tripdata %>% group\_by(ride\_id) %>% filter(n()>1)

## # A tibble: 0 × 13  
## # Groups: ride\_id [0]  
## # … with 13 variables: ride\_id <chr>, rideable\_type <chr>, started\_at <dttm>,  
## # ended\_at <dttm>, start\_station\_name <chr>, start\_station\_id <chr>,  
## # end\_station\_name <chr>, end\_station\_id <chr>, start\_lat <dbl>,  
## # start\_lng <dbl>, end\_lat <dbl>, end\_lng <dbl>, member\_casual <chr>

## Secara logika, seluruh nilai kolom "started\_at" memiliki date-time yang lebih dahulu (besar) dibandingkan dengan nilai pada kolom "ended\_at"  
tripdata %>% filter(started\_at > ended\_at)

## # A tibble: 147 × 13  
## ride\_id ridea…¹ started\_at ended\_at start…² start…³  
## <chr> <chr> <dttm> <dttm> <chr> <chr>   
## 1 FC1EFEF4475D… classi… 2021-01-09 15:42:45 2021-01-09 15:41:02 Montic… KA1504…  
## 2 B1235D38EB2F… electr… 2021-01-06 18:33:12 2021-01-06 18:31:07 Daley … TA1306…  
## 3 F79335E3A77A… electr… 2021-03-29 15:41:21 2021-03-29 15:41:20 Ashlan… 13249   
## 4 5D2797A8FFA7… classi… 2021-03-13 18:02:58 2021-03-13 18:02:57 Dayton… 13058   
## 5 BC53ECCBC762… classi… 2021-04-07 16:11:33 2021-04-07 16:11:26 Ashlan… 13434   
## 6 209C097828F9… electr… 2021-04-27 17:13:44 2021-04-27 17:11:32 <NA> <NA>   
## 7 6E81034B446F… electr… 2021-04-23 09:43:39 2021-04-23 09:43:29 Dayton… 13058   
## 8 318DD838369A… classi… 2021-04-30 10:56:32 2021-04-30 10:56:30 Dayton… 13058   
## 9 8ADD13BD8F6A… classi… 2021-04-17 12:43:36 2021-04-17 12:43:27 Dayton… 13058   
## 10 3EC1B5A4D4B9… classi… 2021-05-05 16:10:04 2021-05-05 16:09:51 Dayton… 13058   
## # … with 137 more rows, 7 more variables: end\_station\_name <chr>,  
## # end\_station\_id <chr>, start\_lat <dbl>, start\_lng <dbl>, end\_lat <dbl>,  
## # end\_lng <dbl>, member\_casual <chr>, and abbreviated variable names  
## # ¹​rideable\_type, ²​start\_station\_name, ³​start\_station\_id

## hasilnya, lebih dari 137 baris yang memiliki nilai "started\_at" yang lebih kecil dari pada "ended\_at". Filter data yang tidak logis  
tripdata2 <- filter(tripdata, started\_at < ended\_at)  
tripdata2 %>% filter(started\_at > ended\_at) ## pengecekan kembali dataframe

## # A tibble: 0 × 13  
## # … with 13 variables: ride\_id <chr>, rideable\_type <chr>, started\_at <dttm>,  
## # ended\_at <dttm>, start\_station\_name <chr>, start\_station\_id <chr>,  
## # end\_station\_name <chr>, end\_station\_id <chr>, start\_lat <dbl>,  
## # start\_lng <dbl>, end\_lat <dbl>, end\_lng <dbl>, member\_casual <chr>

## memisahkan tanggal, tahun, bulan, hari, dan jam  
tripdata2$date <- lubridate::date(tripdata2$started\_at)  
tripdata2$year <- lubridate::year(tripdata2$started\_at)  
tripdata2$month <- lubridate::month(tripdata2$started\_at)  
tripdata2$day <- lubridate::day(tripdata2$started\_at)  
tripdata2$day\_of\_week <- format(as.Date(tripdata2$started\_at), "%A")  
tripdata2$start\_hour <- lubridate::hour(tripdata2$started\_at)  
  
## menghitung durasi perjalanan dalam detik  
tripdata2 <- tripdata2 %>% mutate(ride\_duration\_secs=abs(difftime(tripdata2$ended\_at,tripdata2$started\_at,units="secs")))  
str(tripdata2)

## tibble [5,594,410 × 20] (S3: tbl\_df/tbl/data.frame)  
## $ ride\_id : chr [1:5594410] "E19E6F1B8D4C42ED" "DC88F20C2C55F27F" "EC45C94683FE3F27" "4FA453A75AE377DB" ...  
## $ rideable\_type : chr [1:5594410] "electric\_bike" "electric\_bike" "electric\_bike" "electric\_bike" ...  
## $ started\_at : POSIXct[1:5594410], format: "2021-01-23 16:14:19" "2021-01-27 18:43:08" ...  
## $ ended\_at : POSIXct[1:5594410], format: "2021-01-23 16:24:44" "2021-01-27 18:47:12" ...  
## $ start\_station\_name: chr [1:5594410] "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" "California Ave & Cortez St" ...  
## $ start\_station\_id : chr [1:5594410] "17660" "17660" "17660" "17660" ...  
## $ end\_station\_name : chr [1:5594410] NA NA NA NA ...  
## $ end\_station\_id : chr [1:5594410] NA NA NA NA ...  
## $ start\_lat : num [1:5594410] 41.9 41.9 41.9 41.9 41.9 ...  
## $ start\_lng : num [1:5594410] -87.7 -87.7 -87.7 -87.7 -87.7 ...  
## $ end\_lat : num [1:5594410] 41.9 41.9 41.9 41.9 41.9 ...  
## $ end\_lng : num [1:5594410] -87.7 -87.7 -87.7 -87.7 -87.7 ...  
## $ member\_casual : chr [1:5594410] "member" "member" "member" "member" ...  
## $ date : Date[1:5594410], format: "2021-01-23" "2021-01-27" ...  
## $ year : num [1:5594410] 2021 2021 2021 2021 2021 ...  
## $ month : num [1:5594410] 1 1 1 1 1 1 1 1 1 1 ...  
## $ day : int [1:5594410] 23 27 21 7 23 9 4 14 9 24 ...  
## $ day\_of\_week : chr [1:5594410] "Saturday" "Wednesday" "Thursday" "Thursday" ...  
## $ start\_hour : int [1:5594410] 16 18 22 13 2 14 5 15 9 19 ...  
## $ ride\_duration\_secs: 'difftime' num [1:5594410] 625 244 80 702 ...  
## ..- attr(\*, "units")= chr "secs"

## calculate the distance to meter from the coordinates   
tripdata2$ride\_distance <- distGeo(matrix(c(tripdata2$start\_lng, tripdata2$start\_lat), ncol = 2),  
 matrix(c(tripdata2$end\_lng, tripdata2$end\_lat), ncol = 2))  
tripdata2$ride\_distance <- tripdata2$ride\_distance  
  
## the dataframe contains few entries with ride duration is negative due to wrong entries let's make sure to remove these from the data frame.  
tripdata2 <- tripdata2 %>% filter(ride\_distance > 0 | start\_station\_name!="HQ QR") %>% drop\_na()

## Step 4 - Menganalisis data

## analisis durasi penggunaan user dalam detik  
table(tripdata2$member\_casual) ## menghitung banyaknya perjalanan user

##   
## casual member   
## 2048302 2539802

mean(tripdata2$ride\_duration\_secs)

## Time difference of 1308.715 secs

median(tripdata2$ride\_duration\_secs)

## Time difference of 732 secs

max(tripdata2$ride\_duration\_secs)

## Time difference of 3356649 secs

min(tripdata2$ride\_duration\_secs)

## Time difference of 1 secs

Perjalanan setiap pengendara dalam satu tahun

|  |  |
| --- | --- |
| **Casual** | **Member** |
| 2048302 | 2539802 |

Durasi berkendara dari kedua pengendara selama 1 tahun

* Mean = 1308.715 secs
* Median = 732 secs
* Max = 3356649 secs
* Min = 1 secs

## membandingkan penggunaan Member dan Casual dalam waktu satu tahun  
aggregate(tripdata2$ride\_duration\_secs ~ tripdata2$member\_casual, FUN = mean)

## tripdata2$member\_casual tripdata2$ride\_duration\_secs  
## 1 casual 1950.6091 secs  
## 2 member 791.0405 secs

aggregate(tripdata2$ride\_duration\_secs ~ tripdata2$member\_casual, FUN = median)

## tripdata2$member\_casual tripdata2$ride\_duration\_secs  
## 1 casual 999 secs  
## 2 member 583 secs

aggregate(tripdata2$ride\_duration\_secs ~ tripdata2$member\_casual, FUN = max)

## tripdata2$member\_casual tripdata2$ride\_duration\_secs  
## 1 casual 3356649 secs  
## 2 member 89738 secs

aggregate(tripdata2$ride\_duration\_secs ~ tripdata2$member\_casual, FUN = min)

## tripdata2$member\_casual tripdata2$ride\_duration\_secs  
## 1 casual 1 secs  
## 2 member 1 secs

| Agregat | Member | Casual |
| --- | --- | --- |
| **Mean** | 791.0405 secs | 1950.6091 secs |
| **Median** | 583 secs | 999 secs |
| **Max** | 89738 secs | 3356649 secs |
| **Min** | 1 secs | 1 secs |

## analisis waktu penggunaan dalam meter  
mean(tripdata2$ride\_distance)

## [1] 2129.285

median(tripdata2$ride\_distance)

## [1] 1621.533

max(tripdata2$ride\_distance)

## [1] 33762.61

min(tripdata2$ride\_distance)

## [1] 0

Jarak tempuh kedua pengendara selama 1 tahun

* Mean = 2129.285 meter
* Median = 1621.533 meter
* Max = 33762.61 meter
* Min = 0 meter

Jarak tempuh 0 meter diakibatkan pengendara sepeda menggunakan stasiun yang sama untuk keberangkatan dan stasiun tujuan akhir.

## membandingkan jarak tempuh dalam meter setiap user  
aggregate(tripdata2$ride\_distance ~ tripdata2$member\_casual + tripdata2$day\_of\_week, FUN = mean)

## tripdata2$member\_casual tripdata2$day\_of\_week tripdata2$ride\_distance  
## 1 casual Friday 2166.963  
## 2 member Friday 2052.001  
## 3 casual Monday 2067.298  
## 4 member Monday 2039.135  
## 5 casual Saturday 2283.022  
## 6 member Saturday 2186.894  
## 7 casual Sunday 2244.900  
## 8 member Sunday 2187.047  
## 9 casual Thursday 2130.047  
## 10 member Thursday 2046.676  
## 11 casual Tuesday 2094.318  
## 12 member Tuesday 2054.092  
## 13 casual Wednesday 2119.999  
## 14 member Wednesday 2066.291

aggregate(tripdata2$ride\_distance ~ tripdata2$member\_casual + tripdata2$day\_of\_week, FUN = median)

## tripdata2$member\_casual tripdata2$day\_of\_week tripdata2$ride\_distance  
## 1 casual Friday 1691.085  
## 2 member Friday 1534.696  
## 3 casual Monday 1599.692  
## 4 member Monday 1512.772  
## 5 casual Saturday 1797.699  
## 6 member Saturday 1666.662  
## 7 casual Sunday 1740.317  
## 8 member Sunday 1650.854  
## 9 casual Thursday 1673.203  
## 10 member Thursday 1523.702  
## 11 casual Tuesday 1633.966  
## 12 member Tuesday 1521.386  
## 13 casual Wednesday 1662.328  
## 14 member Wednesday 1535.214

aggregate(tripdata2$ride\_distance ~ tripdata2$member\_casual + tripdata2$day\_of\_week, FUN = max)

## tripdata2$member\_casual tripdata2$day\_of\_week tripdata2$ride\_distance  
## 1 casual Friday 32174.65  
## 2 member Friday 27662.54  
## 3 casual Monday 29011.99  
## 4 member Monday 25208.04  
## 5 casual Saturday 30157.49  
## 6 member Saturday 25221.31  
## 7 casual Sunday 28802.73  
## 8 member Sunday 31531.60  
## 9 casual Thursday 29366.01  
## 10 member Thursday 26318.04  
## 11 casual Tuesday 33762.61  
## 12 member Tuesday 23418.15  
## 13 casual Wednesday 31883.39  
## 14 member Wednesday 24734.28

aggregate(tripdata2$ride\_distance ~ tripdata2$member\_casual + tripdata2$day\_of\_week, FUN = min)

## tripdata2$member\_casual tripdata2$day\_of\_week tripdata2$ride\_distance  
## 1 casual Friday 0  
## 2 member Friday 0  
## 3 casual Monday 0  
## 4 member Monday 0  
## 5 casual Saturday 0  
## 6 member Saturday 0  
## 7 casual Sunday 0  
## 8 member Sunday 0  
## 9 casual Thursday 0  
## 10 member Thursday 0  
## 11 casual Tuesday 0  
## 12 member Tuesday 0  
## 13 casual Wednesday 0  
## 14 member Wednesday 0

* Mean

| No | Pengendara | Day | Jarak (Meter) |
| --- | --- | --- | --- |
| 1 | Casual | Sunday | 2244.900 |
| 2 | Member | Sunday | 2187.047 |
| 3 | Casual | Monday | 2067.298 |
| 4 | Member | Monday | 2039.135 |
| 5 | Casual | Tuesday | 2094.318 |
| 6 | Member | Tuesday | 2054.092 |
| 7 | Casual | Wednesday | 2119.999 |
| 8 | Member | Wednesday | 2066.291 |
| 9 | Casual | Thursday | 2130.047 |
| 10 | Member | Thursday | 2046.676 |
| 11 | Casual | Friday | 2166.963 |
| 12 | Member | Friday | 2052.001 |
| 13 | Casual | Saturday | 2283.022 |
| 14 | Member | Saturday | 2186.894 |

* Median

| No | Pengendara | Day | Jarak (Meter) |
| --- | --- | --- | --- |
| 1 | Casual | Sunday | 1740.317 |
| 2 | Member | Sunday | 1650.854 |
| 3 | Casual | Monday | 1599.692 |
| 4 | Member | Monday | 1512.772 |
| 5 | Casual | Tuesday | 1633.966 |
| 6 | Member | Tuesday | 1521.386 |
| 7 | Casual | Wednesday | 1662.328 |
| 8 | Member | Wednesday | 1535.214 |
| 9 | Casual | Thursday | 1673.203 |
| 10 | Member | Thursday | 1523.702 |
| 11 | Casual | Friday | 1691.085 |
| 12 | Member | Friday | 1534.696 |
| 13 | Casual | Saturday | 1797.699 |
| 14 | Member | Saturday | 1666.662 |

* Max

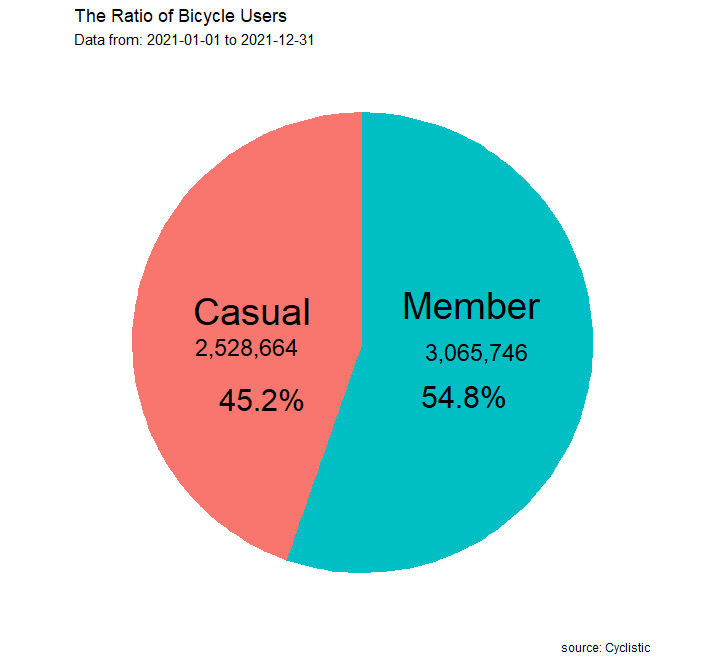
| No | Pengendara | Day | Jarak (Meter) |
| --- | --- | --- | --- |
| 1 | Casual | Sunday | 28802.73 |
| 2 | Member | Sunday | 31531.60 |
| 3 | Casual | Monday | 29011.99 |
| 4 | Member | Monday | 25208.04 |
| 5 | Casual | Tuesday | 33762.61 |
| 6 | Member | Tuesday | 23418.15 |
| 7 | Casual | Wednesday | 31883.39 |
| 8 | Member | Wednesday | 24734.28 |
| 9 | Casual | Thursday | 29366.01 |
| 10 | Member | Thursday | 26318.04 |
| 11 | Casual | Friday | 32174.65 |
| 12 | Member | Friday | 27662.54 |
| 13 | Casual | Saturday | 30157.49 |
| 14 | Member | Saturday | 25221.31 |

* Min

| No | Pengendara | Day | Jarak (Meter) |
| --- | --- | --- | --- |
| 1 | Casual | Sunday | 0 |
| 2 | Member | Sunday | 0 |
| 3 | Casual | Monday | 0 |
| 4 | Member | Monday | 0 |
| 5 | Casual | Tuesday | 0 |
| 6 | Member | Tuesday | 0 |
| 7 | Casual | Wednesday | 0 |
| 8 | Member | Wednesday | 0 |
| 9 | Casual | Thursday | 0 |
| 10 | Member | Thursday | 0 |
| 11 | Casual | Friday | 0 |
| 12 | Member | Friday | 0 |
| 13 | Casual | Saturday | 0 |
| 14 | Member | Saturday | 0 |

**Visualisasi data**

## grafik 1  
## membandingkan ratio jumlah pengendara dengan pie chart  
mindate <- min(tripdata2$date)   
maxdate <- max(tripdata2$date) ## untuk keterangan periode penggunaan waktu di subtitle grafik  
  
ggplot(data = tripdata2)+  
 geom\_bar(mapping=aes(x = "",fill = member\_casual))+  
 coord\_polar("y")+  
 annotate("text", label = "Member", x = 1, y = 900000, size = 10)+  
 annotate("text", label = "3,065,746", x = 1, y = 1200000, size = 6)+  
 annotate("text", label = "54.8%", x = 1, y = 1500000, size = 8)+  
 annotate("text", label = "Casual", x = 1 ,y = 3650000, size = 10)+  
 annotate("text", label = "2,528,664", x = 1, y = 3420000, size = 6)+  
 annotate("text", label = "45.2%", x = 1, y = 3070000, size = 8)+  
 theme(plot.background = element\_blank(),  
 panel.background = element\_blank(),  
 axis.title = element\_blank(),  
 axis.ticks = element\_blank(),  
 axis.text = element\_blank(),  
 legend.position =" none")+  
 labs(title ="The Ratio of Bicycle Users",  
 subtitle = paste0("Data from: ", mindate, " to ", maxdate),  
 caption = "source: Cyclistic")

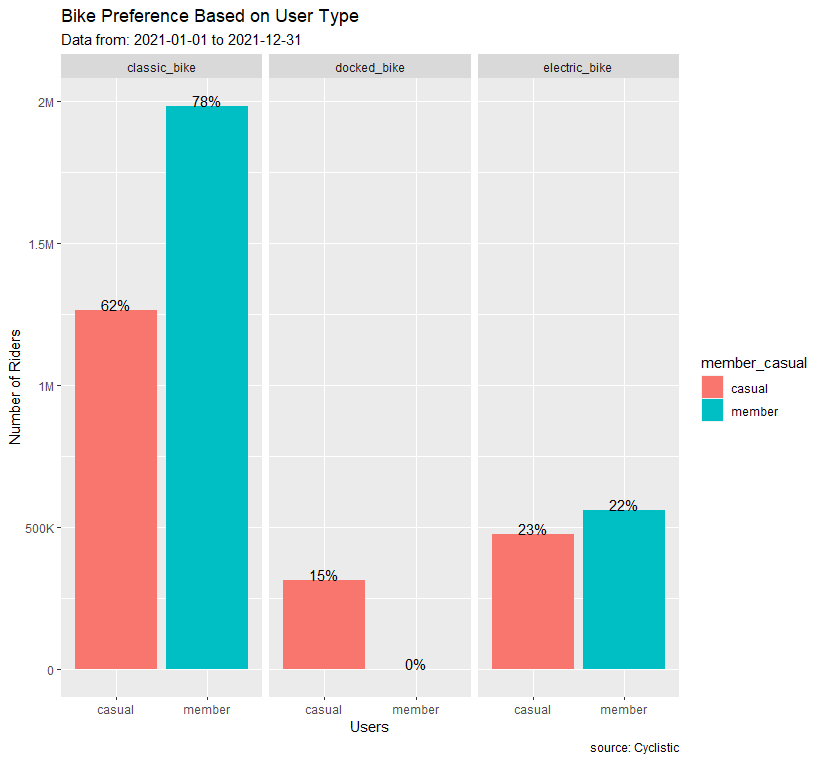


Gambar 6.1 Ratio Pengguna Sepeda dalam Satu Tahun

Penggendara member 54.8% lebih banyak menggunakan sepeda dibandingkan dengan pengendara casual yang hanya 45.2% dalam kurun waktu satu tahun, akan tetapi selisihnya tidak terlalu signifikan. Sehingga perlu merancang strategi pemasaran baru untuk meningkatkan jumlah pengendara casual yang membeli keangotaan tahunan.

## grafik 2  
## menemukan preferensi jenis sepeda yang digunakan kedua user   
ggplot(  
 tripdata2%>%  
 group\_by(member\_casual, rideable\_type) %>%  
 summarise(n = n())%>%  
 mutate(percentage = n/sum(n) \*100),  
 aes(x = member\_casual, y = n, fill = member\_casual)) +   
 geom\_col(position = "dodge") +  
 geom\_text(aes(label = paste0(round(percentage), "%")), position = position\_dodge(0.85), vjust = 0) +  
 labs(x = "Users", y = "Number of Riders", title = "Bike Preference Based on User Type",  
 subtitle = paste0("Data from: ", mindate, " to ", maxdate),  
 caption = "source: Cyclistic")+  
 scale\_y\_continuous(breaks = c(0, 500000,1000000, 1500000, 2000000), labels = c("0", "500K", "1M", "1.5M", "2M"))+  
 facet\_wrap(~rideable\_type)

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



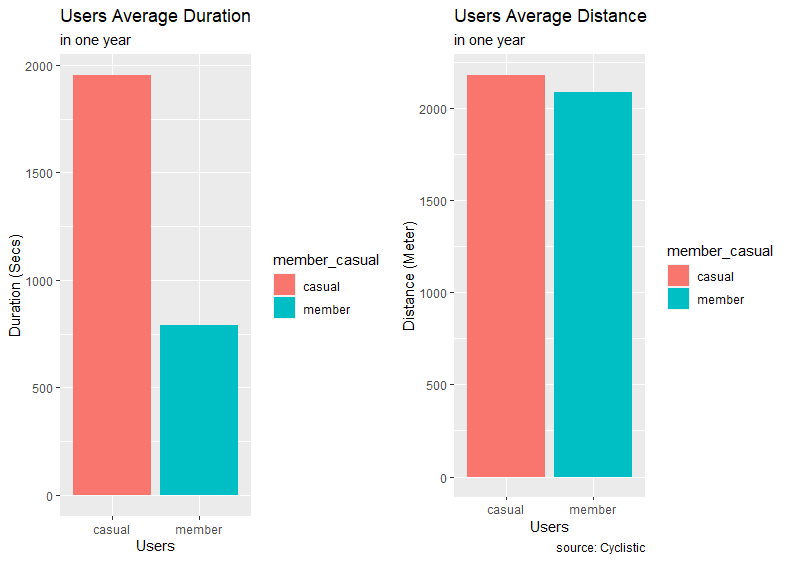
Gambar 6.2 Preferensi Jenis Sepeda yang Digunakan

Terdapat 3 tipe sepeda yang disediakan perusahaan Cyclistic, diantaranya: *classic bike, docked bike,* dan *electric bike*. Tipe sepeda paling banyak digunakan dalam kurun waktu satu tahun, yaitu: *classic bike, electric bike*, dan *docked bike*. Beberapa kemungkinan alasan pengendara dalam memilih sepeda yang digunakan:

| Classic Bike | Docked Bike | Electric Bike |
| --- | --- | --- |
| lebih praktis digunakan dibandingkan electric bike dan docked bike | fasilitas stasiun yang masih kurang karena tipe sepeda docked hanya dapat di parkir pada stasiun tertentu | Fasilitas recharge daya untuk tipe sepeda ini mungkin masih belum baik |
| lebih sehat dibandingkan dengan electric bike | penempatan sepeda kurang stategis | penempatan sepeda kurang stategis |
| cost sewa lebih murah |  | mengayuh ketika baterai habis terasa sulit dan berat |

## grafik 3  
## melakukan evaluasi terhadap jarak dan waktu rata-rata mengendarai sepeda untuk kedua jenis pengendara  
averages <- tripdata2 %>% group\_by(member\_casual) %>%  
 summarise(average\_duration = mean(ride\_duration\_secs), average\_distance = mean(ride\_distance))   
  
# plot grafik untuk rata-rata durasi berkendara   
graph\_avr\_duration <- ggplot(data = averages)+  
 geom\_col(mapping = aes(x= member\_casual, y = average\_duration, fill = member\_casual))+  
 labs(title = "Users Average Duration", x = "Users", y = "Duration (Secs)",  
 subtitle = "in one year")  
  
## plot grafik untuk rata-rata jarak berkenda   
graph\_avr\_distance <- ggplot(data = averages)+  
 geom\_col(mapping = aes(x = member\_casual, y = average\_distance, fill = member\_casual))+  
 labs(title = "Users Average Distance", x = "Users", y = "Distance (Meter)",  
 subtitle = "in one year",  
 caption = "source: Cyclistic")  
  
## gabungkan 2 plot grafik kedalam satu grid   
grid.arrange(graph\_avr\_duration,graph\_avr\_distance, ncol = 2)

## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.



Gambar 6.3 Rata-Rata Waktu dan Jarak Penggunaan Sepeda

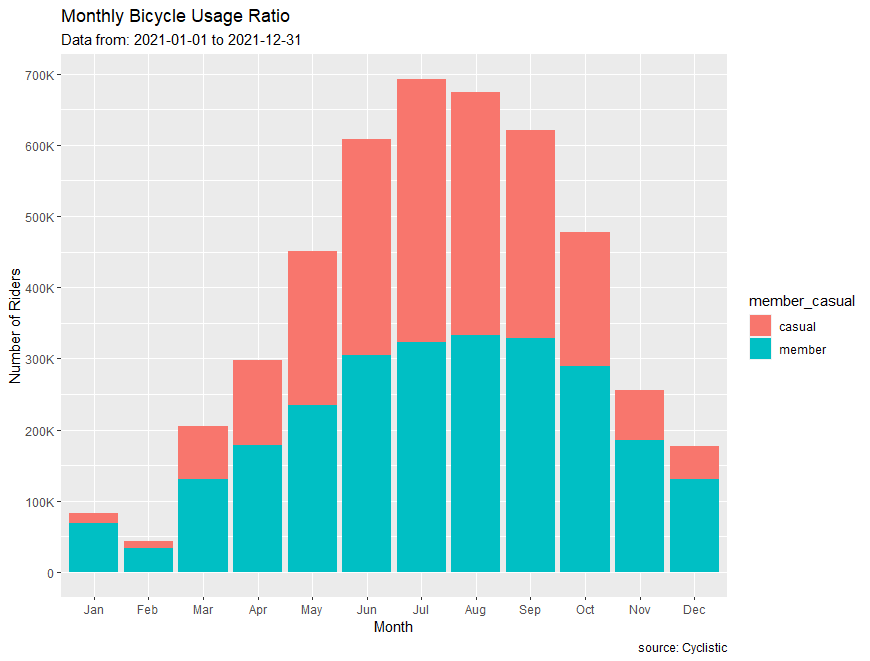
Durasi rata-rata penggunaan sepeda paling banyak digunakan oleh pengendara casual mencapai 1951 detik dan pengendara member 791 detik dalam satu tahun. Jarak tempuh sepeda rata-rata pengendara casual 2181 meter dan untuk pengendara member 2088 meter. Melihat durasi waktu dan jarak berkendara kedua jenis pengendara dapat disimpulkan:

* Pengendara casual, durasi berkendara lebih lama dan jarak tempuh sedikit lebih jauh dari pengendara member. Kemungkinan pengendara casual menggunakan sepeda dalam kondisi santai dan berkendara ke tempat yang lebih jauh seperti untuk berekreasi dan bersantai.
* Pengendara member, durasi berkendara lebih singkat dan jarak tempuh tidak terlalu jauh. Kemungkinan pengendara member menggunakan sepeda dalam kondisi yang lebih cepat untuk menuju ke tempat yang dekat seperti untuk pergi berkerja.

## grafik 4  
## analisis jumlah pengendara sepeda setiap bulannya bedasarkan tipe pengendara   
monthly\_avg <- tripdata2 %>%   
 group\_by(member\_casual, month) %>%   
 summarise(number\_of\_riders = n(), average\_duration = mean(ride\_distance)) %>%  
 arrange(month) %>%   
 ungroup()

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

## mengurutkan bulan   
monthly\_avg$month <- factor(monthly\_avg$month, levels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec" ))  
  
ggplot(data = monthly\_avg)+  
 geom\_col(mapping = aes(x = month, y = number\_of\_riders, fill = member\_casual))+  
 labs(title = "Monthly Bicycle Usage Ratio", x = "Month", y = "Number of Riders",  
 subtitle = paste0("Data from: ", mindate, " to ", maxdate),  
 caption = "source: Cyclistic")+  
 scale\_y\_continuous(breaks = c(0, 100000, 200000, 300000, 400000, 500000, 600000, 700000), labels = c("0", "100K", "200K", "300K", "400K", "500K", "600K", "700K"))



Gambar 6.4 Rasio Pemakaian Sepeda Setiap Bulan

Lokasi stasiun yang berada di daerah Chicago mengakibatkan penggunaan sepeda bergantung dengan musim. Adapun pembagian musim secara umum sebagai berikut:

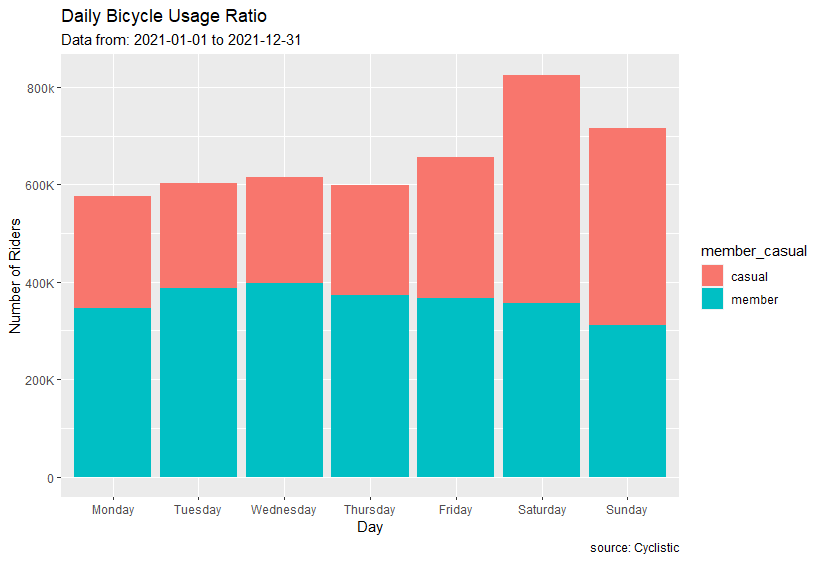
1. Musim Semi (Maret - Mei)
2. Musim Panas (Juni - Agustus)
3. Musim Gugur (September - November)
4. Musim Dingin (Desember - Februari)

Pengendara banyak menggunakan sepeda di musim semi hingga musim panas. Sementara untuk musim gugur penggunaan sepeda mulai menurun karena kondisi cuaca yang mulai dingin. Pada musim dingin penggunaan sepeda sangat sedikit, menggunakan sepeda di cuaca yang bersalju sangat kurang efektif.

## grafik 5  
## analisis jumlah pengendara sepeda setiap hari bedasarkan tipe pengendara  
daily\_avg <- tripdata2 %>%   
 group\_by(member\_casual, day\_of\_week) %>%   
 summarise(number\_of\_riders = n(), average\_duration = mean(ride\_distance)) %>%   
 arrange(day\_of\_week) %>%   
 ungroup()

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

daily\_avg$day\_of\_week <- factor(daily\_avg$day\_of\_week, levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday" ))  
  
ggplot(data = daily\_avg)+  
 geom\_col(mapping = aes(x = day\_of\_week, y = number\_of\_riders, fill = member\_casual))+  
 labs(title = "Daily Bicycle Usage Ratio", x = "Day", y = "Number of Riders",  
 subtitle = paste0("Data from: ", mindate, " to ", maxdate),  
 caption = "source: Cyclistic")+  
 scale\_y\_continuous(breaks = c(0, 200000, 400000, 600000, 800000), labels = c("0", "200K", "400K", "600K", "800k"))



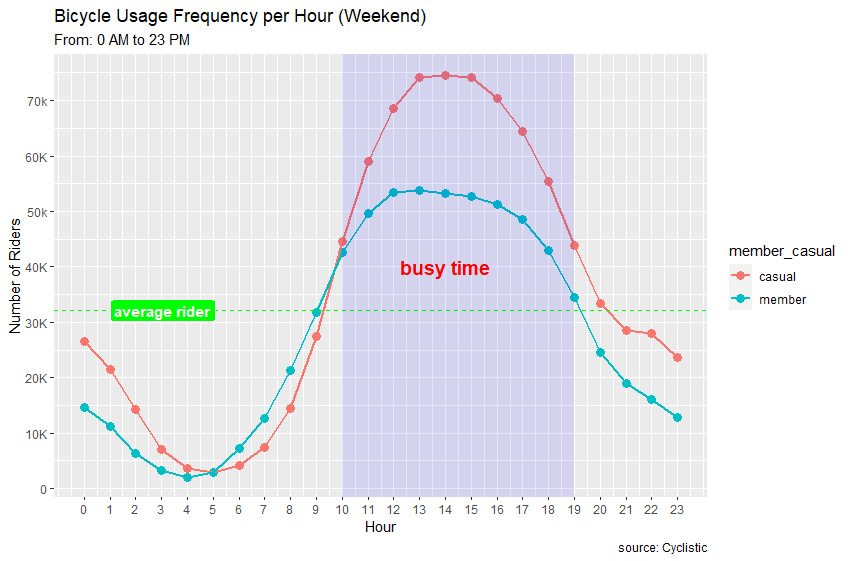
Gambar 6.5 Rasio Pemakaian Sepeda Setiap Hari

Dalam perjalanan per hari, pengendara casual menggunakan sepeda lebih banyak dibandingkan pengendara member setiap harinya. Pada akhir pekan penggunaan sepeda mengalami peningkatan untuk pengendara casual dan mengalami sedikit penurunan pada pengendara member. Hal ini mungkin terjadi karena banyak pengguna sepeda lebih cenderung bersepeda untuk bersantai di akhir pekan sehingga penyewaan sepeda meningkat. Pengendara member mengalami penurunan penggunaan sepeda pada akhir pekan, karena penggunaan untuk pergi ke tempat kerja berkurang.

## grafik 6  
## menganalisis frekuensi penggunaan sepeda setiap jamnya pada akhir pekan  
minhour <- min(tripdata2$start\_hour)   
maxhour <- max(tripdata2$start\_hour)  
  
weekend\_avg <- tripdata2 %>% filter(day\_of\_week %in% c("Sunday","Saturday")) %>%   
 group\_by(member\_casual, start\_hour) %>%  
 summarise(number\_of\_riders = n()) %>%   
 ungroup()

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

## membuat plot untuk merata-ratakan jumlah pengendara sepeda  
weekend\_plot <- mean(weekend\_avg$number\_of\_riders)  
  
ggplot(data = weekend\_avg, mapping = aes(x = start\_hour, y = number\_of\_riders, color = member\_casual, group = member\_casual))+  
 geom\_point(size = 3)+  
 geom\_line(size = 1)+   
 geom\_hline(aes(yintercept = weekend\_plot), colour = "green", linetype = "dashed")+  
 geom\_label(aes(x = 1,y = weekend\_plot+1, label = "average rider", hjust = 0), colour = "white", fill = "green", fontface="bold")+  
 annotate("rect", xmin = 10, xmax = 19, ymin = -Inf, ymax = Inf, alpha = 0.1, fill = "blue")+  
 annotate("text", x = 14, y = 40000, label = "busy time", size = 5, fontface = "bold", color = "red")+  
 labs(title = "Bicycle Usage Frequency per Hour (Weekend)", x = "Hour", y = "Number of Riders",  
 subtitle = paste0("From: ", minhour, " AM to ", maxhour, " PM "),  
 caption = "source: Cyclistic") +  
 scale\_y\_continuous(breaks = c(0, 10000, 20000, 30000, 40000, 50000, 60000, 70000), labels = c("0", "10K", "20K", "30K", "40K", "50k", "60K", "70k")) +  
 scale\_x\_continuous( breaks = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23)) ## atur skala x jadi setiap jam



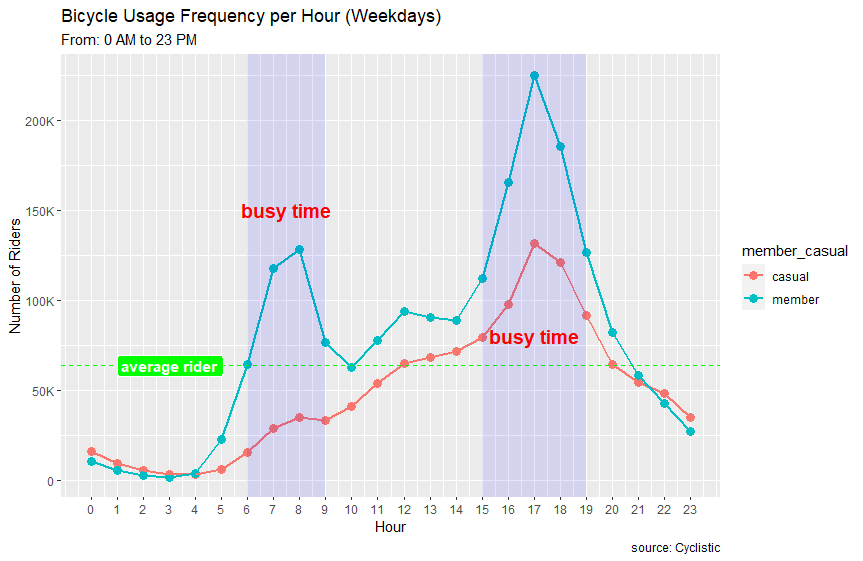
Gambar 6.6 Frekuensi Penggunaan Sepeda Setiap Jam pada Akhir Pekan

Di akhir pekan terjadi lonjakan penggunaan sepeda pada pukul 10.00 pagi hingga 19.00 sore. Pengendara casual paling banyak melakukan perjalanan menggunakan sepeda di jam sibuk tersebut. Jam sibuk tersebut merupakan waktu terbaik untuk menghabiskan waktu untuk bersantai di akhir pekan.

## grafik 7  
## menganalisis frekuensi penggunaan sepeda setiap jamnya pada hari kerja  
weekday\_avg <- tripdata2 %>% filter(day\_of\_week != "Sunday" & day\_of\_week != "Saturday") %>%   
 group\_by(member\_casual, start\_hour) %>%  
 summarise(number\_of\_riders = n()) %>%   
 ungroup()

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

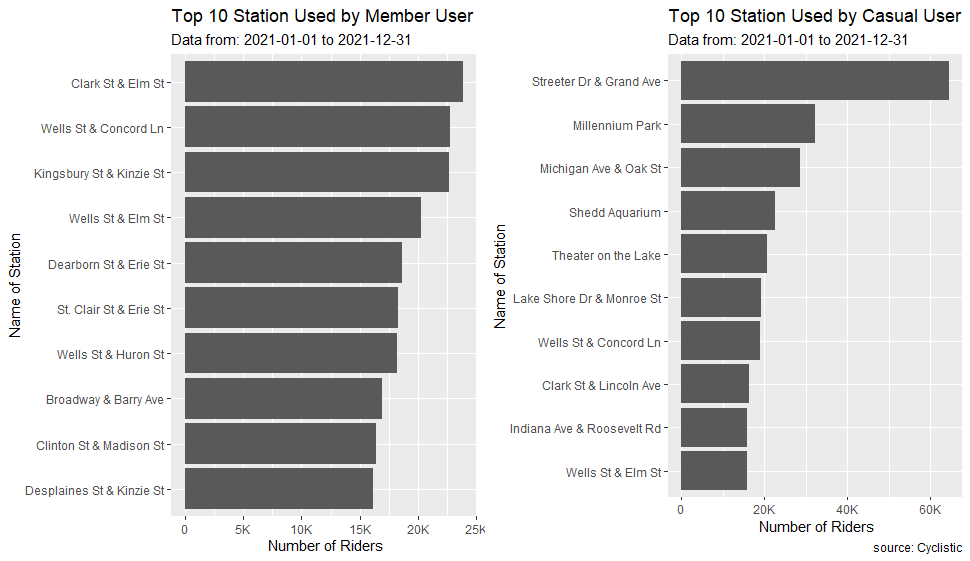
## membuat plot untuk merata-ratakan jumlah pengendara sepeda  
weekday\_plot <- mean(weekday\_avg$number\_of\_riders)  
  
ggplot(data = weekday\_avg, mapping = aes(x = start\_hour, y = number\_of\_riders, color = member\_casual, group = member\_casual))+  
 geom\_point(size = 3)+  
 geom\_line(size = 1)+  
 geom\_hline(aes(yintercept = weekday\_plot),colour = "green",linetype = "dashed")+  
 geom\_label(aes(x = 1, y = weekday\_plot+1, label = "average rider", hjust = 0), colour = "white", fill = "green", fontface = "bold")+  
 annotate("rect", xmin = 6, xmax = 9, ymin = -Inf, ymax = Inf, alpha = 0.1, fill = "blue")+  
 annotate("rect", xmin = 15, xmax = 19, ymin = -Inf, ymax = Inf, alpha = 0.1, fill = "blue")+  
 annotate("text", x = 7.5, y = 150000, label = "busy time", size = 5, fontface = "bold", color = "red")+  
 annotate("text", x = 17, y = 80000, label = "busy time", size = 5, fontface = "bold", color = "red")+  
 labs(title = "Bicycle Usage Frequency per Hour (Weekdays)", x = "Hour", y = "Number of Riders",  
 subtitle = paste0("From: ", minhour, " AM to ", maxhour, " PM "),  
 caption = "source: Cyclistic") +  
 scale\_y\_continuous(breaks = c(0, 50000, 100000, 150000, 200000), labels = c("0", "50K", "100K", "150K", "200K")) +  
 scale\_x\_continuous( breaks = c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23))



Gambar 6.7 Frekuensi Penggunaan Sepeda Setiap Jam pada Hari Kerja

Di hari kerja, pengendara member mendominasi berkendara pada jam sibuk kerja yaitu pukul 6.00-9.00 pagi dan 15.00-19.00 sore yang merupakan waktu untuk berangkat dan pulang kerja. Sekitar pukul 12.00 siang mengalami sedikit lonjakan, waktu tersebut merupakan waktu untuk makan siang. Kemungkinan pengendara member menggunakan sepeda untuk pergi membeli makan siang. Sedangkan pengendara casual terjadi lonjakan berkendara pada pukul 15.00-19.00 sore. Kemungkinan pengendara casual menggunakan sepeda pada jam tersebut sebagai alternatif tranportasi karena kemacetan saat jam pulang kerja ataupun sekedar menggunakan sepeda hanya untuk bersantai.

## grafik 8  
## 10 stasiun yang paling sering digunakan pengendara jenis member   
top\_station\_member <- tripdata2 %>% filter(member\_casual == "member") %>%   
 group\_by(start\_station\_name) %>%  
 summarise(number\_of\_riders = n()) %>%   
 arrange(-number\_of\_riders) %>%   
 head(n = 10) %>%   
 ungroup()  
  
graph\_tsm <- ggplot(data = top\_station\_member)+  
 geom\_col(mapping = aes(x = number\_of\_riders, y = reorder(start\_station\_name, number\_of\_riders)))+  
 labs(title = "Top 10 Station Used by Member User", x = " Number of Riders", y = "Name of Station",  
 subtitle = paste0("Data from: ", mindate, " to ", maxdate))+  
 scale\_x\_continuous(breaks = c(0, 5000, 10000, 15000, 20000, 25000), labels = c("0", "5K", "10K", "15K", "20K", "25K"))  
  
## 10 stasiun yang paling sering digunakan pengendara jenis casual  
top\_station\_casual <- tripdata2 %>% filter(member\_casual == "casual") %>%   
 group\_by(start\_station\_name) %>%  
 summarise(number\_of\_riders = n()) %>%   
 arrange(-number\_of\_riders) %>%   
 head(n = 10) %>%   
 ungroup()  
  
graph\_tsc <- ggplot(data = top\_station\_casual)+  
 geom\_col(mapping = aes(x = number\_of\_riders, y = reorder(start\_station\_name, number\_of\_riders)))+  
 labs(title = "Top 10 Station Used by Casual User", x = " Number of Riders", y = "Name of Station",  
 subtitle = paste0("Data from: ", mindate, " to ", maxdate),  
 caption = "source: Cyclistic")+  
 scale\_x\_continuous(breaks = c(0, 20000, 40000, 60000), labels = c("0", "20K", "40K", "60K"))  
  
## gabungkan 2 plot grafik kedalam satu grid  
grid.arrange(graph\_tsm, graph\_tsc, ncol=2)



Gambar 6.8 10 Stasiun Tersibuk

Stasiun sepeda yang digunakan pengendara casual cenderung terletak di lokasi rekreasi sekitar pantai. Sementara lokasi stasiun sepeda yang digunakan oleh pengendara member cenderung berada di kawasan pertokoan dan perkantoran kecenderungan penggunaan setiap pengendara berbeda, pengendara casual lebih menggunakan sepeda untuk bersantai di tempat rekreasi sementara pengendara member menggunakan sepeda sebagai alat transportasi menuju tempat kerja atau berbelanja. Terdapat 2 stasiun yang paling sibuk yaitu: Wells St & Concord Ln dan Wells St & Elm St, karena lalu lintas tinggi pengendara member dan casual melakukan keberangkatan sepeda dari stasiun tersebut.

## Step 5 - Membagikan

beberapa temuan hasil dari hasil menganalisis data:

1. Sebagaian besar pengendara member menggunakan sepeda untuk bekerja sementara pengendara casual untuk rekreasi,
2. Jenis sepeda yang paling banyak digunakan dari kedua pengendara, yaitu: classic > electric > docked,
3. Penggunaan sepeda paling banyak terjadi di bulan maret hingga agustus,
4. Penggunaan sepeda pada hari kerja cenderung di waktu berangkat dan pulang kerja dan di akhir pekan cenderung di waktu 10.00 hingga 19.00, dan
5. Stasiun yang terletak di sekitar pantai banyak diakses oleh pengendara casual, sementara stasiun yang terletak dikawasan pertokoan dan perkantoran banyak diakses pengendara member.

## Step 6 - Bertindak

1. **Strategi untuk mendorong pengendara casual membeli keangotaan tahunan**

* *Evaluasi untuk pengendara member*
* Keanggotaan tahunan harus terlihat hemat biaya untuk lebih meyakinkan pengendara casual membeli keanggotaan tahunan, dan
* Membuat akses prioritas penggunaan sepeda untuk pengendara member saat di jam sibuk.
* *Evaluasi untuk pengendara casual*
* Tiket sekali jalan dan tiket satu hari penuh terlalu fleksibel dan tampak tidak ada manfaat nyata untuk memiliki keanggotaan tahunan,
* Menaikkan harga harga tiket sekali jalan dan sehari penuh, dan
* Memberikan penawaran membership mingguan atau bulanan yang lebih murah dibandingkan total pembelian tiket perhari selama satu minggu atau satu bulan, mungkin saja pelanggan tidak ingin terikat waktu yang panjang dan biaya yang lebih mahal untuk membeli keanggotaan tahunan.

1. **Strategi promosi kampanye**

* Kampanye pemasaran secara online memalui social media dapat dimulai pada bulan Maret saat mulai memasuki musim semi untuk memperkenalkan promosi keanggotaan tahunan saat pelanggan kembali lagi beraktifitas menggunakan sepeda,
* Perkenalkan kesepakatan promosi keanggotaan mingguan dan bulanan untuk pengendara casual di akhir pekan, diwaktu tersebut pengendara casual dominan menggunakan sepeda,
* Demi meningkatnya minat terhadap sepeda listrik pada kedua jenis pengendara perlu menyertakan promosi dan penawaran khusus terkait penggunaan sepeda listrik, serta lebih banyak meletakan sepeda listrik pada tempat-tempat yang digunakan untuk bersantai, dan
* Iklan luar ruang harus dikonsentrasikan di stasiun awal paling populer dan di stasiun dengan lalu lintas tinggi di Wells St & Concord Ln dan Wells St & Elm St.