

Cyclistic Bike-share Analysis

Based on Internal Data : Jan 2020 – Dec 2021



Presented by : **Ahmad Ridho**
Last Updated : **January 6th, 2022**

Outline



Case and Data Introduction

Summary of Analysis

Conclusions and Recommendations

Outline



Case and Data Introduction

Summary of Analysis

Conclusions and Recommendations

Case Introduction

Cyclistic, a bike-share company in Chicago, has two types of customer. **Casual riders** who purchase single-ride or full-day passes, and **members** who purchase annual membership.

Cyclistic's finance analysts have concluded that **annual members are much more profitable** than casual riders. Based on this insight, the marketing wants to **design marketing strategies aimed at converting casual riders into members**.

Objective:

Identify how annual members and casual riders differ to support marketing strategies.



Data

The data is made available by Motivate International Inc. which operates the city of Chicago's Divvy bicycle sharing service which is powered by Lyft.

5 million records* from [January until December 2021](#) are used for analysis.

The data contain records of each trip, not records of usage of each customer. So that [analysis is limited on how casual riders and members differ based on annual recorded trips](#) where multiple trips by the same user are treated as different data, not as one.

R programming is used for data processing due to its ability in handling huge data effectively.



*after processing

Outline

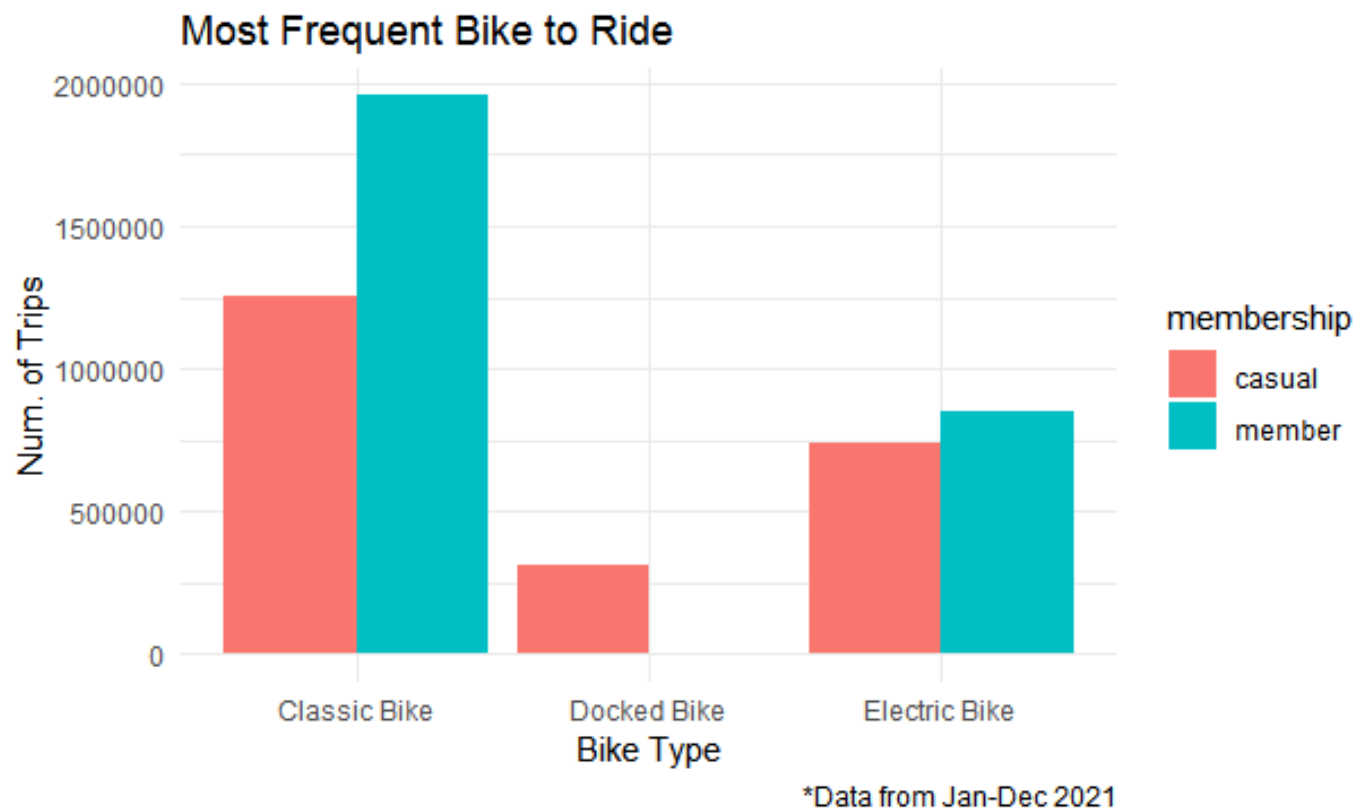


Case and Data Introduction

Summary of Analysis

Conclusions and Recommendations

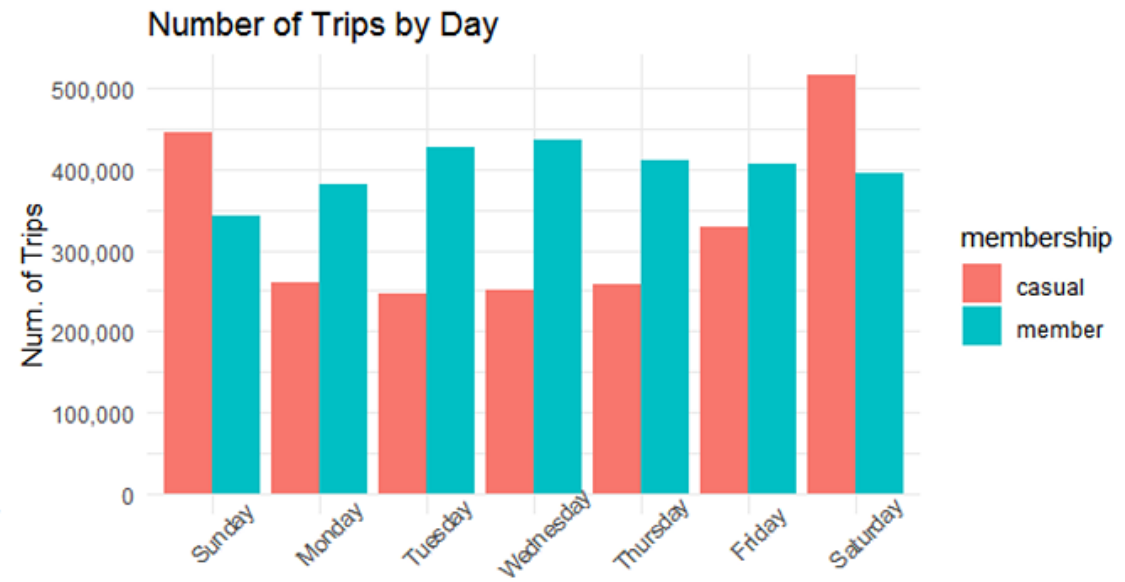
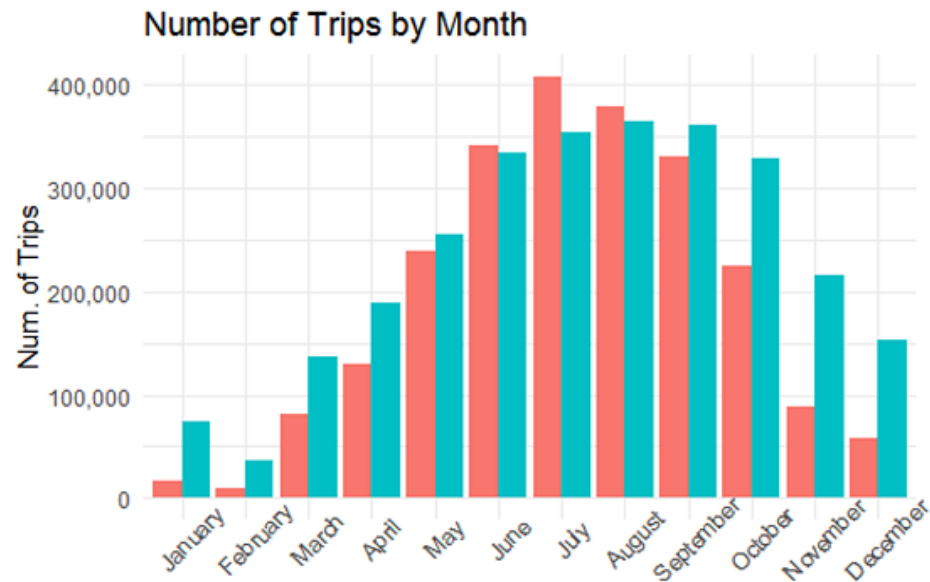
Trips by Bike



- The classic bike is more often to use for trips than docked bike and electric bike, both by casual riders and members.
- With a total of more than 5 million trips in 2021, 3,200,000 trips were done using classic bike with 40% of them being casual riders.

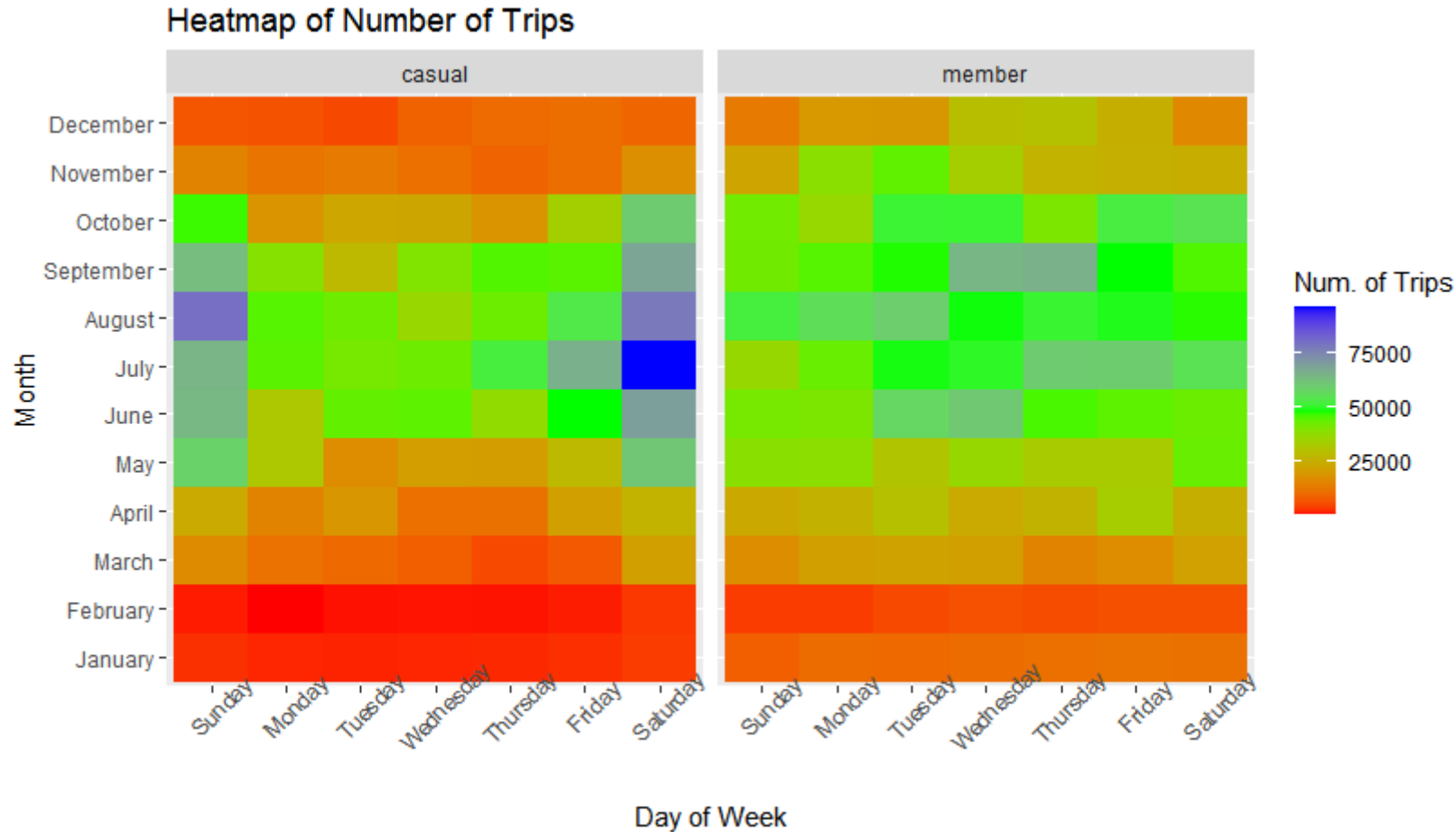
Riding Period

- Cyclistic usage for both casual and member riders is **seasonal** with high demand in **June, July, August, and September**.
- The number of trips done by **casual riders on weekends is more than on weekdays** while it is steady for members throughout the week.



*Data from Jan-Dec 2021

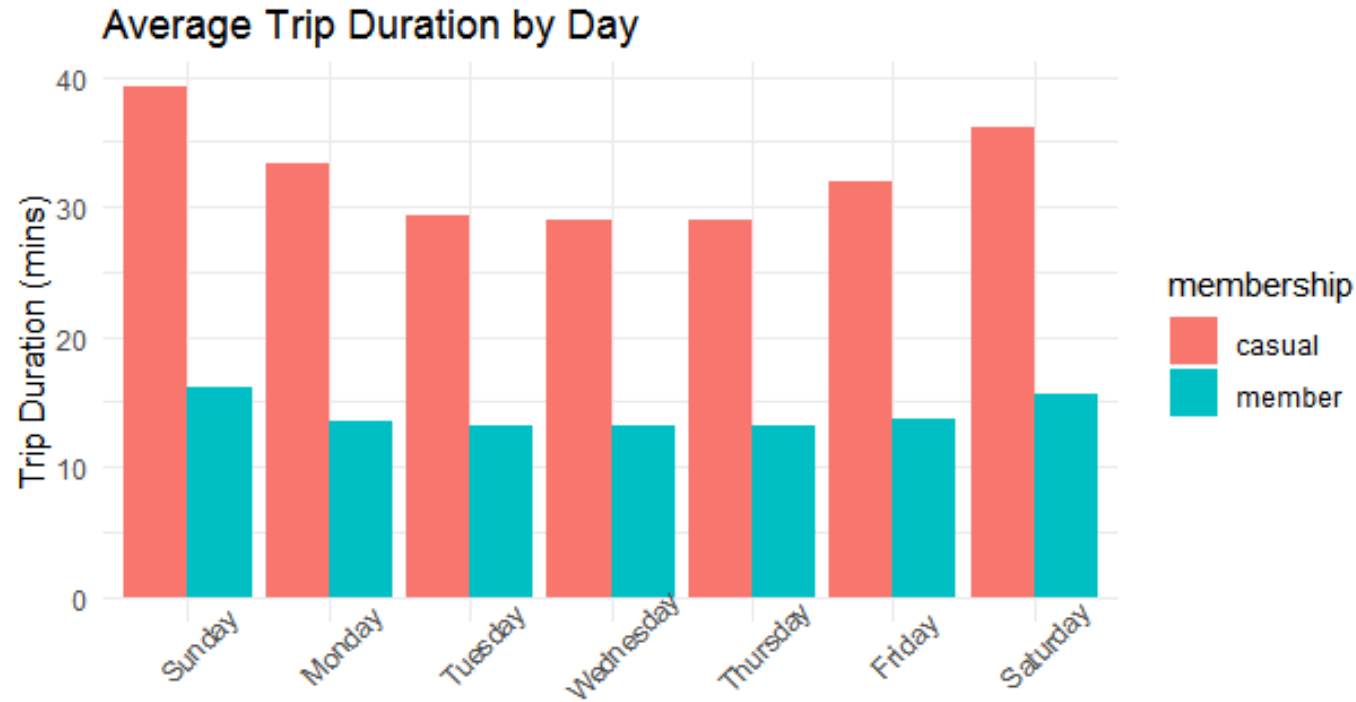
Riding Heatmap



*Data from Jan-Dec 2021

- The number of casual riders' trips on **Saturday and Sunday** were consistently higher than the other days throughout the year, especially in **June, July, August, and September**.
- The number of trips of members was **slightly constant** through the week with more trips in **June to October** compared to other months.

Average Trip Duration



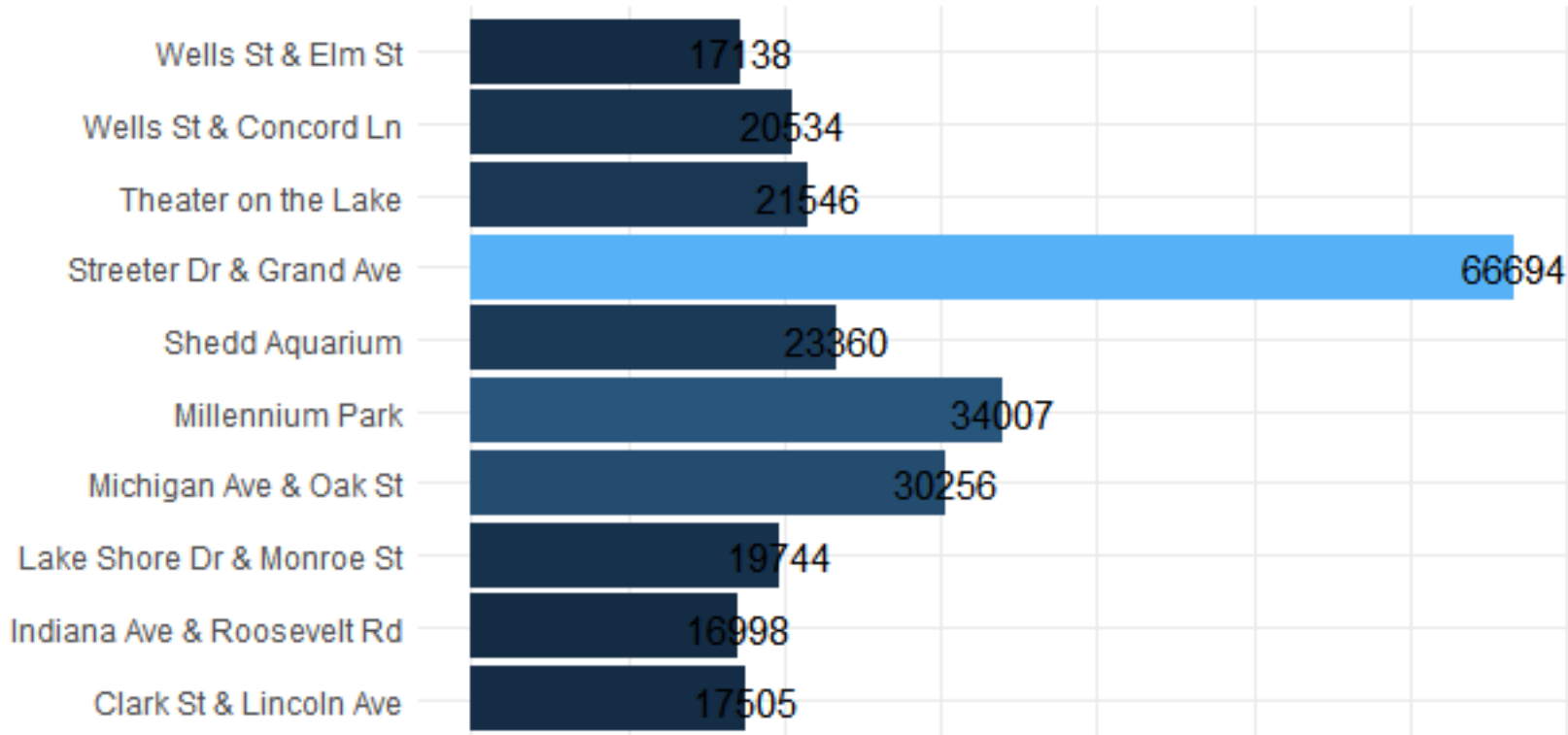
*Data from Jan-Dec 2021

- Although the number of trips by casual riders was fewer than members was, it was longer in duration.
- The average trip duration for casual riders is 33.5 minutes while it is only 14 minutes for members.

Casual Riders Stations

Popular Stations of Casual Riders

Measured by total trips started at the stations.



*Data from Jan-Dec 2021

- The [Streeter Dr & Grand Ave](#), [Millennium Park](#), and [Michigan Ave & Oak St](#) are the three most visited stations where casual riders start the trip from.
- These stations might be a [good consideration for in-person marketing](#).

Outline



Case and Data Introduction

Summary of Analysis

Conclusions and Recommendations

Conclusions



The **Classic bike** is more often to use for **trips** both by casual riders or members.



The bike-share program is used **seasonally with high demand from June to September** for both casual riders and members.



Total trips by casual riders are higher on weekends and steady during weekdays, this is different from members that active on all days of the week. This indicates that **casual riders mostly use bicyclic for recreational purposes.**



Although annual trip by casual riders is fewer than members, their **weekly riding is longer than members.**



Streeter Dr & Grand Ave, Millenium Park, and Michigan Ave & Oak St are the three most stations where casual riders start the trip from.

Recommendations

The recommendation for marketing strategy are:

- **Design more flexible packages** such as monthly or semester subscriptions. These packages give flexibility to those who are not willing for an annual subscription.
- Consider **offering promos or coupons for weekends or seasonal events**. This might be able to increase customer experience at the same time attract new customers. By increasing customer experience, casual riders will be easier to encourage for subscriptions.
- **Design more intimate marketing material for casual riders**. Casual riders already have their trip histories that show their riding habit. Campaign that aligns with their habit will encourage them for subscriptions.
- Besides online marketing, **in-person marketing is worth trying**, especially around popular sites which often visited by casual riders as well as **the stations they usually start the trip from**.

Thank You

Ahmad Ridho

ahmdrdo@gmail.com

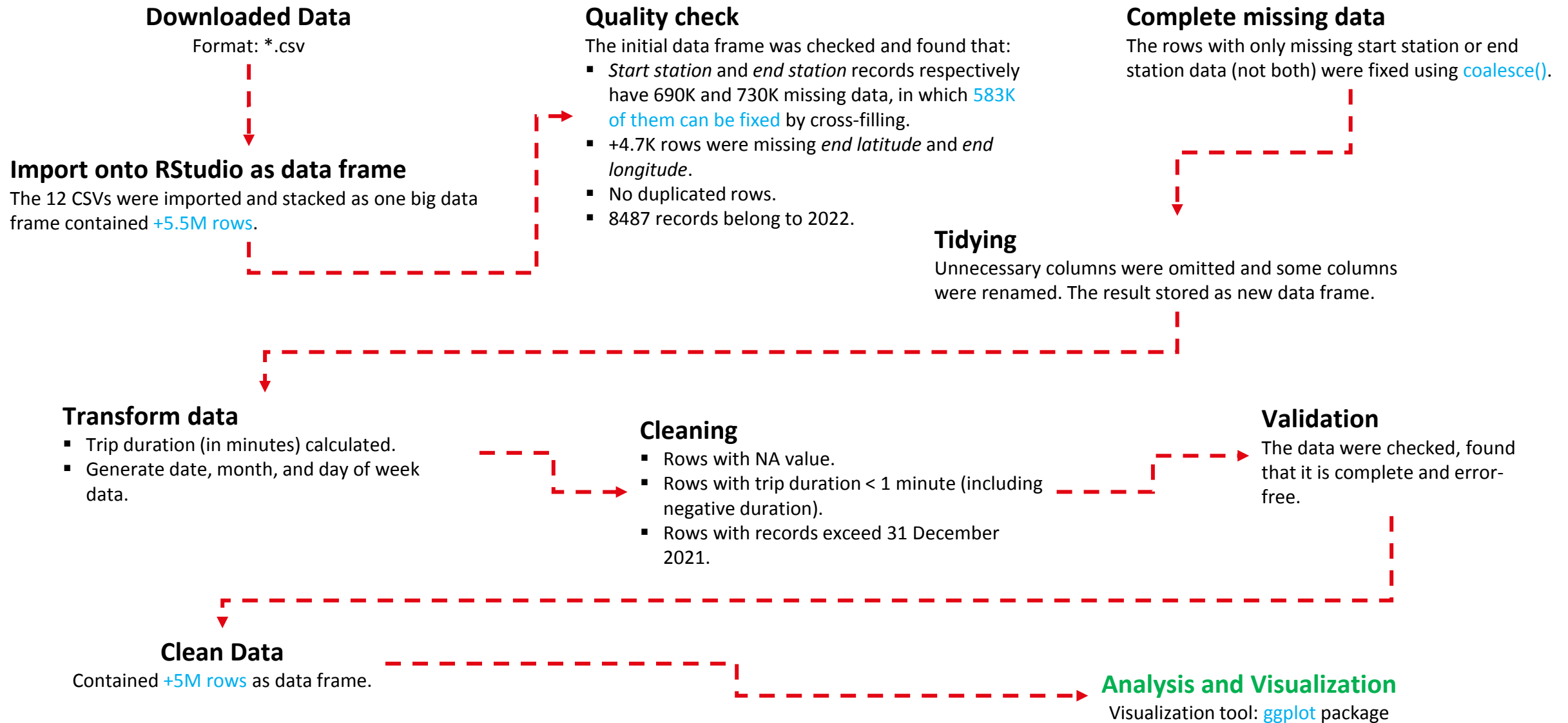
<https://github.com/ahmdrdo>



Appendix



Data Processing Workflow



R Script

```
# -----  
# Author: Ahmad Ridho  
# Email: ahmdrdo@gmail.com  
# -----
```

```
setwd("~/Collection/Portfolio/Cyclistic Bike-sharing")
```

```
# Load packages.  
library(tidyverse)  
library(dplyr)  
library(lubridate)  
library(ggplot2)  
library(skimr)
```

```
# =====  
# IMPORT DATA  
# =====
```

```
data_dir <- "~/Collection/Portfolio/Cyclistic Bike-sharing/Used Dataset"
```

```
# Stack individual files as one.  
trips_raw <- list.files(path = data_dir, pattern = ".csv", full.names = TRUE) %>%  
  lapply(read_csv) %>%  
  bind_rows()
```

R Script (cont...)

```
# =====  
# COMPLETE MISSING DATA  
# =====  
  
trips_clean <- trips_raw  
trips_clean <- trips_clean %>% # Fill missing station name.  
  mutate(start_station_name = coalesce(start_station_name, end_station_name),  
         end_station_name = coalesce(end_station_name, start_station_name))  
  
# =====  
# TIDYING  
# =====  
  
# Remove unnecessary columns.  
trips_clean <- select(trips_clean, -c(start_station_id, end_station_id,  
                                     start_lat, start_lng,  
                                     end_lat, end_lng))  
  
# Rename columns.  
trips_clean <- rename(trips_clean, trip_id = ride_id,  
                     bike_type = rideable_type,  
                     membership = member_casual)
```

R Script (cont...)

```
# =====  
# TRANSFORM DATA  
# =====  
  
# Calculate trip duration (result in minutes).  
trips_clean$trip_duration <- difftime(trips_clean$ended_at, trips_clean$started_at, units = "mins")  
# Reformat trip duration.  
trips_clean$trip_duration <- as.numeric(as.character(trips_clean$trip_duration))  
  
# Generate Date, Month, Day of Week, and Hour.  
trips_clean$date <- as.Date(trips_clean$started_at)  
trips_clean$month <- format(trips_clean$date, "%B")  
trips_clean$day_of_week <- format(trips_clean$date, "%A")  
  
# Fix the order of ordinal data.  
trips_clean$month <- ordered(trips_clean$month, levels = c("January", "February", "March", "April", "May", "June", "July", "August", "September",  
"October", "November", "December"))  
trips_clean$day_of_week <- ordered(trips_clean$day_of_week, levels = c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
```

R Script (cont...)

```
# =====  
# CLEANING  
# =====  
  
# Remove NA and trip duration < 1 min (inc. minus duration).  
trips_clean <- trips_clean %>%  
  na.omit() %>%  
  filter(trip_duration >= 1)  
  
# Filter data only for 2021.  
trips_clean <- filter(trips_clean, date >= "2021-01-01" & date <= "2021-12-31")  
  
# =====  
# VALIDATION  
# =====  
  
skim_without_charts(trips_clean)
```

R Script (cont...)

```
# =====  
# DATA VISUALIZATION  
# =====
```

```
# Bike used by rider type.  
ggplot(data = trips_clean, aes(x = bike_type, fill = membership)) +  
  geom_bar(position = "dodge") +  
  theme_minimal() +  
  labs(title = "Most Frequent Bike to Ride", caption = "*Data from Jan-Dec 2021", x = "Bike Type", y = "Num. of Trips") +  
  scale_x_discrete(labels = c("Classic Bike", "Docked Bike", "Electric Bike"))
```

```
# Number of trips by month.  
ggplot(data = trips_clean, aes(x = month, fill = membership)) +  
  geom_bar(position = "dodge") +  
  theme_minimal() +  
  labs(title = "Number of Trips by Month", caption = "*Data from Jan-Dec 2021", y = "Num. of Trips") +  
  theme(axis.title.x = element_blank(), axis.text.x = element_text(angle = 45)) +  
  scale_y_continuous(labels = scales::comma)
```

```
# Number of trips by day.  
ggplot(data = trips_clean, aes(x = day_of_week, fill = membership)) +  
  geom_bar(position = "dodge") +  
  theme_minimal() +  
  labs(title = "Number of Trips by Day", caption = "*Data from Jan-Dec 2021", y = "Num. of Trips") +  
  theme(axis.title.x = element_blank(), axis.text.x = element_text(angle = 45)) +  
  scale_y_continuous(labels = scales::comma)
```

R Script (cont...)

```
# Heatmap number of Trips
trips_clean %>%
  count(membership, month, day_of_week, name = "count") %>%
  ggplot(aes(x = day_of_week, y = month, fill = count)) +
  geom_tile() +
  facet_wrap(~membership) +
  scale_fill_gradientn(name = "Num. of Trips", colors = rainbow(3)) +
  labs(title = "Heatmap of Number of Trips", caption = "*Data from Jan-Dec 2021",
       x = "Day of Week", y = "Month") +
  theme(axis.text.x = element_text(angle = 45))
```

```
# Average trip duration.
trips_clean %>%
  group_by(membership, day_of_week) %>%
  summarise(avg_duration = mean(trip_duration)) %>%
  ggplot(aes(x = day_of_week, y = avg_duration, fill = membership)) +
  geom_col(position = "dodge") +
  theme_minimal() +
  labs(title = "Average Trip Duration by Day", caption = "*Data from Jan-Dec 2021",
       x = "Day of week", y = "Trip Duration (mins)") +
  theme(axis.title.x = element_blank(), axis.text.x = element_text(angle = 45))
```

R Script (cont...)

```
# Top stations of casual riders.
trips_clean %>%
  filter(membership == "casual") %>%
  count(start_station_name, name = "count") %>%
  top_n(n = 10, wt = count) %>%
  ggplot(aes(x = count, y = start_station_name, fill = count)) +
  geom_col(show.legend = FALSE) +
  theme_minimal() +
  labs(title = "Popular Stations of Casual Riders", subtitle = "Measured by total trips started at the stations.", caption = "*Data from Jan-Dec 2021",
        x = "Num. of Trips", y = "Station Name") +
  theme(axis.text.x = element_blank(), axis.title = element_blank()) +
  geom_text(aes(label = count))
```


Data Scheme*

Field Name	Description
ride_id	Unique trip or ride identification number.
rideable_type	Type of bike used.
started_at	Date-time (UTC) when the trip started.
ended_at	Date-time (UTC) when trip ended.
start_station_name	Station where trip started
start_station_id	Station's ID where trip started
end_station_name	Station where trip ended
end_station_id	Station's ID where trip ended
start_lat	Latitude of start point
start_lng	Longitude of start point
end_lat	Latitude of end point
end_lng	Longitude of end point
member_casual	Membership type. Casual — single-ride or full-day passes, Member — annual membership.

*Initial data scheme before processing