

Analysis of Divvy Trips 2019 Q1

Hanim Sofiah

2024-02-28

Introduction

This document presents an analysis of Divvy Trips data for the first quarter of 2019. The analysis includes descriptive statistics, average ride lengths by user type and day of the week, and the number of rides by day of the week.

Setup Environment

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(lubridate)

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

library(readr)
library(ggplot2)
```

Load Data

```
trips_data <- read_csv("Divvy_Trips_2019_Q1.csv")

## Rows: 365069 Columns: 14
## -- Column specification -----
## Delimiter: ","
## chr  (6): start_time, end_time, from_station_name, to_station_name, usertype...
## dbl  (6): trip_id, bikeid, from_station_id, to_station_id, birthyear, day_of...
## num  (1): tripduration
```

```
## time (1): ride_length
##
## 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.
```

Descriptive Analysis

```
summary(trips_data)
```

```
##      trip_id      start_time      end_time      bikeid
## Min.   :21742443 Length:365069   Length:365069   Min.    :    1
## 1st Qu.:21848765 Class :character Class :character 1st Qu.:1777
## Median :21961829 Mode  :character Mode  :character Median :3489
## Mean   :21960872                      Mean   :3429
## 3rd Qu.:22071823                      3rd Qu.:5157
## Max.   :22178528                      Max.   :6471
##
##      tripduration  from_station_id from_station_name to_station_id
## Min.   :      61   Min.   :  2.0   Length:365069   Min.   :  2.0
## 1st Qu.:     326   1st Qu.: 76.0   Class :character 1st Qu.: 76.0
## Median :     524   Median :170.0   Mode  :character Median :168.0
## Mean   :    1016   Mean   :198.1                      Mean   :198.6
## 3rd Qu.:     866   3rd Qu.:287.0                      3rd Qu.:287.0
## Max.   :10628400   Max.   :665.0                      Max.   :665.0
##
##      to_station_name  usertype      gender      birthyear
## Length:365069       Length:365069   Length:365069   Min.   :1900
## Class :character    Class :character   Class :character 1st Qu.:1975
## Mode  :character    Mode  :character   Mode  :character Median :1985
##                                     Mean   :1982
##                                     3rd Qu.:1990
##                                     Max.   :2003
##                                     NA's   :18023
##
##      ride_length      day_of_week
## Length:365069       Min.   :1.000
## Class1:hms          1st Qu.:3.000
## Class2:difftime     Median :4.000
## Mode :numeric       Mean   :4.145
##                                     3rd Qu.:6.000
##                                     Max.   :7.000
##
```

Mean, Max and Mode

```
mean_ride_length <- mean(trips_data$ride_length, na.rm = TRUE)
max_ride_length <- max(trips_data$ride_length, na.rm = TRUE)

getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
mode_day_of_week <- getmode(trips_data$day_of_week)
```

Average Ride Length by User Type and Day Of Week

```
avg_ride_length_by_type <- trips_data %>%
  group_by(usertype) %>%
  summarise(average_ride_length = mean(ride_length, na.rm = TRUE))

avg_ride_length_by_day_and_type <- trips_data %>%
  group_by(day_of_week, usertype) %>%
  summarise(average_ride_length = mean(ride_length, na.rm = TRUE))

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

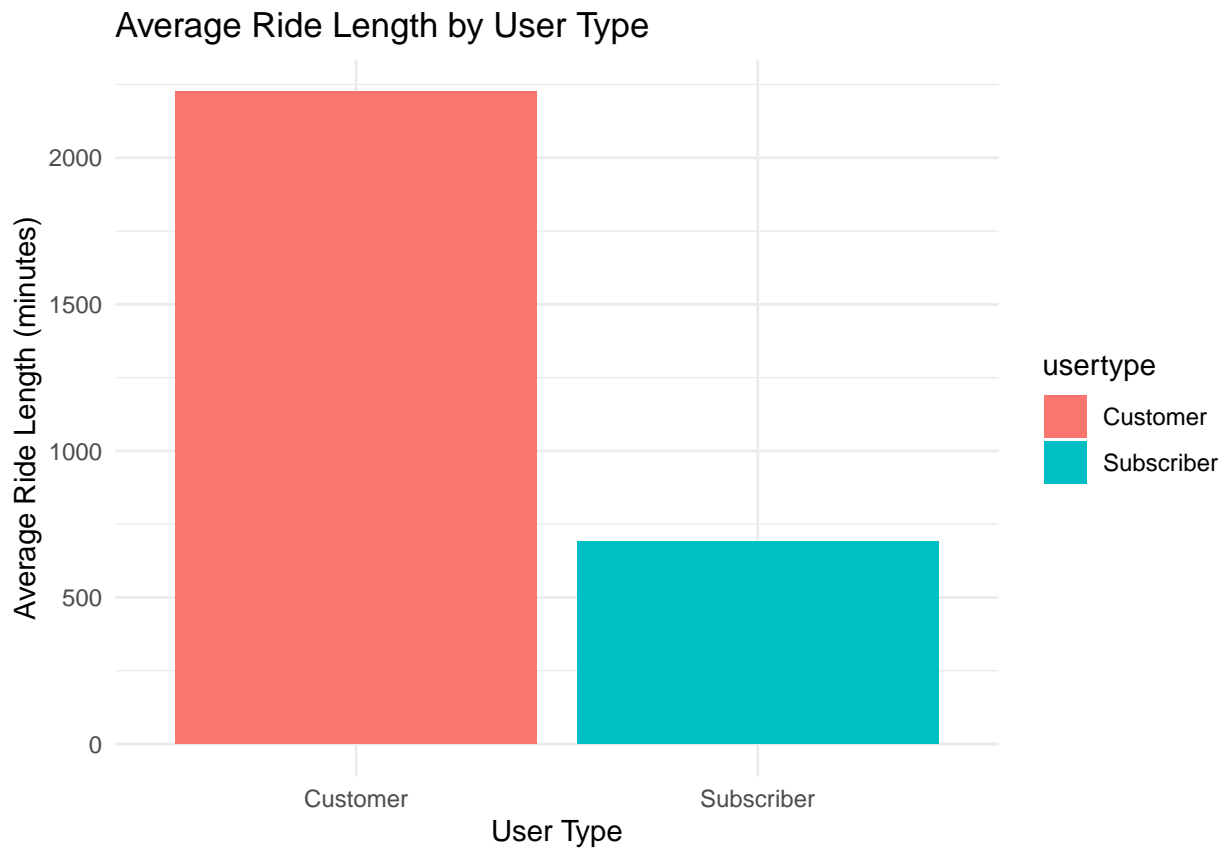
rides_by_day <- trips_data %>%
  group_by(day_of_week) %>%
  summarise(number_of_rides = n())
```

Visualization

Average Ride Length by Type of User

```
ggplot(avg_ride_length_by_type, aes(x = usertype, y = average_ride_length, fill = usertype)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Ride Length by User Type", x = "User Type", y = "Average Ride Length (minutes)") +
  theme_minimal()

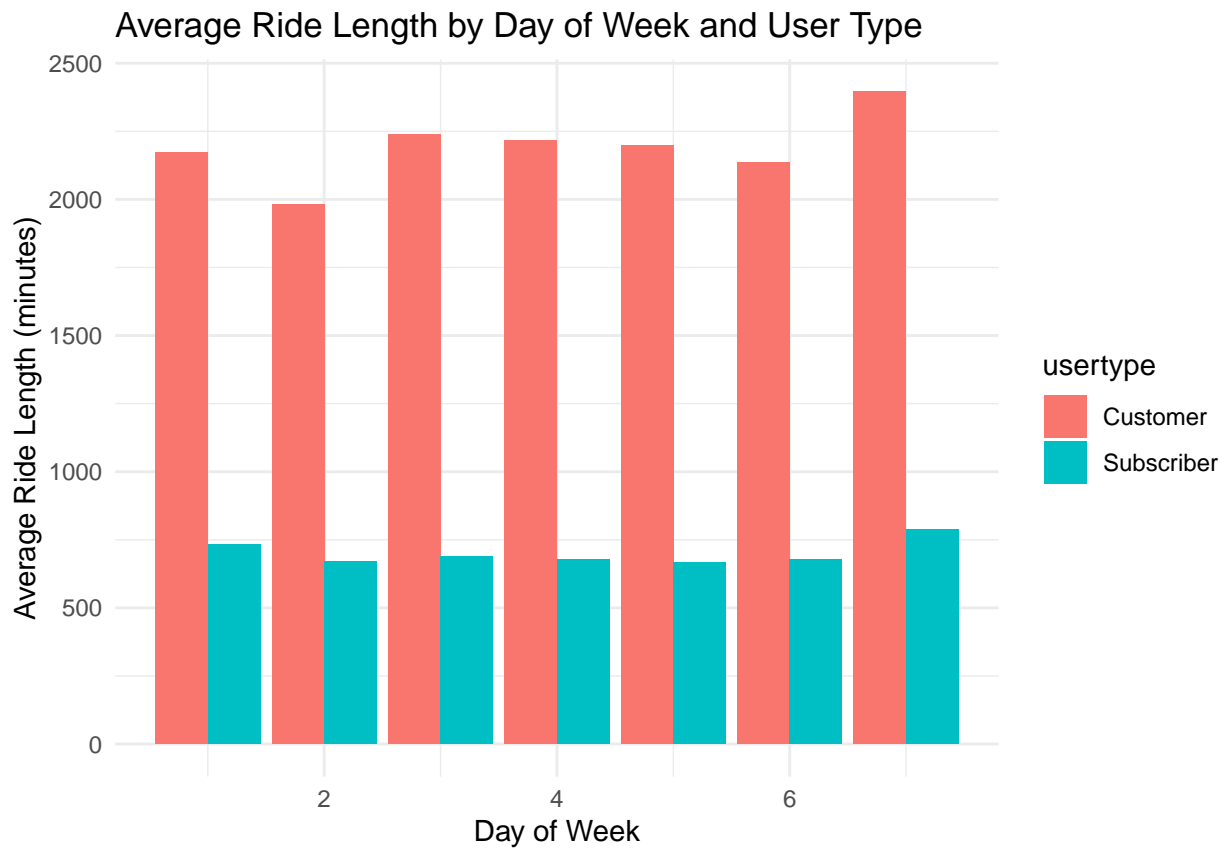
## Don't know how to automatically pick scale for object of type <difftime>.
## Defaulting to continuous.
```



Average Ride Length by Day and Type

```
ggplot(avg_ride_length_by_day_and_type, aes(x = day_of_week, y = average_ride_length, fill = usertype))  
  geom_bar(stat = "identity", position = "dodge") +  
  labs(title = "Average Ride Length by Day of Week and User Type", x = "Day of Week", y = "Average Ride Length")  
  theme_minimal()
```

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



Number of Rides by Day of the Week

```
ggplot(rides_by_day, aes(x = day_of_week, y = number_of_rides)) +
  geom_line(group=1) +
  labs(title = "Number of Rides by Day of the Week", x = "Day of Week", y = "Number of Rides") +
  theme_minimal()
```

