Bikeshare

Ayomide Onayemi

2024-01-16

PROJECT OVERVIEW

This case study is the Google Data Analytics Capstone Project - Case Study 1. The analysis is structured around the 6 steps of data analysis.

INTRODUCTION

ABOUT THE COMPANY

Cyclistic, the bike-share program launched in 2016, has grown to 5,824 bikes across 692 Chicago stations. The marketing strategy focused on broad consumer segments with flexible pricing plans. Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders, prompting a shift in strategy. The goal is to convert casual riders into annual members, rather than creating a marketing campaign that targets all-new customers. To achieve this, historical bike trip data will be analyzed to understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics.

THE GOAL OF THIS CASE STUDY

ASK

Three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

THE BUSINESS TASK

To analyze historical bike trip data and understand the differences between annual members and casual riders. Identified insights and trends from the analysis will inform the design of a new marketing strategy for converting casual riders into annual members. The new strategy seeks to enhance the number of annual members and drive future growth for Cyclistic.

Key Stakeholders

Lily Moreno: The director of marketing.

Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.

Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

PREPARE

Data Source Used

Cyclistic's historical trip data : The datasets have a different name because Cyclistic is a fictional company

The data has been made available by Motivate International Inc. under this license

This is public data, but note that due to data-privacy issues riders PII (personally identifiable information) are prohibited from being used.

Install and load packages

```
#these packages are in tidyverse : dplyr, readr, forcats, stringr, ggplot2, t
ibble, lubridate, tidyr and purrr

library("tidyverse") #for data wrangling and visualizations.
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.
0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflict
s() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library("janitor") #for examining and cleaning dirty data
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library("skimr") #for providing summary statistics about variables in data fr
ames
```

Importing and loading data

```r
#Import and load the dataset for 2023

df1 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202301-divvy-tripdata.csv")
df2 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202302-divvy-tripdata.csv")
df3 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202303-divvy-tripdata.csv")
df4 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202304-divvy-tripdata.csv")
df5 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202305-divvy-tripdata.csv")
df6 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202306-divvy-tripdata.csv")
df7 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202307-divvy-tripdata.csv")
df8 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202308-divvy-tripdata.csv")
df9 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyc
listic Dataset/2023/202309-divvy-tripdata.csv")
df10 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cy
clistic Dataset/2023/202310-divvy-tripdata.csv")
df11 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cy
clistic Dataset/2023/202311-divvy-tripdata.csv")
df12 <- read.csv("C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cy
clistic Dataset/2023/202312-divvy-tripdata.csv")

#merge into one data frame

all_trips <- rbind(df1,df2,df3,df4,df5,df6,df7,df8,df9,df10,df11,df12)

#check the column names

colnames(all_trips)

##  [1] "ride_id"            "rideable_type"      "started_at"
##  [4] "ended_at"           "start_station_name" "start_station_id"
##  [7] "end_station_name"   "end_station_id"     "start_lat"
## [10] "start_lng"          "end_lat"            "end_lng"
## [13] "member_casual"
```

Renaming of columns

```r
all_trips <- rename(all_trips,
                    trip_id = ride_id,
                    bike_type = rideable_type,
                    start_time = started_at,
                    end_time = ended_at,
                    start_station = start_station_name,
                    ss_id = start_station_id,
```

```
                    end_station = end_station_name,
                    es_id = end_station_id,
                    user_type = member_casual)

#statistical summary of the data

summary(all_trips)
```

```
##    trip_id            bike_type           start_time          end_time
##  Length:5719877     Length:5719877     Length:5719877     Length:5719877
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station         ss_id            end_station           es_id
##  Length:5719877     Length:5719877     Length:5719877     Length:5719877
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.63   Min.   :-87.94   Min.   : 0.00   Min.   :-88.16
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
##  Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.07   Max.   :-87.46   Max.   :42.18   Max.   :  0.00
##                                   NA's   :6990    NA's   :6990
##   user_type
##  Length:5719877
##  Class :character
##  Mode  :character
##
##
##
##
```

PROCESS

Drop null, NA values and duplicates

```
#drop null and NA values

all_trips <- all_trips[!apply(is.na(all_trips) | all_trips == "", 1, any), ]

#drop duplicates
```

```
all_trips <- distinct(all_trips)

#statistical summary of the data

summary(all_trips)
```

```
##    trip_id           bike_type          start_time          end_time
##  Length:4331707     Length:4331707     Length:4331707     Length:4331707
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  start_station         ss_id           end_station           es_id
##  Length:4331707     Length:4331707     Length:4331707     Length:4331707
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.65   Min.   :-87.84   Min.   : 0.00   Min.   :-87.84
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
##  Mean   :41.90   Mean   :-87.64   Mean   :41.90   Mean   :-87.64
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :42.06   Max.   :-87.53   Max.   :42.06   Max.   :  0.00
##   user_type
##  Length:4331707
##  Class :character
##  Mode  :character
##
##
##
```

change start_time and end_time from character to date and time data type

```
all_trips$start_time= ymd_hms(all_trips$start_time)
all_trips$end_time= ymd_hms(all_trips$end_time)
```

Calculate the length of each trip in minutes

```
#calculate the length of each trip in minutes

all_trips$trip_length <- as.numeric(difftime(all_trips$end_time, all_trips$st
art_time, units = "mins"))

# Display the first few rows to verify the new column
```

```r
head(all_trips)
```

```
##            trip_id     bike_type          start_time            end_time
## 1 F96D5A74A3E41399 electric_bike 2023-01-21 20:05:42 2023-01-21 20:16:33
## 2 13CB7EB698CEDB88  classic_bike 2023-01-10 15:37:36 2023-01-10 15:46:05
## 3 BD88A2E670661CE5 electric_bike 2023-01-02 07:51:57 2023-01-02 08:05:11
## 4 C90792D034FED968  classic_bike 2023-01-22 10:52:58 2023-01-22 11:01:44
## 5 3397017529188E8A  classic_bike 2023-01-12 13:58:01 2023-01-12 14:13:20
## 6 58E68156DAE3E311 electric_bike 2023-01-31 07:18:03 2023-01-31 07:21:16
##                   start_station        ss_id                    end_statio
n
## 1   Lincoln Ave & Fullerton Ave TA1309000058      Hampden Ct & Diversey Av
e
## 2         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th S
t
## 3        Western Ave & Lunt Ave       RP-005 Valli Produce - Evanston Plaz
a
## 4         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th S
t
## 5         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th S
t
## 6 Lakeview Ave & Fullerton Pkwy TA1309000019      Hampden Ct & Diversey Av
e
##          es_id start_lat start_lng  end_lat   end_lng user_type trip_lengt
h
## 1     202480.0  41.92407 -87.64628 41.93000 -87.64000    member   10.85000
0
## 2 TA1308000002  41.79957 -87.59475 41.80983 -87.59938    member    8.48333
3
## 3          599  42.00857 -87.69048 42.03974 -87.69941    casual   13.23333
3
## 4 TA1308000002  41.79957 -87.59475 41.80983 -87.59938    member    8.76666
7
## 5 TA1308000002  41.79957 -87.59475 41.80983 -87.59938    member   15.31666
7
## 6     202480.0  41.92607 -87.63886 41.93000 -87.64000    member    3.21666
7
```

```r
#statistical summary of the data
```

```r
summary(all_trips)
```

```
##    trip_id           bike_type           start_time
##  Length:4331707     Length:4331707     Min.   :2023-01-01 00:02:06.00
##  Class :character   Class :character   1st Qu.:2023-05-20 13:02:18.00
##  Mode  :character   Mode  :character   Median :2023-07-20 15:12:22.00
##                                        Mean   :2023-07-15 19:09:13.49
##                                        3rd Qu.:2023-09-16 16:19:20.50
##                                        Max.   :2023-12-31 23:58:55.00
```

```
##     end_time                      start_station         ss_id
##  Min.   :2023-01-01 00:07:23.0   Length:4331707     Length:4331707
##  1st Qu.:2023-05-20 13:23:20.5   Class :character   Class :character
##  Median :2023-07-20 15:29:43.0   Mode  :character   Mode  :character
##  Mean   :2023-07-15 19:25:10.5
##  3rd Qu.:2023-09-16 16:39:39.0
##  Max.   :2024-01-01 14:20:23.0
##  end_station           es_id              start_lat       start_lng
##  Length:4331707     Length:4331707     Min.   :41.65   Min.   :-87.84
##  Class :character   Class :character   1st Qu.:41.88   1st Qu.:-87.66
##  Mode  :character   Mode  :character   Median :41.90   Median :-87.64
##                                        Mean   :41.90   Mean   :-87.64
##                                        3rd Qu.:41.93   3rd Qu.:-87.63
##                                        Max.   :42.06   Max.   :-87.53
##     end_lat         end_lng        user_type          trip_length
##  Min.   : 0.00   Min.   :-87.84   Length:4331707     Min.   :  -54.567
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character   1st Qu.:    5.617
##  Median :41.90   Median :-87.64   Mode  :character   Median :    9.800
##  Mean   :41.90   Mean   :-87.64                      Mean   :   15.952
##  3rd Qu.:41.93   3rd Qu.:-87.63                      3rd Qu.:   17.483
##  Max.   :42.06   Max.   :  0.00                      Max.   :12136.300
```

remove values that are <= 0 from trip_length

```
all_trips <- all_trips[all_trips$trip_length > 0, ]
```

create new columns 'day_of_week' and 'month

```
all_trips$day_of_week <- weekdays(all_trips$start_time)

all_trips$month <- months(all_trips$start_time)

# Display the first few rows to verify the new column

head(all_trips)
```

```
##            trip_id     bike_type          start_time            end_time
## 1 F96D5A74A3E41399 electric_bike 2023-01-21 20:05:42 2023-01-21 20:16:33
## 2 13CB7EB698CEDB88  classic_bike 2023-01-10 15:37:36 2023-01-10 15:46:05
## 3 BD88A2E670661CE5 electric_bike 2023-01-02 07:51:57 2023-01-02 08:05:11
## 4 C90792D034FED968  classic_bike 2023-01-22 10:52:58 2023-01-22 11:01:44
## 5 3397017529188E8A  classic_bike 2023-01-12 13:58:01 2023-01-12 14:13:20
## 6 58E68156DAE3E311 electric_bike 2023-01-31 07:18:03 2023-01-31 07:21:16
##                   start_station        ss_id                    end_statio
n
## 1   Lincoln Ave & Fullerton Ave TA1309000058      Hampden Ct & Diversey Av
e
## 2         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th S
t
## 3        Western Ave & Lunt Ave       RP-005 Valli Produce - Evanston Plaz
a
```

```
## 4         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th S
t
## 5         Kimbark Ave & 53rd St TA1309000037        Greenwood Ave & 47th S
t
## 6 Lakeview Ave & Fullerton Pkwy TA1309000019      Hampden Ct & Diversey Av
e
##          es_id start_lat start_lng  end_lat   end_lng user_type trip_lengt
h
## 1     202480.0  41.92407 -87.64628 41.93000 -87.64000    member   10.85000
0
## 2 TA1308000002  41.79957 -87.59475 41.80983 -87.59938    member    8.48333
3
## 3          599  42.00857 -87.69048 42.03974 -87.69941    casual   13.23333
3
## 4 TA1308000002  41.79957 -87.59475 41.80983 -87.59938    member    8.76666
7
## 5 TA1308000002  41.79957 -87.59475 41.80983 -87.59938    member   15.31666
7
## 6     202480.0  41.92607 -87.63886 41.93000 -87.64000    member    3.21666
7
##   day_of_week   month
## 1    Saturday January
## 2     Tuesday January
## 3      Monday January
## 4      Sunday January
## 5    Thursday January
## 6     Tuesday January
```

ANALYZE

Descriptive statistics

The summary statistics for trip_length indicate that the variable has a right-skewed distribution, with a relatively high mean compared to the median.

The minimum trip length is 0.017, signifying very short trips.

The maximum trip length is 12136.300minutes (8 days), suggesting at least one very long trip.

The mean / average trip length is 15.954 minutes, surpassing the median of 9.800, further emphasizing the right-skewed nature of the distribution.

The first quartile (Q1) is 5.617, and the third quartile (Q3) is 17.483. These quartiles reveal that while the majority of trips are relatively short, a notable number of longer trips contribute to the higher mean.

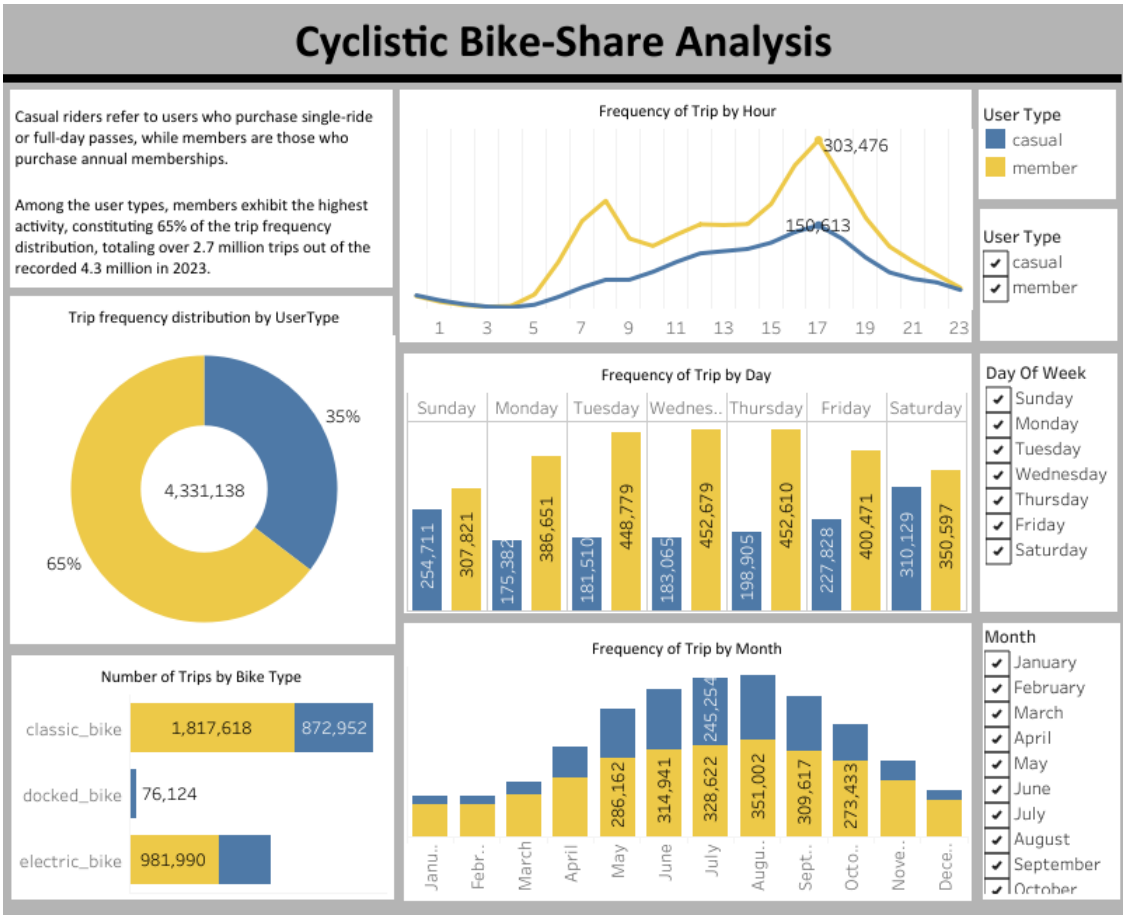
The skewness in the data is likely influenced by user behavior. Trip durations exhibit variability among users who may employ bikes for diverse purposes, such as commuting or leisure. Given that this is fictional data, there are no subject matter experts to consult regarding the specifics of ride start and end.

export clean csv file and upload on Tableau

```
#Get working directory

getwd()

## [1] "C:/Users/ayomide.onayemi/Desktop/Coursera data/Cyclistic/Cyclistic Da
taset/2023"

#And export

write.csv(all_trips, file = "all_trips.csv", row.names = FALSE)
```

SHARE

The data visualization was done using Tableau. Please see the interactive dashboard by clicking this



Cyslistic Bike-Share Dashboard

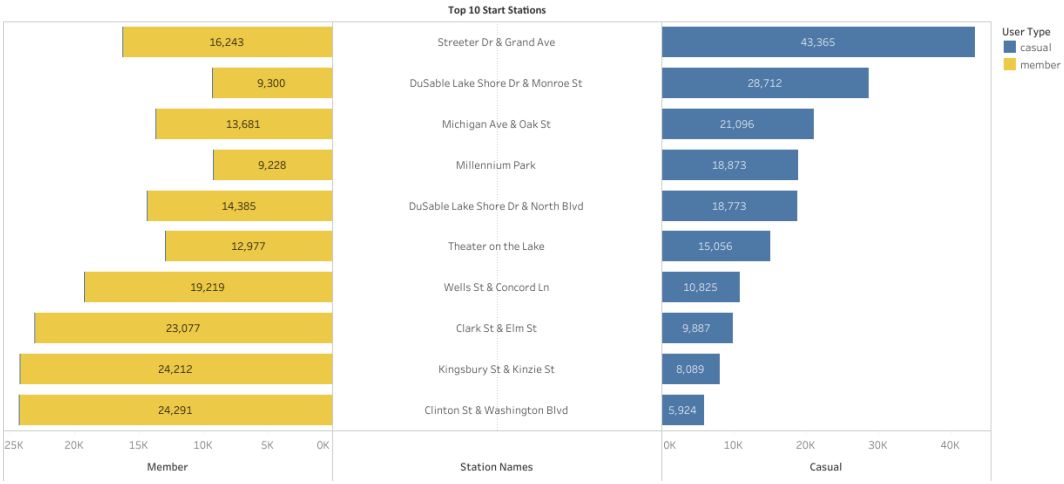
User Type Trip Frequency : The most active user type is “Member,” showing a higher frequency of trips among members compared to casual riders.

Bike Type Usage : The classic bike stands is the most used. Notably, docked bikes are exclusively used by casual riders.

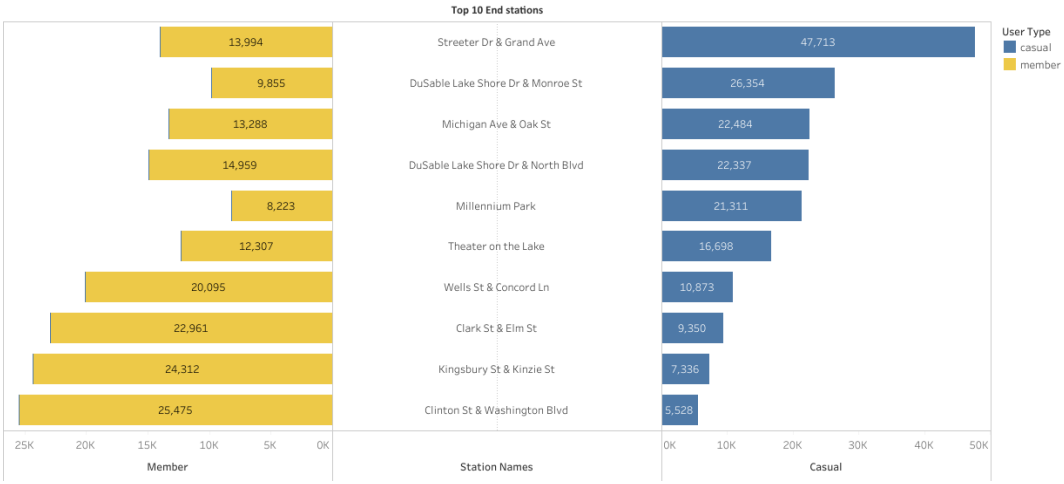
Hourly Trip Frequency : Peak activity for members and casual riders occurs at 17:00, following a notable increase starting from 15:00.

Daily Trip Frequency : Casual riders have higher activity during weekends, particularly on Friday, Saturday and Sunday. In contrast, members show heightened activity on weekdays, notably Tuesday through Thursday. This suggests that members use bikes for commuting, while casual riders engage in leisurely weekend rides.

Monthly Trip Frequency : Both members and casual riders have increased trip frequency from late spring through summer. July marks the peak for casual riders, while members show heightened usage in August.



Top 10 start stations



Top 10 end stations

The above images display the top start and end stations for both casual and member riders, providing additional insights into the most frequently used locations by each rider category.

ACT

Recommendations

1. Seasonal promotions and Campaigns for Casual Riders: Discounts or promotions specifically tailored for weekend rides may be introduced. Special promotions can also be launched during summer or towards the end of spring. Flexible subscription plans can also be introduced, for instance, short-term plans during peak summer months for casual riders.

2. Target digital media campaigns : Since peak activity for casual riders occurs at 17:00, targeted digital media campaigns may be run during peak activity hours to maximize their effectiveness.

3. Promotions on docked bikes : Targeted promotions for annual memberships on docked bikes, as they are primarily used by casual riders.

4. Promote the benefits of annual membership : To effectively communicate the advantages of annual membership to casual riders, targeted advertising strategies(eg. informative flyers) should be implemented at the top stations frequented by casual riders.