

Google Data Analytics Capstone

Noah Borquaye

2024-02-27

CASE STUDY: How does a bike-share navigate Speedy success?

The director of a marketing company believes the company's future success depends on maximizing the number of annual memberships. Therefore, you and your team want to understand how casual riders and annual members use Cyclistic's bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

How Casual Riders and Annual Members Use Cyclistic's Bikes Differently

Data Source and description

The data is a public dataset shared as part of the Google Data Analytics Professional Certificate program. Click [here](https://divvy-tripdata.s3.amazonaws.com/index.html) (<https://divvy-tripdata.s3.amazonaws.com/index.html>) to access the data and the description. I selected the relevant columns for the purpose of this work.

#Reading the data#

```
bike_data <- read.csv("202201-divvy-tripdata.csv")
head(bike_data)
```

```
##           ride_id rideable_type      started_at      ended_at
## 1 C2F7DD78E82EC875 electric_bike 2022-01-13 11:59:47 2022-01-13 12:02:44
## 2 A6CF8980A652D272 electric_bike 2022-01-10 08:41:56 2022-01-10 08:46:17
## 3 BD0F91DFF741C66D classic_bike 2022-01-25 04:53:40 2022-01-25 04:58:01
## 4 CBB80ED419105406 classic_bike 2022-01-04 00:18:04 2022-01-04 00:33:00
## 5 DDC963BFDDA51EEA classic_bike 2022-01-20 01:31:10 2022-01-20 01:37:12
## 6 A39C6F6CC0586C0B classic_bike 2022-01-11 18:48:09 2022-01-11 18:51:31
##           start_station_name start_station_id      end_station_name
## 1      Glenwood Ave & Touhy Ave             525      Clark St & Touhy Ave
## 2      Glenwood Ave & Touhy Ave             525      Clark St & Touhy Ave
## 3 Sheffield Ave & Fullerton Ave    TA1306000016  Greenvew Ave & Fullerton Ave
## 4      Clark St & Bryn Mawr Ave    KA1504000151    Paulina St & Montrose Ave
## 5  Michigan Ave & Jackson Blvd    TA1309000002      State St & Randolph St
## 6      Wood St & Chicago Ave             637      Honore St & Division St
##           end_station_id start_lat start_lng end_lat  end_lng member_casual
## 1              RP-007    42.01280 -87.66591 42.01256 -87.67437      casual
## 2              RP-007    42.01276 -87.66597 42.01256 -87.67437      casual
## 3      TA1307000001    41.92560 -87.65371 41.92533 -87.66580      member
## 4      TA1309000021    41.98359 -87.66915 41.96151 -87.67139      casual
## 5      TA1305000029    41.87785 -87.62408 41.88462 -87.62783      member
## 6      TA1305000034    41.89563 -87.67207 41.90312 -87.67394      member
```

#checking the data structure#

```
str(bike_data)
```

```
## 'data.frame':   103770 obs. of  13 variables:
##  $ ride_id          : chr  "C2F7DD78E82EC875" "A6CF8980A652D272" "BD0F91DFF741C66D" "CBB80ED
419105406" ...
##  $ rideable_type     : chr  "electric_bike" "electric_bike" "classic_bike" "classic_bike" ...
##  $ started_at        : chr  "2022-01-13 11:59:47" "2022-01-10 08:41:56" "2022-01-25 04:53:40"
"2022-01-04 00:18:04" ...
##  $ ended_at          : chr  "2022-01-13 12:02:44" "2022-01-10 08:46:17" "2022-01-25 04:58:01"
"2022-01-04 00:33:00" ...
##  $ start_station_name: chr  "Glenwood Ave & Touhy Ave" "Glenwood Ave & Touhy Ave" "Sheffield
Ave & Fullerton Ave" "Clark St & Bryn Mawr Ave" ...
##  $ start_station_id  : chr  "525" "525" "TA1306000016" "KA1504000151" ...
##  $ end_station_name  : chr  "Clark St & Touhy Ave" "Clark St & Touhy Ave" "Greenview Ave & Fu
llerton Ave" "Paulina St & Montrose Ave" ...
##  $ end_station_id    : chr  "RP-007" "RP-007" "TA1307000001" "TA1309000021" ...
##  $ start_lat         : num  42 42 41.9 42 41.9 ...
##  $ start_lng         : num  -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ end_lat          : num  42 42 41.9 42 41.9 ...
##  $ end_lng          : num  -87.7 -87.7 -87.7 -87.7 -87.6 ...
##  $ member_casual    : chr  "casual" "casual" "member" "casual" ...
```

```
colnames(bike_data)
```

```
## [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"
```

#Selecting relevant columns and removing missing values#

```
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
```

```
bike_df <- bike_data %>%
  select(ride_id,rideable_type, started_at, ended_at, start_station_name, end_station_name,
member_casual) %>%
  na.omit()
```

```
sum(is.na(bike_df))
```

```
## [1] 0
```

#Checking data types and formatting#

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

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

```
new_bike_df <- bike_df %>%
  mutate(started_at = as.POSIXct(started_at, format = "%Y-%m-%d %H:%M:%S")) %>%
  mutate(ended_at = as.POSIXct(ended_at, format = "%Y-%m-%d %H:%M:%S")) %>%
  mutate(trip_duration = as.numeric(difftime(ended_at, started_at, units = "mins")))
```

#Data Aggregation#

```
new_bike_df %>%
  count(name = "Total_rides")
```

```
##   Total_rides
## 1      103770
```

#Number of rides made by each customer type#

```
new_bike_df %>%
  group_by(member_casual ) %>%
  count(name = "number_of_rides")
```

```
## # A tibble: 2 × 2
## # Groups:   member_casual [2]
##   member_casual number_of_rides
##   <chr>          <int>
## 1 casual        18520
## 2 member        85250
```

#Number of rides made ny each bike type#

```
new_bike_df %>%
  group_by(rideable_type) %>%
  count(name = "number_of_rides")
```

```
## # A tibble: 3 × 2
## # Groups:   rideable_type [3]
##   rideable_type number_of_rides
##   <chr>          <int>
## 1 classic_bike    55067
## 2 docked_bike     961
## 3 electric_bike  47742
```

#Average trip duration (minutes) made with each bike type#

```
mean_new_bike <- new_bike_df %>%
  group_by(rideable_type ) %>%
  summarise(mean_trip_duration = mean(trip_duration))%>%
  arrange(desc(mean_trip_duration))
mean_new_bike
```

```
## # A tibble: 3 × 2
##   rideable_type mean_trip_duration
##   <chr>          <dbl>
## 1 docked_bike    223.
## 2 classic_bike   15.4
## 3 electric_bike  10.9
```

```
new_bike_df %>%
  group_by(member_casual, rideable_type ) %>%
  count(name = "total_rideable_type") %>%
  arrange(desc(total_rideable_type))
```

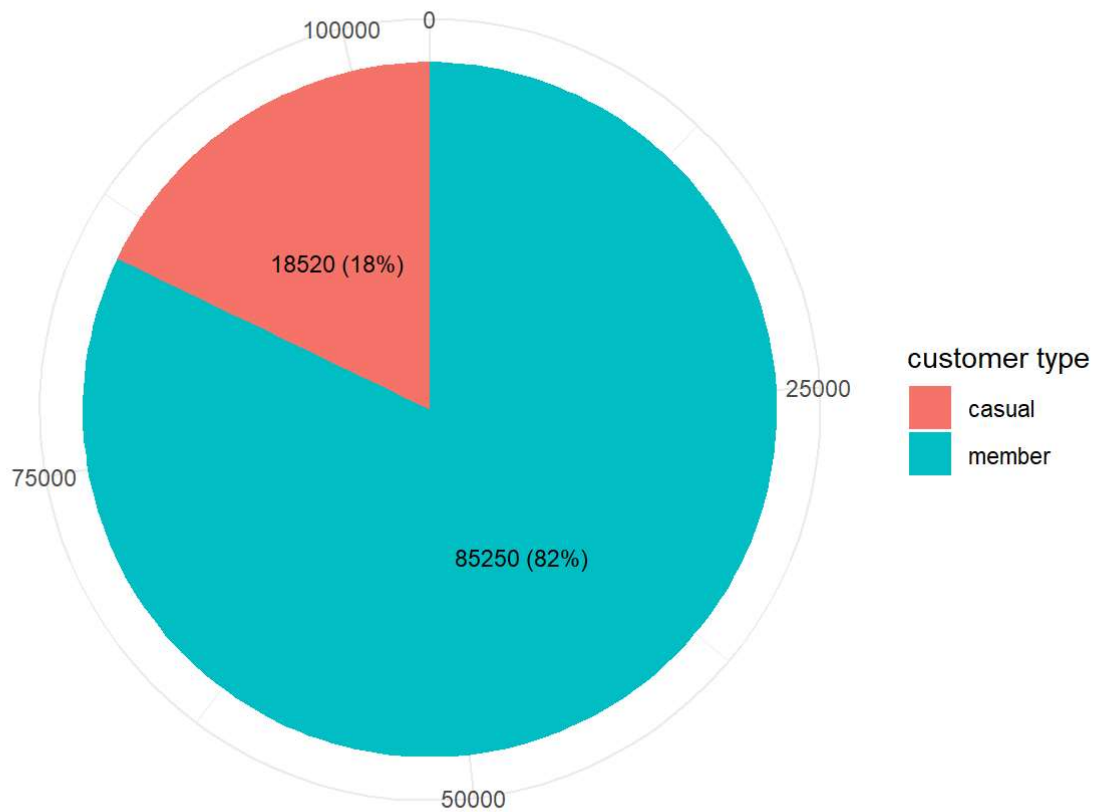
```
## # A tibble: 5 × 3
## # Groups:   member_casual, rideable_type [5]
##   member_casual rideable_type total_rideable_type
##   <chr>         <chr>          <int>
## 1 member        classic_bike    48093
## 2 member        electric_bike   37157
## 3 casual        electric_bike   10585
## 4 casual        classic_bike    6974
## 5 casual        docked_bike     961
```

Visualizations

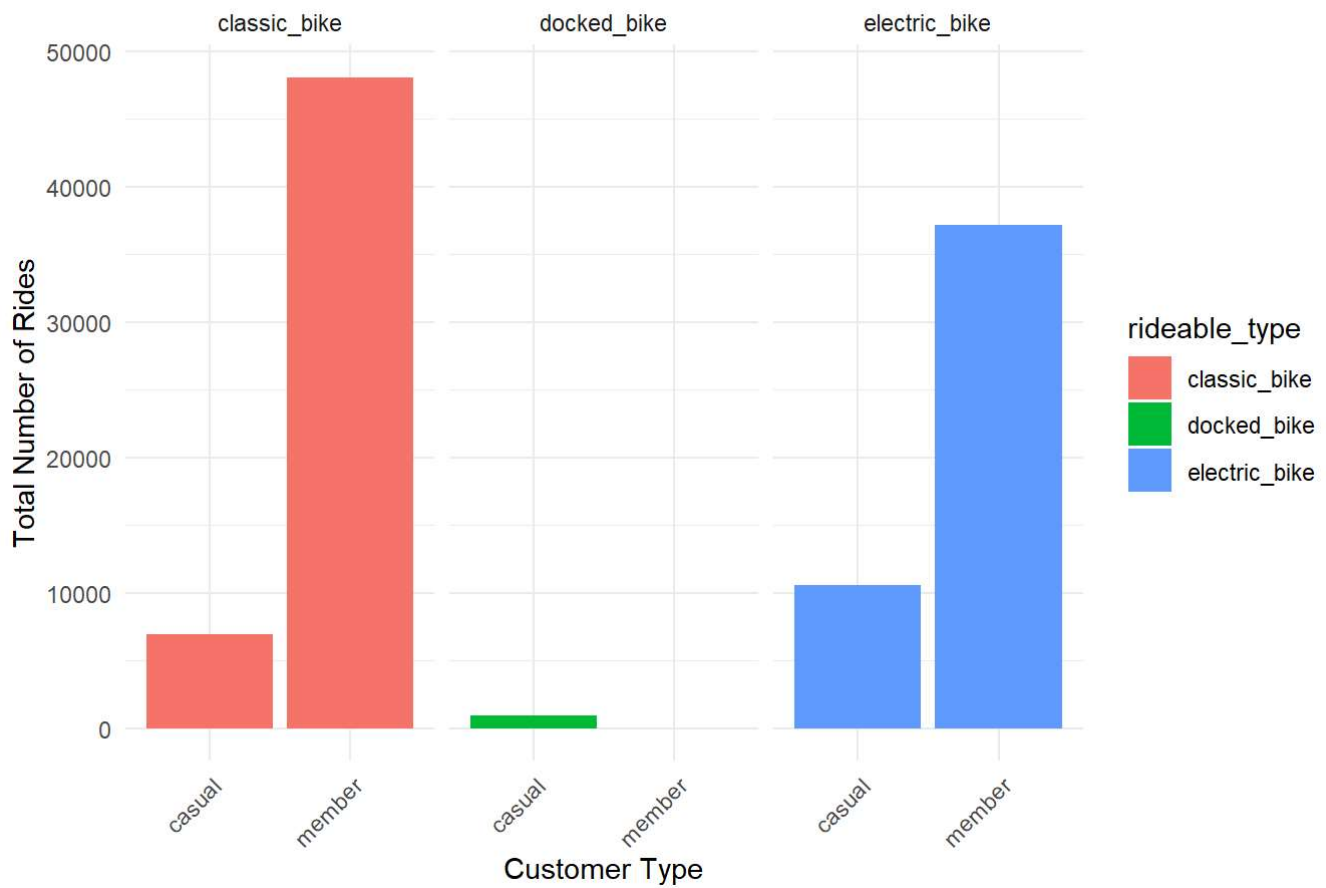
```
library(ggplot2)
```

```
## Coordinate system already present. Adding new coordinate system, which will  
## replace the existing one.
```

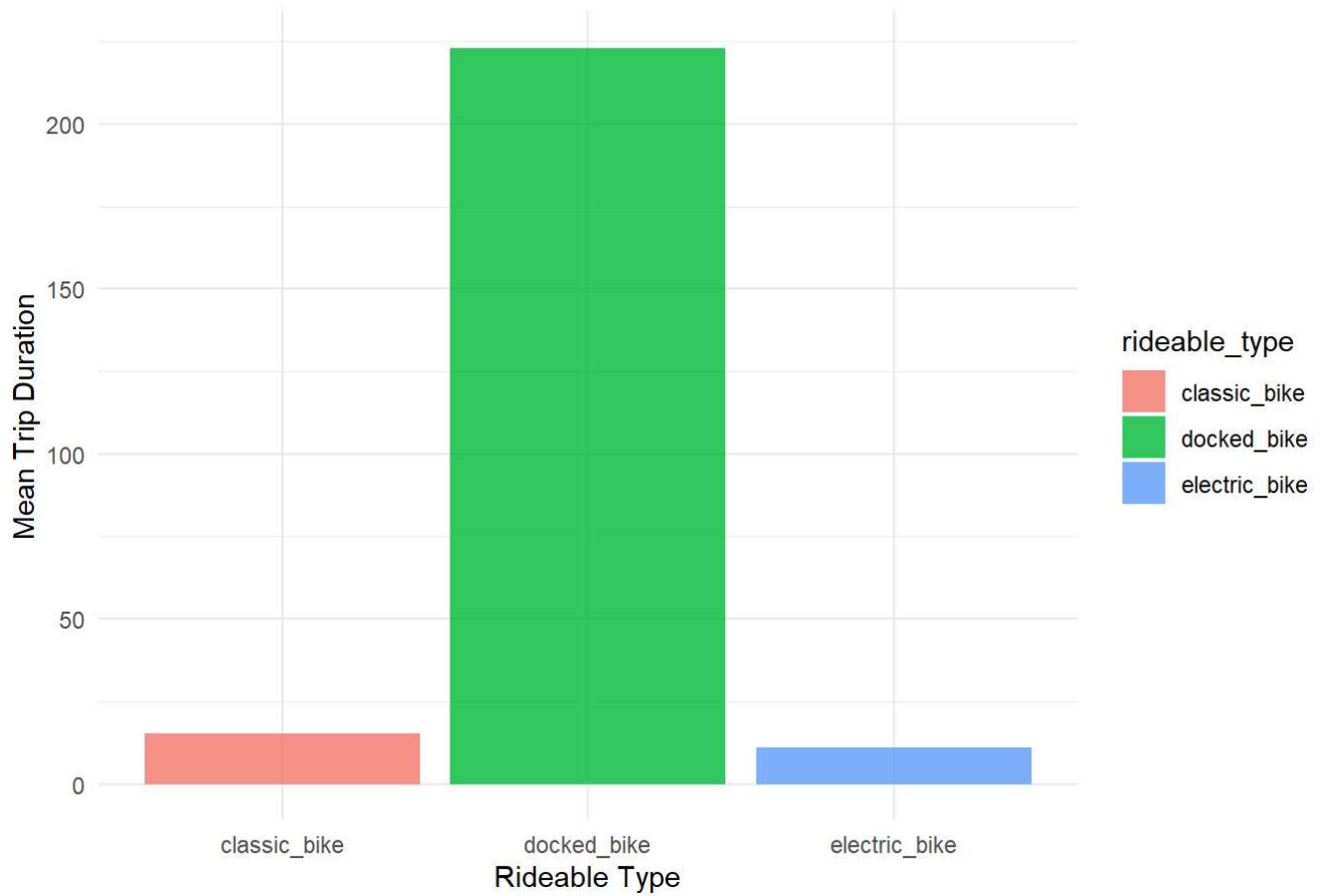
Proportion of Rides by Customer Type



Total Number of Rides by Customer Type and Rideable Type



Mean Trip Duration by Bike Type



Analysis

- 82% of the rides were made by customers who are members while 18% were made by customers who are casual.
- Docked bikes were used by casual customers only, and not often.
- Casual customers used docked bikes for long trips.

Recommendations

- Docked bikes at all stations should be replaced with classic bikes since most customers who are members prefer classic bikes.
- The few customers who use docked bikes should be targeted with promotions
- Since most member customers and high number of casual customers prefer classic bikes, they should be reserved for only member customers. This may compel most classic customers to become members.