Google Capstone Project: Bike Share Speedy Success

Benedict Nyamwanga

2024-05-22

Business Task

The business task is to understand how Cyclistic bikes are used differently by annual members and casual riders. This will enable the management design effective marketing strategies geared to converting casual riders into annual members. By analyzing Cyclistic's historical bike trip data to identify trends and insights will help in having informed decision on how to develop a strategy for targeted marketing campaigns.

Data Sources used

The data used for analysis is Cyclistic's historical bike trip data which is public data from Data License Agreement | Divvy Bikes,Data License Agreement | Divvy Bikes (https://www.divvybikes.com/data-license-agreement) which includes information such as ride ID, rideable type, start and end timestamps, start and end station details, latitude and longitude coordinates, and rider type (member or casual). The data provides insights into how customers utilize Cyclistic bikes for various purposes and durations.

Data Preparation, Processing and Analysis Steps

Technology used

For preparation, processing and analysis, Statistical programming Language R has been used with its IDE(RSudio). ### Set a CRAN mirror

```
options(repos = c(CRAN = "https://cran.rstudio.com"))
```

Data Preparation

Installed Required Packages

```
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

```r
install.packages("lubridate") # to work with dates
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'lubridate' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'lubridate'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Ben\AppData\Local\R\win-library\4.4\00LOCK\lubridate\libs\x64\lubridate.dll
## to
## C:\Users\Ben\AppData\Local\R\win-library\4.4\lubridate\libs\x64\lubridate.dll:
## Permission denied
```

```
## Warning: restored 'lubridate'
```

```
##
## The downloaded binary packages are in
##    C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

```r
install.packages("skimr") # to skim the data
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'skimr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##    C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

```r
install.packages("janitor") # to claen the data
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'janitor' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##    C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

```r
install.packages("readr") # to import the data
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'readr' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'readr'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Ben\AppData\Local\R\win-library\4.4\00LOCK\readr\libs\x64\readr.dll to
## C:\Users\Ben\AppData\Local\R\win-library\4.4\readr\libs\x64\readr.dll:
## Permission denied
```

```
## Warning: restored 'readr'
```

```
##
## The downloaded binary packages are in
##    C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

```
install.packages("dplyr")
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'dplyr'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Ben\AppData\Local\R\win-library\4.4\00LOCK\dplyr\libs\x64\dplyr.dll to
## C:\Users\Ben\AppData\Local\R\win-library\4.4\dplyr\libs\x64\dplyr.dll:
## Permission denied
```

```
## Warning: restored 'dplyr'
```

```
##
## The downloaded binary packages are in
##    C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

```
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/Ben/AppData/Local/R/win-library/4.4'
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\Ben\AppData\Local\Temp\RtmpSS7cdU\downloaded_packages
```

Loaded Packages

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.5
## ✓ forcats   1.0.0     ✓ stringr   1.5.1
## ✓ ggplot2   3.5.1     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.1
## ✓ purrr     1.0.2
## ── Conflicts ──────────────────────────────────── tidyverse_conflicts() ──
## ✖ 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 becom
e errors
```

```
library(lubridate)
library(skimr)
library(janitor)
```

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

```
library(readr)
library(dplyr)
library(ggplot2)
```

Importing Data

```
data1 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202305-divvy-tripdata.cs
v")
```

```
## Rows: 604827 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.
```

```
data2 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202306-divvy-tripdata.csv")
```

```
## Rows: 719618 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.
```

```
data3 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202307-divvy-tripdata.csv")
```

```
## Rows: 767650 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.
```

```
data4 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202308-divvy-tripdata.csv")
```

```
## Rows: 771693 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.
```

```
data5 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202309-divvy-tripdata.csv")
```

```
## Rows: 666371 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.
```

```
data6 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202310-divvy-tripdata.csv")
```

```
## Rows: 537113 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.
```

```
data7 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202311-divvy-tripdata.csv")
```

```
## Rows: 362518 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.
```

```
data8 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202312-divvy-tripdata.cs
v")
```

```
## Rows: 224073 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.
```

```
data9 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202401-divvy-tripdata.cs
v")
```

```
## Rows: 144873 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.
```

```
data10 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202402-divvy-tripdata.cs
v")
```

```
## Rows: 223164 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.
```

```
data11 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202403-divvy-tripdata.cs
v")
```

```
## Rows: 301687 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.
```

```
data12 <-read_csv("E:/BJSchooling/Data Analyst/Case Studies/Capstone-1/202404-divvy-tripdata.cs
v")
```

```
## Rows: 415025 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.
```

Data Processing

Knowing Your Data Structure

```
glimpse(data1)
```

```
## Rows: 604,827
## Columns: 13
## $ ride_id            <chr> "0D9FA920C3062031", "92485E5FB5888ACD", "FB144B3FC8…
## $ rideable_type      <chr> "electric_bike", "electric_bike", "electric_bike", …
## $ started_at         <dttm> 2023-05-07 19:53:48, 2023-05-06 18:54:08, 2023-05-…
## $ ended_at           <dttm> 2023-05-07 19:58:32, 2023-05-06 19:03:35, 2023-05-…
## $ start_station_name <chr> "Southport Ave & Belmont Ave", "Southport Ave & Bel…
## $ start_station_id   <chr> "13229", "13229", "13162", "13196", "TA1308000047",…
## $ end_station_name   <chr> NA, NA, NA, "Damen Ave & Cortland St", "Southport A…
## $ end_station_id     <chr> NA, NA, NA, "13133", "13229", "TA1306000029", "1343…
## $ start_lat          <dbl> 41.93941, 41.93948, 41.85379, 41.89456, 41.95708, 4…
## $ start_lng          <dbl> -87.66383, -87.66385, -87.64672, -87.65345, -87.664…
## $ end_lat            <dbl> 41.93000, 41.94000, 41.86000, 41.91598, 41.93948, 4…
## $ end_lng            <dbl> -87.65000, -87.69000, -87.65000, -87.67733, -87.663…
## $ member_casual      <chr> "member", "member", "member", "member", "member", "…
```

Combine Data

use rbind() or bind_rows() functions to vertically combine data

```
combined_df <- rbind(data1, data2, data3, data4, data5, data6, data7, data8, data9, data10, data
11, data12)
```

Inspection of combined data

you can use str(), glimse(), skim_without_charts() functions

```
skim_without_charts(combined_df)
```

Data summary

Name	combined_df
Number of rows	5738612
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	5738612	0
rideable_type	0	1.00	11	13	0	3	0
start_station_name	885429	0.85	10	64	0	1630	0
start_station_id	885429	0.85	3	35	0	1584	0
end_station_name	939115	0.84	10	64	0	1645	0
end_station_id	939115	0.84	3	36	0	1595	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	41.63	41.88	41.90	41.93	42.07
start_lng	0	1	-87.65	0.03	-87.94	-87.66	-87.64	-87.63	-87.46
end_lat	7610	1	41.90	0.05	0.00	41.88	41.90	41.93	42.18
end_lng	7610	1	-87.65	0.07	-88.16	-87.66	-87.64	-87.63	0.00

Variable type: POSIXct

skim_variable	n_missing	complete_rate	min	max	median	n_unique
started_at	0	1	2023-05-01 00:00:33	2024-04-30 23:59:46	2023-09-01 11:13:33	4836538
ended_at	0	1	2023-05-01 00:04:28	2024-05-02 00:59:33	2023-09-01 11:35:50	4848215

Cleaning Data

Ensure naming consistency

```
combined_df <- clean_names(combined_df)
```

Transforming data

Use mutate() function to create ride_length and day_of_week

```
combined_df <- combined_df %>%
  mutate(
        ride_length = as.numeric(difftime(ended_at, started_at, units = "mins")),
        day_of_week = wday(started_at, label = TRUE))
```

Confirming the changes

```
head(combined_df)
```

```
## # A tibble: 6 × 15
##   ride_id         rideable_type started_at          ended_at
##   <chr>           <chr>         <dttm>              <dttm>
## 1 0D9FA920C3062031 electric_bike 2023-05-07 19:53:48 2023-05-07 19:58:32
## 2 92485E5FB5888ACD electric_bike 2023-05-06 18:54:08 2023-05-06 19:03:35
## 3 FB144B3FC8300187 electric_bike 2023-05-21 00:40:21 2023-05-21 00:44:36
## 4 DDEB93BC2CE9AA77 classic_bike  2023-05-10 16:47:01 2023-05-10 16:59:52
## 5 C07B70172FC92F59 classic_bike  2023-05-09 18:30:34 2023-05-09 18:39:28
## 6 2BA66385DF8F815A classic_bike  2023-05-30 15:01:21 2023-05-30 15:17:00
## # i 11 more variables: 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>,
## #   ride_length <dbl>, day_of_week <ord>
```

Analyse Data

Summary Statistics

```
summary(combined_df)
```

```
##    ride_id          rideable_type        started_at
## Length:5738612     Length:5738612     Min.   :2023-05-01 00:00:33.00
## Class :character   Class :character   1st Qu.:2023-07-06 14:04:16.75
## Mode  :character   Mode  :character   Median :2023-09-01 11:13:33.50
##                                       Mean   :2023-09-23 12:21:14.20
##                                       3rd Qu.:2023-11-16 17:15:54.75
##                                       Max.   :2024-04-30 23:59:46.00
##
##     ended_at                      start_station_name start_station_id
## Min.   :2023-05-01 00:04:28.00   Length:5738612     Length:5738612
## 1st Qu.:2023-07-06 14:28:38.75   Class :character   Class :character
## Median :2023-09-01 11:35:50.00   Mode  :character   Mode  :character
## Mean   :2023-09-23 12:39:36.22
## 3rd Qu.:2023-11-16 17:28:17.50
## Max.   :2024-05-02 00:59:33.00
##
## end_station_name    end_station_id       start_lat        start_lng
## Length:5738612     Length:5738612     Min.   :41.63    Min.   :-87.94
## 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   :-87.65
##                                       3rd Qu.:41.93    3rd Qu.:-87.63
##                                       Max.   :42.07    Max.   :-87.46
##
##     end_lat          end_lng         member_casual       ride_length
## Min.   : 0.00    Min.   :-88.16    Length:5738612     Min.   :-16656.52
## 1st Qu.:41.88    1st Qu.:-87.66    Class :character   1st Qu.:     5.50
## Median :41.90    Median :-87.64    Mode  :character   Median :     9.65
## Mean   :41.90    Mean   :-87.65                       Mean   :    18.37
## 3rd Qu.:41.93    3rd Qu.:-87.63                       3rd Qu.:    17.12
## Max.   :42.18    Max.   : 0.00                        Max.   : 98489.07
## NA's   :7610     NA's   :7610
## day_of_week
## Sun:752173
## Mon:760647
## Tue:833769
## Wed:824172
## Thu:851203
## Fri:822014
## Sat:894634
```

Where ride length was -ve or having N/A was replaced with 0

```
combined_df <- combined_df %>%
  mutate(
    ride_length = ifelse(is.na(ride_length) | ride_length < 0, 0, ride_length)
  )
```

Rechecking summary again for any errors

```
summary(combined_df)
```

```
##     ride_id           rideable_type          started_at
##   Length:5738612     Length:5738612     Min.   :2023-05-01 00:00:33.00
##   Class :character   Class :character   1st Qu.:2023-07-06 14:04:16.75
##   Mode  :character   Mode  :character   Median :2023-09-01 11:13:33.50
##                                         Mean   :2023-09-23 12:21:14.20
##                                         3rd Qu.:2023-11-16 17:15:54.75
##                                         Max.   :2024-04-30 23:59:46.00
##
##       ended_at                       start_station_name start_station_id
##   Min.   :2023-05-01 00:04:28.00   Length:5738612     Length:5738612
##   1st Qu.:2023-07-06 14:28:38.75   Class :character   Class :character
##   Median :2023-09-01 11:35:50.00   Mode  :character   Mode  :character
##   Mean   :2023-09-23 12:39:36.22
##   3rd Qu.:2023-11-16 17:28:17.50
##   Max.   :2024-05-02 00:59:33.00
##
##   end_station_name    end_station_id       start_lat        start_lng
##   Length:5738612     Length:5738612     Min.   :41.63    Min.   :-87.94
##   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   :-87.65
##                                         3rd Qu.:41.93    3rd Qu.:-87.63
##                                         Max.   :42.07    Max.   :-87.46
##
##      end_lat          end_lng        member_casual       ride_length
##   Min.   : 0.00    Min.   :-88.16   Length:5738612     Min.   :    0.00
##   1st Qu.:41.88    1st Qu.:-87.66   Class :character   1st Qu.:    5.50
##   Median :41.90    Median :-87.64   Mode  :character   Median :    9.65
##   Mean   :41.90    Mean   :-87.65                      Mean   :   18.38
##   3rd Qu.:41.93    3rd Qu.:-87.63                      3rd Qu.:   17.12
##   Max.   :42.18    Max.   : 0.00                       Max.   :98489.07
##   NA's   :7610     NA's   :7610
##   day_of_week
##   Sun:752173
##   Mon:760647
##   Tue:833769
##   Wed:824172
##   Thu:851203
##   Fri:822014
##   Sat:894634
```

Statistics of the ride_length

```
combined_df %>%
  group_by(member_casual) %>%
  summarise(mean_ride_length = mean(ride_length),
            median_ride_length = median(ride_length),
            max_ride_length = max(ride_length),
            total_ride_length = sum(ride_length))
```

```
## # A tibble: 2 × 5
##   member_casual mean_ride_length median_ride_length max_ride_length
##   <chr>                    <dbl>              <dbl>           <dbl>
## 1 casual                    28.3               12.0          98489.
## 2 member                    12.9               8.63           1560.
## # i 1 more variable: total_ride_length <dbl>
```

Statistics Number of rides by day of the week

```
combined_df %>%
  count(member_casual, day_of_week)
```

```
## # A tibble: 14 × 3
##    member_casual day_of_week      n
##    <chr>         <ord>        <int>
##  1 casual        Sun         342880
##  2 casual        Mon         244115
##  3 casual        Tue         247474
##  4 casual        Wed         240221
##  5 casual        Thu         260454
##  6 casual        Fri         300901
##  7 casual        Sat         416607
##  8 member        Sun         409293
##  9 member        Mon         516532
## 10 member        Tue         586295
## 11 member        Wed         583951
## 12 member        Thu         590749
## 13 member        Fri         521113
## 14 member        Sat         478027
```

Including Plots for Trends

Plot number of rides by day of the week

```
combined_df %>%
    count(day_of_week, member_casual) %>%
    ggplot(aes(x = day_of_week, y = n, fill = member_casual)) +
    geom_col(position = "dodge") +
    labs(title = "Number of Rides by Day of the Week",
        x = "Day of the Week",
        y = "Number of Rides",
        caption = "Data added by Benedict N")
```

Number of Rides by Day of the Week
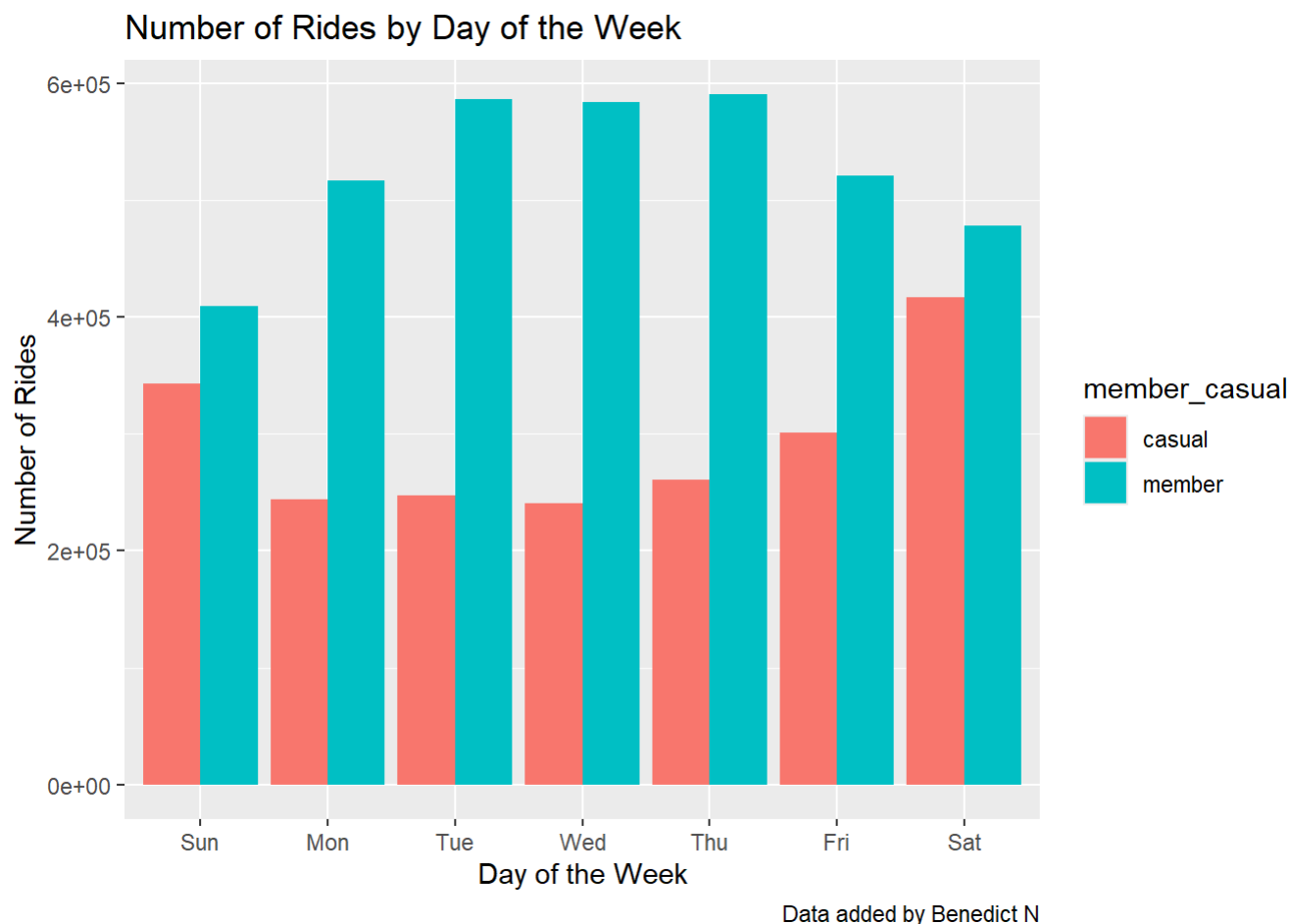
Plot average ride length by day of the week

```
combined_df %>%
    group_by(day_of_week, member_casual) %>%
    summarise(mean_ride_length = mean(ride_length)) %>%
    ggplot(aes(x = day_of_week, y = mean_ride_length, color = member_casual, group = member_casu
al)) +
  # Adjust group aesthetic
    geom_line(size = 1) +
    labs(title = "Average Ride Length by Day of the Week",
        x = "Day of the Week",
        y = "Average Ride Length (mins)",
        caption = "Data added by Benedict N")
```

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

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Average Ride Length by Day of the Week

Data added by Benedict N
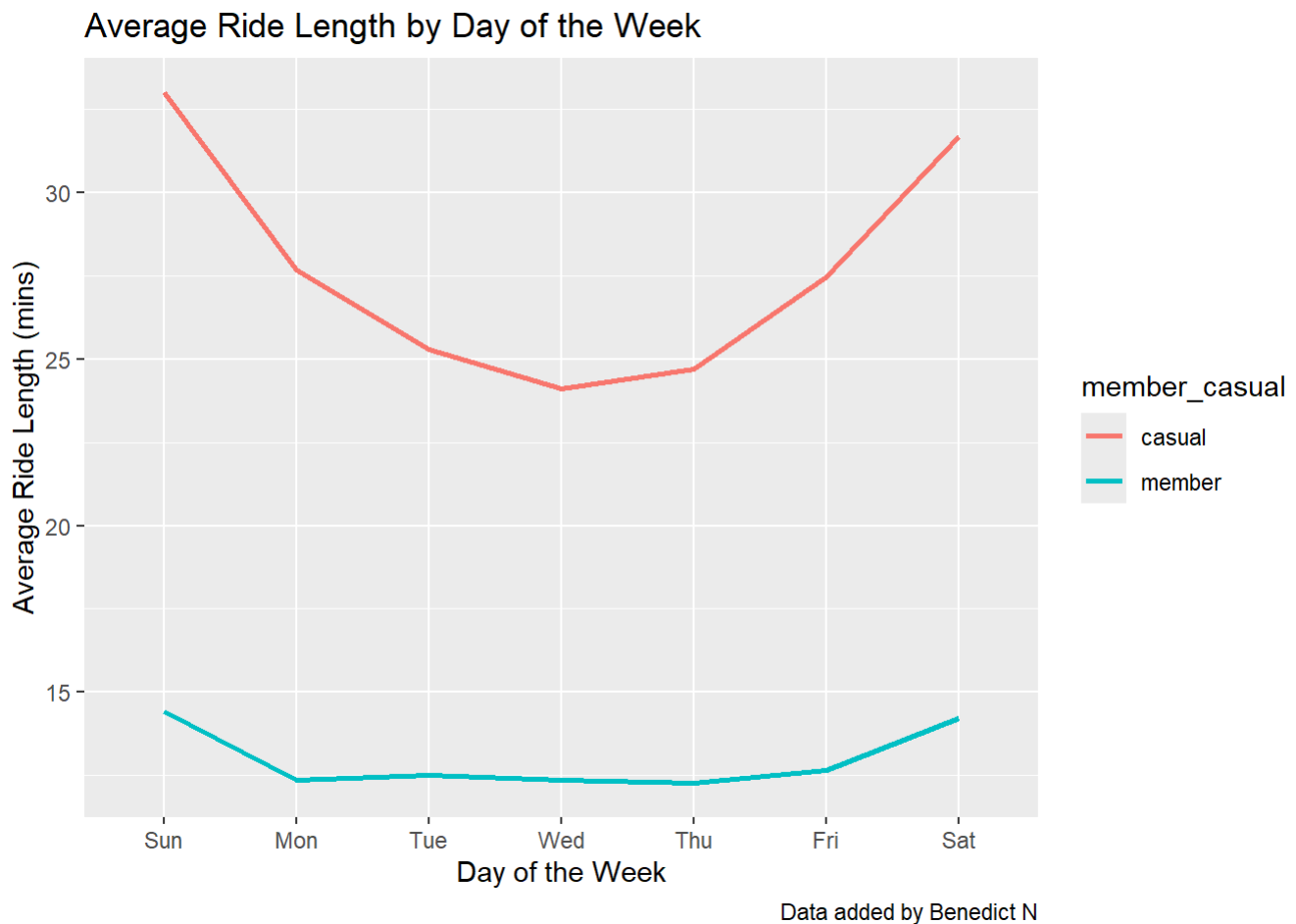
Summary of the analysis and Key Insights

1. Annual members tend to use Cyclistic bikes for shorter durations compared to casual riders.
2. Casual riders, on the other hand, have longer average ride lengths. This may indicate that they use the bikes for leisure or recreational purposes.
3. Throughout the week there are variations in ride patterns, with different usage trends observed for weekdays versus weekends.

Recommendations

1. Target marketing campaigns on weekends when casual riders are more likely to use the service.
2. Promote benefits of membership such as cost savings for frequent riders.
3. Use digital media campaigns highlighting the convenience of annual memberships