Analysis on Cyclistic Trip Data: Casual vs. Annual Members

Andres Machado

01-05-2023

The purpose of this document is to outline the data wrangling, data cleaning, data analysis, and data visualization of Cyclystic's trip data from November 2021 to November 2022.

Data Wrangling

Loading necessary R packages

```
library(tidyverse) # Gives us general data analysis functions
library(lubridate) # Helps with date functions
library(geosphere) # Helps with distance functions
library(ggplot2) # Helps with data visualizations
```

Collecting the data

```
# Reading the csv files from Nov 2021 to Nov 2022
nov_21 <- read_csv("202111-divvy-tripdata.csv")

dec_21 <- read_csv("202112-divvy-tripdata.csv")

jan_22 <- read_csv("202201-divvy-tripdata.csv")

feb_22 <- read_csv("202202-divvy-tripdata.csv")

mar_22 <- read_csv("202203-divvy-tripdata.csv")

apr_22 <- read_csv("202204-divvy-tripdata.csv")

may_22 <- read_csv("202205-divvy-tripdata.csv")

jun_22 <- read_csv("202206-divvy-tripdata.csv")

jul_22 <- read_csv("202207-divvy-tripdata.csv")

aug_22 <- read_csv("202208-divvy-tripdata.csv")

sep_22 <- read_csv("202209-divvy-publictripdata.csv")

oct_22 <- read_csv("202210-divvy-tripdata.csv")

nov_22 <- read_csv("202211-divvy-tripdata.csv")
```

Joining all of the data into a single data frame

In order to combine all of the files together they must have the same column names:

```
colnames(nov_21)
colnames(dec_21)
colnames(jan_22)
colnames(feb_22)
colnames(mar_22)
colnames(apr_22)
colnames(may_22)
colnames(jun_22)
colnames(jul_22)
colnames(aug_22)
colnames(sep_22)
colnames(oct_22)
colnames(nov_22)
```

Output (hidden for cleanliness) showed that all files have the same column names.

We also need check for data type inconsistencies:

```
str(nov_21)
str(dec_21)
str(jan_22)
str(feb_22)
str(mar_22)
str(apr_22)
str(may_22)
str(jun_22)
str(jul_22)
str(aug_22)
str(sep_22)
str(oct_22)
str(nov_22)
```

Output (hidden for cleanliness) shows all data frames have the same underlying structure.

Now we can combine all the data together:

```
alltrips_data <- bind_rows(nov_21, dec_21, jan_22, feb_22, mar_22, apr_22, may_22, jun_22, jul_22, aug_22, sep_22, oct_22, nov_22)
```

Data Cleaning

Inspecting the new table that has been created:

```
colnames(alltrips_data) # Ensure that the column names have remained the same
```

```
##  [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"
```

```
str(alltrips_data) # Checking to see if all data types are the same as before
```

```
## spc_tbl_ [6,093,429 × 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:6093429] "7C00A93E10556E47" "90854840DFD508BA" "0A7D10C
DD144061C" "2F3BE33085BCFF02" ...
##  $ rideable_type     : chr [1:6093429] "electric_bike" "electric_bike" "electric_bik
e" "electric_bike" ...
##  $ started_at        : POSIXct[1:6093429], format: "2021-11-27 13:27:38" "2021-11-27
13:38:25" ...
##  $ ended_at          : POSIXct[1:6093429], format: "2021-11-27 13:46:38" "2021-11-27
13:56:10" ...
##  $ start_station_name: chr [1:6093429] NA NA NA NA ...
##  $ start_station_id  : chr [1:6093429] NA NA NA NA ...
##  $ end_station_name  : chr [1:6093429] NA NA NA NA ...
##  $ end_station_id    : chr [1:6093429] NA NA NA NA ...
##  $ start_lat         : num [1:6093429] 41.9 42 42 41.9 41.9 ...
##  $ start_lng         : num [1:6093429] -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ end_lat           : num [1:6093429] 42 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:6093429] -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ member_casual     : chr [1:6093429] "casual" "casual" "casual" "casual" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
nrow(alltrips_data) # Gives the number of rows in the file
```

```
## [1] 6093429
```

```r
dim(alltrips_data) # Gives the dimensions of the data frame
```

```
## [1] 6093429      13
```

```r
head(alltrips_data) # Gives us first couple rows of data
```

```
## # A tibble: 6 × 13
##   ride_id          rideable_type started_at          ended_at
##   <chr>            <chr>         <dttm>              <dttm>
## 1 7C00A93E10556E47 electric_bike 2021-11-27 13:27:38 2021-11-27 13:46:38
## 2 90854840DFD508BA electric_bike 2021-11-27 13:38:25 2021-11-27 13:56:10
## 3 0A7D10CDD144061C electric_bike 2021-11-26 22:03:34 2021-11-26 22:05:56
## 4 2F3BE33085BCFF02 electric_bike 2021-11-27 09:56:49 2021-11-27 10:01:50
## 5 D67B4781A19928D4 electric_bike 2021-11-26 19:09:28 2021-11-26 19:30:41
## 6 02F85C2C3C5F7D46 electric_bike 2021-11-26 18:34:07 2021-11-26 18:52:49
## # ℹ 9 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>
```

```r
tail(alltrips_data) # Gives us last couple rows of data
```

```
## # A tibble: 6 × 13
##   ride_id          rideable_type started_at          ended_at
##   <chr>            <chr>         <dttm>              <dttm>
## 1 E4F708DDC80A274C electric_bike 2022-11-20 17:38:28 2022-11-20 17:41:01
## 2 C349E243A9BAA6F7 electric_bike 2022-11-25 11:19:52 2022-11-25 11:31:50
## 3 B0B4E85DA43A9194 classic_bike  2022-11-22 16:57:53 2022-11-22 17:31:29
## 4 8D148DD47B59530B classic_bike  2022-11-06 13:04:05 2022-11-06 13:13:33
## 5 0D1170BA18FD33D1 classic_bike  2022-11-06 09:41:29 2022-11-06 15:17:17
## 6 09B20DC75B5EA1E0 electric_bike 2022-11-26 11:59:28 2022-11-26 12:31:04
## # ℹ 9 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>
```

```r
summary(alltrips_data) # Gives the statistical summary of the data
```

```
##    ride_id           rideable_type        started_at                   
##  Length:6093429     Length:6093429     Min.   :2021-11-01 00:00:14.00  
##  Class :character   Class :character   1st Qu.:2022-05-04 16:56:43.00  
##  Mode  :character   Mode  :character   Median :2022-07-07 14:13:08.00  
##                                        Mean   :2022-06-22 09:39:09.57  
##                                        3rd Qu.:2022-09-03 17:21:35.00  
##                                        Max.   :2022-11-30 23:56:11.00  
##                                                                        
##     ended_at                      start_station_name start_station_id  
##  Min.   :2021-11-01 00:04:06.00   Length:6093429     Length:6093429    
##  1st Qu.:2022-05-04 17:09:23.00   Class :character   Class :character  
##  Median :2022-07-07 14:35:34.00   Mode  :character   Mode  :character  
##  Mean   :2022-06-22 09:58:18.41                                        
##  3rd Qu.:2022-09-03 17:43:59.00                                        
##  Max.   :2022-12-01 11:45:53.00                                        
##                                                                        
##  end_station_name   end_station_id       start_lat       start_lng     
##  Length:6093429     Length:6093429     Min.   :41.64   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   :-87.65  
##                                        3rd Qu.:41.93   3rd Qu.:-87.63  
##                                        Max.   :45.64   Max.   :-73.80  
##                                                                        
##     end_lat         end_lng       member_casual     
##  Min.   : 0.00   Min.   :-88.97   Length:6093429    
##  1st Qu.:41.88   1st Qu.:-87.66   Class :character  
##  Median :41.90   Median :-87.64   Mode  :character  
##  Mean   :41.90   Mean   :-87.65                     
##  3rd Qu.:41.93   3rd Qu.:-87.63                     
##  Max.   :42.37   Max.   :  0.00                     
##  NA's   :6065    NA's   :6065
```

There are some issues that need to be dealt with:

1. Verifying that there is only two names for users in the “members_casual” column (casuals and members) and verify that there are three types of bikes in “rideable_type” column.
2. Addition of five new columns to add intricacy to the analysis. The columns will be: date, year, month, day, and day_of_week. This will provide additional opportunities to aggregate the data.
3. Adding a dedicated column for the length of a bike ride titled “ride_length”. Also, one will be made for ride distance titled “ride_distance”. Many calculations will be made comparing the different members.
4. According to Cyclistic, there are some rides taken where the duration of them will be negative and some ride distances that are equal to zero. This because they were taking the bike out of circulation for maintenance and quality control reasons. They need to be removed from the data frame.

Issue 1

Verifying that there is only two names for members in the “members_casual” column (casuals and members) and verify that there are three types of bikes in “rideable_type” column.

```
# Checking member_casual column
table(alltrips_data$member_casual)
```

```
##
##  casual  member
## 2453805 3639624
```

Results show us that we indeed have two types of customers and no other customer types exist.

```
# Checking rideable_type column
table(alltrips_data$rideable_type)
```

```
##
##  classic_bike   docked_bike electric_bike
##       2782538        188091       3122800
```

Results show us that we have three types of bikes. **Special note: Electric bike is the most popular type of bike followed by classic bike.**

Issue 2

Addition of five new columns to add intricacy to the analysis. The columns will be: date, year, month, day, and day_of_week. This will provide additional opportunities to aggregate the data.

```
alltrips_data$date <- as.Date(alltrips_data$started_at) # YYYY-MM-DD

alltrips_data$year <- format(as.Date(alltrips_data$date), "%Y") # Year (####) Column

alltrips_data$month <- format(as.Date(alltrips_data$date), "%m") # Month (00-12) Column

alltrips_data$day <- format(as.Date(alltrips_data$date),"%d") # Day (00-31) Column

alltrips_data$day_of_week <-format(as.Date(alltrips_data$date), "%A") # Day of the Week
(Sunday-Saturday) Column
```

Issue 3

Adding a dedicated column for the length of a bike ride titled “ride_length”. Also, one will be made for ride distance titled “ride_distance”. Many calculations will be made comparing the different members.

```
# Adding the ride_length column in Seconds
alltrips_data$ride_length <- difftime(alltrips_data$ended_at, alltrips_data$started_at)
```

Checking the column data types for all columns:

```
str(alltrips_data)
```

```
## spc_tbl_ [6,093,429 × 19] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ride_id           : chr [1:6093429] "7C00A93E10556E47" "90854840DFD508BA" "0A7D10C
DD144061C" "2F3BE33085BCFF02" ...
##  $ rideable_type     : chr [1:6093429] "electric_bike" "electric_bike" "electric_bik
e" "electric_bike" ...
##  $ started_at        : POSIXct[1:6093429], format: "2021-11-27 13:27:38" "2021-11-27
13:38:25" ...
##  $ ended_at          : POSIXct[1:6093429], format: "2021-11-27 13:46:38" "2021-11-27
13:56:10" ...
##  $ start_station_name: chr [1:6093429] NA NA NA NA ...
##  $ start_station_id  : chr [1:6093429] NA NA NA NA ...
##  $ end_station_name  : chr [1:6093429] NA NA NA NA ...
##  $ end_station_id    : chr [1:6093429] NA NA NA NA ...
##  $ start_lat         : num [1:6093429] 41.9 42 42 41.9 41.9 ...
##  $ start_lng         : num [1:6093429] -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ end_lat           : num [1:6093429] 42 41.9 42 41.9 41.9 ...
##  $ end_lng           : num [1:6093429] -87.7 -87.7 -87.7 -87.8 -87.6 ...
##  $ member_casual     : chr [1:6093429] "casual" "casual" "casual" "casual" ...
##  $ date              : Date[1:6093429], format: "2021-11-27" "2021-11-27" ...
##  $ year              : chr [1:6093429] "2021" "2021" "2021" "2021" ...
##  $ month             : chr [1:6093429] "11" "11" "11" "11" ...
##  $ day               : chr [1:6093429] "27" "27" "26" "27" ...
##  $ day_of_week       : chr [1:6093429] "Saturday" "Saturday" "Friday" "Saturday" ...
##  $ ride_length       : 'difftime' num [1:6093429] 1140 1065 142 301 ...
##   ..- attr(*, "units")= chr "secs"
##  - attr(*, "spec")=
##   .. cols(
##   ..   ride_id = col_character(),
##   ..   rideable_type = col_character(),
##   ..   started_at = col_datetime(format = ""),
##   ..   ended_at = col_datetime(format = ""),
##   ..   start_station_name = col_character(),
##   ..   start_station_id = col_character(),
##   ..   end_station_name = col_character(),
##   ..   end_station_id = col_character(),
##   ..   start_lat = col_double(),
##   ..   start_lng = col_double(),
##   ..   end_lat = col_double(),
##   ..   end_lng = col_double(),
##   ..   member_casual = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Checking if `ride_length` specifically is numeric:

```
is.numeric(alltrips_data$ride_length)
```

```
## [1] FALSE
```

As it is not numeric, this means it is a factor. It has to be numeric in order to run the calculations needed so, converting it to character and then into numeric is the solution:

```
alltrips_data$ride_length <- as.numeric(as.character(alltrips_data$ride_length))

is.numeric(alltrips_data$ride_length) #Verifying that it is now numeric
```

```
## [1] TRUE
```

Now that `ride_length` is numeric and finished, making `ride_distance` is next:

```
alltrips_data$ride_distance <- distGeo(matrix(c(alltrips_data$start_lng, alltrips_data$start_lat), ncol=2), matrix(c(alltrips_data$end_lng, alltrips_data$end_lat), ncol=2))

alltrips_data$ride_distance <- alltrips_data$ride_distance/1000 #distance in km
```

Issue 4

According to Cyclistic, there are some rides taken where the duration of them will be negative and some ride distances that are equal to zero. This because they were taking the bike out of circulation for maintenance and quality control reasons. They need to be removed from the data frame

Since we are removing company data, we are making a new data frame to uphold data integrity:

```
alltrips_data_v2 <- alltrips_data[!(alltrips_data$ride_length <= 0 | alltrips_data$ride_distance == 0),]

alltrips_data_v2 <- na.omit(alltrips_data_v2) # Gets rid of any NA values
dim(alltrips_data_v2)
```

```
## [1] 4454687      20
```

NOTE: This removed 1,638,742 observations from our original data

Data Analysis and Visualization

Descriptive Analysis on Ride Length and Ride Distance

Ride Length

```
# All figures will be in seconds
summary(alltrips_data_v2$ride_length)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
##       1.0     366.0     627.0     978.8    1098.0 2061244.0
```

- Average ride length amongst all riders is 978.8 seconds or about 16.3 minutes

Ride Distance

```
# All figures will be in kilometers
summary(alltrips_data_v2$ride_distance)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
##    0.000    0.973    1.625    2.196    2.801 9817.319 
```

- Average distance of a ride amongst all riders is 2.196 kilometers or about 1.4 miles

Comparing Average Ride Length and Ride Distance between Casual customers and Annual members

Ride Length

Reminder: The average ride length between all users was 978.8 seconds or about 16.3 minutes

```
# Figures will be in Seconds
aggregate(alltrips_data_v2$ride_length ~ alltrips_data_v2$member_casual, FUN = mean)
```

```
##   alltrips_data_v2$member_casual alltrips_data_v2$ride_length
## 1                         casual                    1368.8165
## 2                         member                     735.1251
```

Casual customers take longer rides than annual members

Ride Distance

Reminder: The average ride distance between all users was 2.196 kilometers or about 1.4 miles

```
# Figures will be in Kilometers
aggregate(alltrips_data_v2$ride_distance ~ alltrips_data_v2$member_casual, FUN = mean)
```

```
##   alltrips_data_v2$member_casual alltrips_data_v2$ride_distance
## 1                         casual                       2.319524
## 2                         member                       2.118932
```

Casual customers ride for longer distances than Annual members but it is close

Which days of the week do members and casuals ride the longest and farthest?

```
# Putting the days in order
alltrips_data_v2$day_of_week <- ordered(alltrips_data_v2$day_of_week, levels=c("Sunday",
"Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))

# Average ride time by each day of the week members vs casual users
aggregate(alltrips_data_v2$ride_length ~ alltrips_data_v2$member_casual + alltrips_data_
v2$day_of_week, FUN = mean)
```

```
##    alltrips_data_v2$member_casual alltrips_data_v2$day_of_week
## 1                          casual                       Sunday
## 2                          member                       Sunday
## 3                          casual                       Monday
## 4                          member                       Monday
## 5                          casual                      Tuesday
## 6                          member                      Tuesday
## 7                          casual                    Wednesday
## 8                          member                    Wednesday
## 9                          casual                     Thursday
## 10                         member                     Thursday
## 11                         casual                       Friday
## 12                         member                       Friday
## 13                         casual                     Saturday
## 14                         member                     Saturday
##    alltrips_data_v2$ride_length
## 1                     1551.5971
## 2                      818.6985
## 3                     1395.1766
## 4                      706.8296
## 5                     1211.5366
## 6                      695.1379
## 7                     1180.9124
## 8                      699.2754
## 9                     1230.9636
## 10                     712.4576
## 11                    1286.4690
## 12                     722.7994
## 13                    1540.4711
## 14                     830.9383
```

```
# Average ride distance by each day of the week members vs casual users
aggregate(alltrips_data_v2$ride_distance ~ alltrips_data_v2$member_casual + alltrips_dat
a_v2$day_of_week, FUN = mean)
```

```
##    alltrips_data_v2$member_casual alltrips_data_v2$day_of_week
## 1                          casual                       Sunday
## 2                          member                       Sunday
## 3                          casual                       Monday
## 4                          member                       Monday
## 5                          casual                      Tuesday
## 6                          member                      Tuesday
## 7                          casual                    Wednesday
## 8                          member                    Wednesday
## 9                          casual                     Thursday
## 10                         member                     Thursday
## 11                         casual                       Friday
## 12                         member                       Friday
## 13                         casual                     Saturday
## 14                         member                     Saturday
##    alltrips_data_v2$ride_distance
## 1                        2.408388
## 2                        2.170176
## 3                        2.239965
## 4                        2.050192
## 5                        2.209798
## 6                        2.063733
## 7                        2.307957
## 8                        2.211277
## 9                        2.243973
## 10                       2.089914
## 11                       2.258030
## 12                       2.053922
## 13                       2.448873
## 14                       2.211490
```

As we can see, the most popular days are that of the weekends: Friday, Saturday, and Sunday. With casuals riding preferring Sunday and members preferring Saturday

Which month did members and casuals ride for the longest and furthest amounts?

```
# Ride Length
aggregate(alltrips_data_v2$ride_length ~ alltrips_data_v2$member_casual + alltrips_data_
v2$month +alltrips_data_v2$year, FUN = mean) %>%

# Ride Distance
aggregate(alltrips_data_v2$ride_distance ~ alltrips_data_v2$member_casual + alltrips_dat
a_v2$month +alltrips_data_v2$year, FUN = mean)
```

```
##    alltrips_data_v2$member_casual alltrips_data_v2$month alltrips_data_v2$year
## 1                          casual                     11                  2021
## 2                          member                     11                  2021
## 3                          casual                     12                  2021
## 4                          member                     12                  2021
## 5                          casual                     01                  2022
## 6                          member                     01                  2022
## 7                          casual                     02                  2022
## 8                          member                     02                  2022
## 9                          casual                     03                  2022
## 10                         member                     03                  2022
## 11                         casual                     04                  2022
## 12                         member                     04                  2022
## 13                         casual                     05                  2022
## 14                         member                     05                  2022
## 15                         casual                     06                  2022
## 16                         member                     06                  2022
## 17                         casual                     07                  2022
## 18                         member                     07                  2022
## 19                         casual                     08                  2022
## 20                         member                     08                  2022
## 21                         casual                     09                  2022
## 22                         member                     09                  2022
## 23                         casual                     10                  2022
## 24                         member                     10                  2022
## 25                         casual                     11                  2022
## 26                         member                     11                  2022
##    alltrips_data_v2$ride_distance
## 1                        2.119170
## 2                        1.911230
## 3                        2.055370
## 4                        1.901026
## 5                        1.961231
## 6                        1.739570
## 7                        2.036470
## 8                        1.790109
## 9                        2.326197
## 10                       1.998785
## 11                       2.352637
## 12                       1.992720
## 13                       2.427993
## 14                       2.195223
## 15                       2.387477
## 16                       2.266431
## 17                       2.365507
## 18                       2.267968
## 19                       2.334128
## 20                       2.225028
## 21                       2.294084
## 22                       2.143383
## 23                       2.186665
## 24                       2.011601
## 25                         2.273025
## 26                         2.253723
```

Both groups like the summer months (May, June, July). However, during March 2022 there seems to be an uptick of casual riders.

Comparing the total amount of rides by members and casuals

How many total rides were taken by casual users vs annual members?

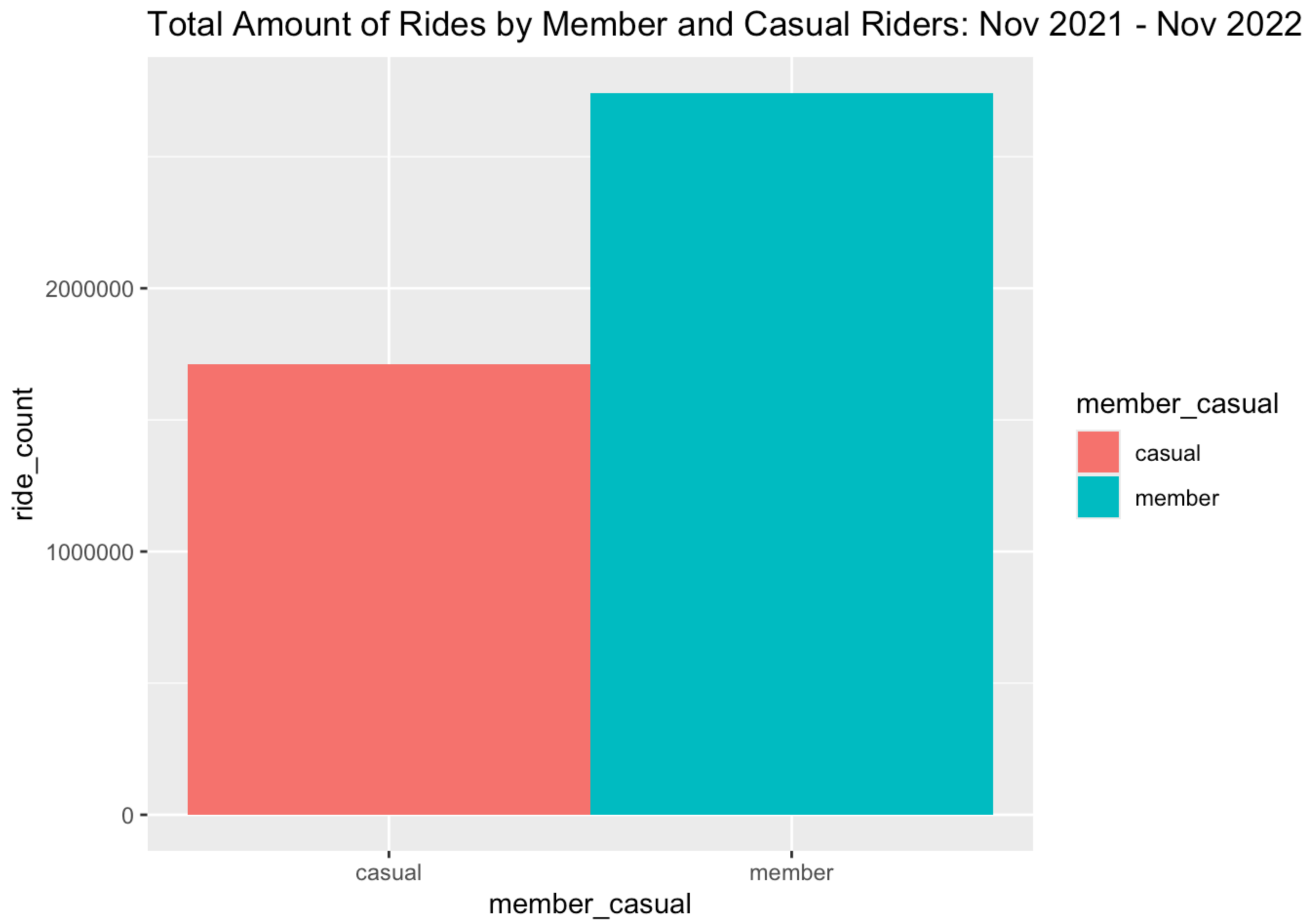
```
alltrips_data_v2 %>%
  group_by(member_casual) %>%
  summarise(ride_count = length(ride_id))
```

```
## # A tibble: 2 × 2
##   member_casual ride_count
##   <chr>              <int>
## 1 casual           1712698
## 2 member           2741989
```

Clearly members rode more times than casuals

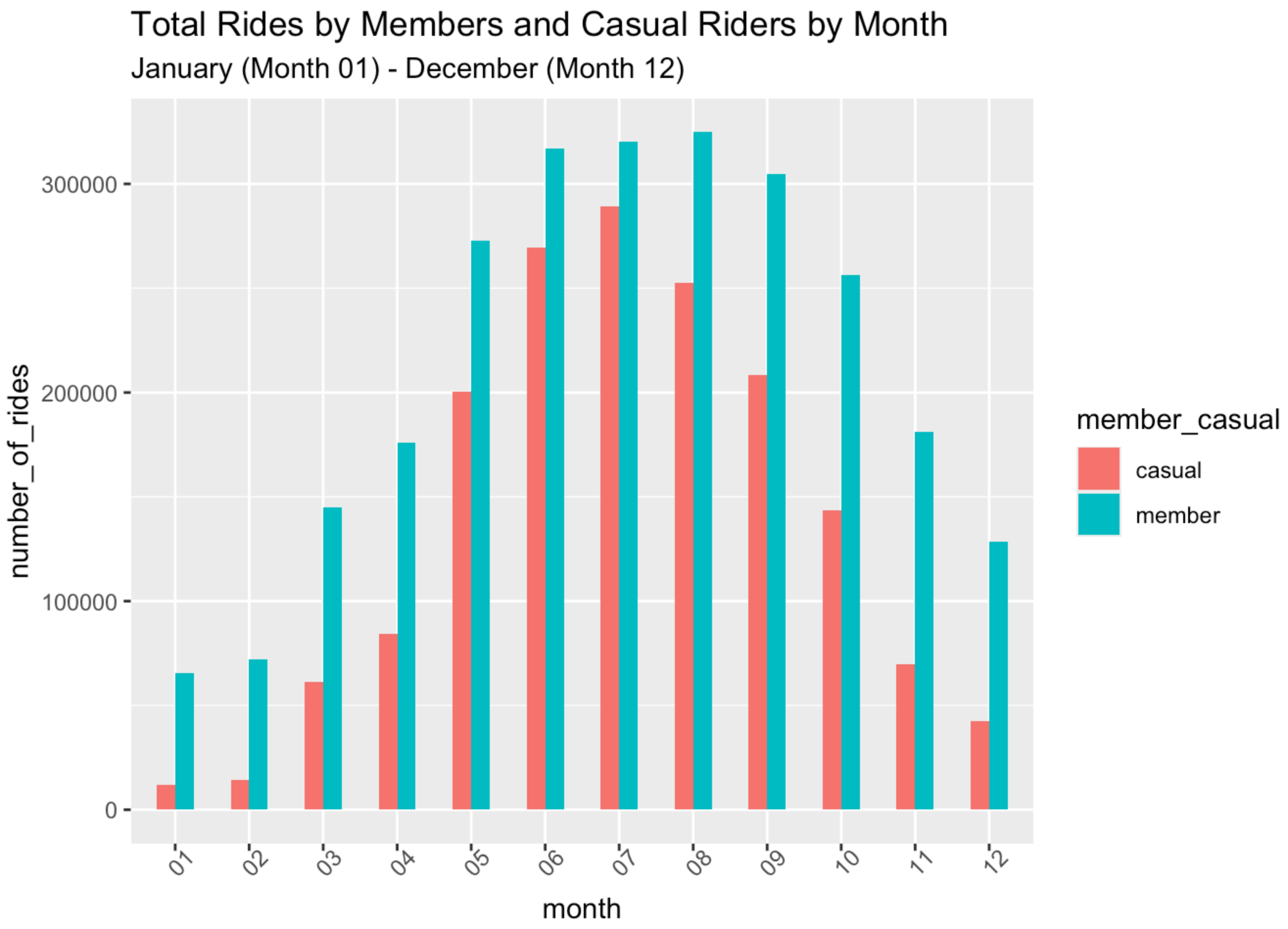
Let’s visualize this:

```
alltrips_data_v2 %>%
  group_by(member_casual) %>%
  summarise(ride_count = length(ride_id)) %>%
  ggplot(aes(x = member_casual, y = ride_count, fill = member_casual)) +
  geom_col(width = 1) +
  labs(title = "Total Amount of Rides by Member and Casual Riders: Nov 2021 - Nov 2022")
+
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```



Visualizing the number of rides each month of the year by each rider type

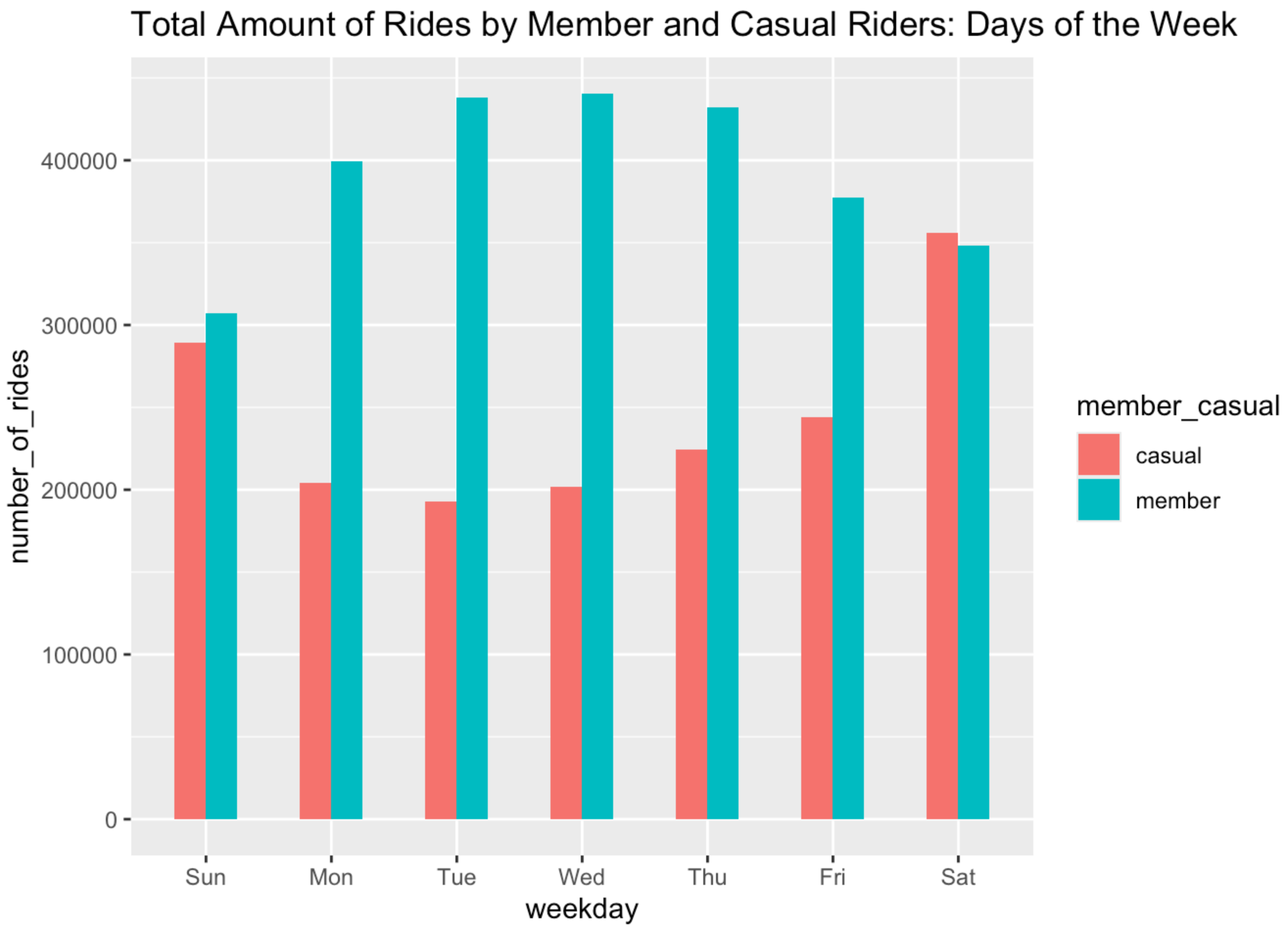
```
alltrips_data_v2 %>%
  group_by(member_casual, month, year) %>%
  summarise(number_of_rides = n(),.groups="drop") %>%
  arrange(member_casual, month, year)  %>%
  ggplot(aes(x = month, y = number_of_rides, fill = member_casual)) +
  labs(title ="Total Rides by Members and Casual Riders by Month", subtitle = "January
(Month 01) - December (Month 12)") +
  theme(axis.text.x = element_text(angle = 45)) +
  geom_col(width=0.5, position = position_dodge(width=0.5)) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```



Both types of users generally follow the same trends: Lower number of rides in the colder months and higher number of rides in the hotter and vacation months. **Members edge out casuals in all months.**

Visualizing the number of rides on days of the week by each rider type

```
alltrips_data_v2 %>%
  mutate(weekday = wday(started_at, label = TRUE)) %>% #Creates weekday field using wday
()
  group_by(member_casual, weekday) %>% # Groups by usertype and weekday
  summarise(number_of_rides = n(), average_duration = mean(ride_length)) %>% #Calculates
the number of rides and average duration
  arrange(member_casual, weekday) %>% #sorts
  ggplot(aes(x = weekday, y = number_of_rides, fill = member_casual)) +
  geom_col(width = 0.5,position = "dodge") +
  labs(title = "Total Amount of Rides by Member and Casual Riders: Days of the Week") +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE))
```



The visual shows that members ride more on the weekdays and casuals ride more on the weekends.