Cyclistic Case Study

Anthony Tetreault

2025-01-22

Cyclistic Trip Data Case Study

Visualizaing Usage

Libraries

```
library(tidyverse)
library(lubridate) #Handling dates/times from char to date/time
library(readxl) #Reading in Excel files
library(scales) #Eases manipulation of axis and their scales
mvc <- read_excel("member_v_casual.xlsx")
```

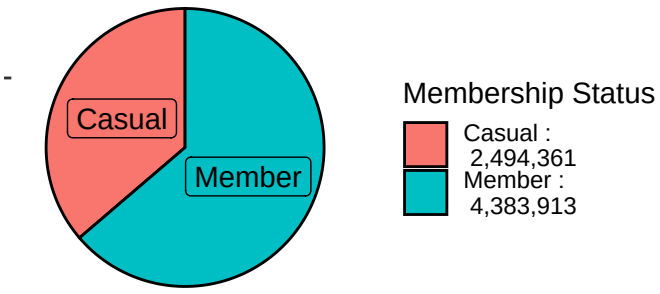
Rider Count by Membership

Setup

```
## Membership Totals
# Separate yearly counts by membership status
memberships <- aggregate(mvc$total_rides, list(membership=mvc$member_casual), FUN=sum)
```

Graph

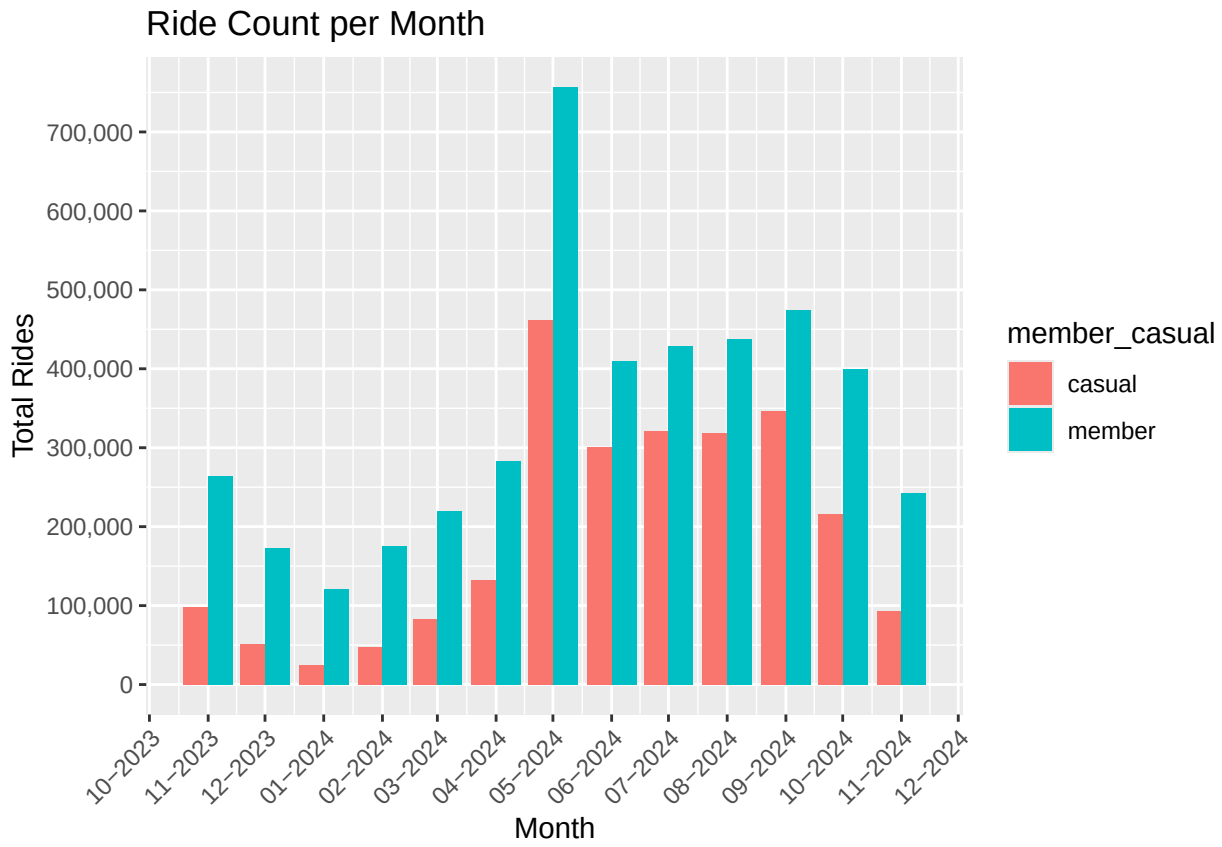
-



Summary Rides associated with Members accounted for approximately 63% of trips and casual 37%. A conservative goal would be to push that split to around 67/33% this year.

Total Rides

Graph



Summary For both membership types, a similar pattern emerges, although casual membership plays a much smaller role during certain portions of the year. January sees the lowest ride count of any month in the year, increasing at a mostly steady rate, with the exception of May, until a downturn around the beginning of Autumn, decreasing until end-of-year.

Due to the climate of the locale, one could infer both the decreased ridership in the Autumn/Winter months and the spike in rides in May might be directed by weather patterns, specifically temperatures (Chicago Yearly Tempeartures).The spike in May overlaps with the first nice days of the year and, regardless of tourism patterns, events celebrating the retreat from hibernation into the warmth of the sun. The same justification, yet in the opposite direction, may explain the ramp up to September and then decline through the end of year, as people are getting the last gasp of Summer, until the chill leaves only the most adventurous of our members riding.

This point, especially when paired with tourism patterns, could explain our casual rider numbers being markedly divergent from member rides during the Winter months, specifically. Not only are less people traveling during these months, but those that are are less likely to choose braving the elements as their primary mode of travel. Although that does not mean there aren't more people out there that could be enticed to brave those elements for environmental or economic reasons.

Average Ride Length

Setup

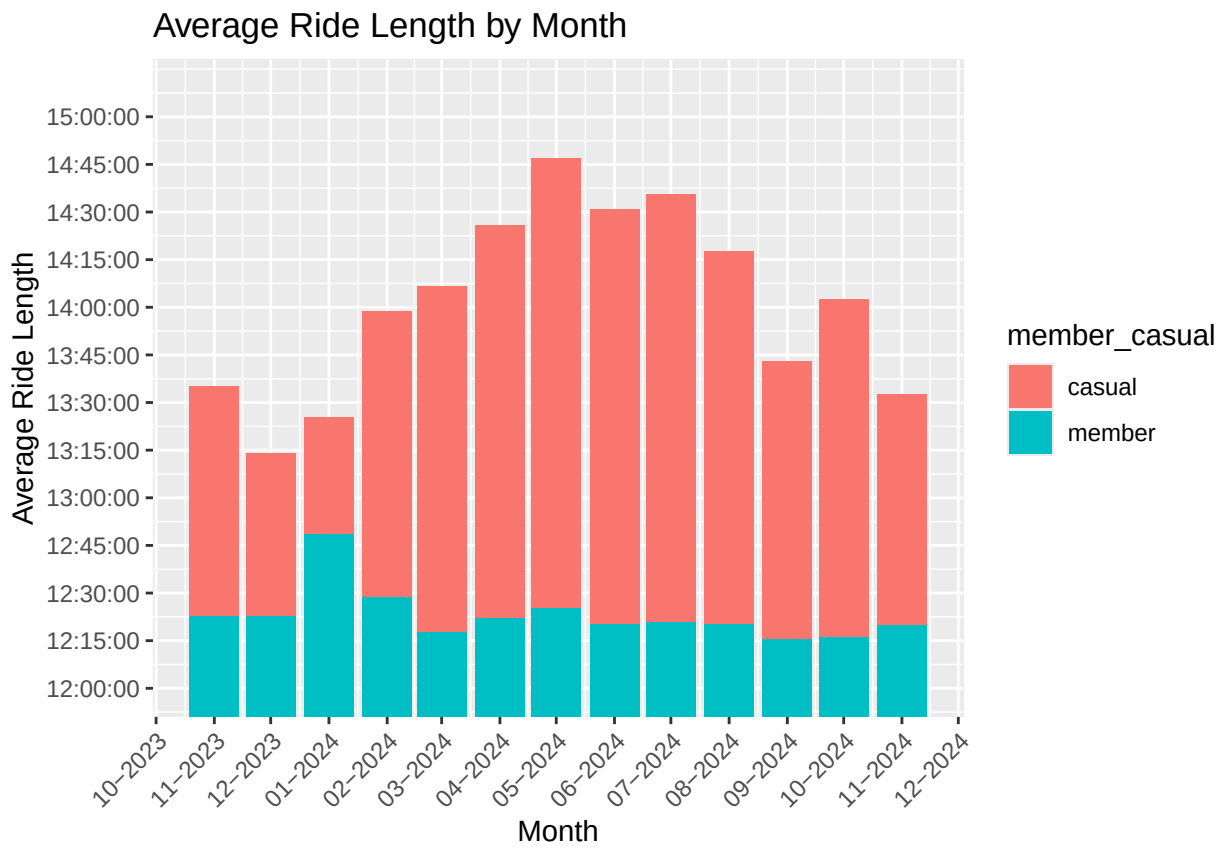
Due to the data types in the SQL language used in exploration (SQLite3), we need to convert from char to hms with lubridate (output in total seconds) and then cast that as.numeric(). With average ride length in seconds, we can better handle our scaling with the "scales" package.

```
# Convert avg_ride_length from HH:MM:SS format to a duration object
mvc$avg_ride_length_duration <- as.numeric(hms(mvc$avg_ride_length))
```

We also need to create a function to format y-axis labels appropriately (HH:MM:SS.)

```
# Custom function to format labels as HH:MM:SS
format_duration <- function(x) {
  sprintf("%02d:%02d:%02d", x %/% 3600, (x %% 3600) %/% 60, x %% 60)
}
```

Graph



Summary Average member ride duration throughout the year is relatively stable at around 12:15:00, although there is a spike in January that would need some further investigation. For casual members, the pattern closely matches our temperature graph cited earlier. We can expect longer trips, possibly more locations or longer outings from tourists trying to maximize experiences, due to the warmer temperature.

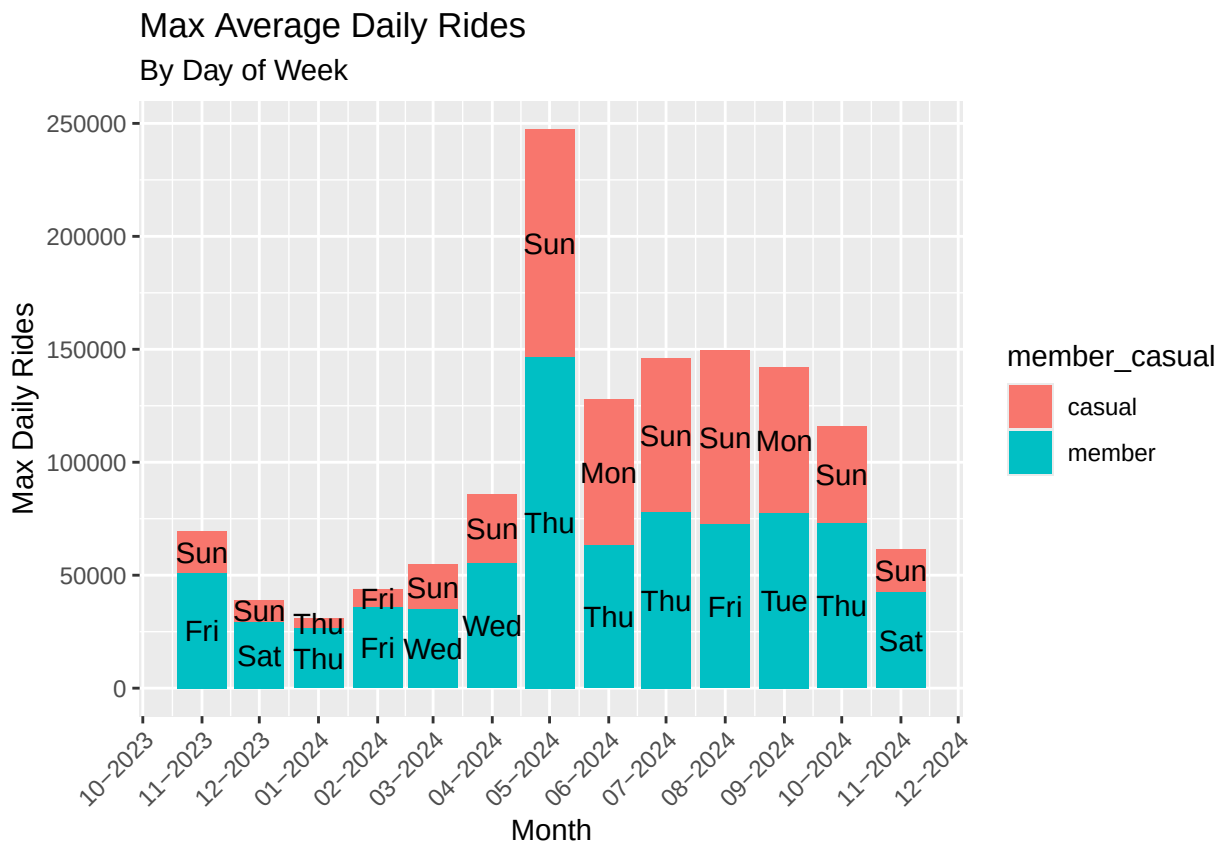
Max Average Daily Rides

Setup

Before graphing, we need to format Day_of_week from numeric to string representation of day to be used as labels on columns.

```
# convert Day_of_week to day abbreviations for labels
mvc$Day_abbr <- factor(mvc$Day_of_week, levels = 1:7, labels = c("Mon", "Tue", "Wed", "Thu", "Fri", "Sa
```

Graph



Summary Casual riders tend to ride the most on the weekend (Sunday) and members tend to ride more during the latter half of the week (Thursday.)