Cyclistic bike-share analysis case study

Sherry_Guirgis

2025-11-13

Introduction

Cyclistic is a bike-share company in Chicago that offers both annual memberships and casual, pay-as-you-go rides. The marketing team for the company believes that the company's future growth depends on converting casual riders into annual members.

As the junior data analyst on the marketing analytics team, the objective of this project is to understand how casual riders and annual members use Cyclistic bikes differently. In this manner, using historical ride data, we analyzed ride counts, trip durations, bike types, weekdays, months of the year, and even seasons of the year in order to put data behind marketing strategic recommendations that would increase the conversion of annual membership.

Executive Summary

This analysis examines the usage patterns of Cyclistic's bike-share system, comparing annual members and casual riders across months, weekdays, bike types, and seasons. The findings reveal that members ride more frequently and consistently throughout the year, primarily for commuting purposes, while casual riders ride less often but take longer trips, particularly on weekends and during summer months. Casual riders also show a strong preference for electric bikes, highlighting their recreational use. These insights indicate clear behavioral differences between members and casual riders, providing actionable guidance for marketing strategies aimed at converting casual riders into annual members, such as targeted seasonal campaigns, promoting electric bikes, and emphasizing the benefits of membership for frequent leisure riders.

Note: Basic data cleaning (e.g., date formatting, duration calculation, column removal) was performed in Excel before importing.

Analysis

Download the needed Packages

```r
library("tidyverse")

library("ggplot2")

library("readr")

library("lubridate")
```

```
library("dplyr")

library("readxl")

library('scales')
```

Data Cleaning Done in Excel

Version 1.0.0 (06-11-2025)

New

```
- Added column "Week_day" to give each day of the week a number starting with Sunday as 1, and set it's
- Added Column "Trip_Length" to calculate the time taken for each ride.
```

Changes

```
- Changed date format to M-D-YYYY h:mm ("started_at" & "ended_at" columns).
- Removed the following columns (start_station_name, start_station_id, end_station_name, end_station_id
```

These changes were done to each month (each work sheet of the project).

Get the needed data and setting the working directory

```
setwd("C:\\Users\\Enter Computer\\OneDrive\\Documents\\binaries")

file112024 <- read_excel("202411-divvy-tripdata.xlsx", sheet = "202411-divvy-tripdata")

file122024 <- read_excel("202412-divvy-tripdata.xlsx", sheet = "202412-divvy-tripdata")

file012025 <- read_excel("202501-divvy-tripdata.xlsx", sheet = "202501-divvy-tripdata")

file022025 <- read_excel("202502-divvy-tripdata.xlsx", sheet = "202502-divvy-tripdata")

file032025 <- read_excel("202503-divvy-tripdata.xlsx", sheet = "202503-divvy-tripdata")

file042025 <- read_excel("202504-divvy-tripdata.xlsx", sheet = "202504-divvy-tripdata")

file052025 <- read_excel("202505-divvy-tripdata.xlsx", sheet = "202505-divvy-tripdata")

file062025 <- read_excel("202506-divvy-tripdata.xlsx", sheet = "202506-divvy-tripdata")

file072025 <- read_excel("202507-divvy-tripdata.xlsx", sheet = "202507-divvy-tripdata")

file082025 <- read_excel("202508-divvy-tripdata.xlsx", sheet = "202508-divvy-tripdata")

file092025 <- read_excel("202509-divvy-tripdata.xlsx", sheet = "202509-divvy-tripdata")

file102025 <- read_excel("202510-divvy-tripdata.xlsx", sheet = "202510-divvy-tripdata")
```

Merging all files into one

```
trips_data <- bind_rows(file112024, file122024, file012025, file022025, file032025, file042025, file052
```

Convert started_at to proper datetime

```
trips_data$started_at <- as.POSIXct(trips_data$started_at, format = "%m-%d-%Y %H:%M")

## Seperate the started_at column to dates & times

trips_data$date <- format(trips_data$started_at, "%m-%d-%Y")

trips_data$time <- format(trips_data$started_at, "%H:%M")

## Remove Trip_Length Column

trips_data$Trip_Length <- NULL

## Calculate the trip length

trips_data$trip_length_mins <- as.numeric(
  difftime(trips_data$ended_at, trips_data$started_at, units = "mins")
)

## Some trips end the next day which gives negative values

trips_data$trip_length_mins <- ifelse(
  trips_data$trip_length_mins < 0,
  trips_data$trip_length_mins + 24*60,  # add 24 hours in minutes
  trips_data$trip_length_mins
)

## Extract month and weekday names

trips_data$month_name <- format(trips_data$started_at, "%B")

trips_data$weekday_name <- format(trips_data$started_at, "%A")

## To make sure data is correct

unique(trips_data$rideable_type)

unique(trips_data$member_casual)

drop_na(trips_data)

### Data shown is "electric_bike" "classic_bike"  for rideable type and "member" "casual" for member_ca

## Filter for trips that are less than 0 minutes

filter(trips_data, trip_length_mins < 0)
```

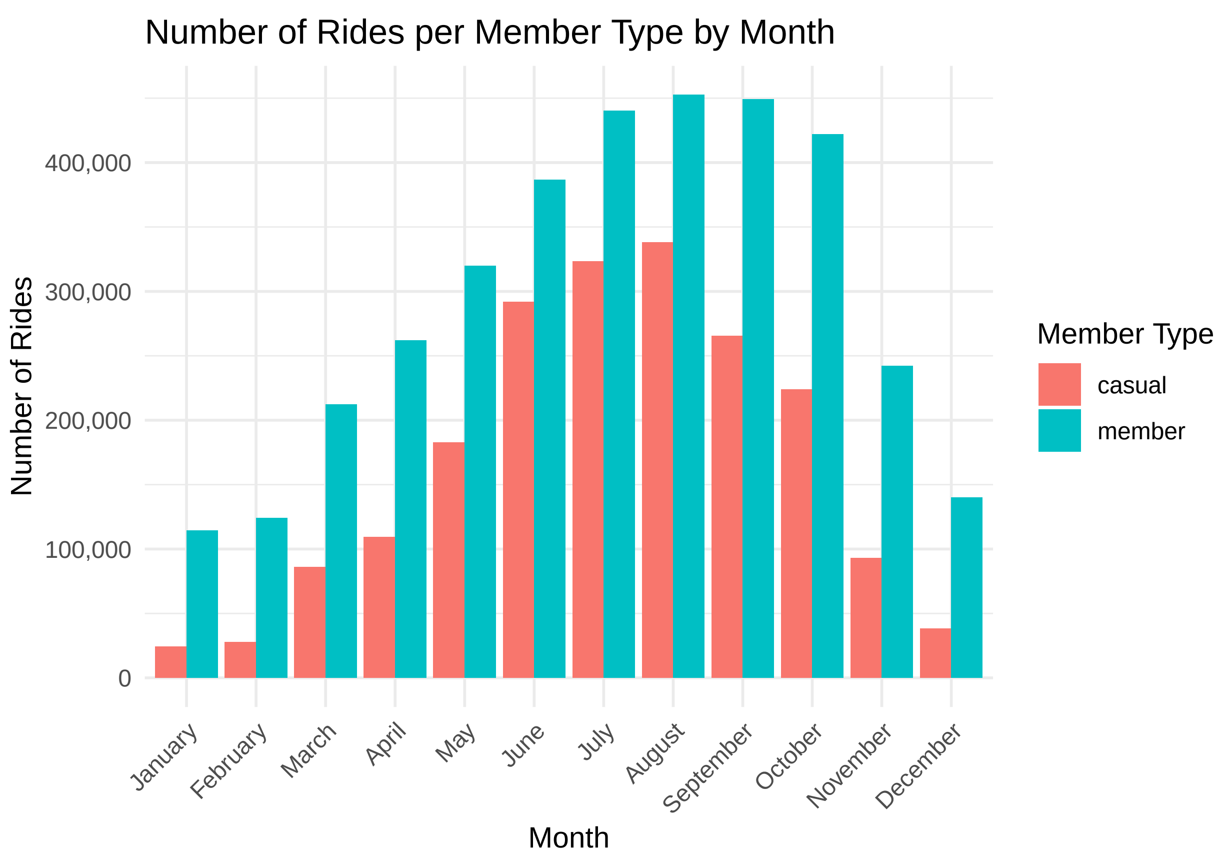
```
### no points were below zero

## Calculate number of rides, Avg & Max trip length per member type for each month

rides_per_month <- trips_data %>%
  group_by(member_casual, month_name) %>%
  summarise(num_rides = n(), max_ride = max(trip_length_mins), average_trip_length = mean(trip_length_mi
```

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

```
## To arrange month names

rides_per_month$month_name <- factor(
  rides_per_month$month_name,
  levels = month.name
)
```

visualizations

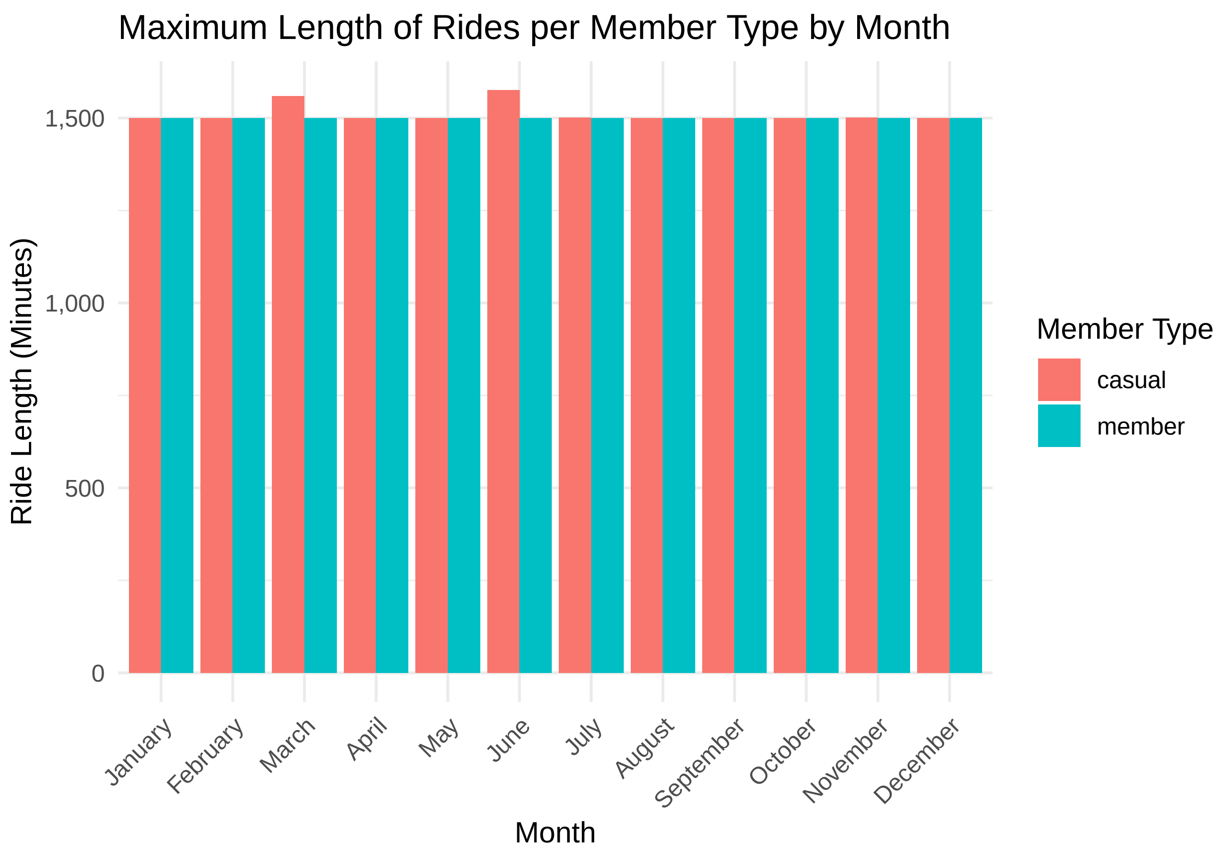
Number of rides



Members take more rides year-round, while casual riders peak in summer (June–August).

Casual ridership drops sharply in winter, showing stronger seasonal dependence.

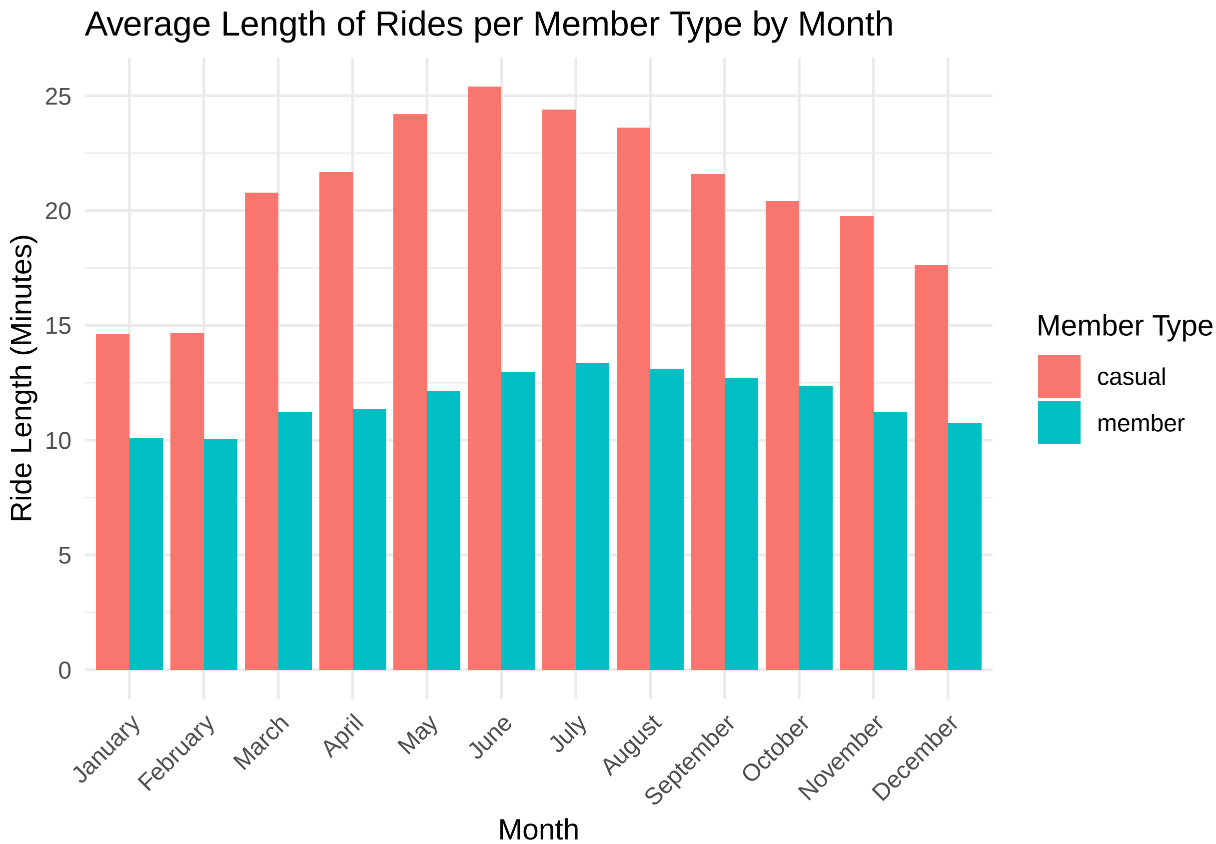
Maximum ride length



Maximum ride lengths are similar for members and casual riders across months.

This suggests that extreme trip durations occur for both groups, so max ride length isn't a strong differentiator.

Average trip length



Casual riders' trips are roughly twice as long as members'.

Members ride shorter, more consistent trips typical of commuting patterns.

Calculating Number of rides per bike ride for members every month

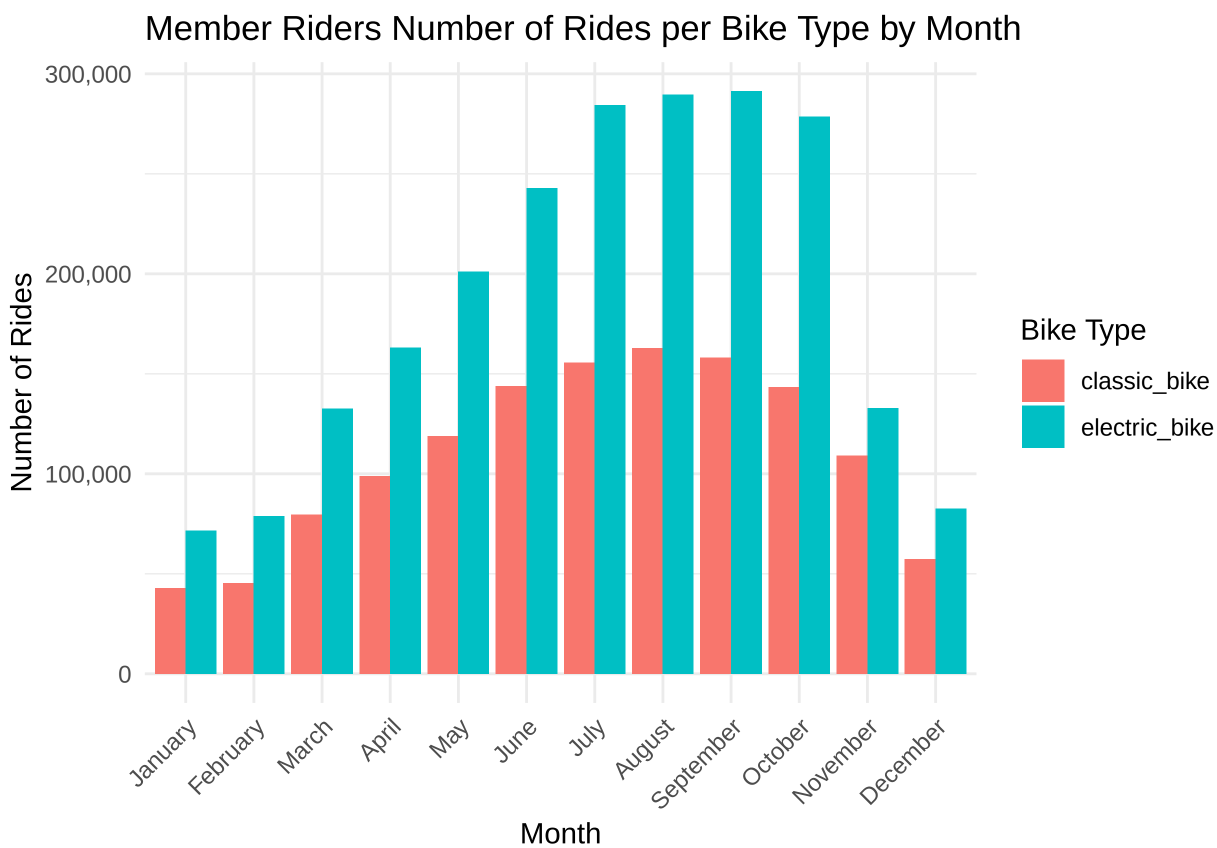
```
bike_type_per_month <- trips_data %>%
  group_by(rideable_type, month_name, member_casual) %>%
  filter(member_casual == 'member') %>%
  summarise(number_per_bike_type = n())
```

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

```
#To arrange month names
bike_type_per_month$month_name <- factor(
  bike_type_per_month$month_name,
  levels = month.name
)
```

visualizations

Number of rides / rideable type for members



Members ride both bike types, with electric bikes slightly more popular. Usage peaks in summer.

Calculating Number of rides per bike ride for casual riders every month

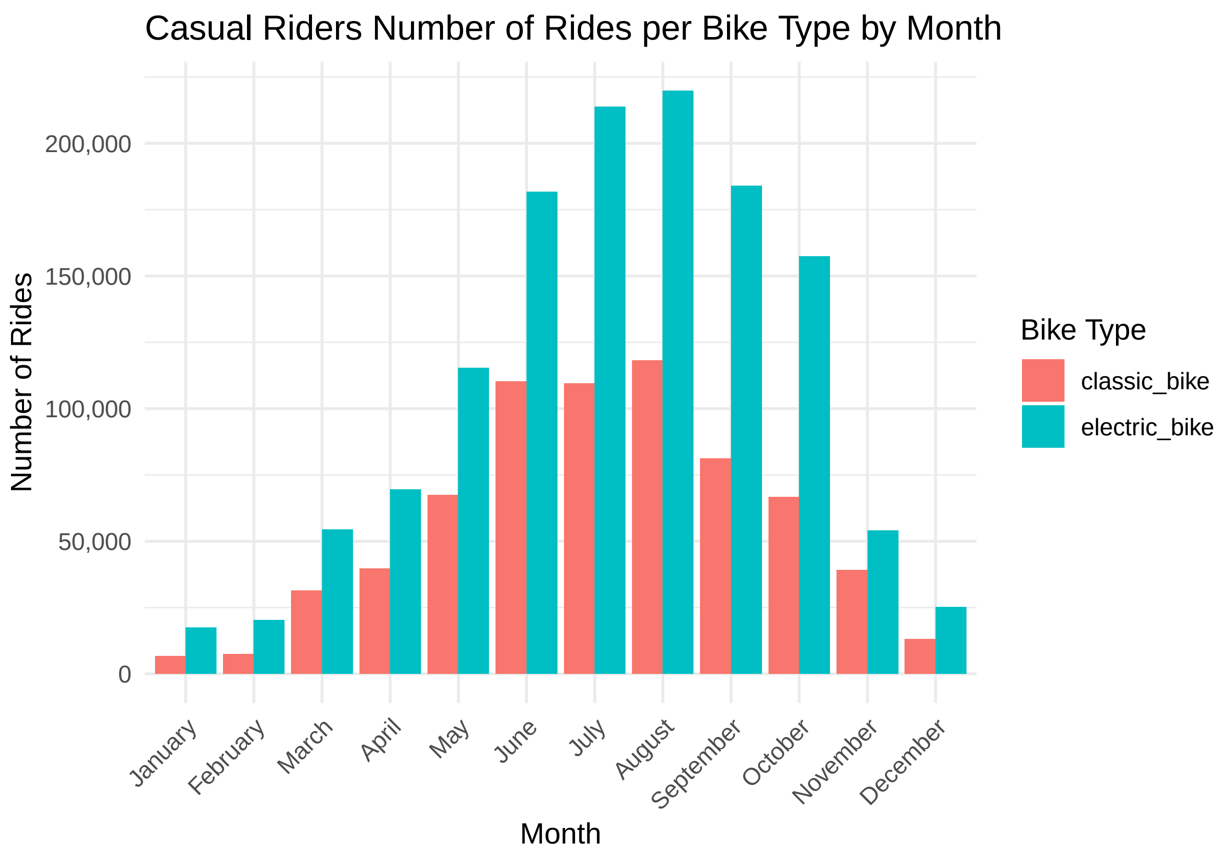
```
bike_type_per_month <- trips_data %>%
  group_by(rideable_type, month_name, member_casual) %>%
  filter(member_casual == 'casual') %>%
  summarise(number_per_bike_type = n())
```

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

```
#To arrange month names
bike_type_per_month$month_name <- factor(
  bike_type_per_month$month_name,
  levels = month.name)
```

visualizations

Number of rides / rideable type for casual riders



Casual riders ride more electric bikes than classic bikes, with usage peaking in summer and dropping in winter.

Insight

Members use bikes more evenly across types and months, reflecting commuting or regular use, whereas casual riders show a strong preference for electric bikes in summer, indicating leisure or recreational use. This difference can help target marketing strategies — for example, promoting seasonal membership options to casual electric bike users.

Stats per weekday

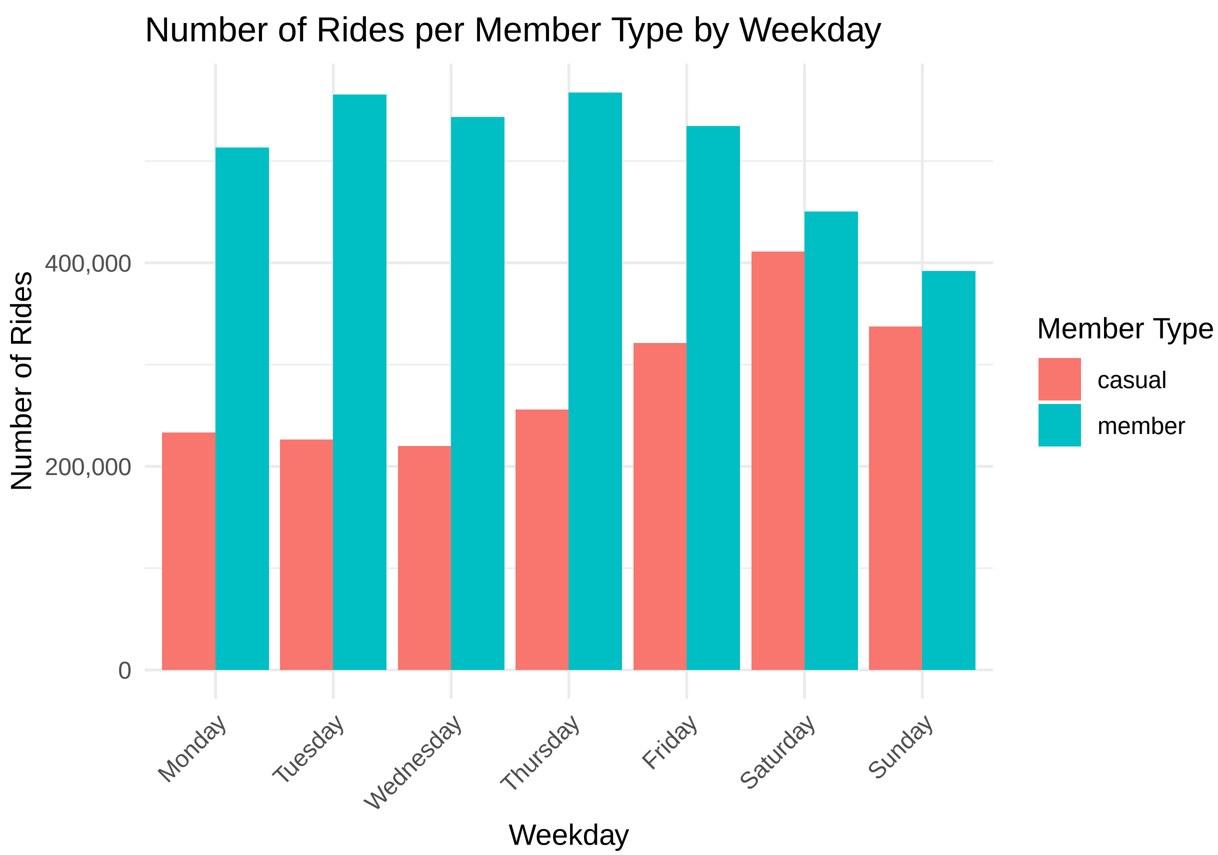
```
weekday_trips <- trips_data %>%
  group_by(weekday_name, member_casual) %>%
  summarise(weekday_trip_no = n(), avg_weekday_trip = mean(trip_length_mins), max_weekday_trip = max(tri
```

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

```
#to arrange weekdays
weekday_trips$weekday_name <- factor(weekday_trips$weekday_name,
                                     levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Sa
```

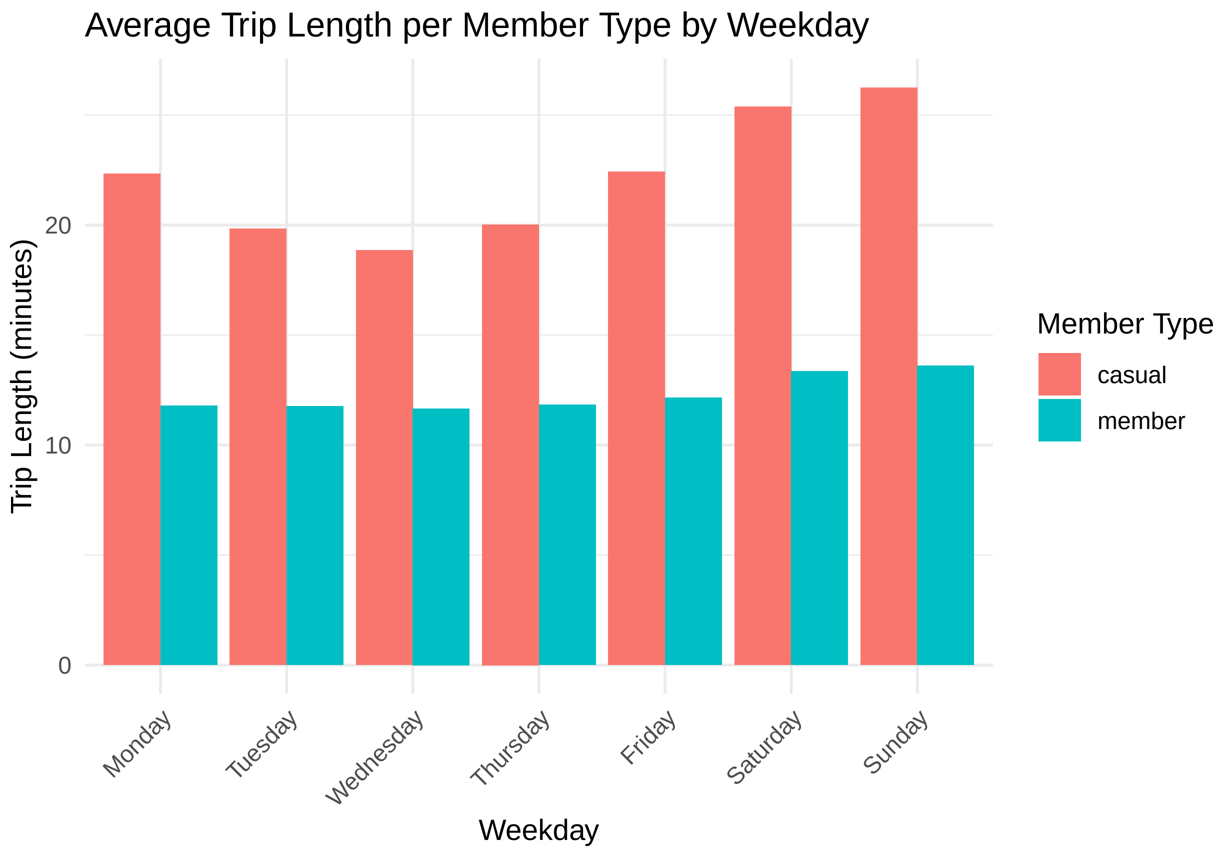
visualizations

Number of Rides per Weekday



Members ride more frequently during weekdays, while casual riders peak on weekends, showing leisure-oriented usage.

Averange trip length



Casual riders take longer trips than members every day, especially on weekends.

Maximum Trip Length

Maximum trip lengths are similar across members and casuals, so extreme durations don't differentiate rider types.

Stats per Season

```
summer_trips <- trips_data %>%
  filter (month_name %in% c("June", "July", "August")) %>%
  group_by(member_casual, month_name) %>%
  summarise(number_of_rides = n(), avg_duration = mean(trip_length_mins), max_trip_duration = max(trip_l
  mutate(season = "Summer")
```

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

```
fall_trips <- trips_data %>%
  filter (month_name %in% c("September", "October", "November")) %>%
  group_by(member_casual, month_name) %>%
  summarise(number_of_rides = n(), avg_duration = mean(trip_length_mins), max_trip_duration = max(trip_l
  mutate(season = "Fall")
```

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

```
winter_trips <- trips_data %>%
  filter (month_name %in% c("December", "January", "February")) %>%
  group_by(member_casual, month_name) %>%
  summarise(number_of_rides = n(), avg_duration = mean(trip_length_mins), max_trip_duration = max(trip_l
  mutate(season = "Winter")
```

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

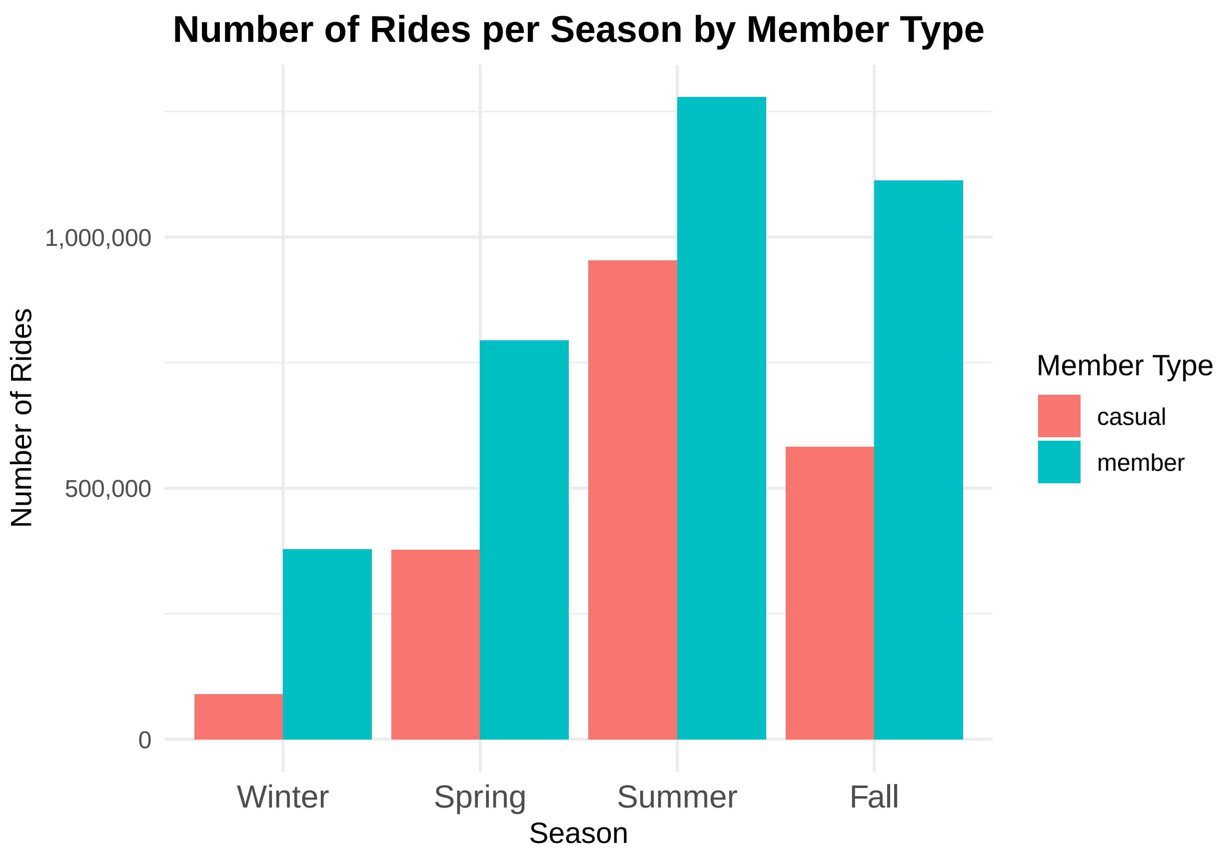
```
spring_trips <- trips_data %>%
  filter (month_name %in% c("March", "April", "May")) %>%
  group_by(member_casual, month_name) %>%
  summarise(number_of_rides = n(), avg_duration = mean(trip_length_mins), max_trip_duration = max(trip_l
  mutate(season = "Spring")
```

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

```
### Combine them all into one dataset

seasonal_summary <- bind_rows(summer_trips, fall_trips, winter_trips, spring_trips)

rides_per_season <- seasonal_summary %>%
  group_by(member_casual, season) %>%
  summarise( number_of_rides = sum(number_of_rides), avg_duration = mean(avg_duration))
```

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

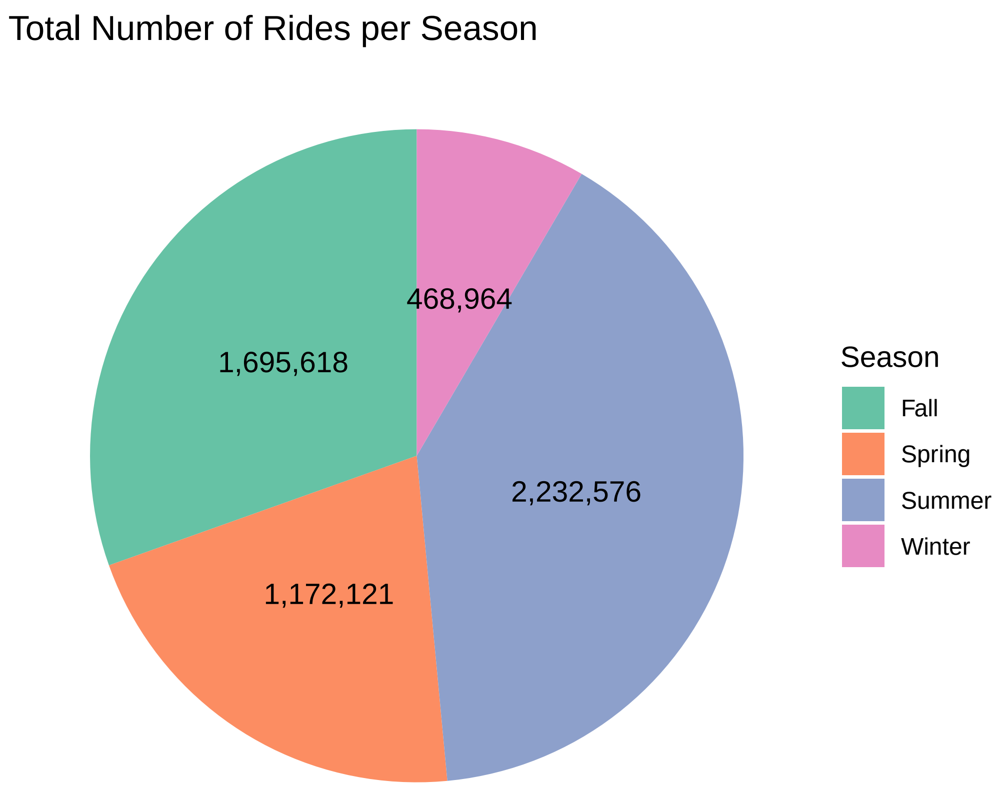
```
## Visualizations

rides_per_season$season <- factor(rides_per_season$season,
                                  levels = c("Winter", "Spring", "Summer", "Fall"))
```



Members ride consistently across all seasons, peaking in summer, while casual riders show strong seasonal variation, with the highest rides in summer and the lowest in winter.

Number of Rides per Season

```
total_rides_per_season <- seasonal_summary %>%
  group_by(season) %>%
  summarise( total_number_of_rides = sum(number_of_rides), avg_duration = mean(avg_duration))
```



Summer accounts for the highest number of rides, followed by fall and spring, while winter has the lowest.

This highlights strong seasonal trends in Cyclistic usage overall.

Summary / Answers to Key Questions

1. How do members and casual riders differ in their overall usage?

Members ride more frequently throughout the year and consistently use the bikes across months and seasons.

Casual riders ride less often, but the trips are generally longer.

Members travel more for commutes, whereas casuals ride much more for recreation/leisure on weekends and during summer.

2. How does the usage vary by month?

Members: The ride counts are high throughout the year, peaking during summer months.

Casuals: The counts peak in summer, drop sharply in winter, showing strong seasonal dependence.

Average trip length: casual riders take longer rides than members in all months.

3. How does usage vary by weekday?

Members: ride most on weekdays, with peaks on Tuesday and Thursday, consistent with commuting behavior.

Casuals: They ride most on weekends, especially Saturday and Sunday, indicating leisure trips.

Average length of trip: Casual rides are about twice as long as member rides, while weekend trips are the longest.

4. How does usage vary by bike type?

Its members ride both classic and electric bikes, but electric bikes are slightly more popular in general.

Casual riders really favor electric bikes, especially during summer months.

This suggests that casual riders take longer, more recreational trips with electric bikes.

5. How does usage vary by season?

Summer: highest number of rides and longest trips for both members and casuals.

Winter: lowest ride counts, in particular for casual riders.

Spring and fall: moderate usage, with casuals taking longer rides than members.

Overall, casual riders' usage is highly seasonal, while members are more consistent.

6. What are the actionable insights one may recommend?

Target casual riders during summer and weekends with membership incentives, such as seasonal or flexible memberships.

Promote electric bicycles in marketing campaigns to casual riders for use in longer, recreational trips.

Emphasize membership cost and convenience benefits to casual riders who frequently take summer rides.

Use weekday versus weekend usage patterns to tailor marketing messages: commute-focused benefits for members, leisure/recreational benefits for casuals.