

Cyclistic Case Study: Ride Analysis for Stakeholders

Gabriela Olivera

2025-12-03

Executive Summary

The purpose of this case study is to analyse **how annual members and casual riders use Cyclistic bikes differently**, providing actionable insights for stakeholders.

Key questions addressed:

- What are the differences in ride duration and frequency between **Casual** and **Annual** users?
- How does usage vary by **weekday vs weekend**, day of the week, month, and hour of day?
- What recommendations can be made for **fleet management, marketing, and operational planning**?

Data Overview

The dataset combines Cyclistic trip data for Q1 2019 and Q1 2020.

- Rows: Each row represents a single bike ride.
- Columns: Ride ID, start/end times, user type, ride duration, day of week, month, start hour.
- **Notes:**
 - Some columns contain NA values (e.g., `rideable_type` or columns only present in 2019). These do not affect the main analysis.
 - Data is cleaned to remove invalid or extreme rides (ride length 0 or >24 hours).

Load Dataset

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.6
## v forcats    1.0.1      v stringr    1.6.0
## v ggplot2    4.0.1      v tibble     3.3.0
## v lubridate  1.9.4      v tidyr      1.3.1
```

```
## v purrr      1.2.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# Load the full cleaned dataset
cyclistic_full <- read_csv("cyclistic_cleaned_Q1_2019_2020.csv")
```

```
## Rows: 20000 Columns: 31
## -- Column specification -----
## Delimiter: ","
## chr  (12): from_station_name, to_station_name, gender, ride_id, rideable_type...
## dbl  (13): bikeid, tripduration, from_station_id, to_station_id, birthyear, ...
## dtm   (2): started_at, ended_at
## date  (2): start_date, end_date
## time  (2): start_time_only, end_time_only
##
## 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.
```

```
# Create a smaller sample for knitting (20k rows)
cyclistic <- cyclistic_full %>% sample_n(20000)
```

```
# Optional: save the sample dataset for reference
write_csv(cyclistic, "cyclistic_sample.csv")
```

Notes for stakeholders:

- We are using a subset of 20,000 rides to make the report efficient and readable.
- The full dataset contains all rides from Q1 2019 and Q1 2020.
- All analyses and visualisations use this sample but reflect the overall trends.

Data Cleaning & Preparation

The following transformations were applied to prepare the dataset for analysis:

- Column names were standardised for consistency (ride_id, started_at, ended_at, member_casual).
- Ride length was calculated in minutes for comparison purposes.
- Day of week, month, and start hour were extracted to analyse trends over time.
- Days were classified as **Weekday** or **Weekend**.
- Users were grouped into **Casual** (Customer + casual) and **Annual** (member + Subscriber) for clearer comparison.

Quick Summary by User Type Group

This summary shows **average ride length and number of rides per user type**, giving a high-level view of usage patterns.

```
# Quick summary of ride counts and average duration by user type
cyclistic %>%
  group_by(user_type_group) %>%
  summarise(
    num_rides = n(),
    avg_ride_length = mean(ride_length),
    .groups = "drop"
  )
```

```
## # A tibble: 2 x 3
##   user_type_group num_rides avg_ride_length
##   <chr>           <int>         <dbl>
## 1 Annual           18216           11.5
## 2 Casual           1784           36.8
```

Notes / Stakeholder interpretation:

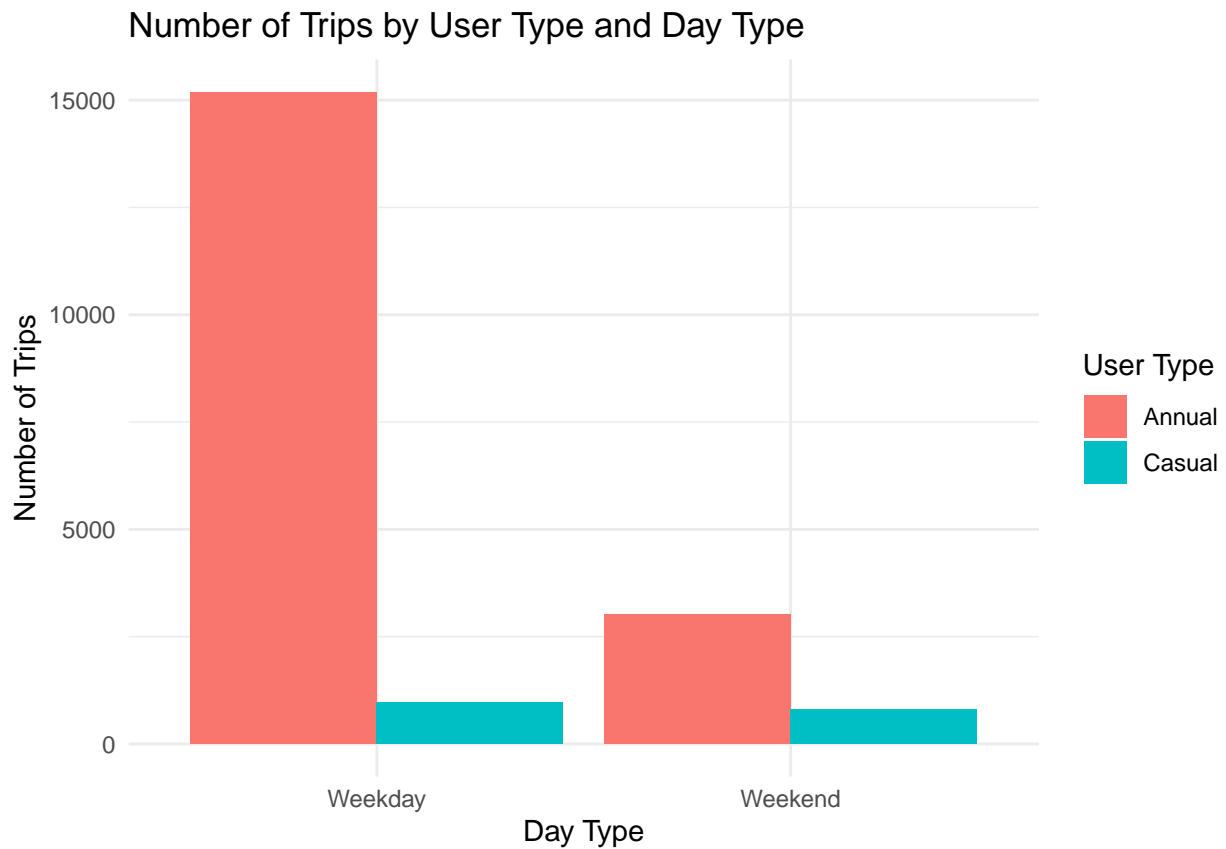
- **Annual** users tend to take more frequent but shorter rides.
 - **Casual** users take fewer rides but for longer durations.
-

Exploratory Data Analysis (EDA)

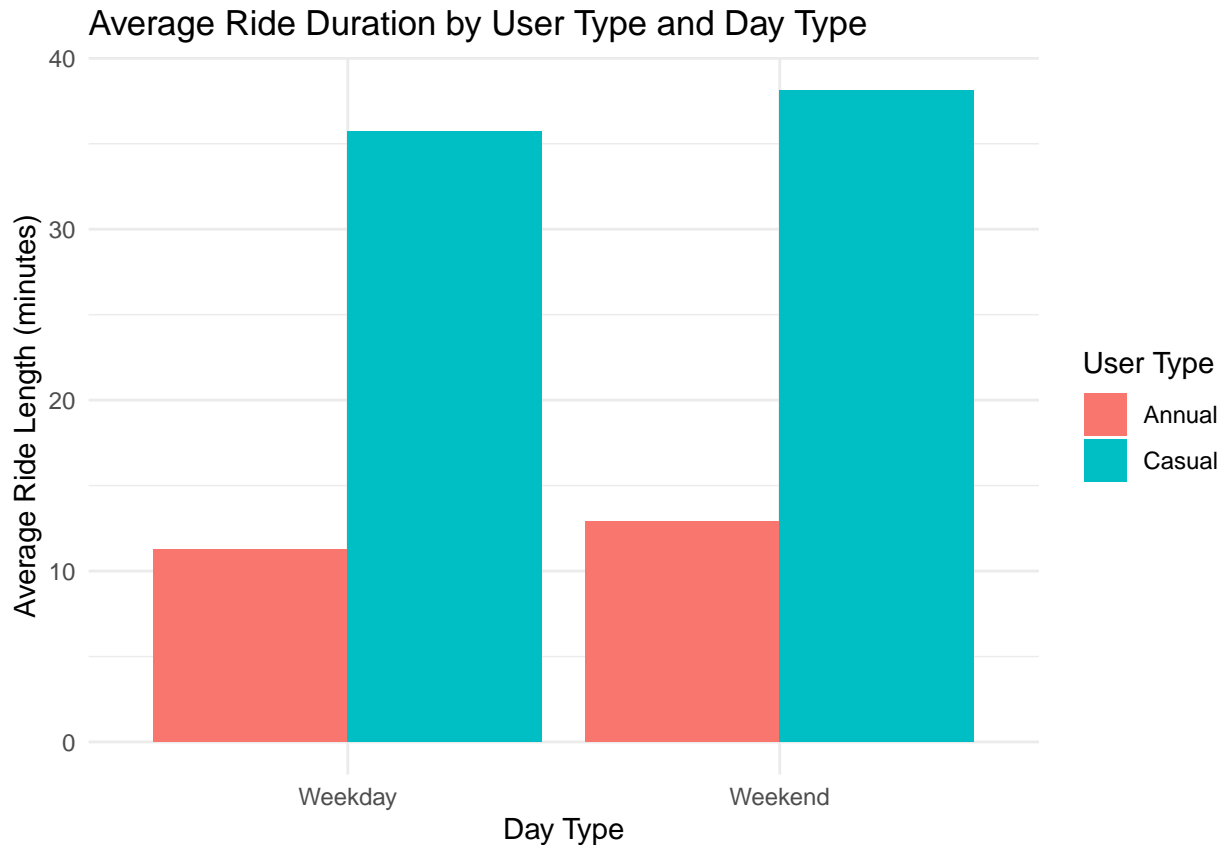
1. Trips by Day Type (Weekday vs Weekend)

Objective: Compare usage patterns between **Casual** and **Annual** users on weekdays vs weekends.

```
# Number of trips by user type and day type
ggplot(cyclistic, aes(x = day_type, fill = user_type_group)) +
  geom_bar(position = "dodge") +
  labs(
    title = "Number of Trips by User Type and Day Type",
    x = "Day Type",
    y = "Number of Trips",
    fill = "User Type"
  ) +
  theme_minimal()
```



```
# Average ride duration by user type and day type
ggplot(cyclistic, aes(x = day_type, y = ride_length, fill = user_type_group)) +
  stat_summary(fun = mean, geom = "bar", position = "dodge") +
  labs(
    title = "Average Ride Duration by User Type and Day Type",
    x = "Day Type",
    y = "Average Ride Length (minutes)",
    fill = "User Type"
  ) +
  theme_minimal()
```



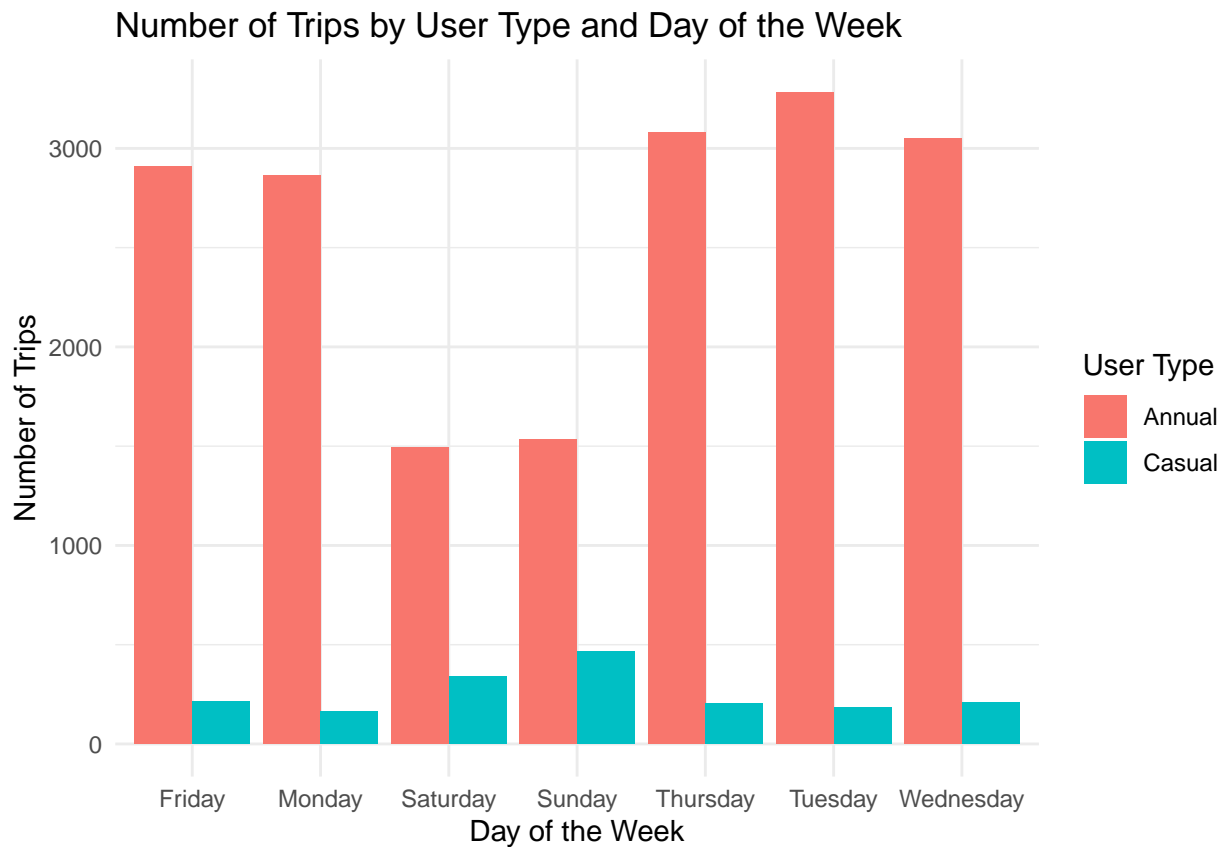
Stakeholder Observations:

- **Casual riders** show higher activity on weekends and tend to take **longer rides**.
- **Annual riders** ride mostly on weekdays, reflecting **commuting patterns**.

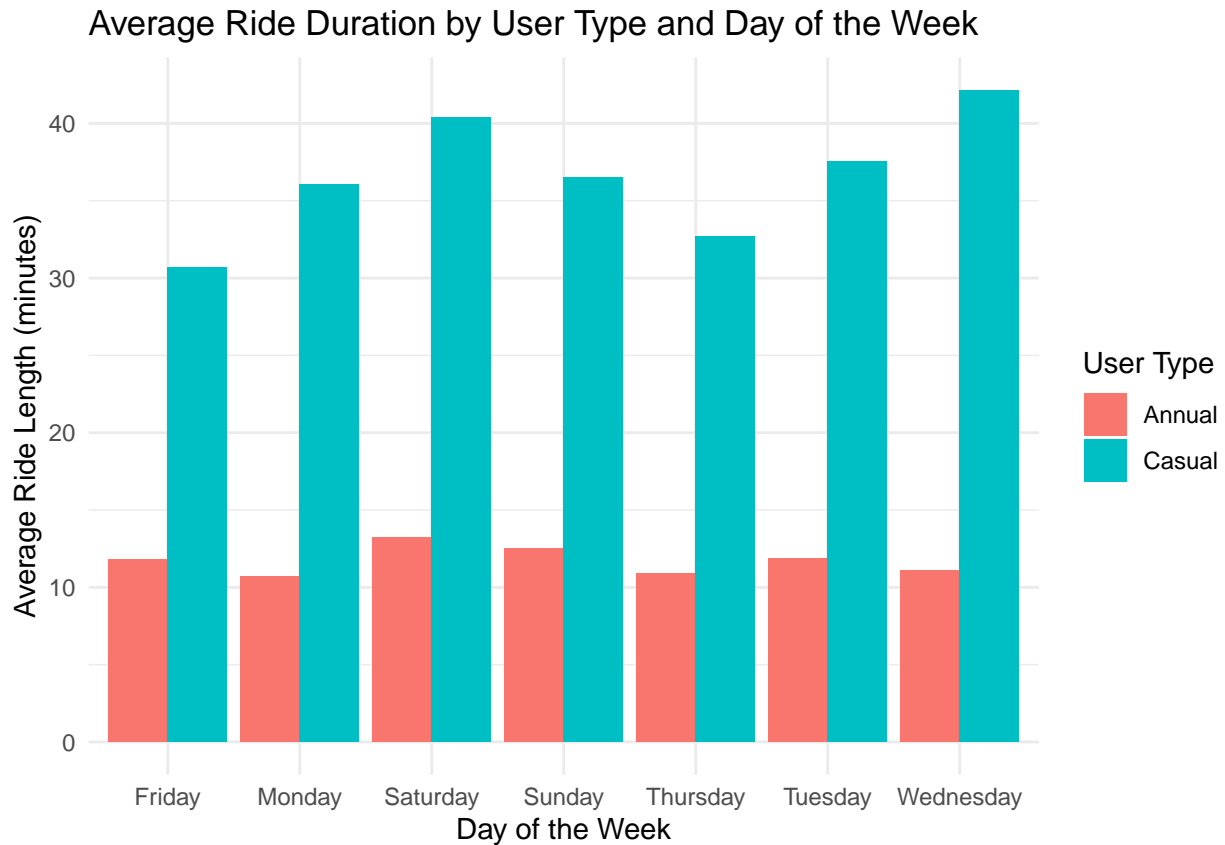
2. Trips by Day of the Week

Objective: Understand how bike usage varies across the week for **Casual** vs **Annual** users.

```
ggplot(cyclistic, aes(x = day_of_week, fill = user_type_group)) +
  geom_bar(position = "dodge") +
  labs(
    title = "Number of Trips by User Type and Day of the Week",
    x = "Day of the Week",
    y = "Number of Trips",
    fill = "User Type"
  ) +
  theme_minimal()
```



```
ggplot(cyclistic, aes(x = day_of_week, y = ride_length, fill = user_type_group)) +
  stat_summary(fun = mean, geom = "bar", position = "dodge") +
  labs(
    title = "Average Ride Duration by User Type and Day of the Week",
    x = "Day of the Week",
    y = "Average Ride Length (minutes)",
    fill = "User Type"
  ) +
  theme_minimal()
```

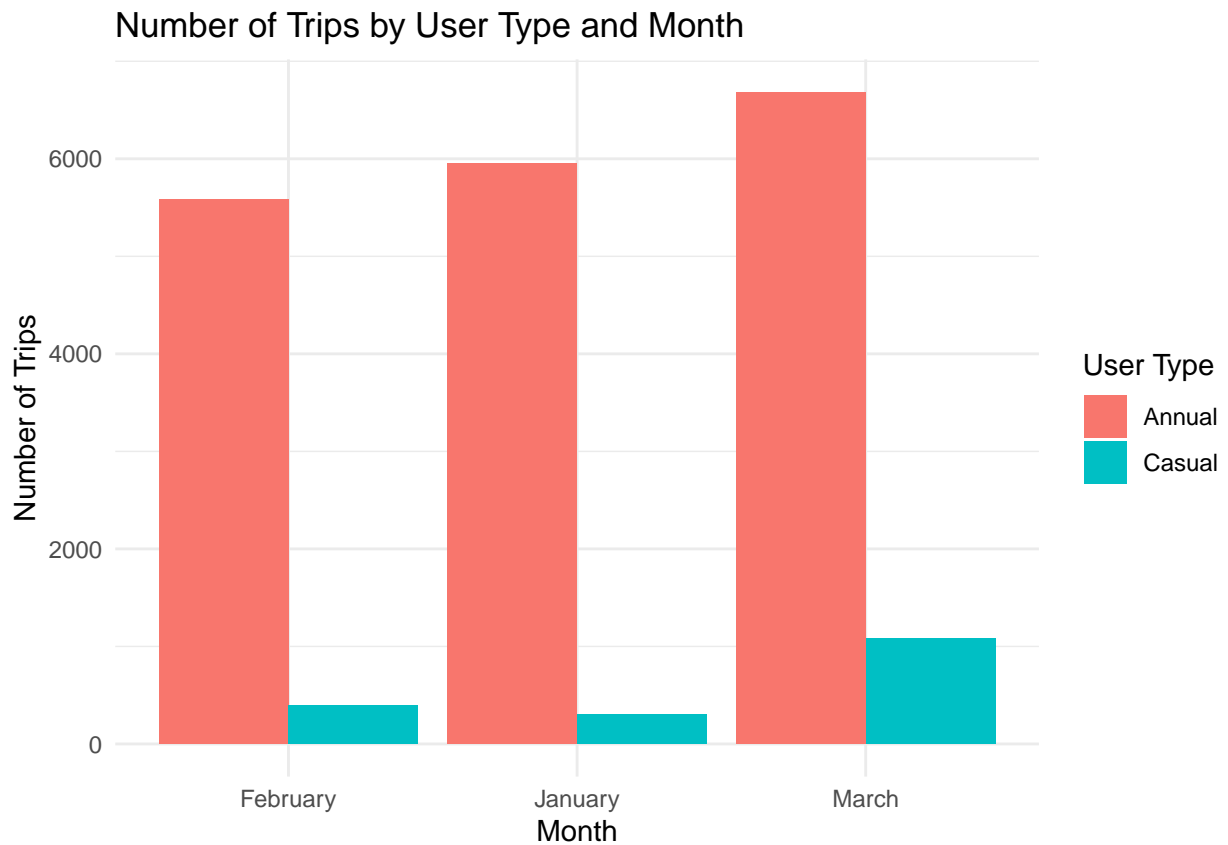


Insights:

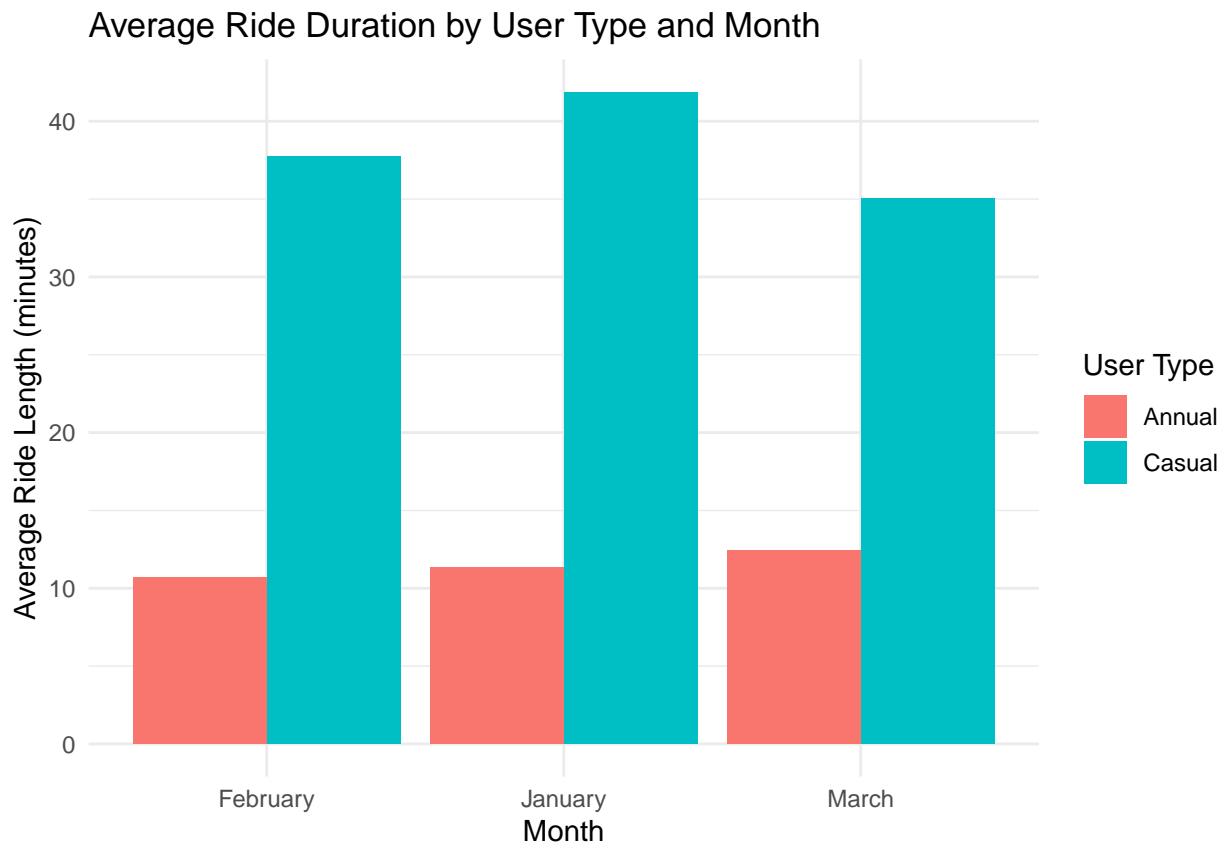
- Casual riders peak on **Saturday** and **Sunday**.
- Annual riders show steady **weekday** usage with **slight peaks mid-week**.

3. Monthly Trends

```
ggplot(cyclistic, aes(x = month, fill = user_type_group)) +
  geom_bar(position = "dodge") +
  labs(
    title = "Number of Trips by User Type and Month",
    x = "Month",
    y = "Number of Trips",
    fill = "User Type"
  ) +
  theme_minimal()
```



```
ggplot(cyclistic, aes(x = month, y = ride_length, fill = user_type_group)) +  
  stat_summary(fun = mean, geom = "bar", position = "dodge") +  
  labs(  
    title = "Average Ride Duration by User Type and Month",  
    x = "Month",  
    y = "Average Ride Length (minutes)",  
    fill = "User Type"  
  ) +  
  theme_minimal()
```

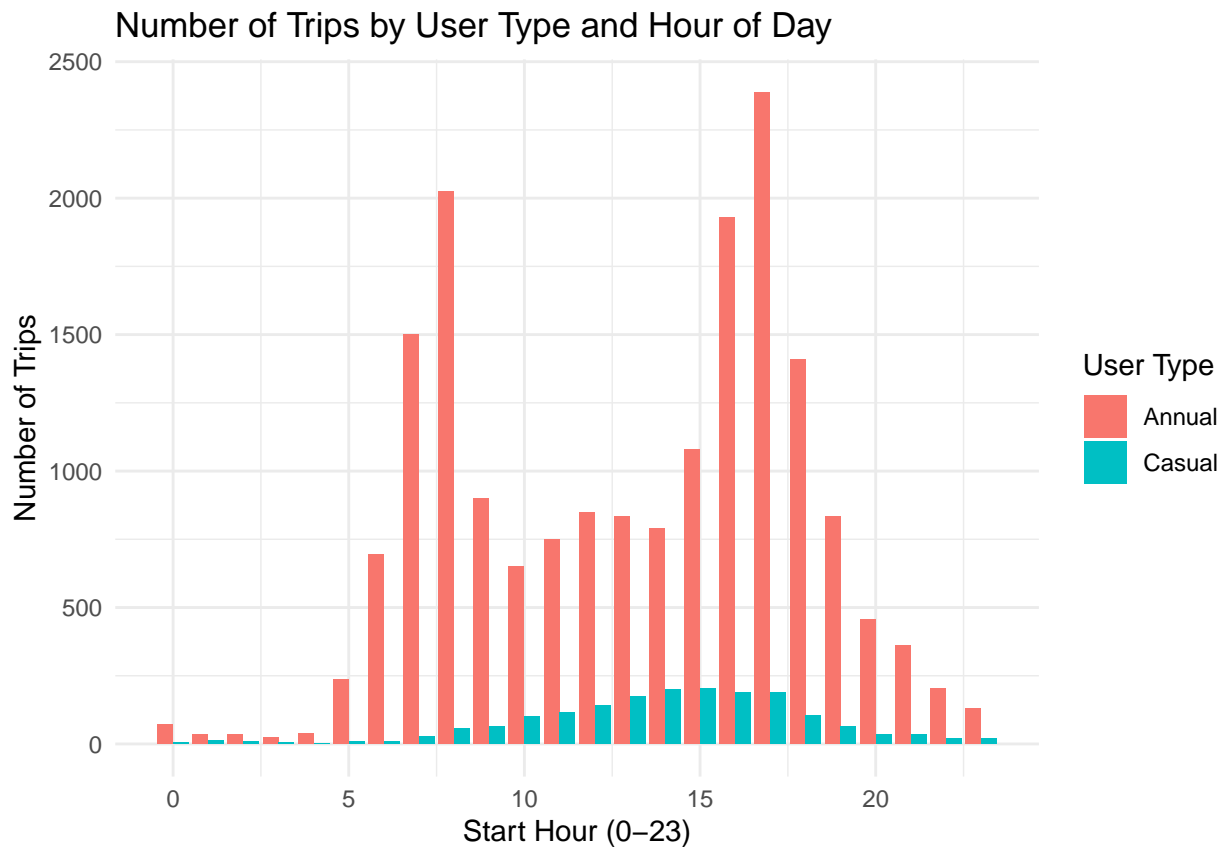



Observations:

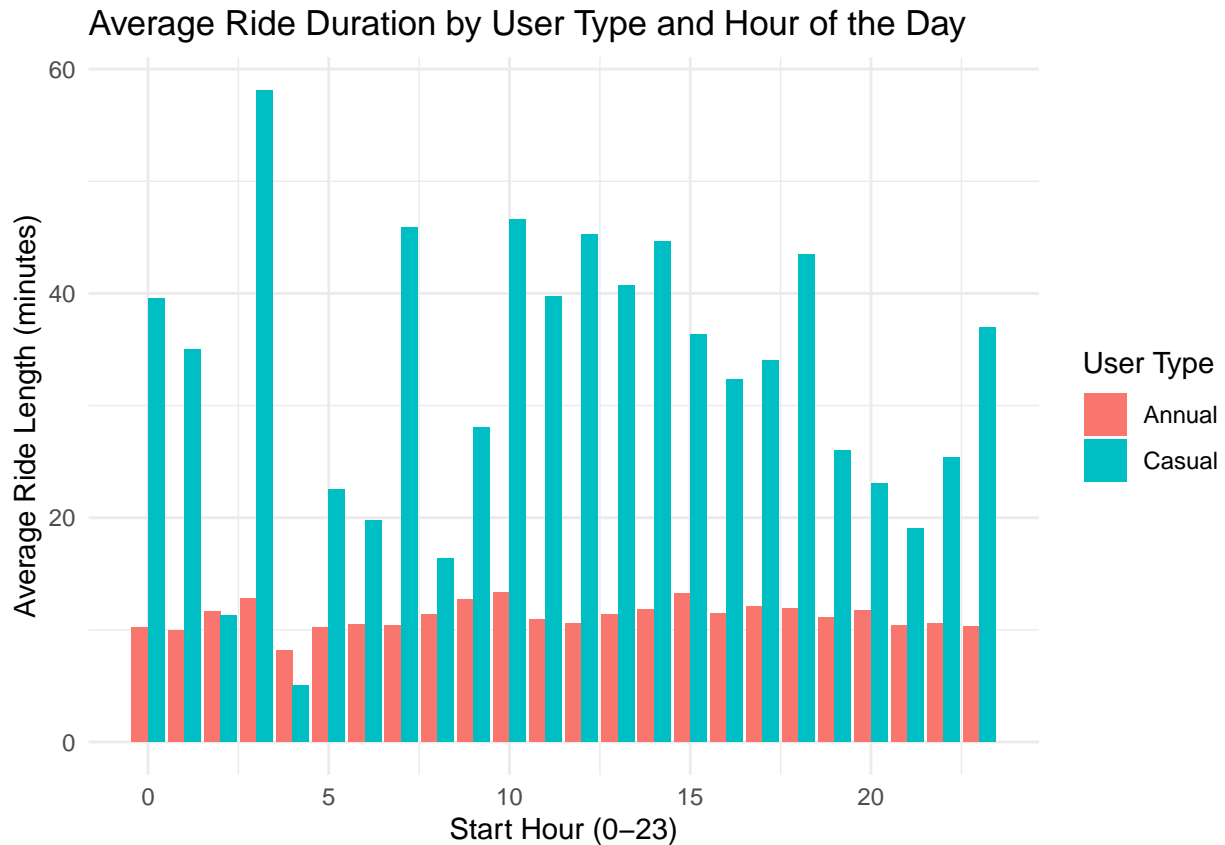
- Both groups ride more in March.
- Casual riders show stronger seasonal increase.

4. Hour-of-Day Analysis

```
# Number of trips by user type and hour of the day
ggplot(cyclistic, aes(x = start_hour, fill = user_type_group)) +
  geom_bar(position = "dodge") +
  labs(
    title = "Number of Trips by User Type and Hour of Day",
    x = "Start Hour (0-23)",
    y = "Number of Trips",
    fill = "User Type"
  ) +
  theme_minimal()
```



```
# Average ride duration by user type and hour of the day
ggplot(cyclistic, aes(x = start_hour, y = ride_length, fill = user_type_group)) +
  stat_summary(fun = mean, geom = "bar", position = "dodge") +
  labs(
    title = "Average Ride Duration by User Type and Hour of the Day", # ¡CORRECCIÓN AQUÍ!
    x = "Start Hour (0-23)",
    y = "Average Ride Length (minutes)",
    fill = "User Type"
  ) +
  theme_minimal()
```



Stakeholder Insights:

- **Casual riders** peak on **Saturday and Sunday**, indicating weekend recreational use.
- **Annual riders** maintain **steady weekday usage** with slight mid-week peaks, consistent with commuting patterns.

Key Insights

1. Usage Differences:

- Annual riders: more frequent rides, shorter duration (~11–12 min).
- Casual riders: fewer rides, longer duration (~36–41 min).

2. Weekday vs Weekend:

- Annual riders: mainly weekdays.
- Casual riders: mainly weekends.

3. Day-of-Week Patterns:

- Annual: steady weekdays.
- Casual: peaks on Saturday/Sunday.

4. Monthly Trends:

- Increase in spring/summer, especially for Casual riders.
- Hour-of-Day Trends:

5. Annual: commute peaks.

- Casual: late morning and afternoon recreational use.

Recommendations

Fleet Management:

- Allocate more bikes to **central hub stations** during Annual riders' peak commuting hours: 7:00-9:00 a.m. and 4:00-6:00 p.m. (Monday to Friday).

Marketing:

- Shift focus from general promotion to **Annual Membership Conversion**.
- Incentivise **Casual riders** by promoting the cost savings and convenience of short, routine weekday commutes.
- Offer special weekend perks (e.g., discounted day passes) to convert current weekend users.

Maintenance:

- Schedule maintenance and operational work during the lowest-use hours: **10:00 p.m. to 5:00 a.m.** and during the **middle of the week** (Tuesday/Wednesday) to maximize fleet availability during peak demand.

Seasonal Planning:

Prepare for the strong seasonal surge (Q1 trend analysis). **Initiate conversion campaigns and fleet rotation in March** to capitalise on the increasing number of Casual riders transitioning into the spring season.

Data Availability

The fully cleaned and analysed dataset is available as:

```
write_csv(cyclistic, "cyclistic_cleaned_Q1_2019_2020.csv")
```

Interactive Data Visualisation

The final recommendations are supported by the interactive Tableau Dashboard, which allows stakeholders to filter the usage data by user type, day, and time.

Click here to view the full interactive dashboard: [<https://public.tableau.com/app/profile/gabriela.olivera4636/viz/CyclisticQ1UsageStrategyDashboard/ExecutiveSummary>]
