

Divvy Bike Share Analysis Report

Zheyang

July 21, 2025

Contents

Introduction	1
Load Data	1
1. Average Ride Duration by User Type	2
2. Bike Share Peak Hours of Different Users	3
Peak Usage Hours by User Type	3
3. Ride Activity by Day of Week	5
Average Duration by Day and Rider Type	5
Number of Ride by Day of Week	7
4. Trip Count by Bike Type	8
5. Average Trip Duration by Bike Type	9
6. Top Start Stations	11
Top Station Across Different Users	11
Top 50 Station by both customer and Member	12
Findings	13
Recommendation	13

Introduction

The purpose of this report is to explore **how annual members and customers use Divvy bikes differently**. The insights aim to support marketing decisions and identify patterns in ridership behavior. Customers are those riders who purchase single-ride or full-day passes

Load Data

```
divvy_2024 <- read_csv("D:/Julianne/Google Data Analytics/Portfolio/Capstone/Data/processed/All_Divvy_Trip_Data.csv")
colnames(divvy_2024)
```

```
## [1] "trip_id"           "bike_type"           "start_time"
## [4] "end_time"          "start_station"       "start_station_id"
## [7] "end_station"       "end_station_id"     "start_lat"
## [10] "start_lng"         "end_lat"            "end_lng"
## [13] "user_type"         "ride_duration_minutes" "day_of_week"
## [16] "weekend_flag"
```

1. Average Ride Duration by User Type

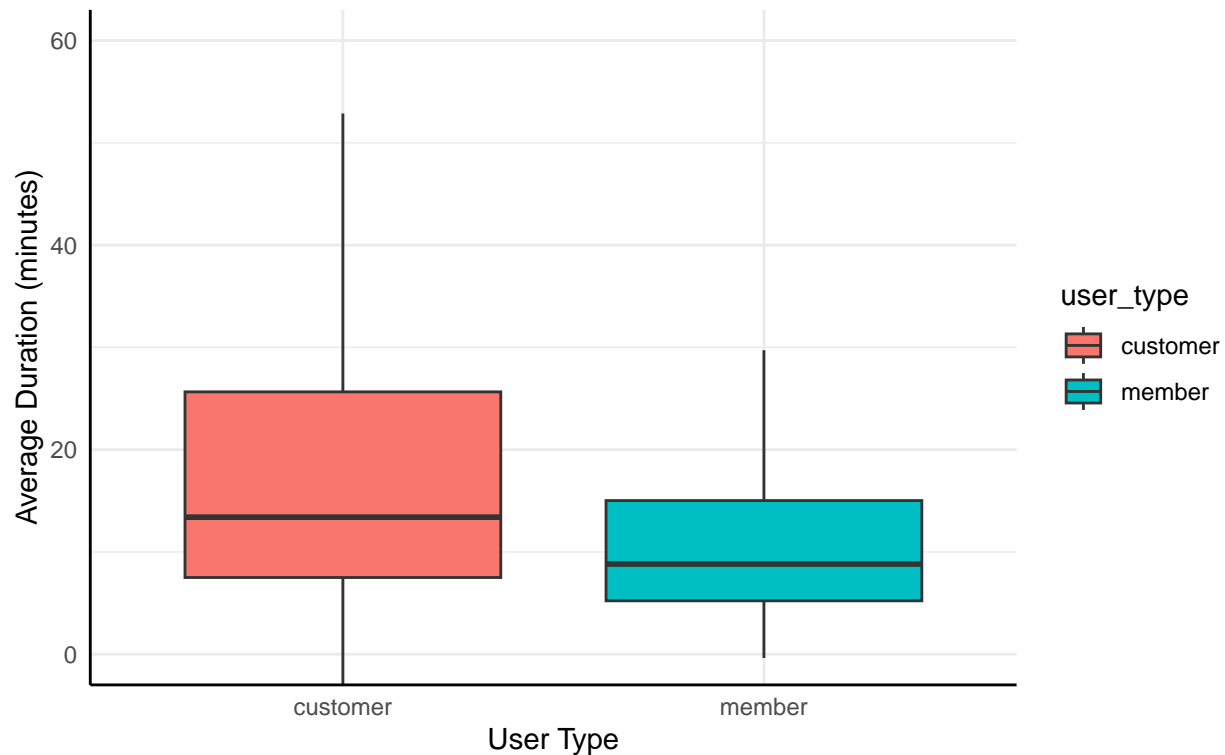
To begin the analysis, we explore whether customers and annual members differ in the amount of time they spend on Cylcistic bikes. Understanding this behavior helps identify their **intentions**—whether for leisure or commuting. This section uses a **boxplot** (excluding outliers) to visualize and **compare the typical ride durations between user types**.

```
average_duration <- divvy_2024 %>%
  group_by(user_type) %>%
  summarise(average_duration = mean(ride_duration_minutes), total_trip_count = n())
head(average_duration)
```

```
## # A tibble: 2 x 3
##   user_type average_duration total_trip_count
##   <chr>         <dbl>         <int>
## 1 customer      24.0         1521552
## 2 member       12.5         2686636
```

```
ggplot(divvy_2024, aes(x = user_type, y = ride_duration_minutes, fill = user_type)) +
  geom_boxplot(outlier.shape = NA) +
  coord_cartesian(ylim = c(0, 60)) +
  labs(title = "Fig 1: customers Rides Are 2x Longer Than Member Rides on Average",
       subtitle = "Median ride duration by user type (outliers excluded)",
       y = "Average Duration (minutes)", x = "User Type") +
  theme_minimal() +
  theme(axis.line = element_line(color = "black"))
```

Fig 1: customers Rides Are 2x Longer Than Member Rides on Average
Median ride duration by user type (outliers excluded)



Insight: The longer median duration for customers (after excluding outliers) supports the assumption that customers value experience over efficiency. Members likely ride frequently but briefly, emphasizing convenience. The plot shows a wider spread for customers' duration than member's average duration, suggesting customers are more flexible on travel time, while members might be more utility focused.

2. Bike Share Peak Hours of Different Users

Peak Usage Hours by User Type

Understanding when during the day users ride the most is essential for resource allocation and planning. This section analyzes the distribution of ride start times to **identify peak usage hours** for both customers and annual members. The goal is to **differentiate time-based behavior** that reflects each group's unique motivations.

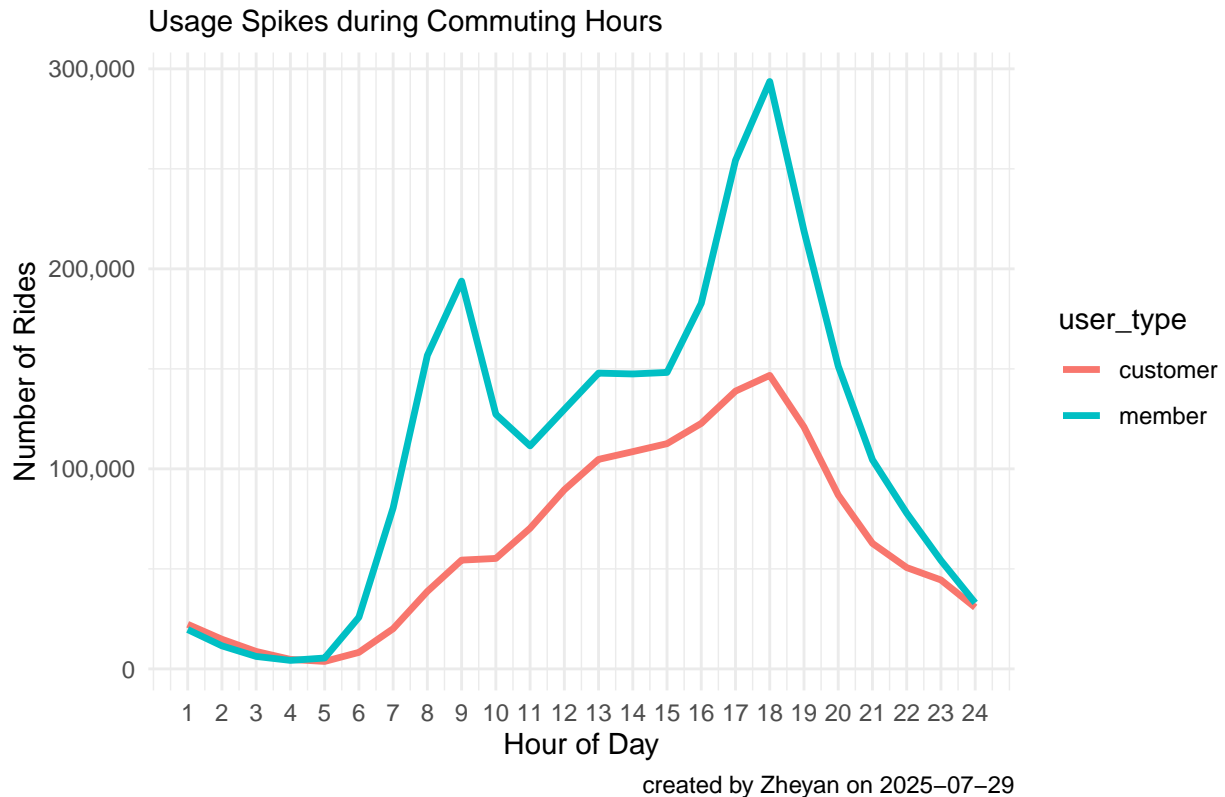
```
trip_by_hour <- divvy_2024 %>%
  mutate(hour = lubridate::hour(start_time) + 1) %>%
  group_by(user_type, hour) %>%
  summarise(trip_count = n(), .groups = "drop")
ggplot(trip_by_hour, aes(x = hour, y = trip_count, color = user_type)) +
  geom_line(size = 1.2) +
  scale_x_continuous(breaks = 1:24) +
  scale_y_continuous(labels = comma) +
  labs(title = "Fig 2.1: Bike Share Peak Hours of Different Users",
```

```

subtitle = "Usage Spikes during Commuting Hours",
caption = paste0("created by Zheyang on ", today()),
x = "Hour of Day", y = "Number of Rides") +
theme_minimal()

```

Fig 2.1: Bike Share Peak Hours of Different Users



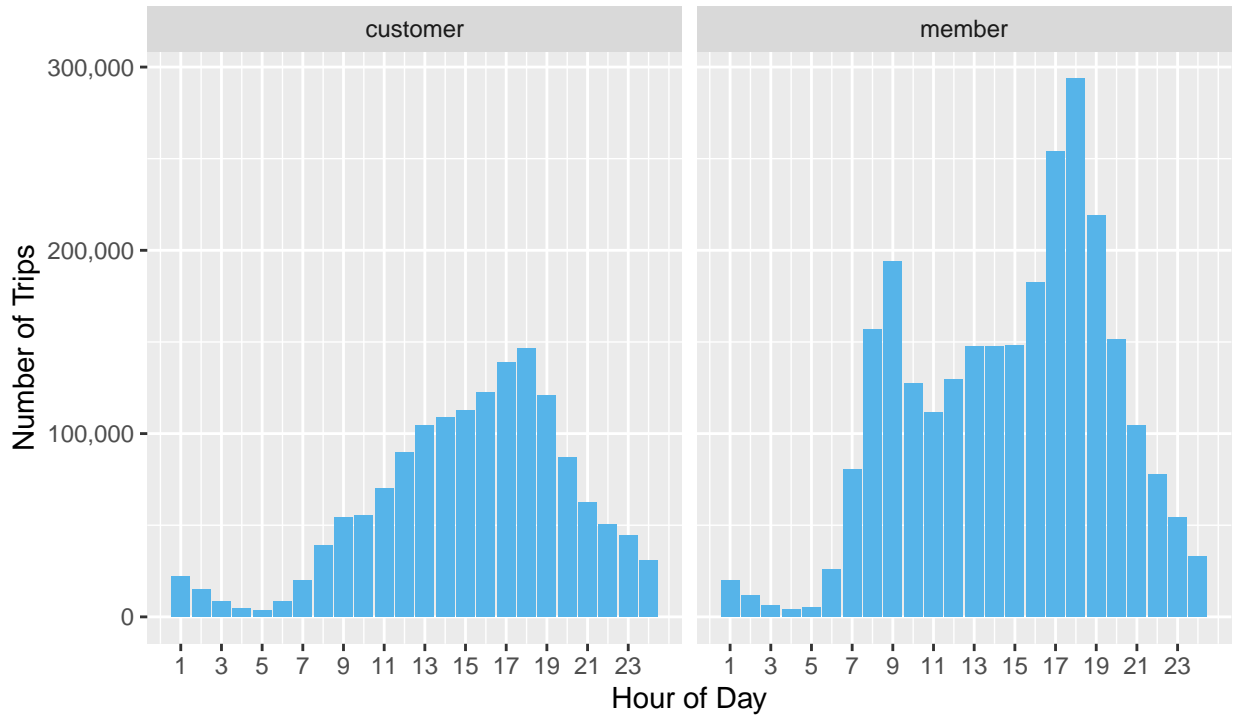
By plotting the hourly ride volume, we can observe that member rides cluster around **traditional commuting times** (e.g., 9 AM and 6 PM), while customer activities increase during **mid-afternoon** or **daylight hours** (2 PM to 6 PM). **Hourly Ride Distribution (Faceted)** To further distinguish **usage patterns by time of day**, break down the hourly ride distribution and visualize it separately for customers and members. Faceted bar charts allow us to examine how **trip frequency**** varies **across each hour** for both groups in parallel. I want to highlight the customers preferred hours and whether members have distinct rush hour peaks.

```

divvy_2024 %>%
  mutate(hour = hour(start_time) + 1) %>%
  group_by(user_type, hour) %>%
  summarise(trip_count = n(), .groups = "drop") %>%
  ggplot(aes(x = hour, y = trip_count)) +
  geom_col(fill = "#56B4E9") +
  facet_wrap(~ user_type) +
  scale_x_continuous(breaks = seq(1, 24, 2)) +
  scale_y_continuous(labels = comma) +
  labs(title = "Fig 2.2: Ride Distribution of User by Hour",
        subtitle = "Hourly Ride Patterns Reveal Key Usage Differences",
        x = "Hour of Day", y = "Number of Trips",
        caption = paste0("created by Zheyang on ", today()))

```

Fig 2.2: Ride Distribution of User by Hour
Hourly Ride Patterns Reveal Key Usage Differences



created by Zheyang on 2025-07-29

This supports differentiated planning—members need consistent supply across **rush hours**, while customers require **flexible access on afternoons**. These temporal patterns show clearly distinct ride purposes: **work vs leisure**.

3. Ride Activity by Day of Week

Following the analysis of hourly usage patterns, we now examine how **daily trends** further differentiate ride behavior between customers and members. While hourly data revealed that **members cluster around weekday commute hours** and **customers prefer afternoon rides**, this section shifts focus to explore weekly rhythms.

By comparing **average ride duration** and **trip counts** across each **day of the week**, we investigate whether members follow a **workday commute cycle**, and whether customers are more **weekend-oriented**, riding for **leisure or recreation**.

Average Duration by Day and Rider Type

First, we can explore the average ride duration by day of week for each user type.

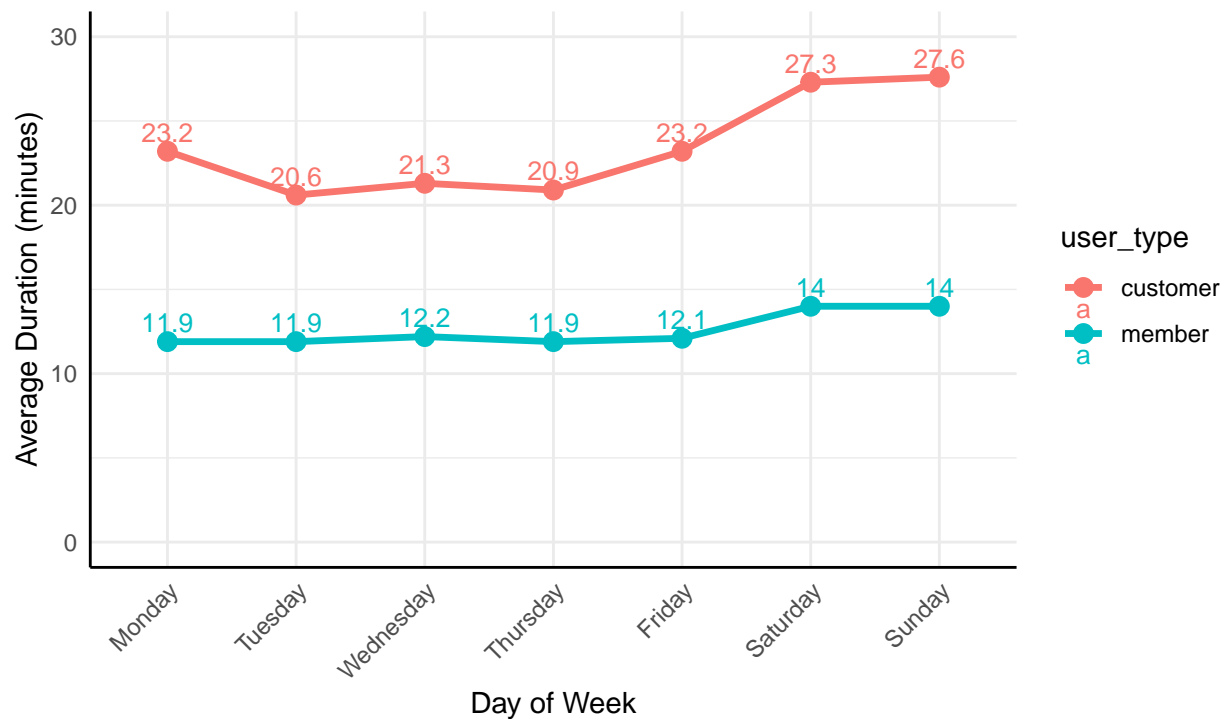
```
avg_duration_day <- divvy_2024 %>%
  mutate(day_of_week = factor(day_of_week,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"),
    ordered = TRUE)) %>%
  group_by(user_type, day_of_week) %>%
  summarise(avg_duration = round(mean(ride_duration_minutes), 1), .groups = "drop")
```

```

avg_duration_day %>%
  ggplot(aes(x = day_of_week, y = avg_duration, color = user_type, group = user_type)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  geom_text(aes(label = avg_duration), vjust = -0.5, size = 3.5) + # Correct column name
  scale_y_continuous(limits = c(0, 30)) +
  labs(
    title = "Fig 3.1: Average Ride Duration by Day and Rider Type",
    subtitle = "The Weekend Effect: Customer Demand Surges",
    x = "Day of Week", y = "Average Duration (minutes)",
    caption = paste0("created by Zheyang on ", today())
  ) +
  theme_minimal() +
  theme(
    axis.line = element_line(color = "black"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

```

Fig 3.1: Average Ride Duration by Day and Rider Type
The Weekend Effect: Customer Demand Surges



created by Zheyang on 2025-07-29

It can be observed that regardless of the day of the week, on average, customers ride for **approximately 2 times longer** than members. We can reasonably assume that members probably utilize the bike rental service mainly for **commuting purposes**. In contrast, customers rent the bike for a **one or few time uses** and have various **riding purposes**. If we have rider_id, further analysis can be conducted to look at number of rentals per customers vs members.

Number of Ride by Day of Week

Furthermore, we can validate our assumption regarding the purpose of rides for different users by analyzing the number of trips per day per user.

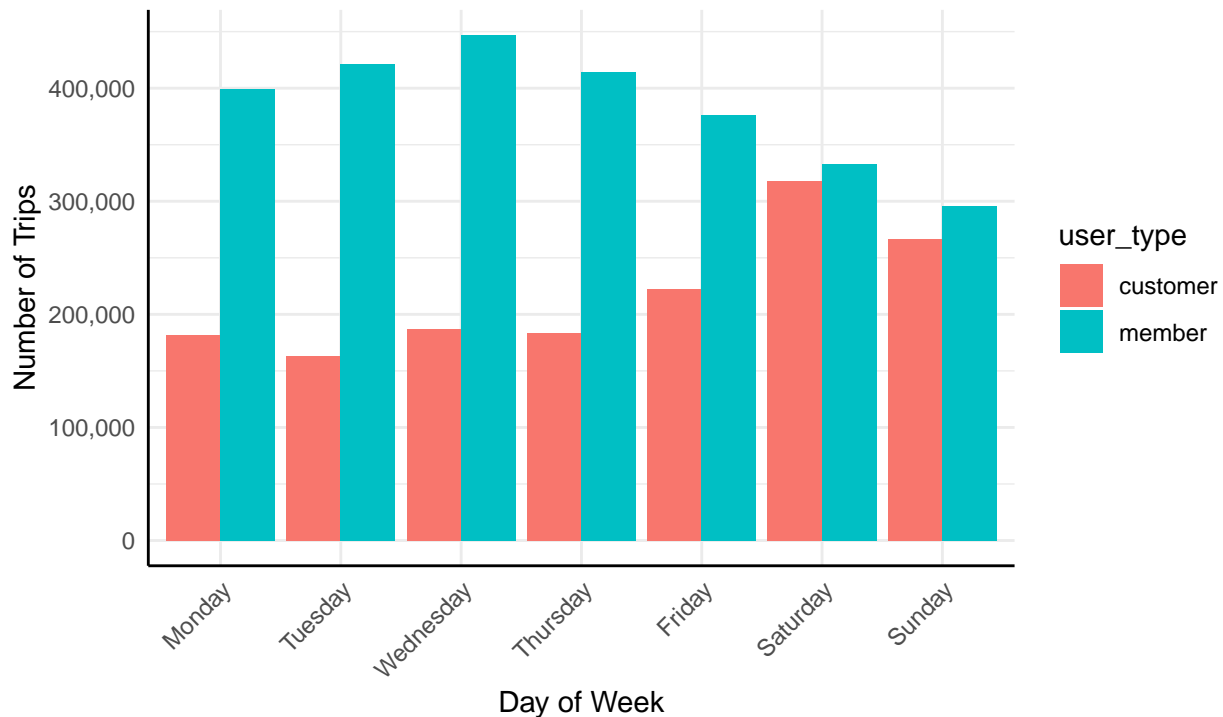
```
trip_by_day <- divvy_2024 %>%
  mutate(day_of_week = factor(day_of_week,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"),
    ordered = TRUE)) %>%
  group_by(user_type, day_of_week) %>%
  summarise(trip_count = n(), .groups = "drop")
top3_trip_day <- trip_by_day %>%
  arrange(user_type, desc(trip_count)) %>% # sort trip_count descending within user_type
  group_by(user_type) %>%
  slice_max(trip_count, n = 3) # keep top 3 per user_type
head(top3_trip_day)
```

```
## # A tibble: 6 x 3
## # Groups:   user_type [2]
##   user_type day_of_week trip_count
##   <chr>      <ord>         <int>
## 1 customer  Saturday      317848
## 2 customer  Sunday        266194
## 3 customer  Friday        222575
## 4 member   Wednesday     446905
## 5 member   Tuesday       421424
## 6 member   Thursday      414260
```

The results now indicate that, for customers, the most common day for bike rentals is Sunday, while for members, it is Wednesday.

```
trip_by_day %>%
  ggplot(aes(x = day_of_week, y = trip_count, fill = user_type)) +
  geom_col(position = "dodge") +
  scale_y_continuous(labels = comma) +
  labs(title = "Fig 3.2: Ride Count by Day of Week for Different Users",
    subtitle = "The Weekend Effect: Customer Demand Surges",
    x = "Day of Week", y = "Number of Trips",
    caption = paste0("created by Zheyao on ", today())) +
  theme_minimal() +
  theme(axis.line = element_line(color = "black"),
    axis.text.x = element_text(angle = 45, hjust = 1))
```

Fig 3.2: Ride Count by Day of Week for Different Users
The Weekend Effect: Customer Demand Surges



created by Zheyang on 2025-07-29

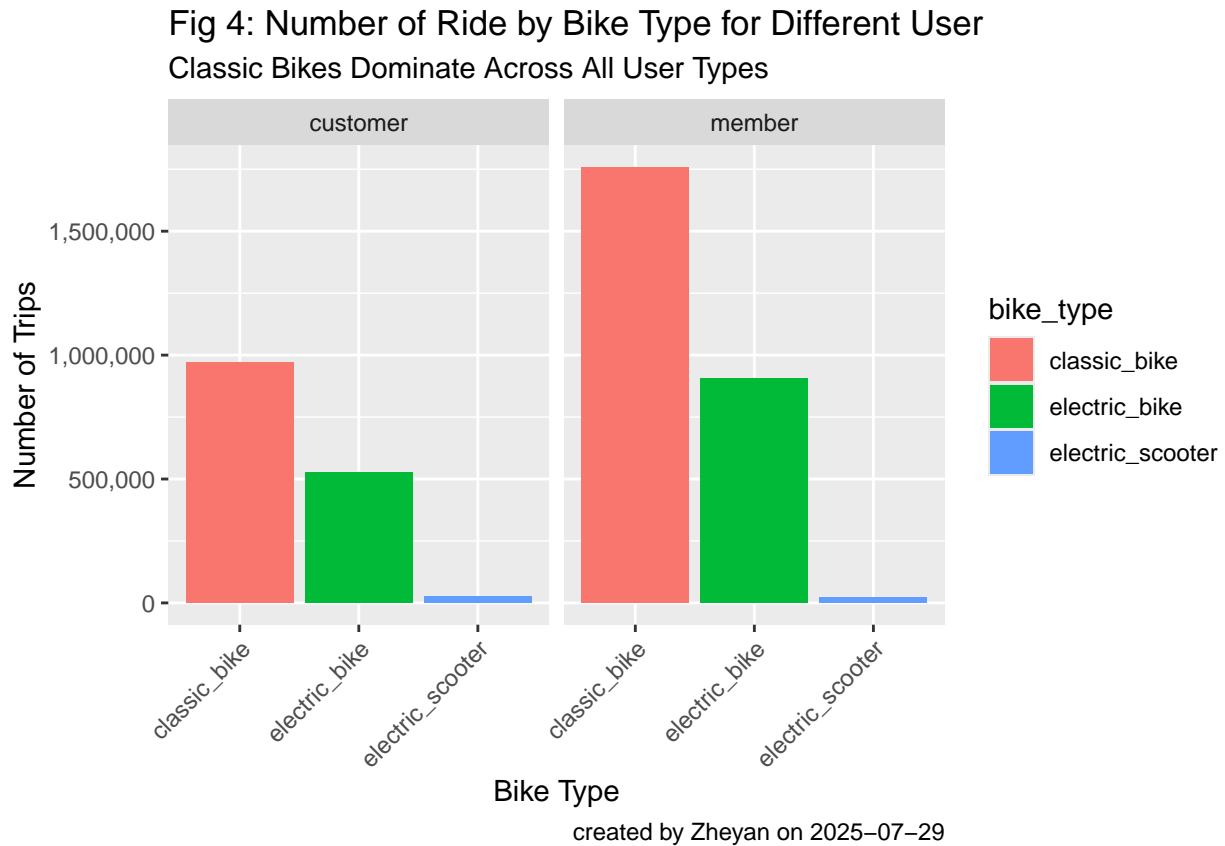
The visual data indicates that customers predominantly ride on **days approaching the weekend**. Members, on the other hand, tend to ride mostly during **weekdays**. Notably, the ridership shows a **decline** as the weekend approaches. This confirms members are **routine commuters**, while customers are **weekend explorers**. Targeting promotions based on day-of-week patterns can increase engagement.

4. Trip Count by Bike Type

Different rider types may favor different bike types depending on convenience, speed, or comfort. This section compares the **number of rides** taken by each **user group** using each **bike type**—classic bikes, electric bikes, and electric scooters.

```
trip_by_bike_type <- divvy_2024 %>%
  group_by(bike_type, user_type) %>%
  summarise(trip_count = n())

ggplot(trip_by_bike_type, aes(x = bike_type, y = trip_count, fill = bike_type)) +
  geom_col() +
  facet_wrap(~ user_type) +
  scale_y_continuous(labels = comma) +
  labs(title = "Fig 4: Number of Ride by Bike Type for Different User",
       subtitle = "Classic Bikes Dominate Across All User Types",
       x = "Bike Type", y = "Number of Trips",
       caption = paste0("created by Zheyang on ", today())) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Classic bikes (blue bars) account for the overwhelming **majority of rides** across both user types. Classic bikes should **remain the core fleet component**. Stations should maintain classic bike dominance while gradually introducing electric options.

5. Average Trip Duration by Bike Type

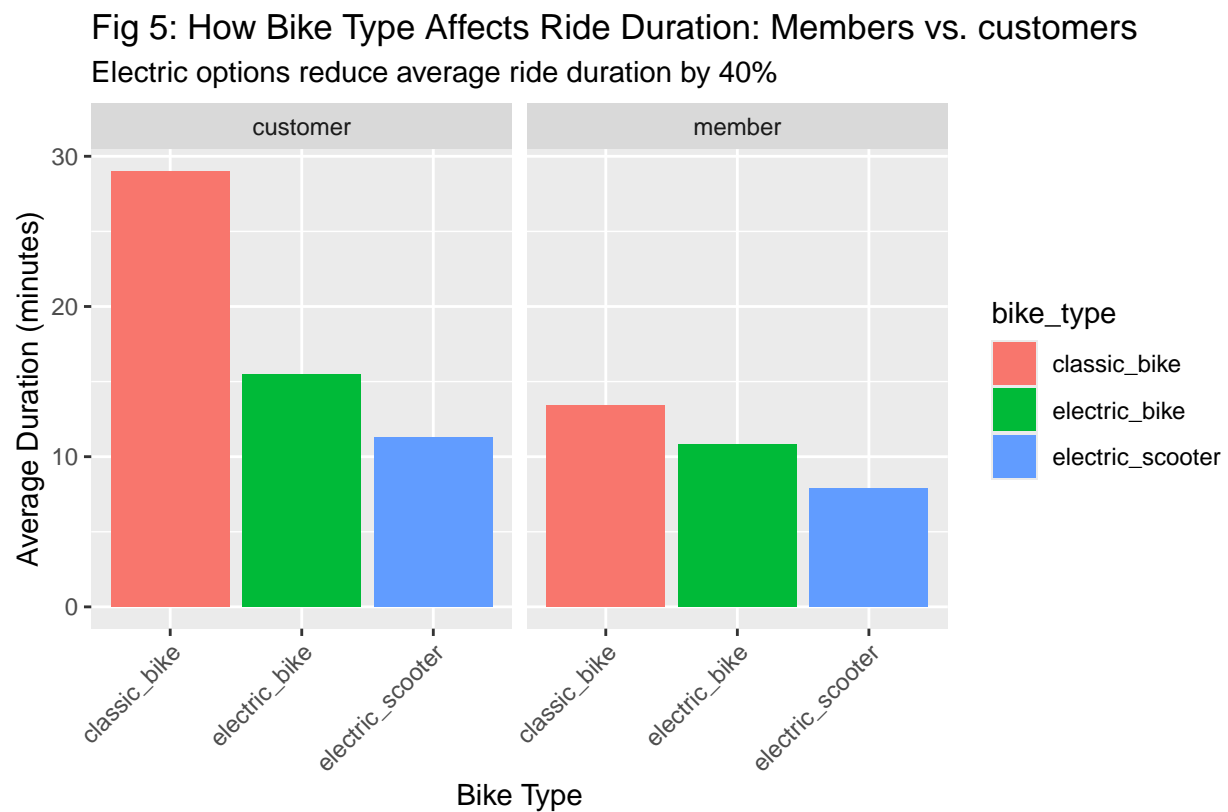
While ride count tells us about popularity, ride duration can offer insights into how each bike type is used. This section explores **how long, on average, each user type rides different bike types**.

```
trip_duration_bike_type <- divvy_2024 %>%
  group_by(bike_type, user_type) %>%
  summarise(avg_duration_bike = mean(ride_duration_minutes))

duration_by_bike <- trip_duration_bike_type %>%
  pivot_wider(
    names_from = user_type,
    values_from = avg_duration_bike
  ) %>%
  arrange(bike_type)
print(duration_by_bike)
```

```
## # A tibble: 3 x 3
## # Groups:   bike_type [3]
##   bike_type      customer member
##   <chr>          <dbl>  <dbl>
## 1 classic_bike      29.0   13.4
## 2 electric_bike     15.5   10.8
## 3 electric_scooter  11.3    7.88
```

```
ggplot(trip_duration_bike_type, aes(x = bike_type, y = avg_duration_bike, fill = bike_type)) +
  geom_col() +
  facet_wrap(~ user_type) +
  scale_y_continuous(labels = comma) +
  labs(title = "Fig 5: How Bike Type Affects Ride Duration: Members vs. customers",
       subtitle = "Electric options reduce average ride duration by 40%",
       x = "Bike Type", y = "Average Duration (minutes)",
       caption = paste0("created by Zheyang on ", today())) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



created by Zheyang on 2025-07-29

Customers spend nearly **twice as long** on classic bikes compared to electric bikes (16 min) and scooters (12 min). This suggests **leisure-oriented usage**, such as scenic rides, exercise, or relaxed exploration, where speed is less critical.

Annual members spend **consistent ride times** across all bike types. Analyzing the **number of rides** and the **average time** per bike for members indicates that members probably use bikes for **short, practical trips** (such as commuting or running errands), where efficiency is more crucial than the riding experience.

6. Top Start Stations

Identifying the **top starting stations** helps uncover key locations that serve different types of riders.

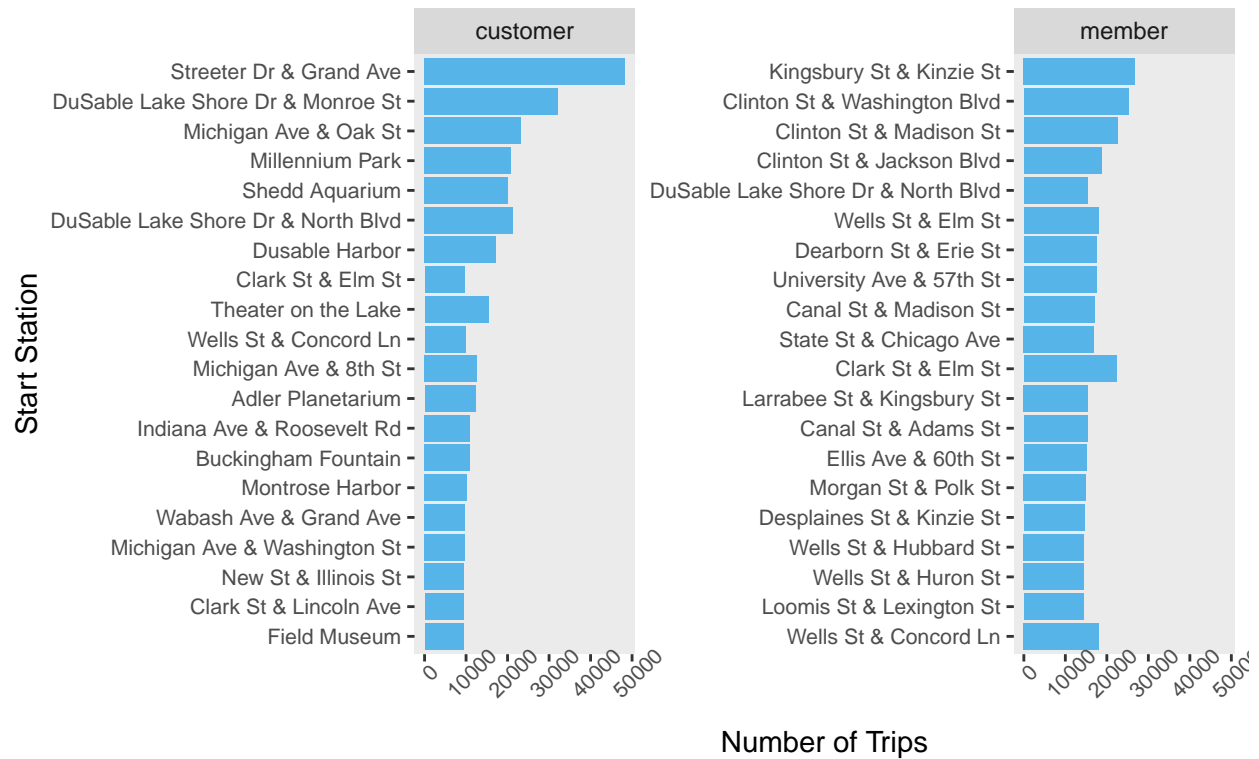
Top Station Across Different Users

Analyze the top 20 start stations for each user group to understand where demand is highest

```
top_stations_by_user <- divvy_2024 %>%
  group_by(user_type, start_station) %>%
  summarise(trip_count = n(), .groups = "drop") %>%
  group_by(user_type) %>%
  slice_max(trip_count, n = 20) %>%
  ungroup()

ggplot(top_stations_by_user, aes(x = reorder(start_station, trip_count), y = trip_count)) +
  geom_col(fill = "#56B4E9") +
  coord_flip() +
  facet_wrap(~ user_type, scales = "free_y") +
  labs(title = "Top 20 Start Stations by User Type",
       x = "Start Station", y = "Number of Trips",
       caption = paste0("created by Zheyang on ", today())) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.text.y = element_text(size = 8),
        axis.text.x = element_text(size = 8, angle = 45))
```

Top 20 Start Stations by User Type



created by Zheyang on 2025-07-29

Top 50 Station by both customer and Member

By comparing these with **member usage** and **total activity**, we can identify **prime locations for outreach** aimed at converting customers into members

```
top_50_station <- divvy_2024 %>%
  filter(!is.na(start_station)) %>%
  group_by(user_type, start_station) %>%
  summarise(trip_count = n(), .groups = "drop") %>%
  pivot_wider(names_from = user_type, values_from = trip_count, values_fill = 0) %>%
  mutate(total_trips = customer + member) %>%
  arrange(desc(customer)) %>%
  slice_head(n = 50)
head(top_50_station, 10)
```

```
## # A tibble: 10 x 4
##   start_station      customer member total_trips
##   <chr>              <int>  <int>    <int>
## 1 Streeter Dr & Grand Ave    48315  13961    62276
## 2 DuSable Lake Shore Dr & Monroe St  32199   9295    41494
## 3 Michigan Ave & Oak St     23296  13014    36310
## 4 DuSable Lake Shore Dr & North Blvd  21360  15303    36663
## 5 Millennium Park          20840   9703    30543
## 6 Shedd Aquarium           20114   4557    24671
## 7 Dusable Harbor           17264   5990    23254
```

## 8 Theater on the Lake	15398	11995	27393
## 9 Michigan Ave & 8th St	12556	6628	19184
## 10 Adler Planetarium	12273	4868	17141

Findings

- **Members** ride most during **rush hours** (8 AM & 5–6 PM), reflecting commuting habits. Customers are more active in the **afternoons** (2–6 PM), suggesting leisure or recreational use.
- **Customers** peak on **weekends**, especially **Saturday**, showing a preference for leisure. Members ride steadily on **weekdays**, especially Tuesday to Thursday, indicating work-related use.
- **Customers** take trips **3 times longer** than members, pointing to **non-commuting purposes**. Members have **shorter, consistent duration**, reflecting daily, functional use.
- Both groups prefer **classic bikes** in usage, but in addition to that, customers still ride twice as long, suggesting **leisure-focused usage** like scenic or relaxed rides. Members show more consistent **usage count** and **duration** across bike types, reflecting **practical, efficiency-driven** trips such as commuting or errands.

Recommendation

- Offer a **cheaper subscription** designed for consumers who are leisurely: the “Weekend Rider” plan, which corresponds with their current **usage patterns** and grants access for **three individual day passes** or **six single ride passes**. Customers may enjoy **membership benefits** on the days they ride the most with our three-day pass model, which also reduces the **barrier** to **commitment**.
- Offer a **credit system** where customers earn points from rides or referrals, which can be **redeemed for discounts** on annual memberships. This rewards frequent use and **encourages group-based, social riders** to upgrade by benefiting from a “Refer a Friend” program.
- Deploy **membership awareness campaigns** at the **top customer-used stations** and during **peak afternoon hours**. Highlight key benefits—like **unlimited rides**, **faster checkout**, and **priority access**.