Cyclistic Bike-Share Analysis

Sarim Rizvi

2025-05-29

Business Task

Cyclistic is a fictional bike-share company, and wants more people to join its annual membership programme. This report analyses ride data from April 2022 to March 2023 to understand the difference in bike usage between casual riders and annual members, which can help in getting insights to develop a marketing strategy aimed at converting casual riders into members.

Data Understanding

- Data source: 12 months of trip data from April 2022 to March 2023.

- Each ride includes timestamps, location coordinates, bike type, and user type.

- Over 5.7 million valid rides included after cleaning.

- Spatial station data is retained but not directly analyzed, except for heatmap visualization.

Data Preprocessing

- Combining monthly CSVs into a single file.

- Removing rides with negative or zero duration.

- Filtering out trips over 24 hours.

- Creating derived features: ride_length, day_of_week, hour, month.

- Converting categorical columns to factors

```r
all_trips_clean <- read_csv("data/clean/cyclistic_cleaned.csv")
```

```r
all_trips_clean <- all_trips_clean %>%
  mutate(
    member_casual = factor(member_casual, levels = c("member", "casual")),
    rideable_type = factor(rideable_type),
    month = factor(month, levels = month.abb[c(4:12, 1:3)]),
    day_of_week = factor(day_of_week, levels = c("Mon", "Tue", "Wed", "Thu", "Fri", "Sat", "Sun"))
  )
```

# Data Findings and Visualisation

Summary

```r
all_trips_clean %>%
  group_by(member_casual) %>%
  summarise(
    avg_ride = mean(ride_length),
    median_ride = median(ride_length),
    count = n(),
    .groups = "drop"
  )
```

```
## # A tibble: 2 x 4
##   member_casual avg_ride median_ride   count
##   <fct>            <dbl>       <dbl>   <int>
## 1 member            12.2        8.67 3465251
## 2 casual            19.2       12.0  2160821
```

Ride Patterns by Day of Week

```r
all_trips_clean %>%
  group_by(day_of_week, member_casual) %>%
  summarise(count = n(), .groups = "drop") %>%
  ggplot(aes(x = day_of_week, y = count, fill = member_casual)) +
  geom_col(position = "dodge") +
  labs(title = "Ride Volume by Day", x = "Day", y = "Rides", fill = "Rider Type") +
  theme_minimal()
```

Ride Volume by Day
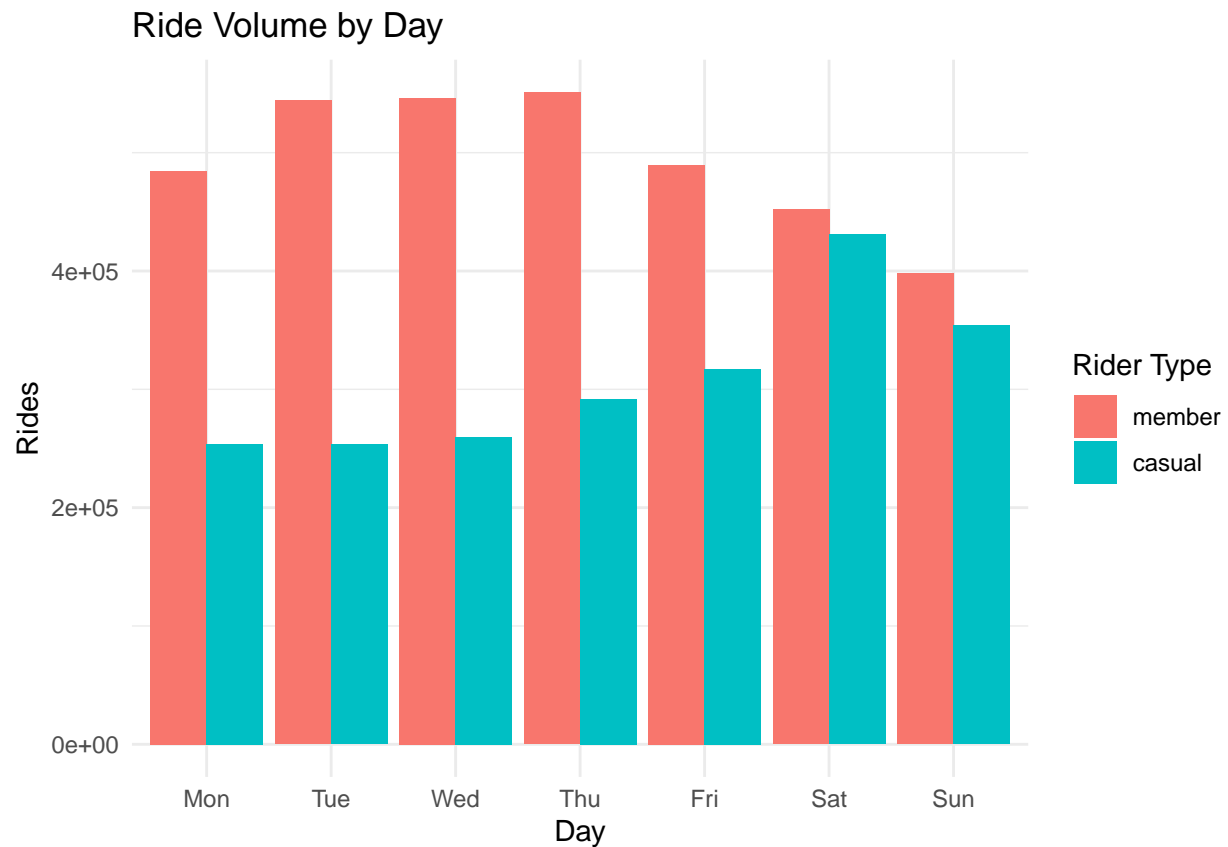
Hourly Ride Trends

```
all_trips_clean %>%
  group_by(hour, member_casual) %>%
  summarise(count = n(), .groups = "drop") %>%
  ggplot(aes(x = hour, y = count, fill = member_casual)) +
  geom_col(position = "dodge") +
  scale_x_continuous(breaks = seq(0, 23, by = 4)) +
  labs(title = "Ride Frequency by Hour", x = "Hour", y = "Rides", fill = "Rider Type") +
  theme_minimal() +
  theme(legend.position = "inside",
        legend.position.inside = c(0.9, 0.9),
        legend.background = element_rect(fill = "white", color = "gray"),
        legend.title = element_text(size = 9),
        legend.text = element_text(size = 8))
```

Ride Frequency by Hour
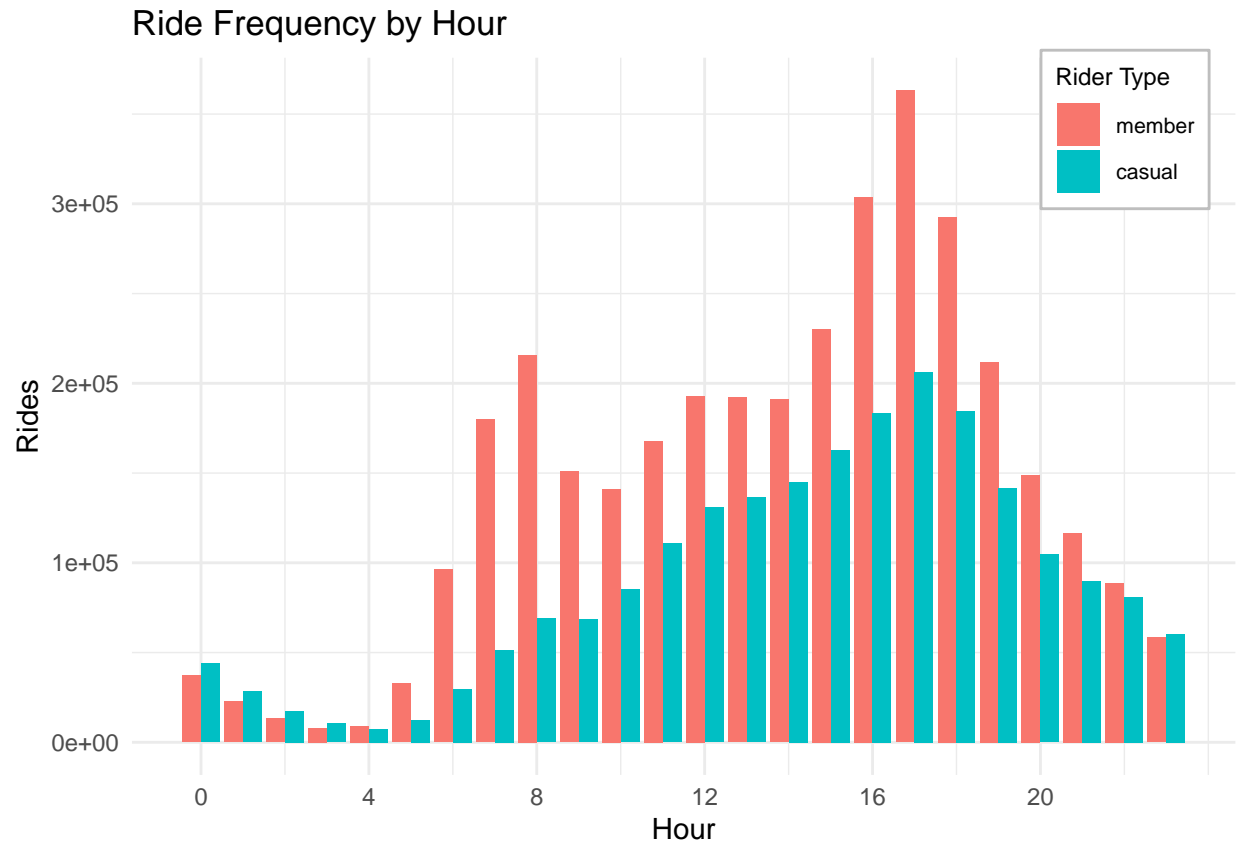
Ride Length Distribution

```
ggplot(all_trips_clean, aes(x = ride_length, fill = member_casual)) +
  geom_density(alpha = 0.4) +
  coord_cartesian(xlim = c(0, 60)) +
  labs(title = "Ride Length Density (Under 60 min)", x = "Ride Duration (min)", y = "Density", fill = "
  theme_minimal() +
  theme(legend.position = "inside",
        legend.position.inside = c(0.9, 0.9),
        legend.background = element_rect(fill = "white", color = "gray"),
        legend.title = element_text(size = 9),
        legend.text = element_text(size = 8))
```

Ride Length Density (Under 60 min)
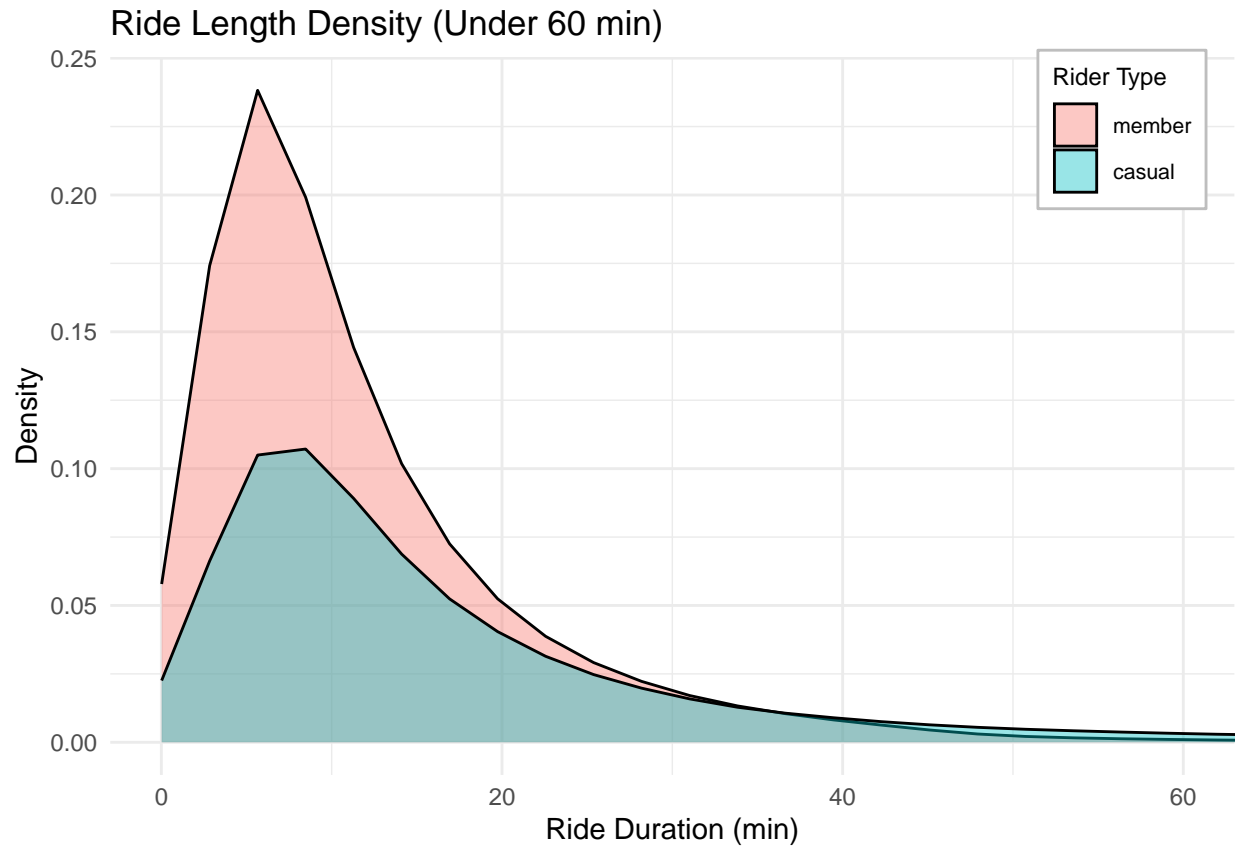
Spatial Distribution

```r
start_points <- all_trips_clean %>%
  filter(!is.na(start_lat), !is.na(start_lng)) %>%
  sample_n(100000)

ggplot(start_points, aes(x = start_lng, y = start_lat)) +
  stat_density_2d(aes(fill = after_stat(level)), geom = "polygon", contour = TRUE, alpha = 0.7) +
  facet_wrap(~ member_casual) +
  scale_fill_viridis_c(option = "magma") +
  labs(
    title = "Heatmap of Ride Starting Points",
    subtitle = "Comparison of Casual vs. Member Riders",
    x = "Longitude", y = "Latitude", fill = "Density Level"
  ) +
  theme_minimal()
```

Heatmap of Ride Starting Points
Comparison of Casual vs. Member Riders
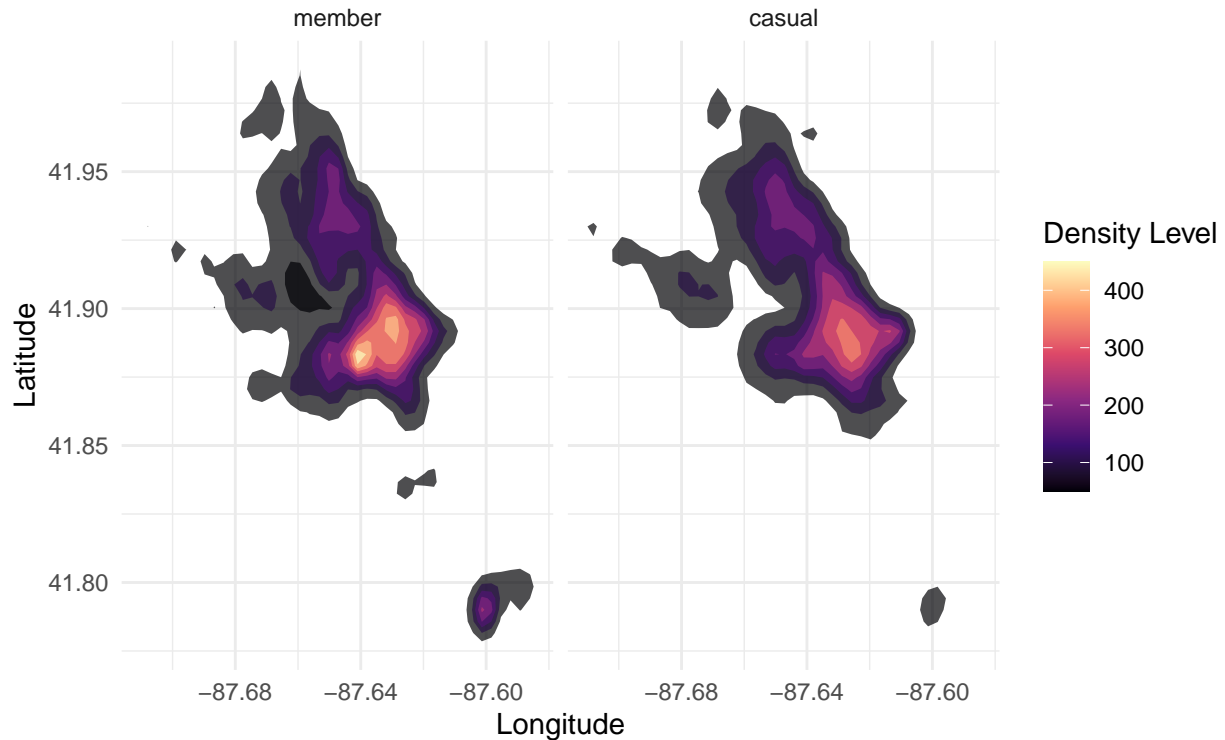
Data Insights

- Casual riders take longer rides on average and mostly ride on weekends and afternoons.

- Members ride more consistently during weekdays and commute hours.

- Usage peaks in summer months for both groups, but especially for casual users.

- Members mostly ride classic bikes, while casual riders also use electric bikes. frequently.

- Casual and member riders start and end trips in the same popular areas.

Recommendations

- Promote memberships to casual weekend riders via app or email offers, especially in peak summer months.

- Highlight membership value for commuting to casual users who ride during weekday mornings.

- Run targeted digital ads on weekends to convert high-duration casual riders into members.