Erica Yang
BUS446
Final Project

Citi-Bike Ride Sharing Analytics Report

Purpose:

To explore the correlation between the preferred days of the week for bike rides between members vs casual riders.

Is there a correlation between the preferred days of the week for bike rides between members vs. casual riders?

Task Plan:

Load dataset with 'started_at' and 'member_casual.'

Filter for 'casual' and 'member' riders

Group and count by day of the week and 'member_casual'

Visualization and Conclusion:

Create a bar chart to show the relationship.

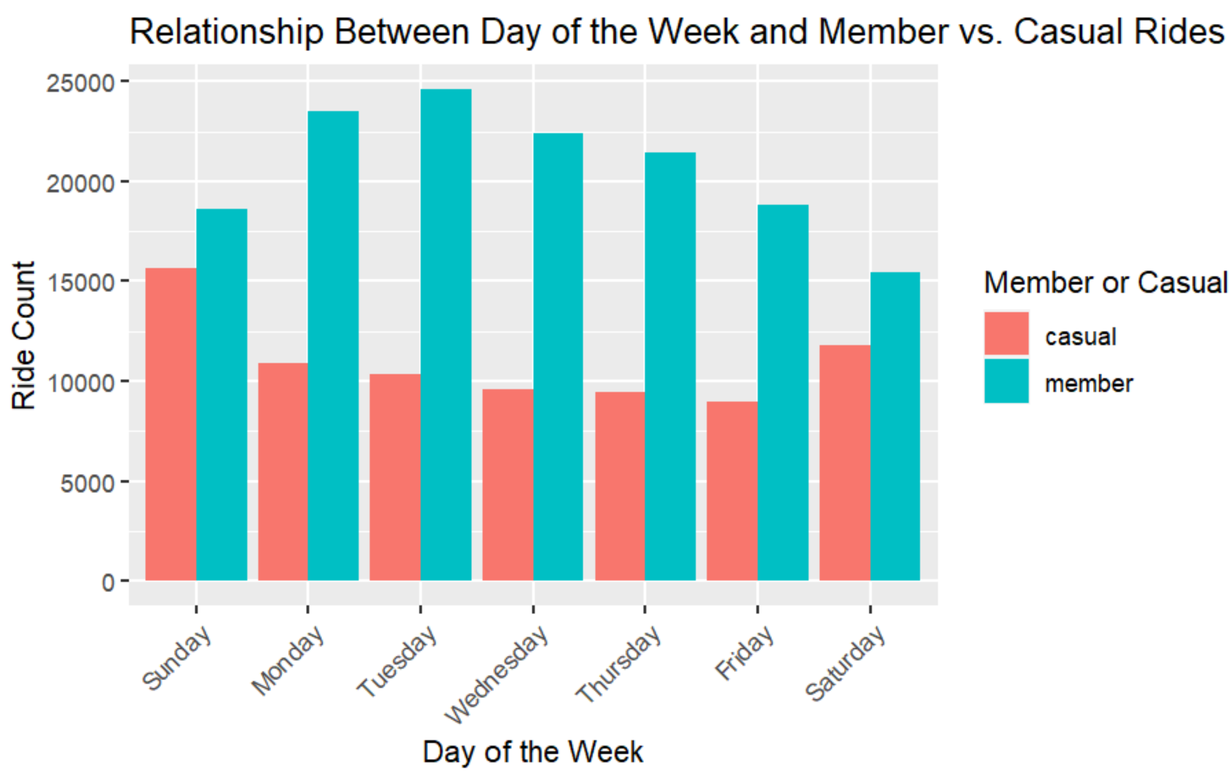
Analyze findings and draw conclusions.

Code:

```
# Filter for 'casual' and 'member' riders

filtered_data <- data_all %>%

  filter(member_casual %in% c('casual', 'member'))


# Group and count by day of the week and 'member_casual'

ride_count <- filtered_data %>%

  mutate(day_of_week = factor(weekdays(started_at), levels = c("Sunday", "Monday",
"Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))) %>%

  group_by(day_of_week, member_casual, .drop = TRUE) %>%
```

```
  summarize(count = n(), .groups = 'drop_last')

# Create a bar chart to visualize the relationship
ggplot(ride_count, aes(x = day_of_week, y = count, fill = member_casual)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Relationship Between Day of the Week and Member vs. Casual Rides",
       x = "Day of the Week", y = "Ride Count",
       fill = "Member or Casual") +
  scale_x_discrete(drop = FALSE) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Graph:



Erica Yang
BUS446
Final Project

How can this be used in a business situation? For what type of job?

Note: Based on the graph details, Sunday and Saturday have the most active casual riders. Therefore, if a business desires to convert the casual riders to upgrade to members, then it would make sense to advertise heavier on those days' vs the other days.

There are patterns in bike usage that can be based on geography and location. Bike usage might only be for those commuting to and from work on weekdays vs a spike in bike activity during weekends when people go out for a longer period of time. Understanding these patterns can be valuable in business situations for roles related to urban planning, transportation management, or even for businesses directly involved in the bike-sharing industry.

Examples:

- *Urban Planners:* Knowing when and where people tend to use bikes can inform the development of bike lanes, parking facilities, or even the placement of bike-sharing stations.
- *Transportation Companies:* Businesses offering bike-sharing services can optimize their operations by deploying more bikes or adjusting pricing based on peak usage times and popular locations.
- *Marketing and Sales:* Retailers selling bikes or bike-related accessories can tailor their marketing campaigns, discounts, or promotions to align with the days where bike usage is highest.
- *Data Analysts and Researchers:* Professionals in these roles can use this information to conduct deeper analysis, identify trends, and make predictions about future bike usage patterns.
- *Tourism Industry*: In tourist-heavyf areas, understanding bike usage patterns can help tour companies plan routes and schedules for bike tours more effectively.

Purpose:

To study bike usage patterns across locations at different times aims to reveal geographic influences on daily bicycle utilization.

Are there specific patterns in the usage of bikes based on their geography during different times of the day?

Task Plan:

Load relevant dataset columns: 'start_station_name,' 'end_station_name,' 'start_lat,' 'start_lng,' 'end_lat,' 'end_lng,' 'started_at,' 'ended_at.'

Group bike stations by their geographic coordinates.

Categorize rides by time of day (e.g., morning, afternoon, evening) based on timestamps.

Calculate metrics for station usage and times of the day.

Visualization and Insights:

Create a scatter plot to display station usage patterns in relation to geographic location and time of day.

Analyze scatter plot for trends and insights.

This concise plan outlines the steps to explore bike station usage patterns based on location and time.

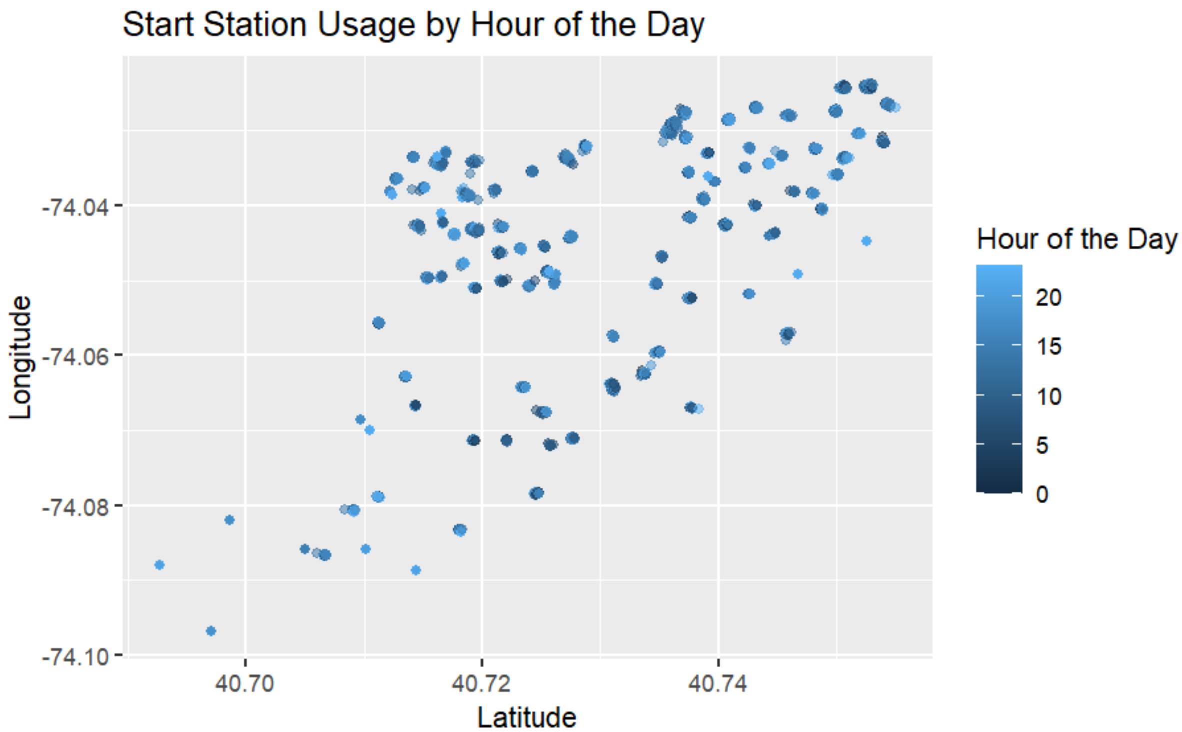
Code:

```
# Load necessary libraries

library(dplyr)

library(ggplot2)

library(lubridate)


# You have a dataset named 'data_all' with columns 'start_station_name', 'end_station_name',
'start_lat', 'start_lng', 'end_lat', 'end_lng', 'started_at', and 'ended_at'


# Convert 'started_at' and 'ended_at' to datetime format

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

data_all$ended_at <- as.POSIXct(data_all$ended_at, format = "%Y-%m-%d %H:%M:%S")


# Extract the hour of the day as a numeric value

data_all$start_hour_of_day <- hour(data_all$started_at)

data_all$end_hour_of_day <- hour(data_all$ended_at)


# Group and count by start station, hour of the day

start_station_usage <- data_all %>%
```

```
  select(start_station_name, start_lat, start_lng, start_hour_of_day) %>%
  group_by(start_station_name, start_lat, start_lng, start_hour_of_day) %>%
  summarise(start_count = n(), .groups = "drop")

# Group and count by end station, hour of the day
end_station_usage <- data_all %>%
  select(end_station_name, end_lat, end_lng, end_hour_of_day) %>%
  group_by(end_station_name, end_lat, end_lng, end_hour_of_day) %>%
  summarise(end_count = n(), .groups = "drop")

# Scatter plot example for start stations
ggplot(start_station_usage, aes(x = start_lat, y = start_lng, color = start_hour_of_day)) +
  geom_point(alpha = 0.5) +
  labs(title = "Start Station Usage by Hour of the Day",
       x = "Latitude", y = "Longitude",
       color = "Hour of the Day")
```

Erica Yang
BUS446
Final Project

Graph:



How can this be used in a business situation? For what type of job?

Note: The considerable increase in starting riders between 12 PM and 12 AM, compared to the period from 12 AM to 12 PM, presents an opportunity for businesses to optimize services based on these usage trends.

Understanding this spike can guide bike-sharing companies to strategically allocate more resources, such as deploying additional bikes or staffing, during these peak hours to meet heightened demand. It also allows for targeted marketing efforts, directing promotional campaigns toward these specific time frames to attract more users and maximize business potential.

Examples:

- *Data Analysts*: Analyzing bike usage patterns can help these professionals identify trends, peak hours, and popular routes. This information can be used to optimize bike-sharing services, plan maintenance schedules, or forecast demand for bikes and accessories.
- *Market Researchers*: They can utilize this data to understand consumer behavior related to biking, informing marketing strategies for bike-related products or services. Insights into when and where people use bikes most frequently can guide targeted advertising campaigns.

- *Business Development Managers*: Understanding bike usage patterns can help in forming partnerships or collaborations, such as working with local businesses to provide bike-related services or integrating biking facilities with other transportation modes.
- *Entrepreneurs in the Transportation Sector*: Entrepreneurs looking to start bike-sharing or bike rental businesses can use this data to identify underserved areas or peak times for potential expansion and marketing efforts.

Purpose:

To investigate if members and casual riders differ in ride durations based on the day and starting station.

Do members and casual riders tend to take longer, or shorter rides based on the day of the week and the starting station?

Task Plan:

Load the dataset with key columns: 'started_at,' 'ended_at,' 'start_station_name,' 'member_casual.'

Calculate ride durations by finding the time difference.

Group data by 'day_of_week,' 'start_station_name,' and 'member_casual.'

Calculate the mean ride duration for each group.

Create Visualization:

Create a bar chart to compare the average ride durations for members and casual riders on different days of the week and at different starting stations.

Analyze the bar chart to spot any trends or differences in ride durations between rider types.

This simplified plan helps you understand ride duration patterns for members and casual riders, considering the day of the week and starting stations.

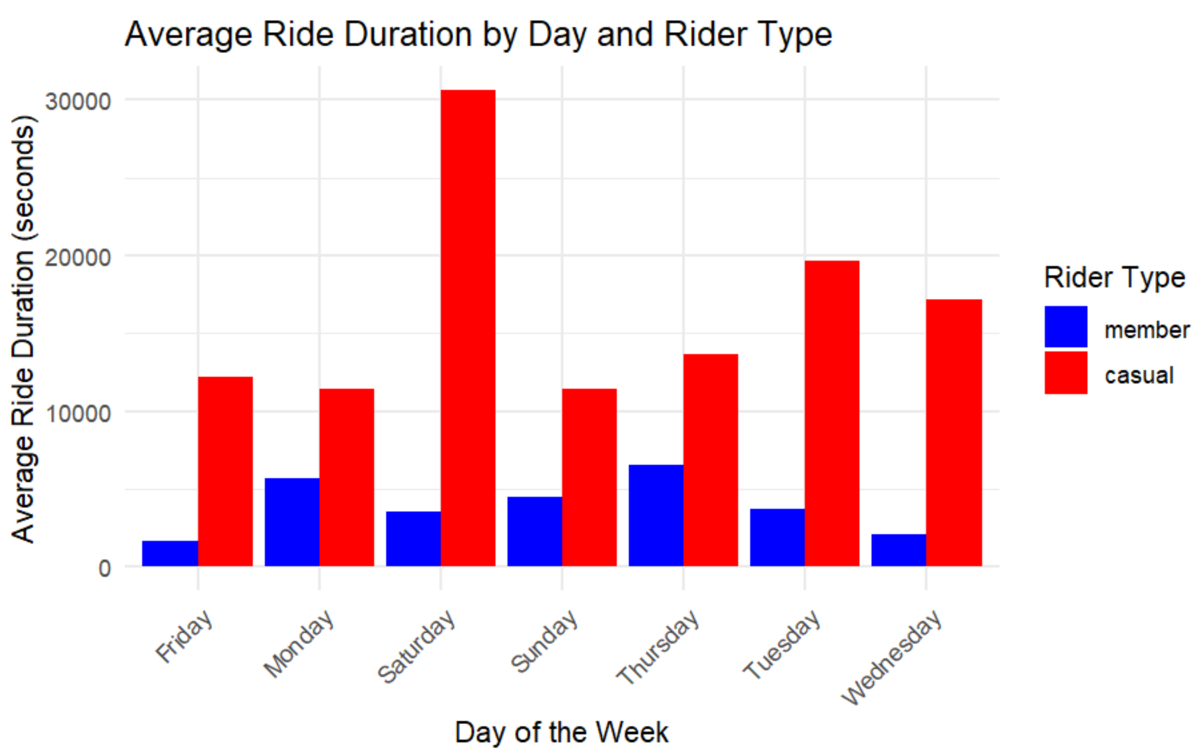
Code:

```
# Load necessary libraries

library(dplyr)

library(ggplot2)
```

```
# Data Preparation

data_all$start_hour <- as.POSIXlt(data_all$started_at)$hour # Extract the hour from 'started_at'
timestamp


# Calculate ride duration as the time difference in seconds

data_all$ride_duration <- as.numeric(difftime(data_all$ended_at, data_all$started_at, units =
"secs"))


# Filter out rows with missing values in member_casual column

data_all <- data_all %>%

  filter(!is.na(member_casual))


# Ensure member_casual column contains only 'member' and 'casual' levels

data_all$member_casual <- factor(data_all$member_casual, levels = c("member", "casual"))


# Group by day of the week, starting station, and rider type

ride_duration_by_day_station <- data_all %>%

  group_by(day_of_week = weekdays(as.Date(started_at)),

           start_station_name,

           member_casual) %>%

  summarise(mean_ride_duration = mean(ride_duration, na.rm = TRUE), .groups = "drop")


# Remove rows with NA values if any

ride_duration_by_day_station <- na.omit(ride_duration_by_day_station)


# Create a plot

ggplot(ride_duration_by_day_station, aes(x = day_of_week, y = mean_ride_duration, fill =
member_casual)) +

  geom_bar(stat = "identity", position = "dodge") +
```

```
labs(title = "Average Ride Duration by Day and Rider Type",
    x = "Day of the Week", y = "Average Ride Duration (seconds)",
    fill = "Rider Type") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
scale_fill_manual(values = c("member" = "blue", "casual" = "red"))
```

Graph:



How can this be used in a business situation? For what type of job?

Note: In general, casual riders seem to have longer ride duration than members on all days of the week.

Understanding whether members and casual riders take longer, or shorter rides based on the day of the week and starting station can provide insights into user behavior and preferences.

Examples:

- *Bike Share Operations Managers*: Knowing the ride duration based on the day of the week and starting station can help in optimizing bike distribution. If certain stations see longer rides on weekends, they might need more bikes available during those times to meet demand.
- *Marketing and Sales Teams*: Understanding the differences in ride durations can help in targeted marketing campaigns. If casual riders tend to take longer rides on specific days or from starting stations, marketing efforts can be tailored to attract more casual riders during those times or locations.
- *Product Development Teams*: Insights into ride durations can inform the development of subscription plans or pricing structures. If members tend to take consistently shorter rides on weekdays from specific stations, there could be opportunities to offer targeted subscription packages incentivizing shorter rides for regular commuters.
- *Data Analysts and Researchers*: Professionals in these roles can delve deeper into this data to identify patterns and correlations that might not be immediately evident. This can lead to more nuanced insights, helping businesses make data-driven decisions in various aspects of operations and strategy.
- *Customer Experience and Support Teams*: Understanding user behavior regarding ride durations can also help in improving user experiences. For instance, if there are stations or days where riders tend to face issues or longer wait times due to high demand, customer support strategies can be adapted accordingly.