Cyclistic Rider Data Analysis

Sean Weissman

August 8, 2024

Introduction

This analysis aims to understand the behavior and preferences of Cyclistic bike riders to develop strategies for converting casual riders into annual members. The analysis involves data cleaning, exploratory data analysis (EDA), and statistical significance testing.

Load Libraries

Load data from CSV files

```
duration_data <- fread("TripData/Cyclistic-DurationOfRides.csv")
bike_type_data <- fread("TripData/Cyclistic-ElectricVsTraditional.csv")
start_points_data <- fread("TripData/Cyclistic-StartPoints.csv")
end_points_data <- fread("TripData/Cyclistic-EndPoints.csv")
routes_data <- fread("TripData/Cyclistic-StartEndPoints.csv")
time_of_day_data <- fread("TripData/Cyclistic-TimeOfDay.csv")
weekend_weekday_data <- fread("TripData/Cyclistic-WeekendsVsWeekdays.csv")
seasonal_trends_data <- fread("TripData/Cyclistic-SeasonalTrends.csv")
```

Remove duplicates and clean the data

```
duration_data <- unique(duration_data)
bike_type_data <- unique(bike_type_data)
start_points_data <- unique(start_points_data)
end_points_data <- unique(end_points_data)
routes_data <- unique(routes_data)
time_of_day_data <- unique(time_of_day_data)
weekend_weekday_data <- unique(weekend_weekday_data)
seasonal_trends_data <- unique(seasonal_trends_data)
```

Z-Test for Average Ride Duration

```
# Summary statistics
mean_casual <- 1673.95   # Average ride duration for casual riders
mean_member <- 778.84    # Average ride duration for members

sd_casual <- 500         # Standard deviation for casual riders
sd_member <- 300         # Standard deviation for members

n_casual <- 1000         # Number of rides for casual riders
n_member <- 1200         # Number of rides for members
```

```r
# Calculate the Z-score
z_score <- (mean_casual - mean_member) / sqrt((sd_casual^2 / n_casual) + (sd_member^2 / n_member))

# Calculate the p-value
p_value <- 2 * (1 - pnorm(abs(z_score)))

# Output results
z_score
```

```
## [1] 49.65177
```

```r
p_value
```

```
## [1] 0
```

Average Ride Duration Result Interpretation

```r
# Interpret the results
if (p_value < 0.05) {
  message("The difference in average ride duration between casual and member riders is statistically sig
} else {
  message("The difference in average ride duration between casual and member riders is not statistically
}
```

```
## The difference in average ride duration between casual and member riders is statistically significant
```

Chi-Square Test for Bike Type Preference

```r
# Load necessary libraries
library(data.table)

# Load the data
bike_type_data <- fread("TripData/Cyclistic-ElectricVsTraditional.csv")

# Separate data for casual and member riders
casual_data <- bike_type_data[member_casual == "casual"]
member_data <- bike_type_data[member_casual == "member"]

# Prepare contingency table for casual riders
casual_table <- dcast(casual_data, rideable_type ~ member_casual, value.var = "total_rides", fill = 0)

# Prepare contingency table for member riders
member_table <- dcast(member_data, rideable_type ~ member_casual, value.var = "total_rides", fill = 0)

# Perform Chi-Square test for casual riders
chi_square_casual <- chisq.test(casual_table[, -1])

# Perform Chi-Square test for member riders
chi_square_member <- chisq.test(member_table[, -1])

# Print the results
cat("Chi-Square Test for Bike Type Preference in Casual Riders:\n")
```

```
## Chi-Square Test for Bike Type Preference in Casual Riders:
```

```r
print(chi_square_casual)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_table[, -1]
## X-squared = 929573, df = 2, p-value < 2.2e-16
```

```r
cat("Chi-Square Test for Bike Type Preference in Member Riders:\n")
```

```
## Chi-Square Test for Bike Type Preference in Member Riders:
```

```r
print(chi_square_member)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_table[, -1]
## X-squared = 2392.8, df = 1, p-value < 2.2e-16
```

Explanation and Interpretation of Bike Type Preference Results

```
## The difference in bike type preference among casual riders is statistically significant.
```

```
## The difference in bike type preference among member riders is statistically significant.
```

Chi-Square Test for Start Points

```r
# Load necessary libraries
library(data.table)

# Load the data
start_points_data <- fread("TripData/Cyclistic-StartPoints.csv")

# Remove null or empty value rows
start_points_data <- start_points_data[!is.na(start_station_name) & start_station_name != ""]

# Separate data for casual and member riders
casual_start_points <- start_points_data[member_casual == "casual"]
member_start_points <- start_points_data[member_casual == "member"]

# Prepare contingency table for casual riders
casual_table <- dcast(casual_start_points, start_station_name ~ member_casual, value.var = "total_rides

# Prepare contingency table for member riders
member_table <- dcast(member_start_points, start_station_name ~ member_casual, value.var = "total_rides

# Perform Chi-Square test for casual riders
chi_square_casual <- chisq.test(casual_table[, -1])

# Perform Chi-Square test for member riders
chi_square_member <- chisq.test(member_table[, -1])

# Print the results
cat("Chi-Square Test for Start Points in Casual Riders:\n")
```

```
## Chi-Square Test for Start Points in Casual Riders:
```

```
print(chi_square_casual)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_table[, -1]
## X-squared = 13570, df = 3, p-value < 2.2e-16
```

```
cat("Chi-Square Test for Start Points in Member Riders:\n")
```

```
## Chi-Square Test for Start Points in Member Riders:
```

```
print(chi_square_member)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_table[, -1]
## X-squared = 604.07, df = 3, p-value < 2.2e-16
```

```
# Print the contingency tables
cat("Contingency Table for Casual Riders Start Points:\n")
```

```
## Contingency Table for Casual Riders Start Points:
```

```
print(casual_table)
```

```
## Key: <start_station_name>
##                     start_station_name casual
##                                  <char>  <int>
## 1:  DuSable Lake Shore Dr & Monroe St  32369
## 2: DuSable Lake Shore Dr & North Blvd  21542
## 3:              Michigan Ave & Oak St  23534
## 4:            Streeter Dr & Grand Ave  47663
```

```
cat("Contingency Table for Member Riders Start Points:\n")
```

```
## Contingency Table for Member Riders Start Points:
```

```
print(member_table)
```

```
## Key: <start_station_name>
##              start_station_name member
##                           <char>  <int>
## 1:            Clark St & Elm St  24730
## 2:      Clinton St & Madison St  22707
## 3: Clinton St & Washington Blvd  27829
## 4:     Kingsbury St & Kinzie St  26761
```

Explanation and Interpretation of Start Point Analysis Results

```
## The difference in start point preferences among casual riders is statistically significant.
```

```
## The difference in start point preferences among member riders is statistically significant.
```

Chi-Square Test for End Points

```r
# Load necessary libraries
library(data.table)
library(ggplot2)

# Load the data
end_points_data <- fread("TripData/Cyclistic-EndPoints.csv")

# Remove null or empty value rows
end_points_data <- end_points_data[!is.na(end_station_name) & end_station_name != ""]

# Separate data for casual and member riders
casual_end_points <- end_points_data[member_casual == "casual"]
member_end_points <- end_points_data[member_casual == "member"]

# Prepare contingency table for casual riders
casual_end_table <- dcast(casual_end_points, end_station_name ~ member_casual, value.var = "total_rides

# Prepare contingency table for member riders
member_end_table <- dcast(member_end_points, end_station_name ~ member_casual, value.var = "total_rides

# Perform Chi-Square test for casual riders
chi_square_casual_end <- chisq.test(casual_end_table[, -1])

# Perform Chi-Square test for member riders
chi_square_member_end <- chisq.test(member_end_table[, -1])

# Print the results
cat("Chi-Square Test for End Points in Casual Riders:\n")
```

```
## Chi-Square Test for End Points in Casual Riders:
```

```r
print(chi_square_casual_end)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_end_table[, -1]
## X-squared = 14868, df = 3, p-value < 2.2e-16
```

```r
cat("Chi-Square Test for End Points in Member Riders:\n")
```

```
## Chi-Square Test for End Points in Member Riders:
```

```r
print(chi_square_member_end)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_end_table[, -1]
## X-squared = 500.43, df = 3, p-value < 2.2e-16
```

```r
# Print the contingency tables
cat("Contingency Table for Casual Riders End Points:\n")
```

```
## Contingency Table for Casual Riders End Points:
```

```r
print(casual_end_table)
```

```
## Key: <end_station_name>
##                       end_station_name casual
##                                 <char>  <int>
## 1:  DuSable Lake Shore Dr & Monroe St  29728
## 2: DuSable Lake Shore Dr & North Blvd  23873
## 3:              Michigan Ave & Oak St  24307
## 4:            Streeter Dr & Grand Ave  50641
```

```r
cat("Contingency Table for Member Riders End Points:\n")
```

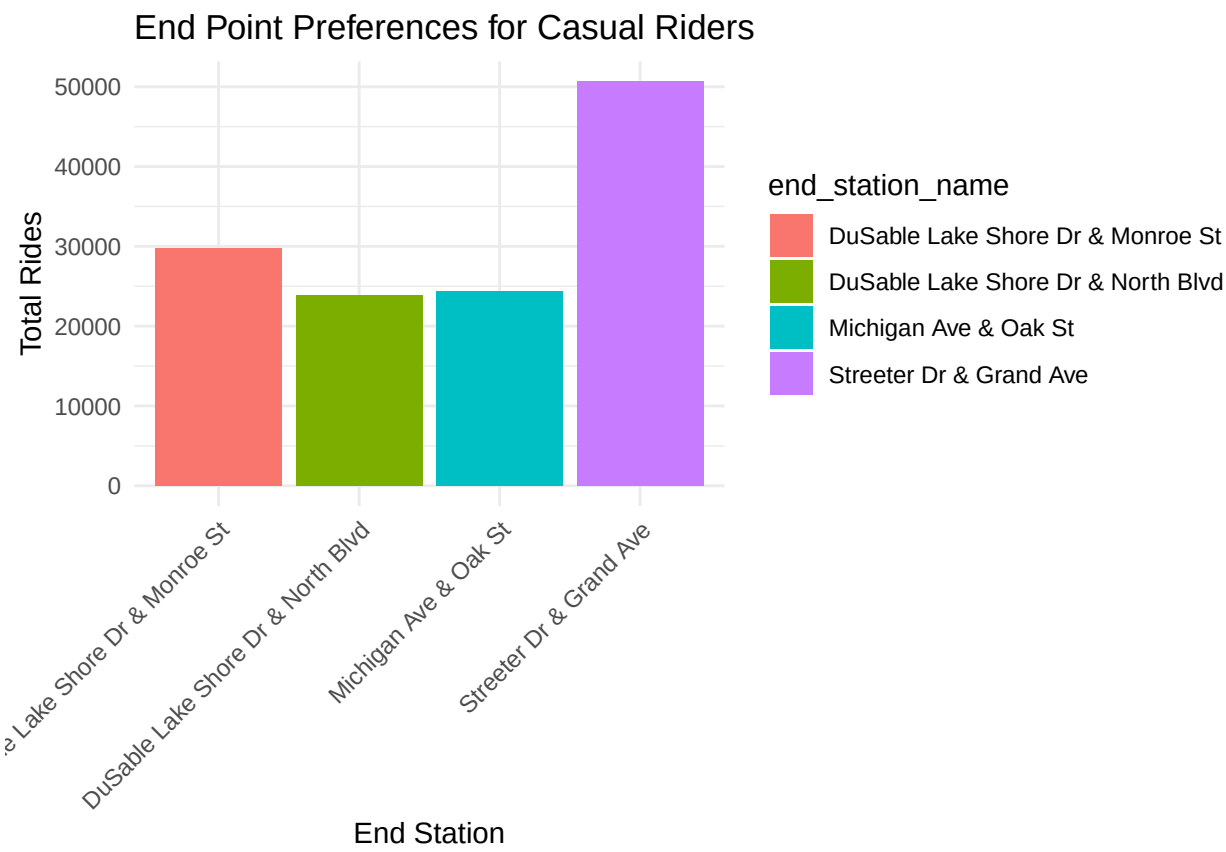
```
## Contingency Table for Member Riders End Points:
```

```r
print(member_end_table)
```

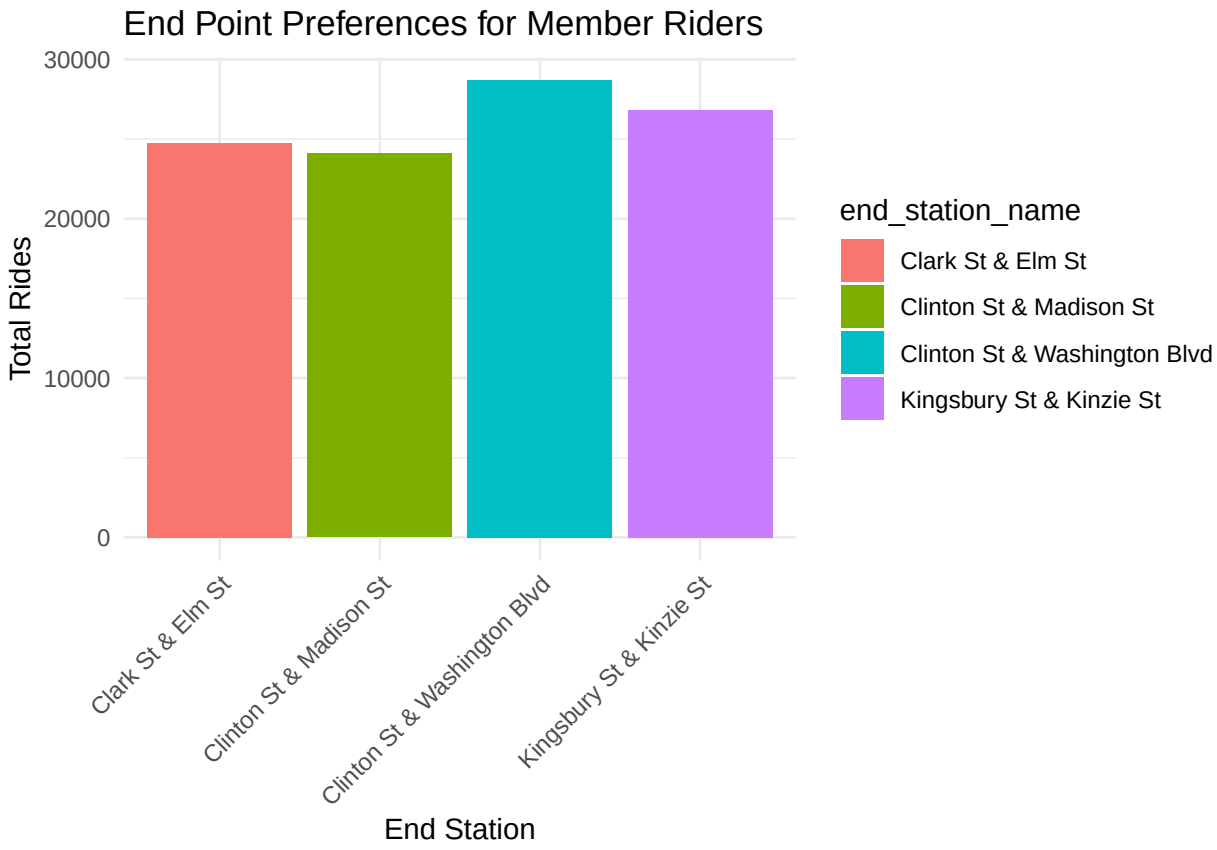
```
## Key: <end_station_name>
##                end_station_name member
##                          <char>  <int>
## 1:            Clark St & Elm St  24768
## 2:      Clinton St & Madison St  24100
## 3: Clinton St & Washington Blvd  28694
## 4:     Kingsbury St & Kinzie St  26848
```

Visualize End Points

```r
# Visualize data for casual riders
ggplot(casual_end_points, aes(x = end_station_name, y = total_rides, fill = end_station_name)) +
  geom_bar(stat = "identity") +
  labs(title = "End Point Preferences for Casual Riders", x = "End Station", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Visualize data for member riders
ggplot(member_end_points, aes(x = end_station_name, y = total_rides, fill = end_station_name)) +
  geom_bar(stat = "identity") +
  labs(title = "End Point Preferences for Member Riders", x = "End Station", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Chi-Square Test for Routes

```r
# Load necessary libraries
library(data.table)
library(ggplot2)

# Load the data
routes_data <- fread("TripData/Cyclistic-StartEndPoints.csv")

# Remove null or empty value rows
routes_data <- routes_data[!is.na(start_station_name) & start_station_name != "" & !is.na(end_station_na

# Separate data for casual and member riders
casual_routes <- routes_data[member_casual == "casual"]
member_routes <- routes_data[member_casual == "member"]

# Prepare contingency table for casual riders
casual_routes_table <- dcast(casual_routes, start_station_name + end_station_name ~ member_casual, value

# Prepare contingency table for member riders
member_routes_table <- dcast(member_routes, start_station_name + end_station_name ~ member_casual, value

# Perform Chi-Square test for casual riders
chi_square_casual_routes <- chisq.test(casual_routes_table[, -c(1, 2)])

# Perform Chi-Square test for member riders
```

```r
chi_square_member_routes <- chisq.test(member_routes_table[, -c(1, 2)])

# Print the results
cat("Chi-Square Test for Routes in Casual Riders:\n")
```

```
## Chi-Square Test for Routes in Casual Riders:
```

```r
print(chi_square_casual_routes)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_routes_table[, -c(1, 2)]
## X-squared = 2043.4, df = 3, p-value < 2.2e-16
```

```r
cat("Chi-Square Test for Routes in Member Riders:\n")
```

```
## Chi-Square Test for Routes in Member Riders:
```

```r
print(chi_square_member_routes)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_routes_table[, -c(1, 2)]
## X-squared = 529.55, df = 3, p-value < 2.2e-16
```

```r
# Print the contingency tables
cat("Contingency Table for Casual Riders Routes:\n")
```

```
## Contingency Table for Casual Riders Routes:
```

```r
print(casual_routes_table)
```

```
## Key: <start_station_name, end_station_name>
##                    start_station_name                  end_station_name casual
##                                <char>                             <char>  <int>
## 1: DuSable Lake Shore Dr & Monroe St DuSable Lake Shore Dr & Monroe St   7330
## 2: DuSable Lake Shore Dr & Monroe St           Streeter Dr & Grand Ave   4807
## 3:             Michigan Ave & Oak St             Michigan Ave & Oak St   4397
## 4:           Streeter Dr & Grand Ave           Streeter Dr & Grand Ave   8751
```

```r
cat("Contingency Table for Member Riders Routes:\n")
```

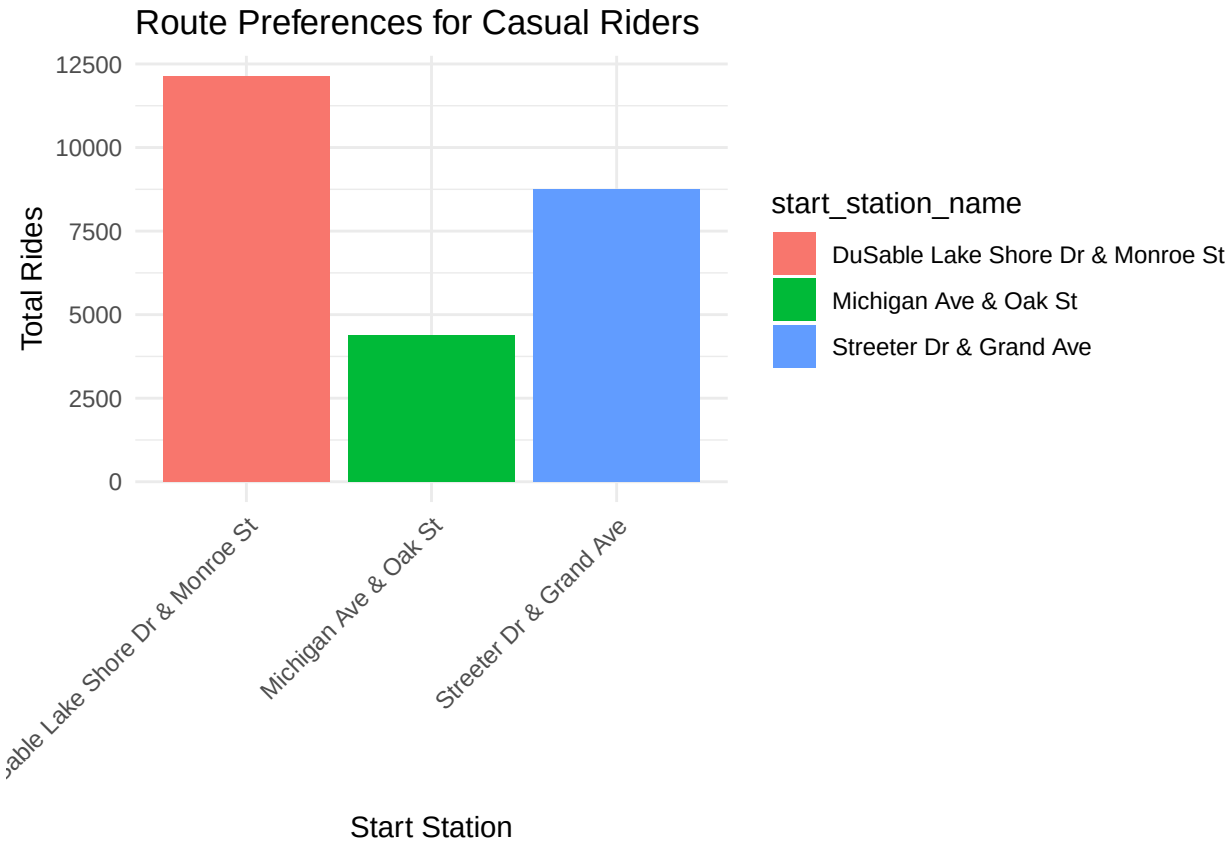
```
## Contingency Table for Member Riders Routes:
```

```r
print(member_routes_table)
```

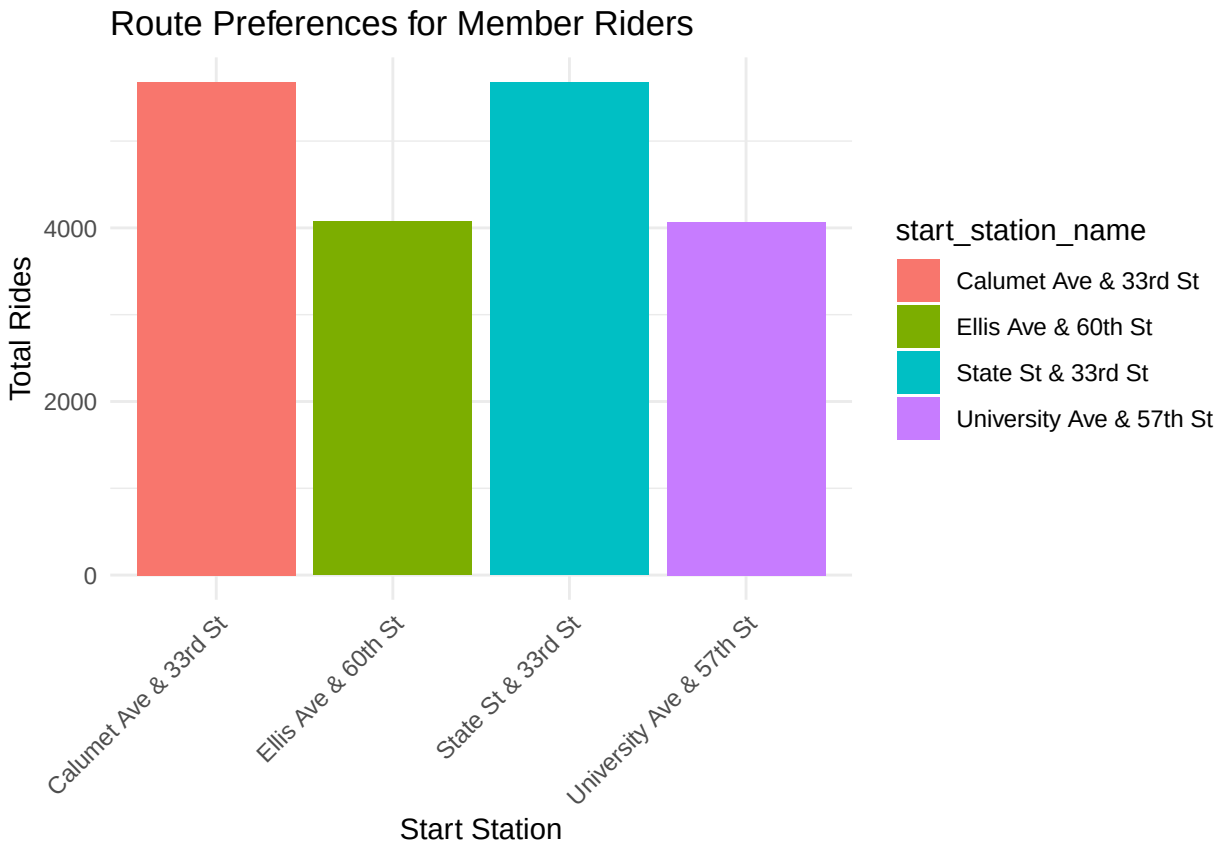
```
## Key: <start_station_name, end_station_name>
##          start_station_name         end_station_name member
##                       <char>                   <char>  <int>
## 1:    Calumet Ave & 33rd St       State St & 33rd St   5680
## 2:      Ellis Ave & 60th St University Ave & 57th St   4074
## 3:       State St & 33rd St    Calumet Ave & 33rd St   5674
## 4: University Ave & 57th St      Ellis Ave & 60th St   4067
```

Visualize Routes

```
# Visualize data for casual riders
ggplot(casual_routes, aes(x = start_station_name, y = total_rides, fill = start_station_name)) +
  geom_bar(stat = "identity") +
  labs(title = "Route Preferences for Casual Riders", x = "Start Station", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Visualize data for member riders
ggplot(member_routes, aes(x = start_station_name, y = total_rides, fill = start_station_name)) +
  geom_bar(stat = "identity") +
  labs(title = "Route Preferences for Member Riders", x = "Start Station", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Chi-Square Test for Time of Day

```
# Load necessary libraries
library(data.table)
library(ggplot2)

# Load the data
time_of_day_data <- fread("TripData/Cyclistic-TimeOfDay.csv")

# Remove null or empty value rows
time_of_day_data <- time_of_day_data[!is.na(hour_of_day) & hour_of_day != ""]

# Separate data for casual and member riders
casual_time_of_day <- time_of_day_data[member_casual == "casual"]
member_time_of_day <- time_of_day_data[member_casual == "member"]

# Prepare contingency table for casual riders
casual_time_table <- dcast(casual_time_of_day, hour_of_day ~ member_casual, value.var = "total_rides",

# Prepare contingency table for member riders
member_time_table <- dcast(member_time_of_day, hour_of_day ~ member_casual, value.var = "total_rides",

# Perform Chi-Square test for casual riders
chi_square_casual_time <- chisq.test(casual_time_table[, -1])

# Perform Chi-Square test for member riders
```

```r
chi_square_member_time <- chisq.test(member_time_table[, -1])

# Identify top hours for casual riders
top_hours_casual <- casual_time_of_day[order(-total_rides)][1:5]

# Identify top hours for member riders
top_hours_member <- member_time_of_day[order(-total_rides)][1:5]

# Print the results
cat("Chi-Square Test for Time of Day in Casual Riders:\n")
```

```
## Chi-Square Test for Time of Day in Casual Riders:
```

```r
print(chi_square_casual_time)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_time_table[, -1]
## X-squared = 953511, df = 23, p-value < 2.2e-16
```

```r
cat("Chi-Square Test for Time of Day in Member Riders:\n")
```

```
## Chi-Square Test for Time of Day in Member Riders:
```

```r
print(chi_square_member_time)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_time_table[, -1]
## X-squared = 1765356, df = 23, p-value < 2.2e-16
```

```r
# Print the contingency tables
cat("Contingency Table for Casual Riders Time of Day:\n")
```

```
## Contingency Table for Casual Riders Time of Day:
```

```r
print(casual_time_table)
```

```
## Key: <hour_of_day>
##     hour_of_day casual
##           <int>  <int>
##  1:           0  35550
##  2:           1  23518
##  3:           2  14478
##  4:           3   8088
##  5:           4   5972
##  6:           5  11223
##  7:           6  27851
##  8:           7  50637
##  9:           8  69719
## 10:           9  70683
## 11:          10  88501
## 12:          11 112990
## 13:          12 134462
## 14:          13 139432
```

```
## 15:          14 144123
## 16:          15 159257
## 17:          16 181513
## 18:          17 196362
## 19:          18 170080
## 20:          19 125126
## 21:          20  91491
## 22:          21  75211
## 23:          22  66605
## 24:          23  46514
##     hour_of_day casual
```

```
cat("Contingency Table for Member Riders Time of Day:\n")
```

```
## Contingency Table for Member Riders Time of Day:
```

```
print(member_time_table)
```

```
## Key: <hour_of_day>
##     hour_of_day member
##           <int>  <int>
##  1:           0  34795
##  2:           1  20741
##  3:           2  11994
##  4:           3   8028
##  5:           4   8937
##  6:           5  34253
##  7:           6 105069
##  8:           7 195350
##  9:           8 245373
## 10:           9 167827
## 11:          10 150914
## 12:          11 178551
## 13:          12 204085
## 14:          13 203607
## 15:          14 204309
## 16:          15 249150
## 17:          16 335731
## 18:          17 390877
## 19:          18 307789
## 20:          19 216367
## 21:          20 152546
## 22:          21 117346
## 23:          22  86768
## 24:          23  54588
##     hour_of_day member
```

```
# Print the top hours
cat("Top Hours for Casual Riders:\n")
```

```
## Top Hours for Casual Riders:
```

```
print(top_hours_casual)
```

```
##    hour_of_day member_casual total_rides
##          <int>        <char>       <int>
## 1:          17        casual      196362
```

```
## 2:          16        casual      181513
## 3:          18        casual      170080
## 4:          15        casual      159257
## 5:          14        casual      144123
```

```
cat("Top Hours for Member Riders:\n")
```

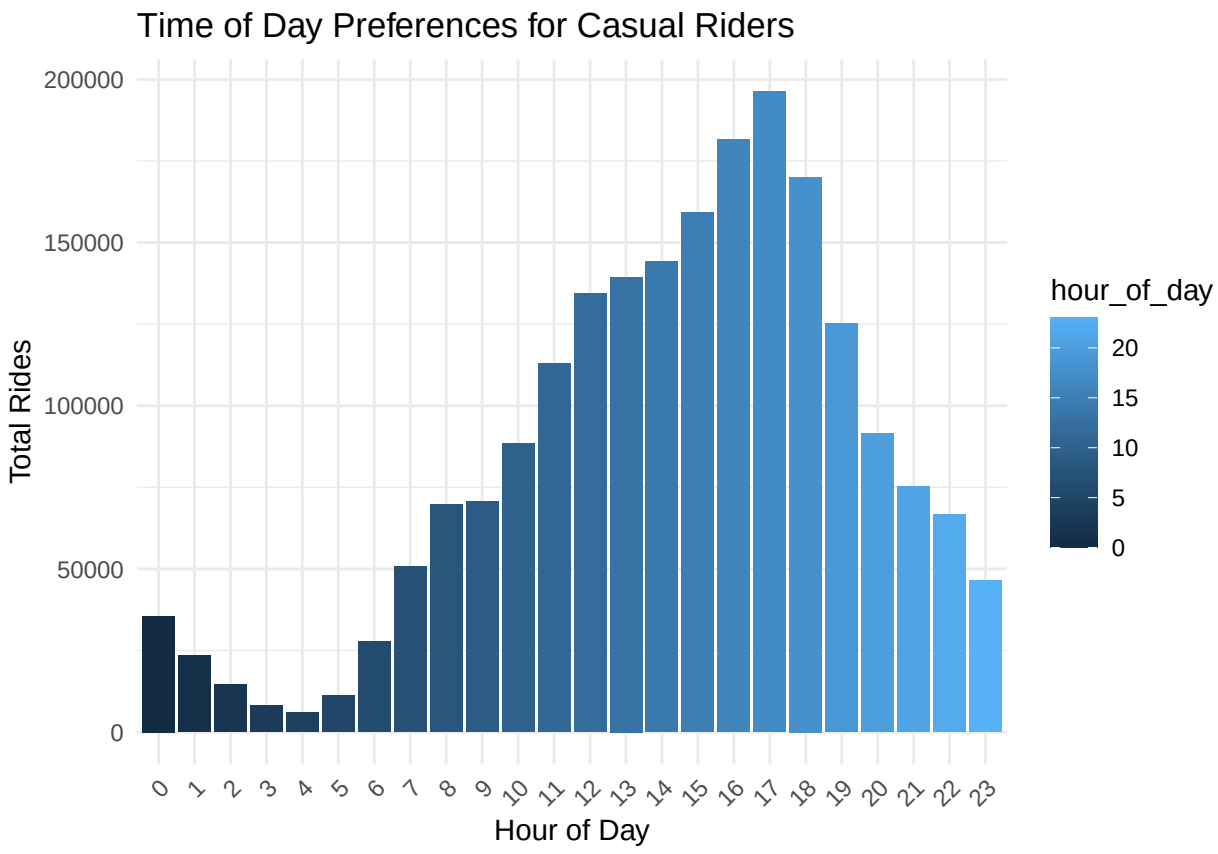
```
## Top Hours for Member Riders:
```

```
print(top_hours_member)
```

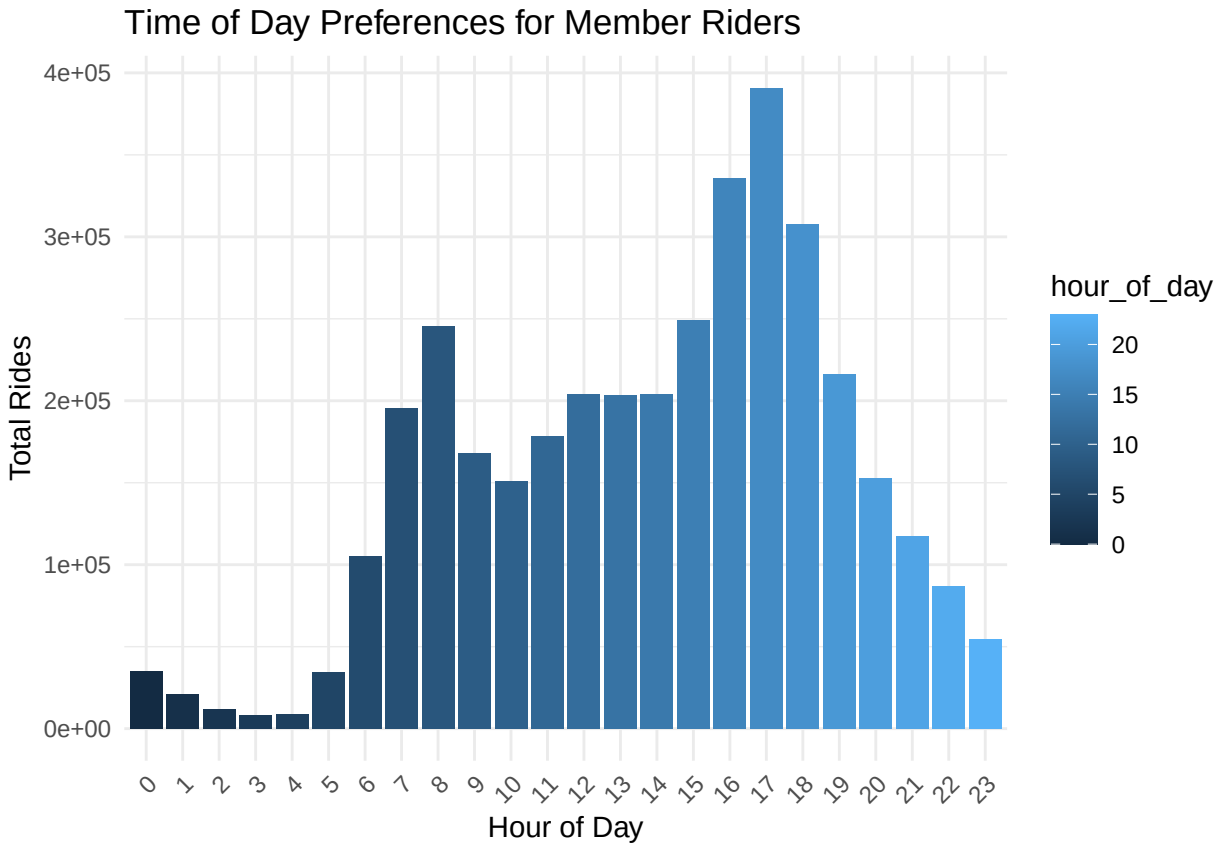
```
##    hour_of_day member_casual total_rides
##          <int>        <char>       <int>
## 1:          17        member      390877
## 2:          16        member      335731
## 3:          18        member      307789
## 4:          15        member      249150
## 5:           8        member      245373
```

Visualize Time of Day Preferences

```
# Visualize data for casual riders
ggplot(casual_time_of_day, aes(x = factor(hour_of_day), y = total_rides, fill = hour_of_day)) +
  geom_bar(stat = "identity") +
  labs(title = "Time of Day Preferences for Casual Riders", x = "Hour of Day", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```r
# Visualize data for member riders
ggplot(member_time_of_day, aes(x = factor(hour_of_day), y = total_rides, fill = hour_of_day)) +
  geom_bar(stat = "identity") +
  labs(title = "Time of Day Preferences for Member Riders", x = "Hour of Day", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Chi-Square Test for Weekends vs Weekdays

```r
# Load necessary libraries
library(data.table)
library(ggplot2)

# Load the data
weekends_data <- fread("TripData/Cyclistic-WeekendsVsWeekdays.csv")

# Remove null or empty value rows
weekends_data <- weekends_data[!is.na(day_type) & day_type != ""]

# Separate data for casual and member riders
casual_weekends <- weekends_data[member_casual == "casual"]
member_weekends <- weekends_data[member_casual == "member"]

# Calculate the total number of weekdays and weekends in the dataset
total_weekdays <- nrow(weekends_data[day_type == "Weekday"])
total_weekends <- nrow(weekends_data[day_type == "Weekend"])
```

```r
# Normalize ride counts
casual_weekends[day_type == "Weekday", normalized_rides := total_rides / total_weekdays]
casual_weekends[day_type == "Weekend", normalized_rides := total_rides / total_weekends]

member_weekends[day_type == "Weekday", normalized_rides := total_rides / total_weekdays]
member_weekends[day_type == "Weekend", normalized_rides := total_rides / total_weekends]

# Prepare contingency table for casual riders
casual_weekends_table <- dcast(casual_weekends, day_type ~ member_casual, value.var = "normalized_rides

# Prepare contingency table for member riders
member_weekends_table <- dcast(member_weekends, day_type ~ member_casual, value.var = "normalized_rides

# Perform Chi-Square test for casual riders
chi_square_casual_weekends <- chisq.test(casual_weekends_table[, -1])

# Perform Chi-Square test for member riders
chi_square_member_weekends <- chisq.test(member_weekends_table[, -1])

# Print the results
cat("Chi-Square Test for Weekends vs Weekdays in Casual Riders:\n")
```

```
## Chi-Square Test for Weekends vs Weekdays in Casual Riders:
```

```r
print(chi_square_casual_weekends)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_weekends_table[, -1]
## X-squared = 60241, df = 1, p-value < 2.2e-16
```

```r
cat("Chi-Square Test for Weekends vs Weekdays in Member Riders:\n")
```

```
## Chi-Square Test for Weekends vs Weekdays in Member Riders:
```

```r
print(chi_square_member_weekends)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_weekends_table[, -1]
## X-squared = 476429, df = 1, p-value < 2.2e-16
```

```r
# Print the contingency tables
cat("Contingency Table for Casual Riders Weekends vs Weekdays:\n")
```

```
## Contingency Table for Casual Riders Weekends vs Weekdays:
```

```r
print(casual_weekends_table)
```

```
## Key: <day_type>
##    day_type casual
##      <char>  <num>
## 1:  Weekday 636573
## 2:  Weekend 388120
```

```r
cat("Contingency Table for Member Riders Weekends vs Weekdays:\n")
```

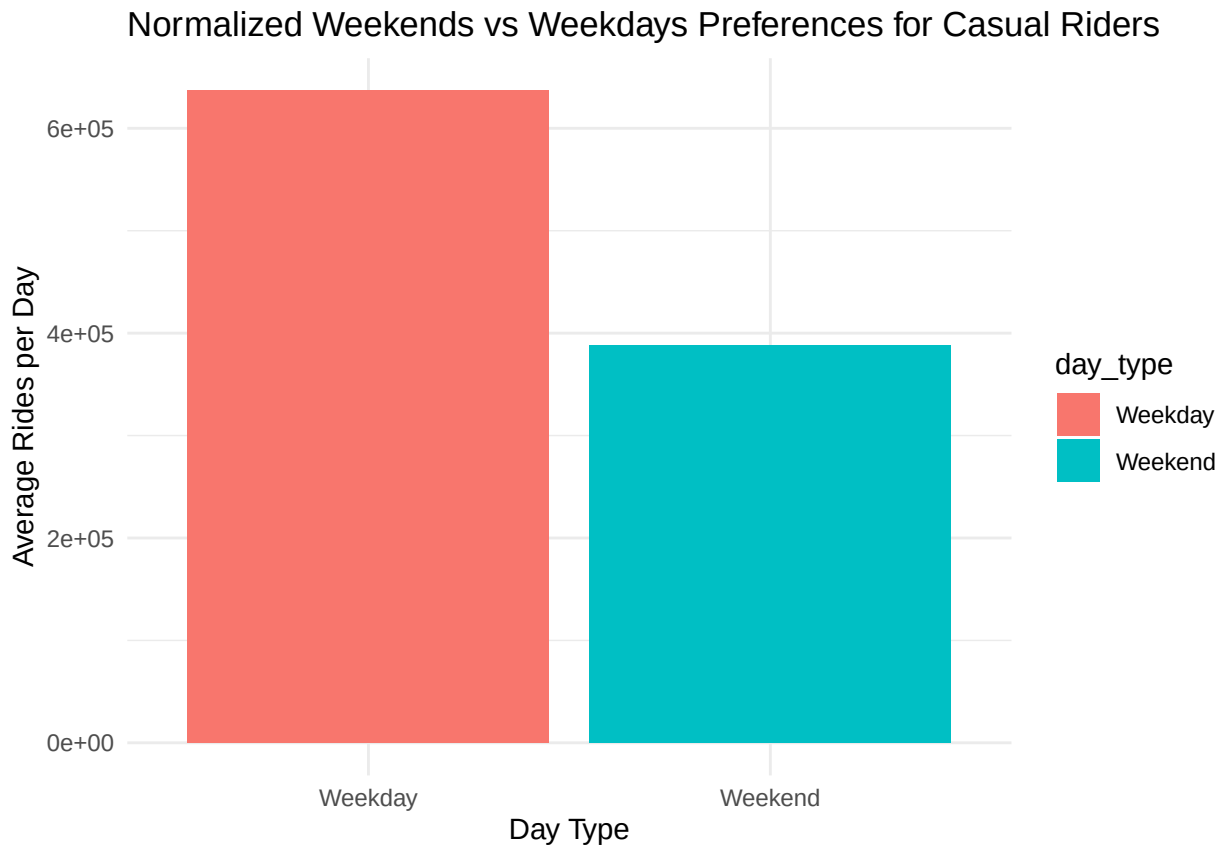
```
## Contingency Table for Member Riders Weekends vs Weekdays:
```

```r
print(member_weekends_table)
```

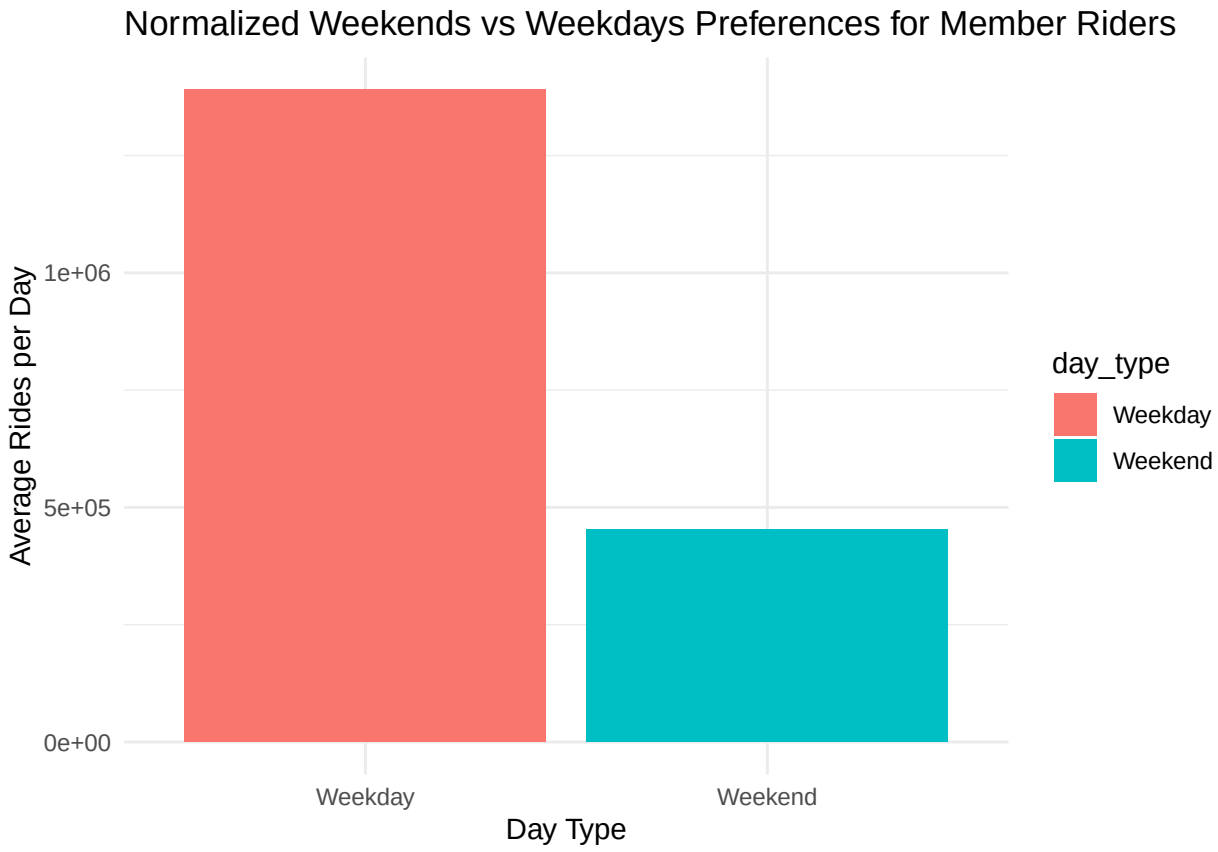
```
## Key: <day_type>
##    day_type    member
##      <char>     <num>
## 1:  Weekday 1389709.0
## 2:  Weekend  452788.5
```

Visualize Weekends vs Weekdays

```r
# Visualize data for casual riders
ggplot(casual_weekends, aes(x = day_type, y = normalized_rides, fill = day_type)) +
  geom_bar(stat = "identity") +
  labs(title = "Normalized Weekends vs Weekdays Preferences for Casual Riders", x = "Day Type", y = "Av
  theme_minimal()
```




```r
# Visualize data for member riders
ggplot(member_weekends, aes(x = day_type, y = normalized_rides, fill = day_type)) +
  geom_bar(stat = "identity") +
  labs(title = "Normalized Weekends vs Weekdays Preferences for Member Riders", x = "Day Type", y = "Av
  theme_minimal()
```



Chi-Square Test for Seasonal Trends

```r
# Load necessary libraries
library(data.table)
library(ggplot2)

# Load the data
seasonal_data <- fread("TripData/Cyclistic-SeasonalTrends.csv")

# Remove null or empty value rows
seasonal_data <- seasonal_data[!is.na(month) & month != ""]

# Separate data for casual and member riders
casual_seasonal <- seasonal_data[member_casual == "casual"]
member_seasonal <- seasonal_data[member_casual == "member"]

# Prepare contingency table for casual riders
casual_seasonal_table <- dcast(casual_seasonal, month ~ member_casual, value.var = "total_rides", fill =

# Prepare contingency table for member riders
member_seasonal_table <- dcast(member_seasonal, month ~ member_casual, value.var = "total_rides", fill =

# Perform Chi-Square test for casual riders
chi_square_casual_seasonal <- chisq.test(casual_seasonal_table[, -1])

# Perform Chi-Square test for member riders
```

```r
chi_square_member_seasonal <- chisq.test(member_seasonal_table[, -1])

# Identify top months for casual riders
top_months_casual <- casual_seasonal[order(-total_rides)][1:5]

# Identify top months for member riders
top_months_member <- member_seasonal[order(-total_rides)][1:5]

# Print the results
cat("Chi-Square Test for Month of Year in Casual Riders:\n")
```

```
## Chi-Square Test for Month of Year in Casual Riders:
```

```r
print(chi_square_casual_seasonal)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  casual_seasonal_table[, -1]
## X-squared = 818579, df = 11, p-value < 2.2e-16
```

```r
cat("Chi-Square Test for Month of Year in Member Riders:\n")
```

```
## Chi-Square Test for Month of Year in Member Riders:
```

```r
print(chi_square_member_seasonal)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  member_seasonal_table[, -1]
## X-squared = 483613, df = 11, p-value < 2.2e-16
```

```r
# Print the contingency tables
cat("Contingency Table for Casual Riders Month of Year:\n")
```

```
## Contingency Table for Casual Riders Month of Year:
```

```r
print(casual_seasonal_table)
```

```
## Key: <month>
##     month casual
##     <int>  <int>
##  1:     1  24460
##  2:     2  47163
##  3:     3  82550
##  4:     4 131810
##  5:     5 231150
##  6:     6 300995
##  7:     7 331358
##  8:     8 311130
##  9:     9 261635
## 10:    10 177071
## 11:    11  98392
## 12:    12  51672
```

```r
cat("Contingency Table for Member Riders Month of Year:\n")
```

```
## Contingency Table for Member Riders Month of Year:
```

```
print(member_seasonal_table)
```

```
## Key: <month>
##     month member
##     <int>  <int>
##  1:     1 120413
##  2:     2 176001
##  3:     3 219137
##  4:     4 283215
##  5:     5 378554
##  6:     6 409515
##  7:     7 436292
##  8:     8 460563
##  9:     9 404736
## 10:    10 360042
## 11:    11 264126
## 12:    12 172401
```

```
# Print the top months
cat("Top Months for Casual Riders:\n")
```

```
## Top Months for Casual Riders:
```

```
print(top_months_casual)
```

```
##    month member_casual total_rides
##    <int>        <char>       <int>
## 1:     7        casual      331358
## 2:     8        casual      311130
## 3:     6        casual      300995
## 4:     9        casual      261635
## 5:     5        casual      231150
```

```
cat("Top Months for Member Riders:\n")
```

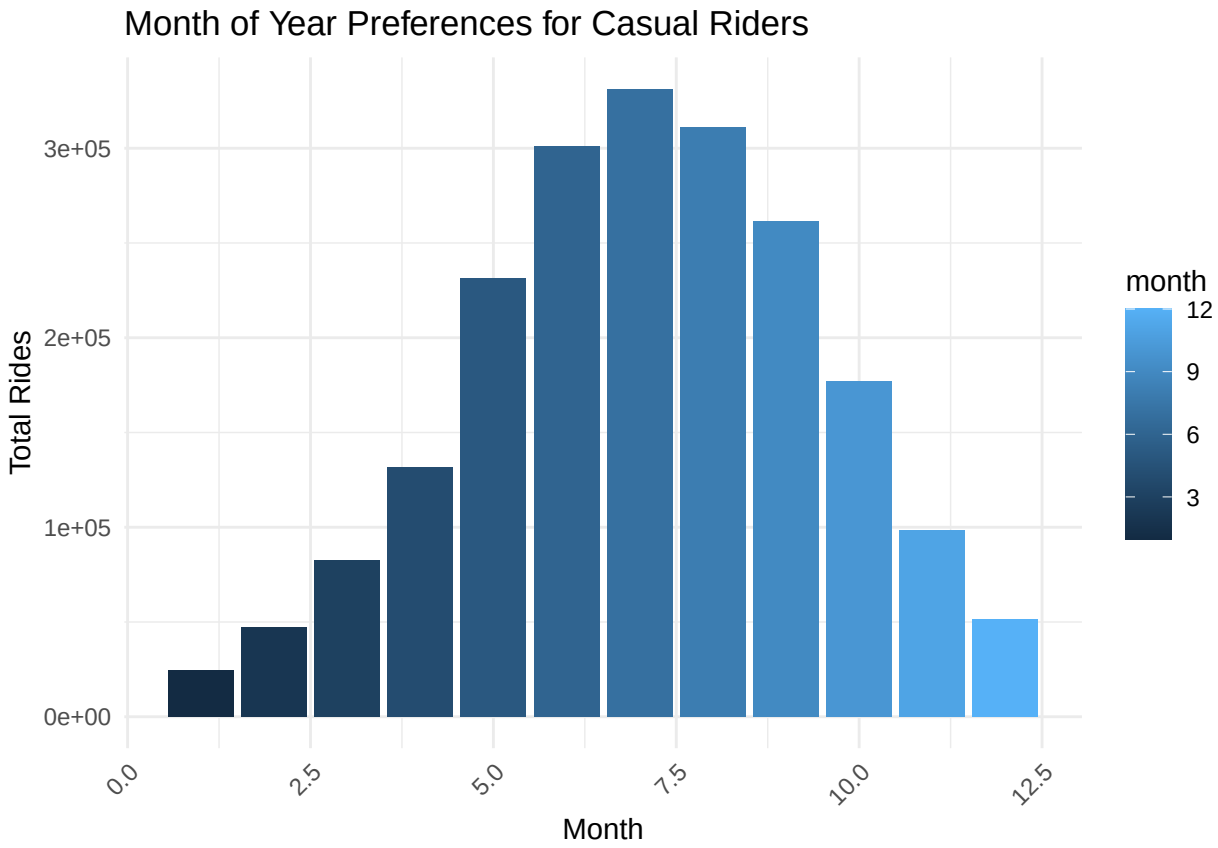
```
## Top Months for Member Riders:
```

```
print(top_months_member)
```

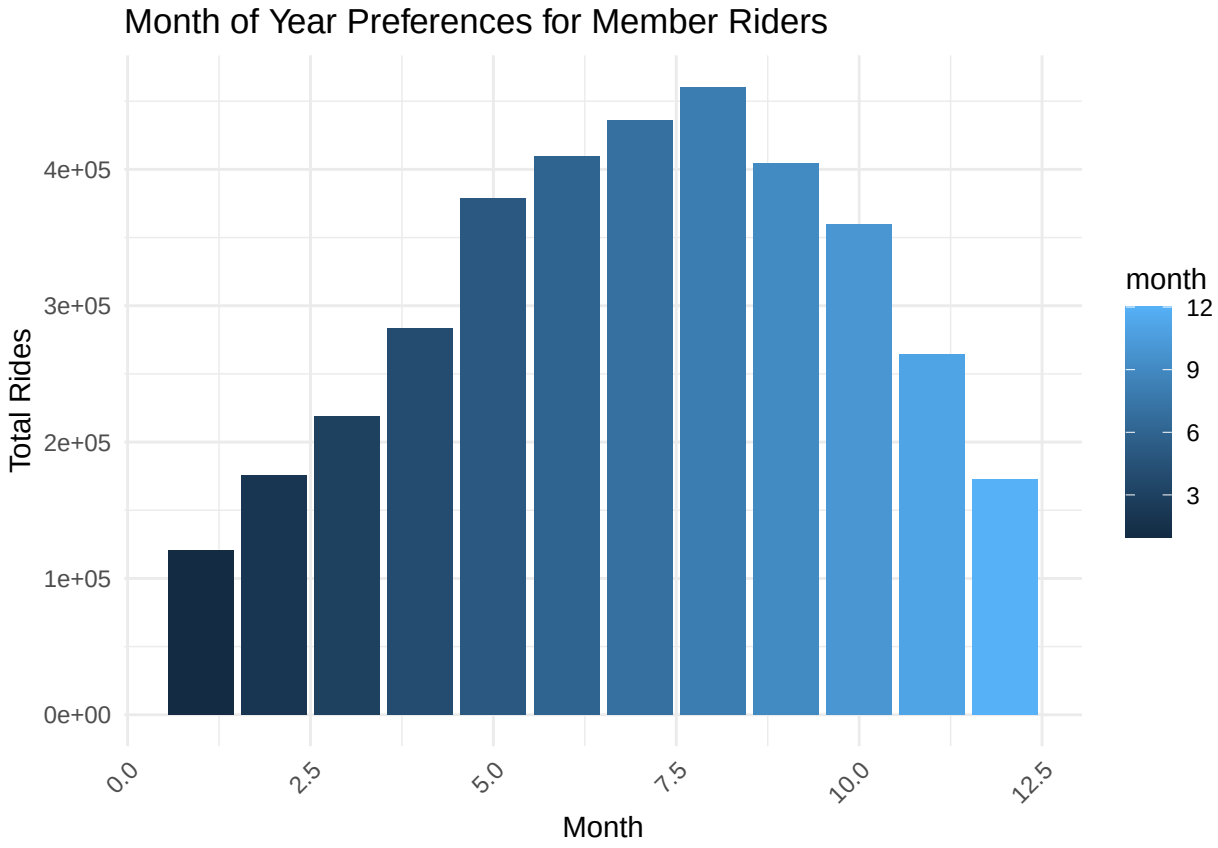
```
##    month member_casual total_rides
##    <int>        <char>       <int>
## 1:     8        member      460563
## 2:     7        member      436292
## 3:     6        member      409515
## 4:     9        member      404736
## 5:     5        member      378554
```

Visualize Seasonal Trends

```
# Visualize data for casual riders
ggplot(casual_seasonal, aes(x = month, y = total_rides, fill = month)) +
  geom_bar(stat = "identity") +
  labs(title = "Month of Year Preferences for Casual Riders", x = "Month", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Visualize data for member riders
ggplot(member_seasonal, aes(x = month, y = total_rides, fill = month)) +
  geom_bar(stat = "identity") +
  labs(title = "Month of Year Preferences for Member Riders", x = "Month", y = "Total Rides") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Conclusion

The analysis reveals significant differences in ride duration, bike type preference, start/end points, routes, time of day, and seasonal trends between casual and member riders. These insights provide actionable recommendations for marketing strategies aimed at converting casual riders to members.