Report: How do annual members and casual riders use Cyclistic bikes dierently?

Iliya GLAZUNOV

2024-12-05

Introduction

About the Company

Cyclistic is a bike-share program in Chicago, offering over 5,800 bikes and 600 stations. In addition to standard bikes, the company offers reclining bikes, tricycles, and cargo bikes, making bike-sharing more inclusive.

Participants and Teams

- **Lily Moreno**: Director of Marketing, developing campaigns to promote the program.
- **Cyclistic Marketing Analytics Team**: Data analysts who collect, analyze, and report data to guide marketing strategy.
- **Cyclistic Executive Team**: Makes decisions on approving marketing programs.

Company Goals

The main goal is to maximize the number of annual subscriptions, as they are more profitable than casual riders. *Cyclistic* aims to understand the differences in bike usage between these two groups to develop effective marketing strategies.

Ask

Problem

The differences in the usage of *Cyclistic* bikes between casual riders and annual members are not fully understood. This hinders the company from developing targeted marketing strategies to increase the number of annual members, which is more profitable for the company.

Task

Analyze trip data to identify significant differences in the behavior of casual riders and annual members. This includes analyzing trip frequency, trip duration, popular routes, time of usage, and other factors.

Goal

My goal as a junior data analyst is to analyze trip data to identify differences in the behavior of casual riders and annual members. Based on this analysis, I will provide recommendations for developing marketing strategies to increase the company's profitability and improve customer satisfaction.

Prepare

Data organization

To perform a comprehensive analysis, I need to acquire the Cyclistic historical trip data from December 2023 to November 2024. This data is provided by Motivate International Inc. under their license.

After downloading the datasets from the specified website, I will organize them into a dedicated folder named "202206–202305_Cyclistic" in .csv format.

Summary of Initial Data Analysis

As a result of the initial data analysis, the following findings were obtained:

- Total Number of Rows: 5,906,269
- Columns and their Description:
  - ride_id: Unique identifier of the trip.
  - rideable_type: Type of bike (e.g., standard bike, tricycle, cargo bike).
  - started_at: Start time of the trip.
  - ended_at: End time of the trip.
  - start_station_name: Name of the start station.
  - start_station_id: ID of the start station.
  - end_station_name: Name of the end station.
  - end_station_id: ID of the end station.
  - start_lat: Latitude of the start station.
  - start_lng: Longitude of the start station.
  - end_lat: Latitude of the end station.
  - end_lng: Longitude of the end station.
  - member_casual: Type of user (annual member or casual rider).

At this stage, I have a good understanding of the data structure and key parameters that will be used for further analysis. This allows me to delve into the details and identify significant differences in user behavior to develop effective marketing strategies.

Process

Choosing the Working Environment

To clean, analyze, and aggregate the large amount of monthly data stored in the folder, I will be using Visual Studio Code (VSCode) — an integrated development environment suitable for the programming language R. I chose VSCode because I have repeatedly used it during the implementation of other projects of mine.

Data Processing

During the data processing stage, I completed the following steps for cleaning and analyzing the data:

1. **Installing and loading packages**: I installed and loaded the necessary packages tidyverse and lubridate, which provide functions for data processing and working with dates.

2. **Setting the working directory**: I set the working directory to the path of the folder where the data files are stored, making it easy to load and save files.

3. **Loading and combining data**: I loaded all the CSV files from the specified folder and combined them into a single dataset for analysis.

4. **Removing missing values**: I removed rows with missing values in important columns such as started_at and ended_at to avoid issues with analysis.

5. **Transforming data**: I transformed the timestamps from string format to date-time format using functions from the lubridate package.

6. **Removing duplicates**: I removed all duplicate rows to ensure each unique trip identifier was represented only once.

7. **Checking for anomalies**: I checked the data for anomalies, such as negative trip durations, and removed such rows.

8. **Creating new columns and sorting data**: I created new columns to simplify the analysis, such as calculating trip duration in minutes. Additionally, I sorted the data and removed invalid trips with durations of less than one minute or more than one day.

9. **Saving the cleaned data**: After all the cleaning and preparation steps, I saved the cleaned dataset to a new CSV file.

Summary

Thanks to the steps described above, I have obtained a clean and ready-to-analyze dataset, which will allow for detailed analysis and informed conclusions in the subsequent sections.

After cleaning the data, the dataset now contains:

- Rows: 4,202,450
- Columns: 14 (two columns were renamed and one new column was added)

Here are the details of the changes:

- started_at was renamed to start_time: The timestamp indicating when the trip started.

- ended_at was renamed to end_time: The timestamp indicating when the trip ended.

- (New) trip_duration: The duration of the trip in minutes, calculated as the difference between end_time and start_time.

The entire code used for these steps is available in the file `data_process.r`.

Analyze & Share

Introduction

In this part, I will analyze the use of bicycles by different user groups (annual members and casual riders) and draw conclusions based on the obtained data. The study covers various aspects of bicycle usage, including the number of trips per month, days of the week, time of day, trip duration, return to the starting station, and popular stations. My findings will help understand the behavioral characteristics of each group and suggest improvements for the bike-sharing system.

Data Analysis and Conclusions

Comparison of the Number of Trips per Month

First, I examined the months during which both groups had the most rides. The data clearly shows that the highest popularity for both groups is during the months from May to September, with a significant increase in the number of people wanting to ride bikes during this period. Conversely, in the colder months (November-March), there are significantly fewer people willing to use bicycles.

These trends are particularly evident among the group that does not have an annual subscription. In this group, there are much fewer people willing to ride bicycles in the cold months compared to those with an annual subscription. However, when the warmer months arrive, the number of non-subscribers wanting to ride increases dramatically.

Comparison of Bicycle Usage by Days of the Week

Next, I investigated the days of the week when people from different groups ride the most. Here, I discovered the following pattern:

- People with a subscription most often use bicycles on weekdays.
- People without a subscription most often use bicycles on weekends.

This indicates that people with a subscription use bicycles to commute to work, while those from the other group use them for leisure.

Comparison of Bicycle Usage at Different Times of the Day

The next step was to find out which hours of the day the bicycles are in particular demand among people from different groups. The data shows:

- People without a subscription most often use bicycles in the afternoon.

- People with a subscription most often use them in the morning and evening hours.

Based on the previous point, it is evident that the peak usage for the subscription group is related to commuting to work/home. As for the non-subscription group, it is harder to pinpoint the exact reason for their preference for this mode of transport at that time. It may be due to avoiding traffic jams or crowded public transport.

Average and Median Trip Duration

After examining the different time periods, I decided to find out the average trip duration and the median. The data shows that people without a subscription spend more time on a trip on average compared to those with a subscription. This may be because non-subscribers use bicycles for leisure and enjoy the moment, while subscribers primarily use bicycles for commuting, and thus spend less time on trips as they want to reach their destination quickly.

Analysis of Returning the Bicycle to the Starting Station

I also analyzed how often the starting point becomes the endpoint within a single trip for both groups. The data shows:

- People without a subscription return to the starting point in about 8% of cases.
- People with a subscription return to the starting point in about 2% of cases.

This can be explained by the fact that non-subscribers take bicycles for leisure and often return to where they started (usually near their home), while subscribers use bicycles for commuting to work, so the starting and ending points do not coincide.

Top 5 Starting and Ending Stations

Finally, I examined the top 5 popular stations among different groups. The graphs show that both casual riders and members have their own top 5 stations. However, it is interesting to note that the starting and ending stations for each group are almost identical.

This indicates that although specific stations may differ between the groups, such consistency suggests that certain stations serve as popular and convenient hubs for bicycle usage within the network.

Conclusion

Based on the analysis conducted, several conclusions can be drawn about the behavior of bike-sharing system users:

1. Annual members use bicycles mainly for daily commuting to work, as evidenced by their activity on weekdays and during peak hours.
2. Casual riders are more likely to rent bicycles for leisure on weekends and during warm months.
3. The average trip duration for casual riders is longer, indicating a more relaxed approach to using bicycles.
4. The return of bicycles to the starting station is characteristic of casual riders, who use bicycles for leisure.
5. Certain stations are popular with both user groups, making them key hubs in the bike-sharing network.

These conclusions can help optimize the bike-sharing system's operations, improve service, and plan new routes and stations.

Act

Based on the data analysis of bicycle usage, I have made several key observations regarding the behavioral patterns of different user groups (annual members and casual riders). These insights help understand how to optimize the bike-sharing system and attract more casual riders to annual memberships.

Recommendations

Based on the identified user behavior patterns and preferences, I propose the following strategic recommendations:

1. Seasonal Campaigns:
   - Launch marketing campaigns in spring and early summer offering discounts on annual memberships to attract casual riders.
   - Provide seasonal discounts and promotions during colder months to maintain user interest and activity.
2. Targeted Weekend Promotions:
   - Develop special weekend offers for casual riders, such as "Ride on weekends and get a discount on annual membership."
   - Include bonuses for annual members that can be used on weekends to motivate sign-ups.
3. Flexible Commuter Plans:
   - Introduce flexible plans for annual members that include morning and evening commute benefits to encourage daily use.
4. Enhanced User Experience:
   - Promote annual memberships as a cost-effective way to make regular short trips, such as commuting to work.
   - Offer casual riders bonuses and discounts for signing up for an annual membership, highlighting convenience and cost savings.
5. Return Incentives:
   - Implement a reward program for casual riders who return bikes to the starting station, encouraging them to consider annual memberships.
6. Station-Based Marketing:
   - Place advertising materials and special offers at popular stations to attract casual riders' attention to annual memberships.
   - Create additional amenities and services at these stations for annual members to enhance their biking experience.

Conclusion

These strategic recommendations aim to optimize the bike-sharing system's operations, improve user experience, and successfully convert casual riders to annual members through targeted marketing efforts. Implementing these initiatives will help increase overall activity and loyalty among users, ultimately enhancing the efficiency and popularity of the bike-sharing system.

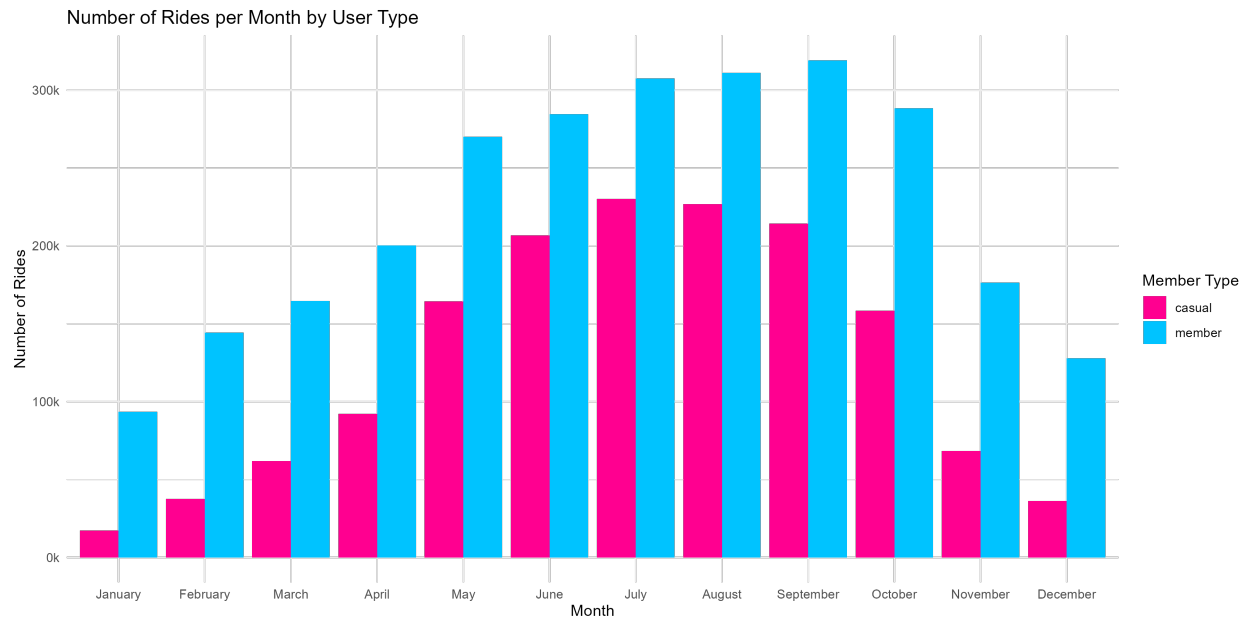# Appendix

Number of Rides per Month by User Type

Figure 1: Graph of the number of trips per month for different user groups
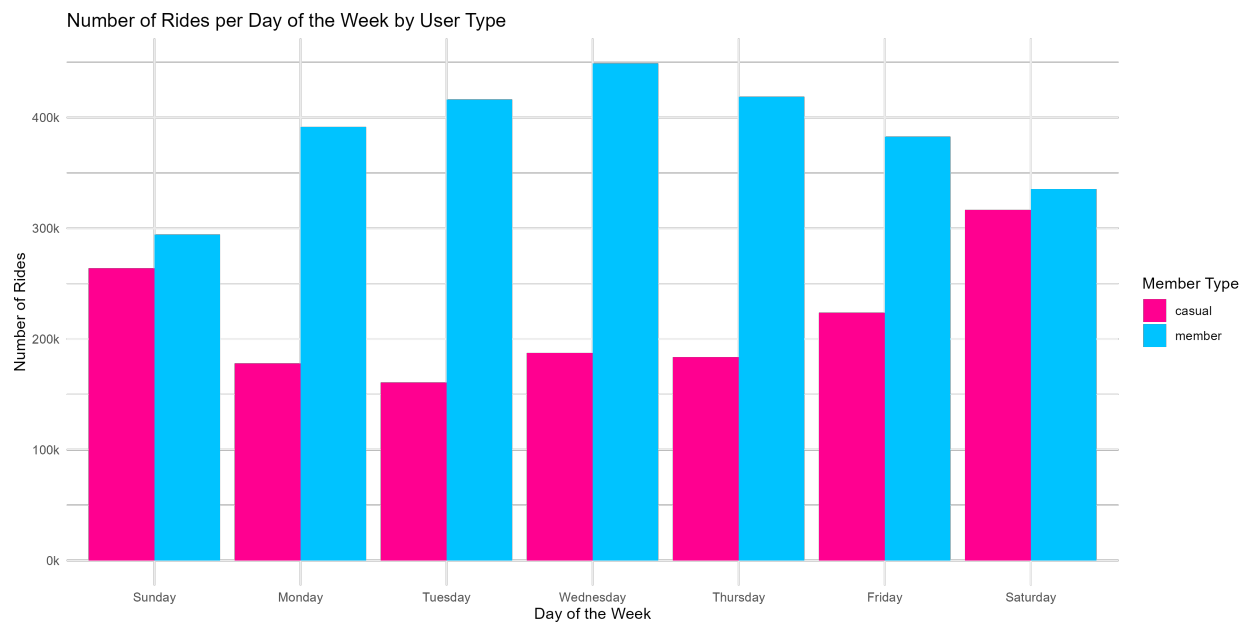
Figure 2: Graph of bicycle usage by days of the week for different user groups
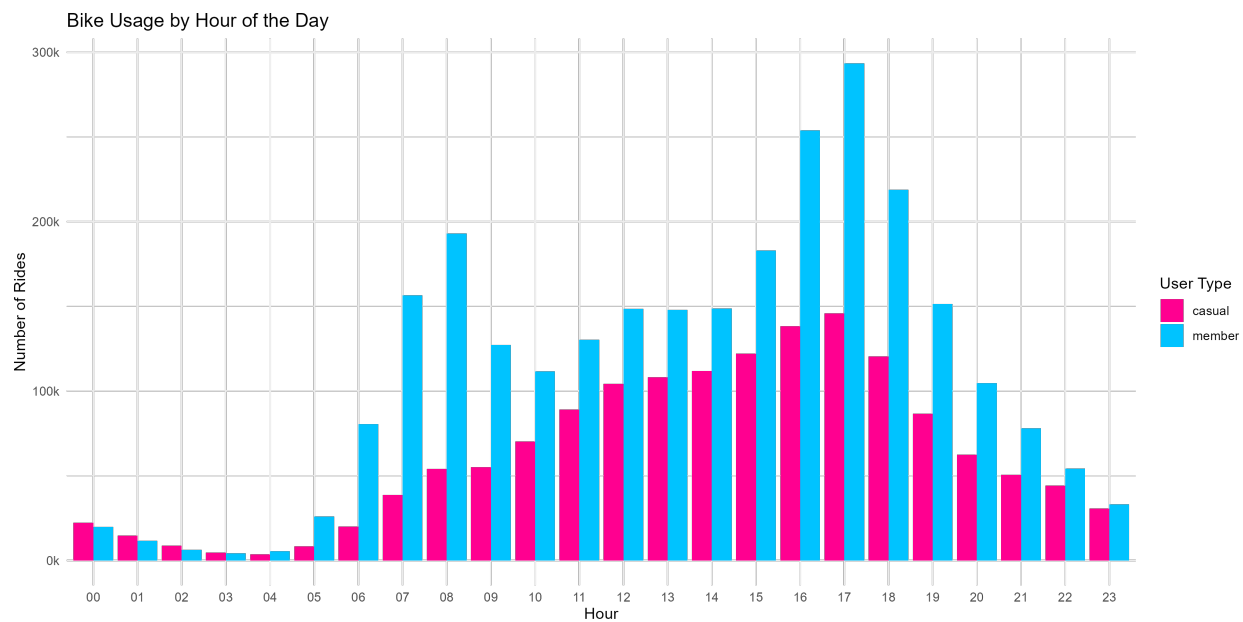

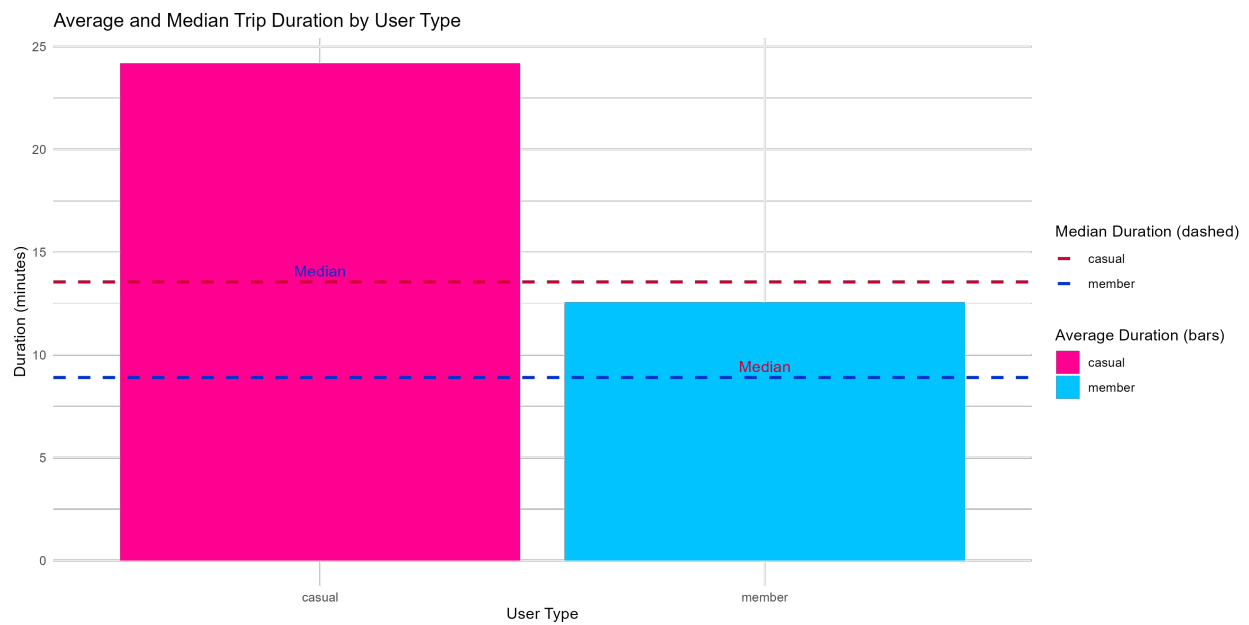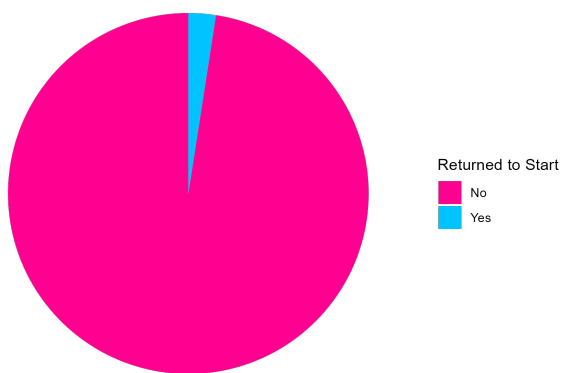Figure 3: Graph of bicycle usage at different times of the day for different user groups

Figure 4: Graph of the average and median trip duration for different user groups
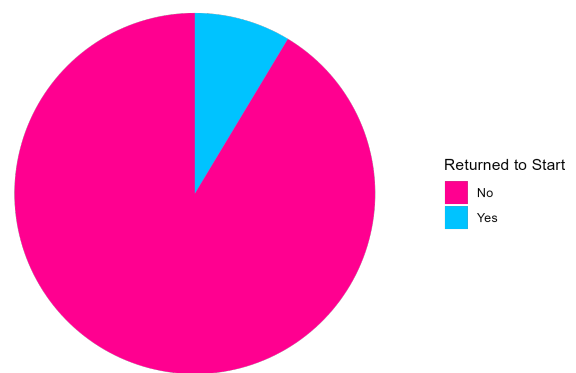
Figure 5: Graph of the analysis of returning the bicycle to the starting station for different user groups
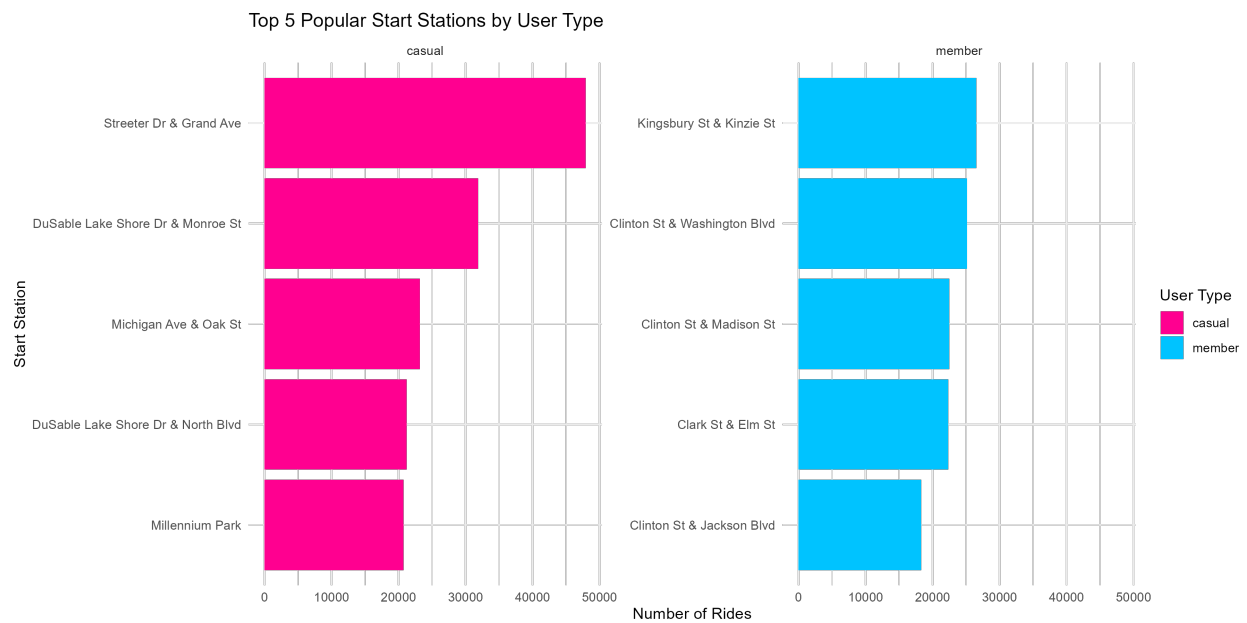
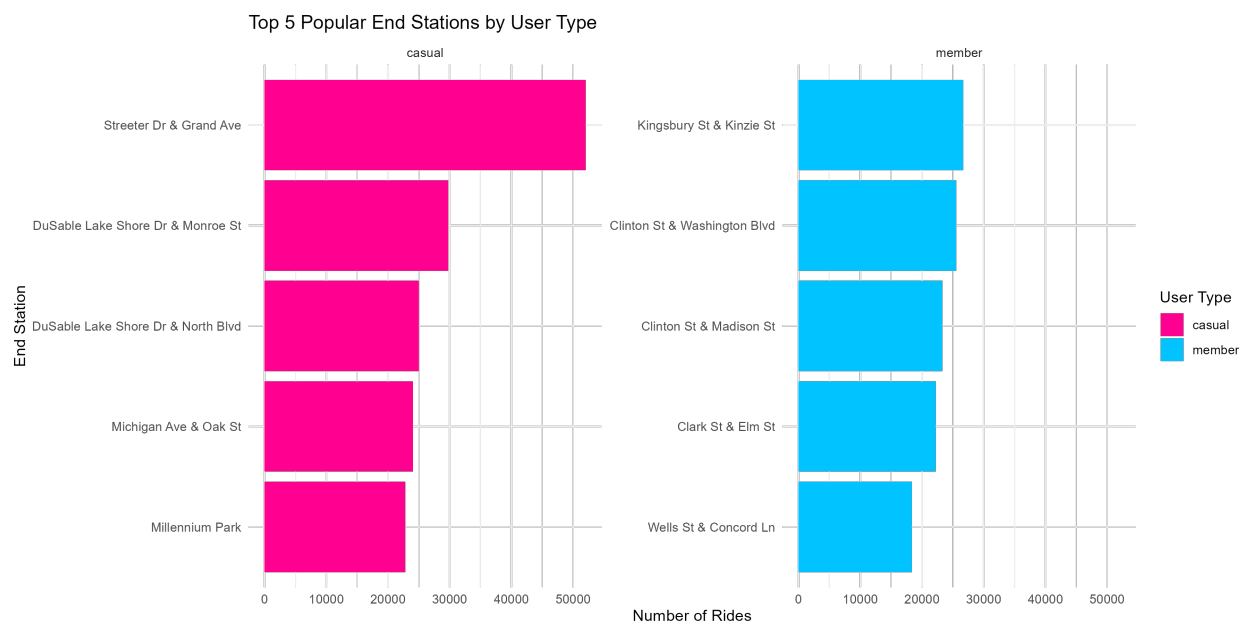Figure 6: Graph of the top 5 starting stations for different user groups

Figure 7: Graph of the top 5 starting ending for different user groups