Analysis of Cyclistic Ridership Historical Data: Casual Riders vs. Members

A Capstone Project in Google Data Analytics Professional Certificate

Analyst: Shahin Khazai

Date: October 2025

Table of Contents

EXECUTIVE SUMMARY

Historical data of Cyclistic ridership from September 2024 through August 2025 was analyzed to discover how casual riders and members use the bike-share services differently to develop a marketing campaign with the goal of converting casual riders to members.

Summary of key findings:

- Members take more rides per month and year, but the average ride duration and maximum ride duration of casual members are longer.
- Both casual riders and members prefer classic bikes over electric bikes, especially in warmer months of the year.
- Casual riders tend to ride around the touristic areas and lakefront, while members routes are concentrated in the business district and at transit hubs.

Based on the results of the analysis, the top three recommendations for converting casual riders to members are:

1. The marketing campaign should be ready to be launched in February/March as the number of casual riders starts to increase, especially on the weekends.
2. Focus on raising awareness about the variety of rides (classic bikes, electric bikes, etc.) as well as the number and location of available stations.
3. Offer promotions, discounts, or payment flexibility as part of onboarding to encourage purchasing annual memberships by casual riders.

INTRODUCTION

Project Background

To increase profitability and growth, Cyclistic, a bike-share company in Chicago, has decided to launch a marketing campaign with the goal of converting casual riders to members. Casual riders are referred to the riders who purchase single-day or full-day passes, while members are defined as those who pay annual membership fees. In this report, historical ridership data from September 2024 through August 2025 (12 months) was analyzed in order to understand how casual riders and members differ in their use of bike-share services. Using Microsoft Excel (Excel) and Google BigQuery (BigQuery), a total of 5.6 million records were processed, transformed, and analyzed to compare the behaviour of casual riders to members. The results of the analysis were then used to generate recommendations for marketing strategies and presented to the executive team for approval.

Report Structure

This report is organized as per the six-phased data analysis cycle: Ask, Prepare, Process, Analyze, Share, Act.

1. **Ask:** defining the business problem/goal
2. **Prepare:** data collection and review
3. **Process:** data cleaning and feature engineering
4. **Analysis:** exploration of data to identify trends, relationships, and anomalies
5. **Share:** communication of the findings to stakeholders
6. **Act:** recommendation of data-driven strategies to be implemented to achieve the business goal(s)

To access the GitHub repository for this project, including appendices and other pertinent documents please visit: https://github.com/insightstories/cyclistic-capstone-project/

ASK

Purpose

The purpose of this data analysis project is to analyze historical data from Cyclistic ridership to find out how casual riders and members use Cyclistic bikes differently. The results of the analysis are used to generate marketing recommendations, which are presented to the executive team for approval. Upon approval, the marketing team will develop a campaign accordingly for converting casual riders to members to increase profitability and growth.

Work Statement

Details of the project, including the scope, deliverables, stakeholders, and a tentative schedule are outlined below in **Table 1**.

Table 1 Project Details: Scope, Stakeholders, Deliverables, and Suggested Schedule

Scope/Major Project Activities	
Activity	**Description/Details**
Data collection	Ridership data from September 2024 through August 2025 is collected.
Data processing/cleaning	Data undergoes EDA, processing, and cleaning.
Data analysis	Data is analyzed to discover how casual riders and members use Cyclistic bikes differently.
Make recommendations	Marketing recommendations are made and presented to the executive team.
Deliver final report	A comprehensive final report, including auxiliary documents will be uploaded as a GitHub repository.
This project does **not** include: • Recommendations for new customer acquisition • Identifying reasons why casual riders buy annual memberships • Design of the marketing campaign • Analysis of other Cyclistic operations/departments	
Stakeholders	
• Cyclistic Marketing Analytics Team • Lily Moreno, Director of Marketing • Cyclistic Executive Team	
Deliverables	
Items	**Description/Details**
Analysis report (current document)	A report outlining the entire process of analysis, including methods, results, and visualizations
Appendices	Auxiliary tables and other documentation that are kept out of the main report for coherence and continuity
Presentation slides	Summary of the project, findings, and recommendations along with the relevant visualizations (plot) to be presented to the executive team
GitHub repository	URL

PREPARE

Summary

Raw data was downloaded from the internet and working copies were made for downstream processing and analysis. Data quality was also assessed, and a raw data dictionary was created.

Data Acquisition

The original (raw) ridership data pertaining to September 2024 through August 2025 was accessed online (https://divvy-tripdata.s3.amazonaws.com/index.html). The data was made available by Motivate International Inc. under the license found here: https://divvybikes.com/data-license-agreement. Each month's data was downloaded as a .zip file, extracted as .csv format, and converted into .xlsx. For data processing and subsequent analysis, copies of the .xlsx files were made and renamed as shown for September 2024 as an example in **Figure 1**. All raw and working files were stored in dedicated folders and subfolders in a local password-protected computer drive.

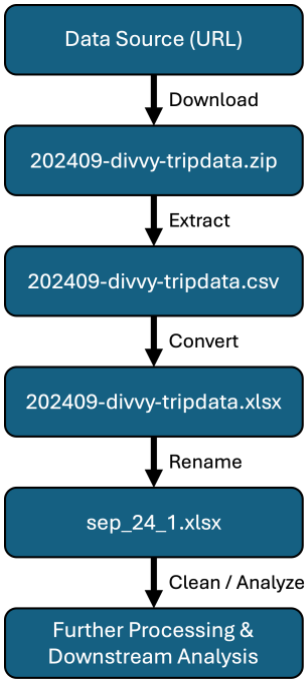


Figure 1 Schematic representation of the project workflow: data acquisition, file name management, and downstream processing and analysis.

Data Assessment

This primary data has the advantage of originality and reliability considering that it is published by the operating company. It is easily accessible, which facilitates verification of the analysis and results by other parties as needed. The data is current and therefore relevant and useful in deriving insights and answering the business question set out in this project. While the data is comprehensive, the riders' privacy is well protected by having removed all personally identifiable information (PII) such as name, address, phone number, email address, and credit card information. Instead of using PIIs, each ride is assigned a unique ride ID, which ensures privacy and does not impede the process of analysis.

Raw Data Inspection and Dictionary

All datasets were organized into the same thirteen columns with the same labels and data types. With no duplications or blanks, the `ride_id` column was chosen as the primary key. Except for instances of blank values in `start_station_name`, `start_station_id`, `end_station_name`, and `end_station_id` columns, there were no other blank values. Below is a data dictionary for the raw data (**Table 2**). After data processing and feature engineering, an updated dictionary was created again, which can be found under the **Process:Processed Data Dictionary** below.

Table 2 Raw Data Dictionary

Column Name	Description	Data Type
`ride_id`	Ride unique ID (PK)[a]	String (hexadecimal identifier)
`rideable_type`	Type of bike	String (categorical)
`started_at`	Date and time of the ride start	Datetime
`ended_at`	Date and time of the ride end	Datetime
`start_station_name`	Name of the ride start station	String
`start_station_id`	ID of the ride start station	String (alphanumerical)
`end_station_name`	Name of the ride end station	String
`end_stations_id`	Name of the ride end ID	String (alphanumerical)
`start_lat`	Latitude of the ride start station	Float
`start_lng`	Longitude of the ride start station	Float
`end_lat`	Latitude of the ride end station	Float
`end_lng`	Longitude of the ride end station	Float
`member_casual`	Type of rider	String (categorical)

[a] Primary key

PROCESS

Summary

Raw data was processed and cleaned using Excel and BigQuery in tandem. Processing involved deleting rows with blank cells as well as deleting some of the original data columns that did not contain valuable information for the analysis. A few new columns were added as engineered features, and a new data dictionary was created.

Please see **Appendix A** under project's GitHub repository for additional information data processing at https://github.com/insightstories/cyclistic-capstone-project.

Data Processing and Cleaning in Excel

The following steps were carried out for all twelve data files from September 2024 through August 2025 in Excel:

1. The dataset was converted into table for easier handling and processing in Excel.
2. The dataset size (columns x rows) was determined (please refer to **Appendix A: Dataset Sizes Before and After Cleaning**).
3. The following columns were deleted as their data could not be used for the analysis: `start_station_id`, `end_station_id`, `start_lat`, `start_lng`, `end_lat`, and `end_lng`
4. `member_casual` column was renamed to `rider`.
5. `start_station_name` was renamed to `start_station`.
6. `end_station_name` was renamed to `end_station`.
7. Columns were filtered for blank cells, and the rows with blank cells were removed.
8. `started_at` and `ended_at` columns were custom-formatted as `yyyy-mm-dd h:mm:ss`.
9. For all columns except `started_at` and `ended_at`, `=TRIM()` function was used to eliminate any leading, lagging, or inter-word extra spaces within the cells. The trimmed columns were copied and pasted as values. The original and trimmed columns were then deleted.

10. Three new columns were added as engineered features for downstream analysis. To see the formulas that were used to create these columns, please refer to **Appendix A: Feature Engineering**. The new columns were:
    10.1. `ride_day`, which indicates the day of the week on which the ride took place,
    10.2. `ride_length_sec`, showing the duration of each ride in seconds, and
    10.3. `ride_length_min`, showing the duration of each ride in minutes.
11. `started_at` and `ended_at` columns were deleted.
12. `ride_id` column was formatted as *Text*, `ride_length_sec` and `ride_length_min` columns were formatted as *Number* with two decimal places, and the rest of the columns were formatted as *General*.
13. The cleaned .xlsx files were saved as .csv and uploaded into BigQuery for further cleaning, e.g., `sep_24_1.xlsx` → `sep_24_1.csv`.

Data Processing and Cleaning with SQL in BigQuery

To further clean the data for analysis, rides lasting 180 seconds or less were deleted. Additionally, rides lasting 300 seconds or less with the same start and end station were also removed. This approach assumed these short trips were errors, maintenance records, or otherwise too brief to be considered meaningful.

To do this in BigQuery using SQL, the dataset `cyclistic` was created under `sk-repo` project and the twelve files (`sep_24_1.csv` through `aug_25_1.csv`) were added to `cyclistic` as tables.

Below is an example query used to create the new table for September 2024 with the criteria mentioned above. For the full set of executable queries, please refer to **Appendix A: SQL Queries**.

```
CREATE TABLE `sk-repo.cyclistic.sep_24_2` AS
SELECT
    *
FROM
    `sk-repo.cyclistic.sep_24_1`
WHERE
    NOT (
        (ride_length_sec <= 300 AND start_station = end_station)
        OR ride_length_sec <= 180
    );
```

After cleaning the data in BigQuery, the clean tables were downloaded as .csv files and converted to .xlsx for the analysis, e.g., `sep_24_2.csv` → `sep_24_2.xlsx`.

Processed Data Dictionary

After processing and feature engineering, a new data dictionary was created for the data tables that were used for the analysis (**Table 3**).

Table 3 Data Dictionary for Processed Data

Column Name	Description	Data Type
`ride_id`	Ride unique ID (PK)[a]	String (hexadecimal identifier)
`rideable_type`	Type of bike	String (categorical)
`start_station`	Name of the ride start station	String
`end_station`	Name of the ride end station	String
`rider`	Type of rider	String (categorical)
`ride_day`	Day of the week on which the ride took place	String
`ride_length_sec`	The length of ride in seconds	Float
`ride_length_min`	The length of ride in minutes	Float

[a] Primary key

ANALYSIS

Summary

Cleaned data in Excel was used to create pivot tables for analysis and visualization. Monthly number and duration of rides, rideable choice (classic vs. electric bike), and top starting stations for casual riders and members were considered for the analysis as outlined below.

Please see **Appendix B** under project's GitHub repository for additional information, including data tables at https://github.com/insightstories/cyclistic-capstone-project.

Number and Duration of Monthly Rides for Casual Riders and Members

For each month, the number or rides for casual riders and members as well as the average and maximum duration of the rides were tabulated as seen in **Table 4**.

Table 4 Monthly Number of Rides, Average Ride Duration, and Maximum Ride Duration for Casual Riders and Members from September 2024 Through August 2025

Month & Year	Casual Rider	Member	Casual Rider Average Ride Duration (Minutes)	Member Average Ride Duration (Minutes)	Casual Rider Maximum Ride Duration (Minutes)	Member Maximum Ride Duration (Minutes)
September 2024	205,759	297,030	23.3	13.1	1,498.4	1,484.8
August 2025	210,849	273,130	23.4	12.7	1,495.1	1,495.9
July 2025	198,675	264,990	18.8	11.9	1,478.2	1,457.9
June 2025	184,410	235,584	15.9	11.4	1,447.9	1,418.8
October 2024	152,105	266,004	14.1	10.9	1,410.0	1,394.5
May 2025	119,339	196,601	14.3	10.8	1,433.7	1,277.1
April 2025	72,591	164,555	20.8	12.1	1,468.1	1,448.7
November 2024	64,886	160,419	21.9	12.4	1,492.8	1,478.4
March 2025	57,969	133,643	24.5	13.1	1,499.6	1,497.5
December 2024	26,228	92,414	25.8	13.9	1,493.6	1,483.3
February 2025	18,095	79,972	25.3	14.4	1,498.0	1,495.0
January 2025	15,637	74,856	25.4	14.0	1,496.3	1,487.9

As seen above, although the number of rides taken by members are larger than casual riders', the average duration as well as the maximum duration of rides taken by casual riders are bigger than members'. The data also shows that during the colder months of the year, the number of casual riders decreases more drastically compared to that of members, probably due to the fact that a portion of the members use the bike-share services to commute to work. The number of casual riders starts to go up again in March.

Using the data above, the total number of rides per month were plotted in descending order segmented into casual riders and members share of the total. The highest number of rides were taken in September 2024, August 2025, and July 2025. The lowest number of rides corresponded to December 2024, February 2025, and January 2025 as shown in **Figure 2**.

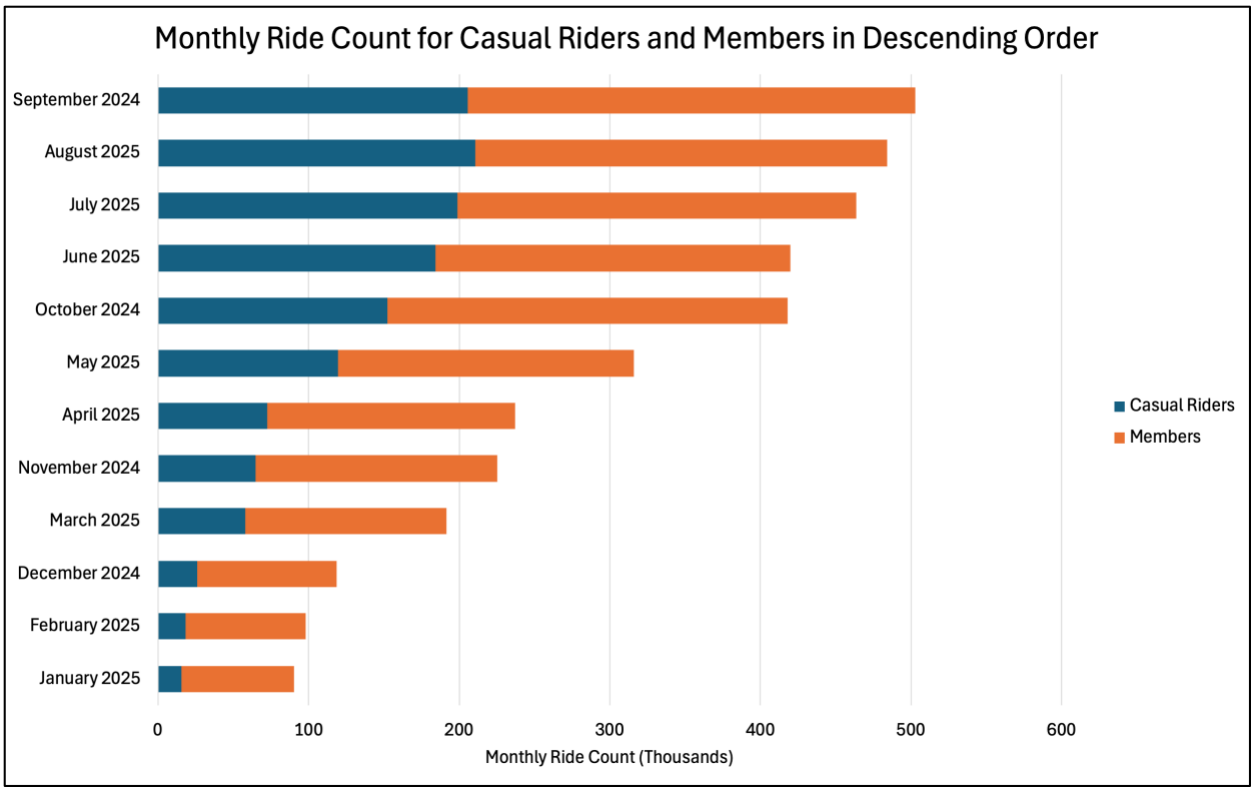


Figure 2 Total number of rides per month and the share of casual riders and member from September 2024 through August 2025 organized in descending order.

The average ride length for casual riders and members were plotted in chronological order from September 2024 through August 2025. As shown in **Figure 3**, the average ride length of casual riders showed a significant drop in the colder months of the year, but members' monthly ride length average did not change as much. Both casual riders and members took longer rides in the warmer months of the year.

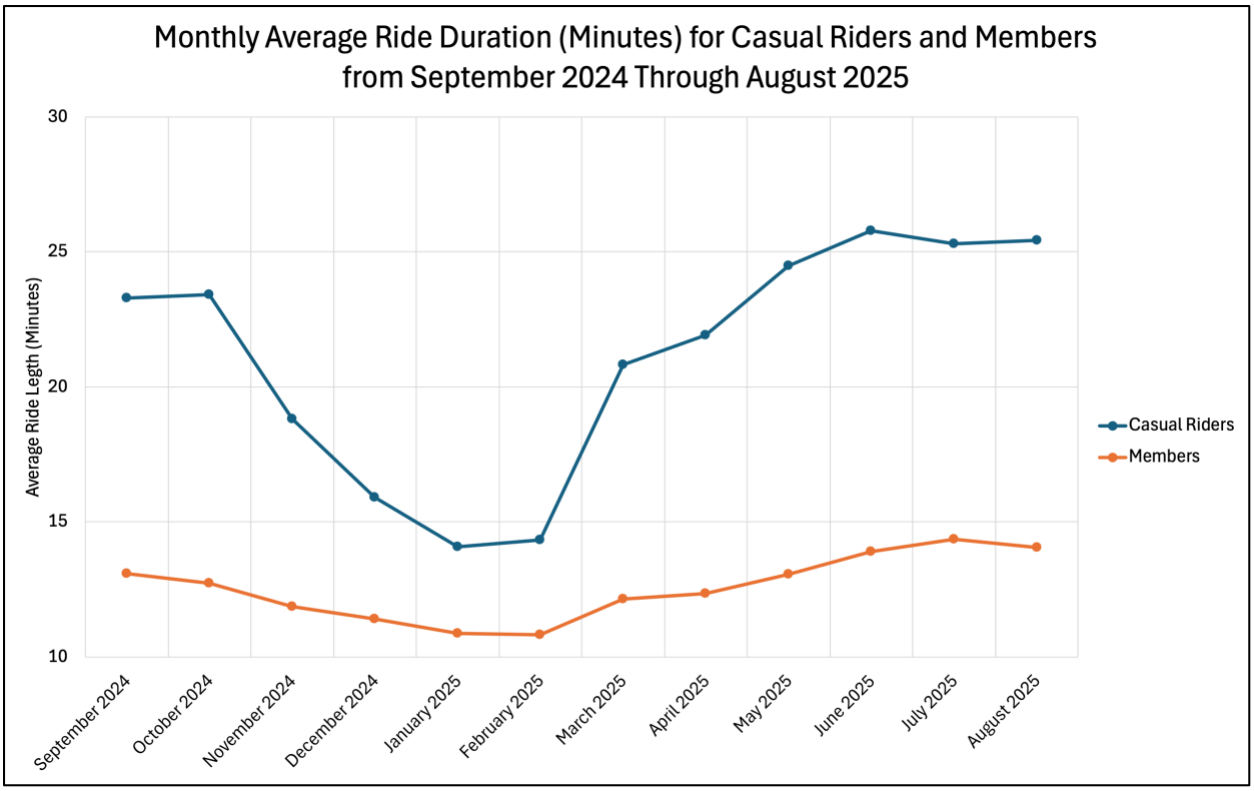


Figure 3 The monthly average ride length (minutes) for casual riders and members from September 2024 through August 2025 in chronological order. Casual riders' average duration of rides drops in colder months while that of members stays more consistent.

Number and Duration of Rides per Days of the Week

To further investigate the differences between casual riders and members, the total number of rides and length of rides per each weekday were plotted for both groups. As **Figure 4** shows, during the workweek and in the colder months of the year, the members' number of rides are bigger than that of the casual riders. However, in warmer weather and on or around the weekends, casual riders take more rides. These instances are indicated with black arrows in **Figure 4**.

Similar to the trend above, during the warmer days and on or around the weekends, the lengths of rides of casual members also tend to be longer as seen in **Figure 5**. Such days tend to be clustered together, which are demarked by dotted round rectangles in the plot.

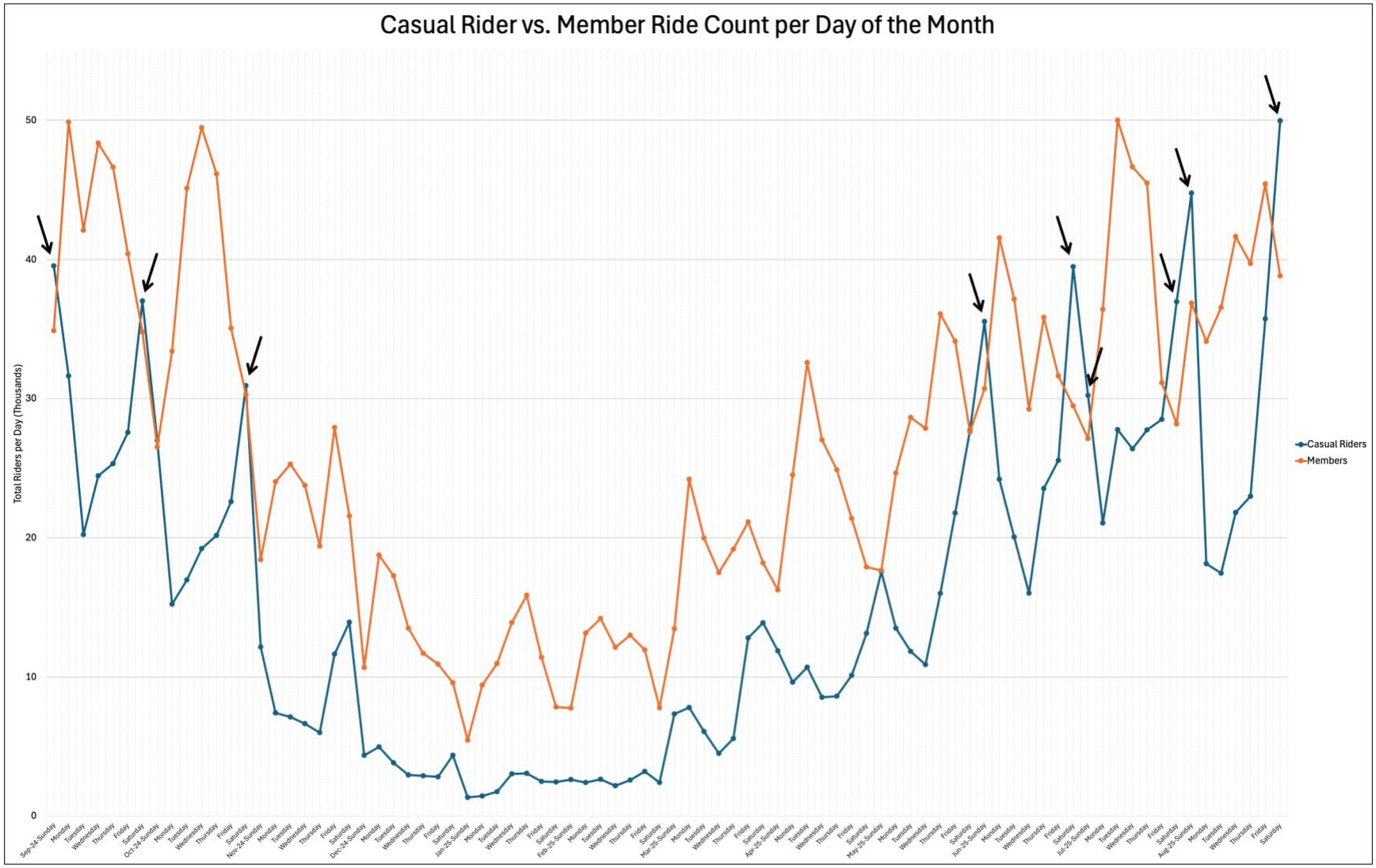


Figure 4 Total rides per days of the week were plotted for each month for both casual riders and members. Although members use the service more during the weekday, during and around the weekends in warmer weather, we see a reversal of the trend, indicated by black arrows in the figure.

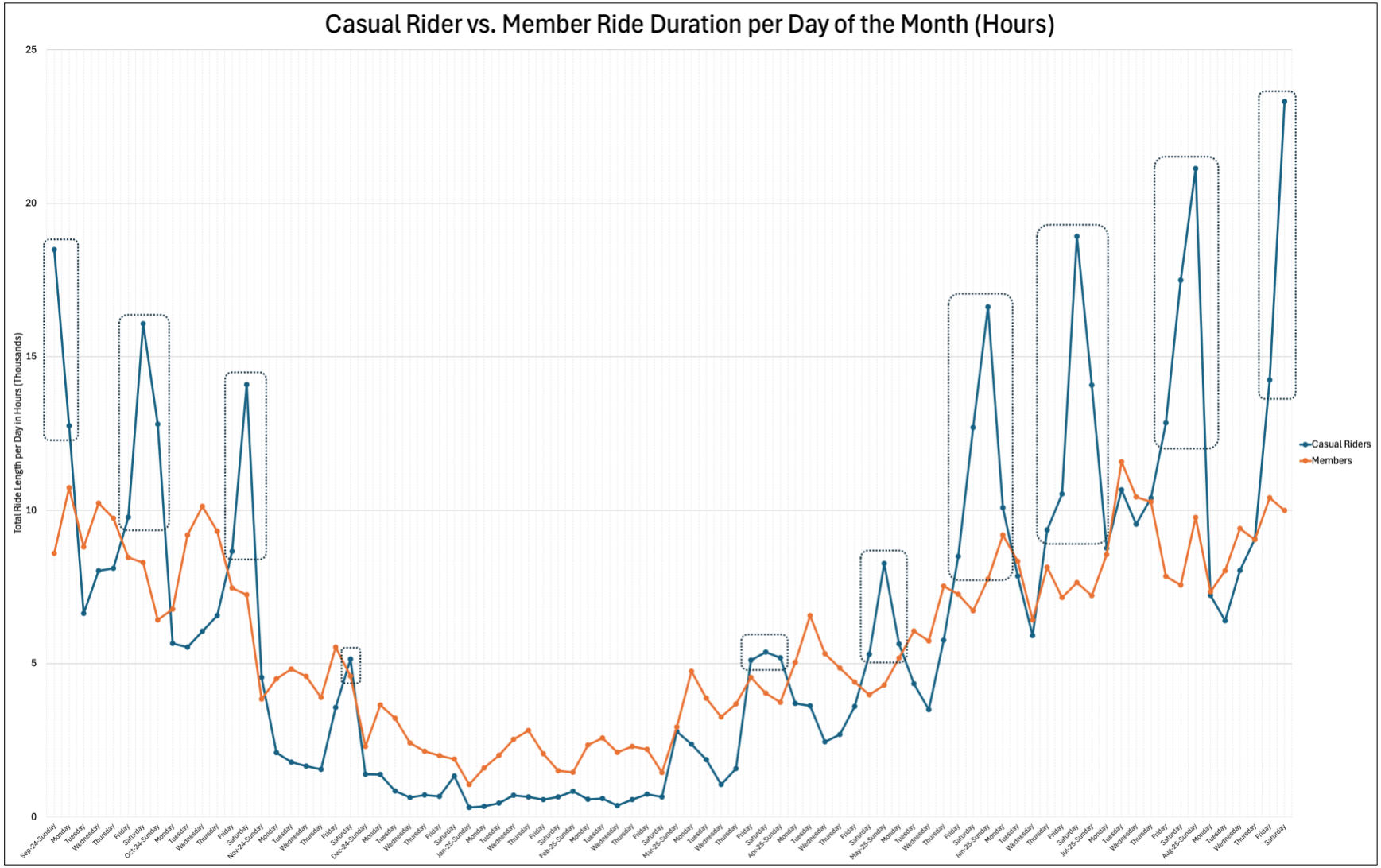


Figure 5 The total length of rides (in hours) taken by casual riders and members were plotted against days of the week of each month. Like the previous results, clusters on or around weekends were identified where casual riders used bikes longer than members, even though the trend for the weekdays or colder times of the year shows members taking longer rides than casual riders. The clusters are demarcated in dotted round rectangles.

Preference for Classic vs. Electric Bike Among Casual Riders and Members

To understand the differences between casual riders and members in terms of their preference for classic vs. electric bikes, the number of monthly rides for each category were organized in **Table 5** and plotted to demonstrate the differences visually in **Figure 6**. It was found that both groups prefer classic bikes over electric ones, although the casual riders showed a slight preference for electric bikes, especially in December, January, and February.

Table 5 Number of Rides Taken with Classic and Electric Bikes by Casual Riders and Members

Month & Year	Casual Riders (Classic Bike)	Casual Riders (Electric Bike)	Members (Classic Bike)	Members (Electric Bike)
September 2024	116,545	65,788	180,548	97,258
October 2024	94,339	57,766	168,265	97,739
November 2024	37,209	27,677	99,034	61,385
December 2024	12,387	13,841	51,939	40,475
January 2025	6,246	9,391	38,600	36,256
February 2025	7,019	11,076	40,696	39,276
March 2025	29,980	27,989	72,425	61,218
April 2025	37,931	34,660	90,149	74,406
May 2025	64,608	54,731	109,100	87,501
June 2025	105,786	78,624	134,211	101,373
July 2025	104,877	93,798	144,981	120,009
August 2025	113,390	97,459	151,693	121,437

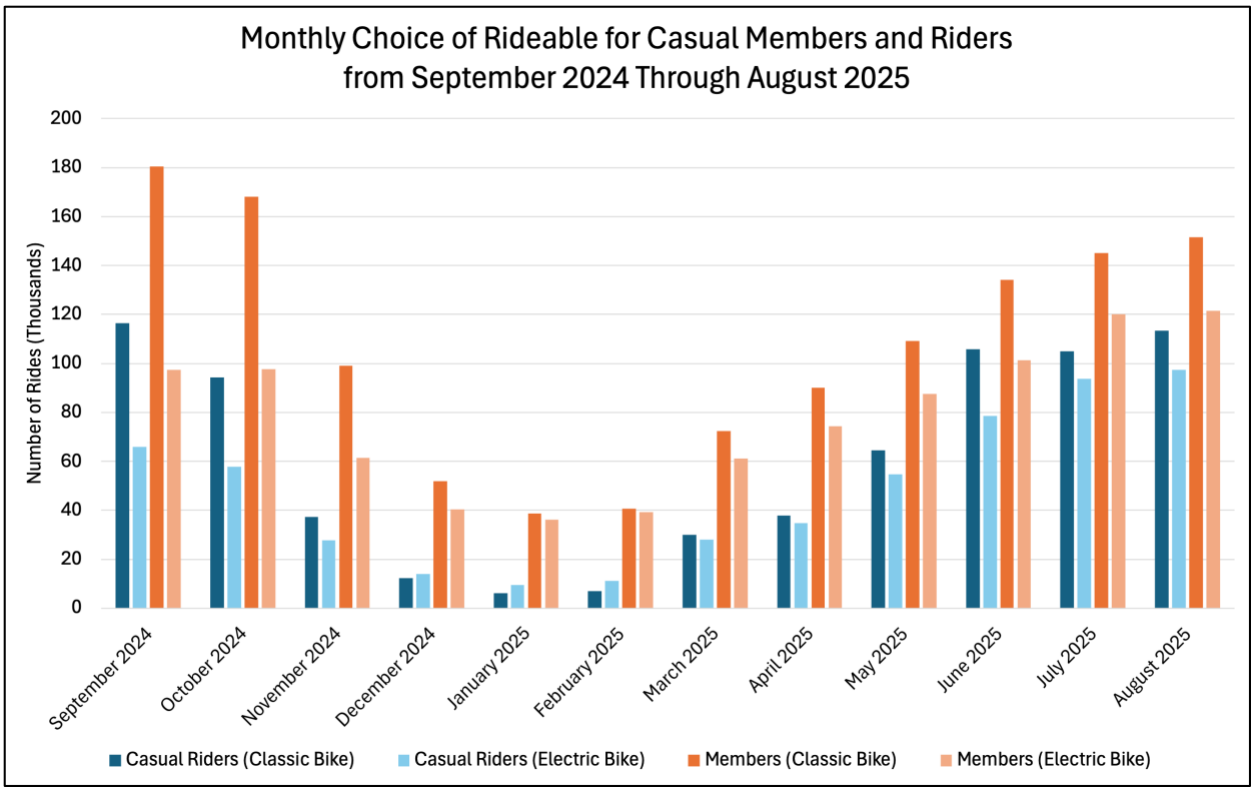


Figure 6 The choice of rideable (classic or electric) for casual riders and members were plotted. As seen here, both groups prefer classic bikes over electric bikes, although the casual riders showed a preference for the electric bike during the colder months of the year.

The Twenty Most Popular Starting Stations for Casual Riders and Members

To further investigate the differences between casual riders and members, pivot tables were created to identify the most popular starting stations for each group. The data was then tabulated and plotted as found in **Table 6** and **Figure 7** and **Figure 8**.

Table 6 Top Twenty Starting Stations for Casual Riders and Members

Station Name	Casual Rider	Station Name	Member
Streeter Dr & Grand Ave	33,281	**Kingsbury St & Kinzie St**	25,372
DuSable Lake Shore Dr & Monroe St	28,485	**Clinton St & Washington Blvd**	21,567
Michigan Ave & Oak St	20,053	**Clark St & Elm St**	19,131
Millennium Park	18,084	**Clinton St & Madison St**	18,872
DuSable Lake Shore Dr & North Blvd	17,500	**Canal St & Madison St**	18,437
Shedd Aquarium	16,723	**Clinton St & Jackson Blvd**	16,345
DuSable Harbor	14,535	**State St & Chicago Ave**	15,179
Theater on the Lake	13,873	**Wells St & Elm St**	15,054
Navy Pier	11,405	**Wells St & Concord Ln**	14,799
Michigan Ave & 8th St	10,799	**Canal St & Adams St**	14,232

Adler Planetarium	9,656	**Larrabee St & Kingsbury St**	13,618
Indiana Ave & Roosevelt Rd	9,125	**Daley Center Plaza**	13,534
Montrose Harbor	9,059	**University Ave & 57th St**	13,351
Field Museum	8,960	**Dearborn St & Erie St**	13,205
Michigan Ave & Washington St	8,854	**Wells St & Huron St**	13,138
Wabash Ave & Grand Ave	8,615	**DuSable Lake Shore Dr & North Blvd**	12,631
Clark St & Armitage Ave	8,436	**Wells St & Hubbard St**	12,517
Kingsbury St & Kinzie St	8,383	**Orleans St & Merchandise Mart Plaza**	11,871
Clark St & Elm St	8,257	**Dearborn Pkwy & Delaware Pl**	11,780
Clark St & Lincoln Ave	8,122	**Desplaines St & Kinzie St**	11,739

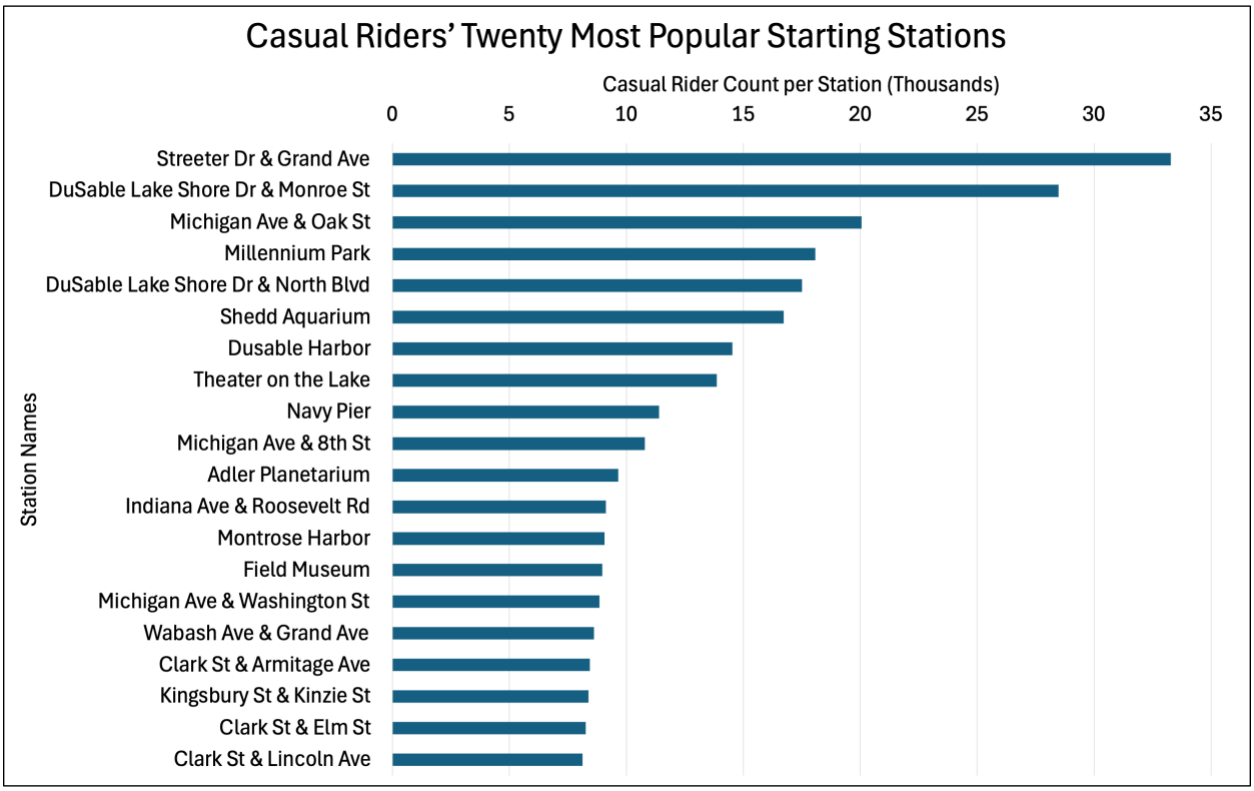


Figure 7 The casual riders' twenty most popular starting stations

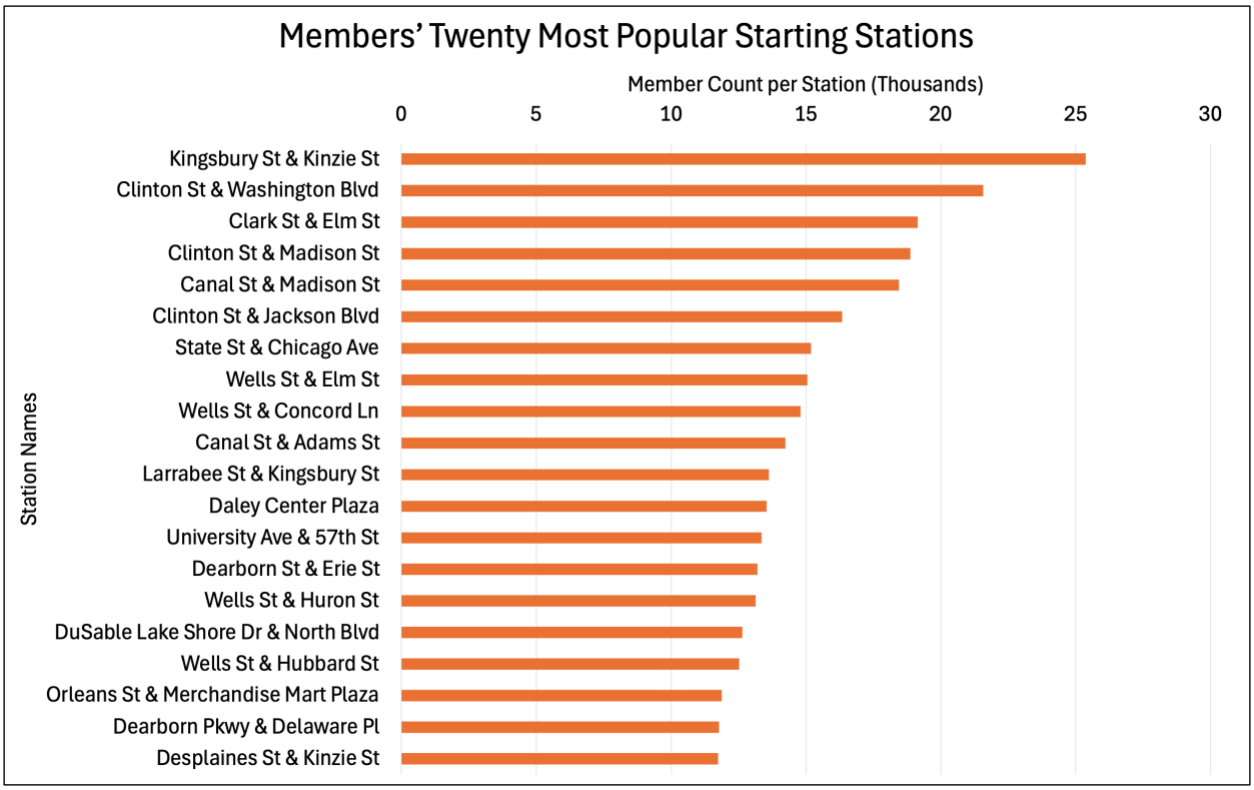
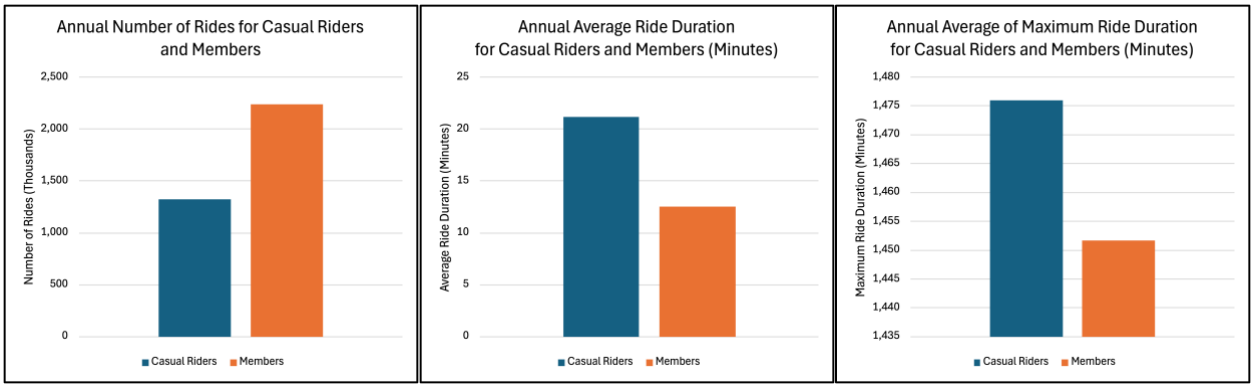


Figure 8 The members' twenty most popular starting stations

SHARE

Comparing the Annual Ride Count, Average, and Maximum Duration Between Casual Riders and Members

Although members take more rides annually, the average ride duration and maximum ride duration of casual members are larger than those of members as summarized in **Figure 9**.



(a) ***(b)*** ***(c)***

Figure 9*(a)* The annual number of rides taken by members is larger than the total number of casual riders but ***(b)*** the average duration and ***(c)*** the average maximum duration of casual riders' rides is bigger than members' rides.

Overall Rideable Preference for Casual Riders and Members

Using the data in **Table 5**, a plot was generated to show the preference of casual riders and members for each type of rideable (classic vs. electric bike) as shown in **Figure 10**. Both groups prefer classic bikes, although the casual riders show a preference for electric bikes during the colder months of the year as we saw previously.

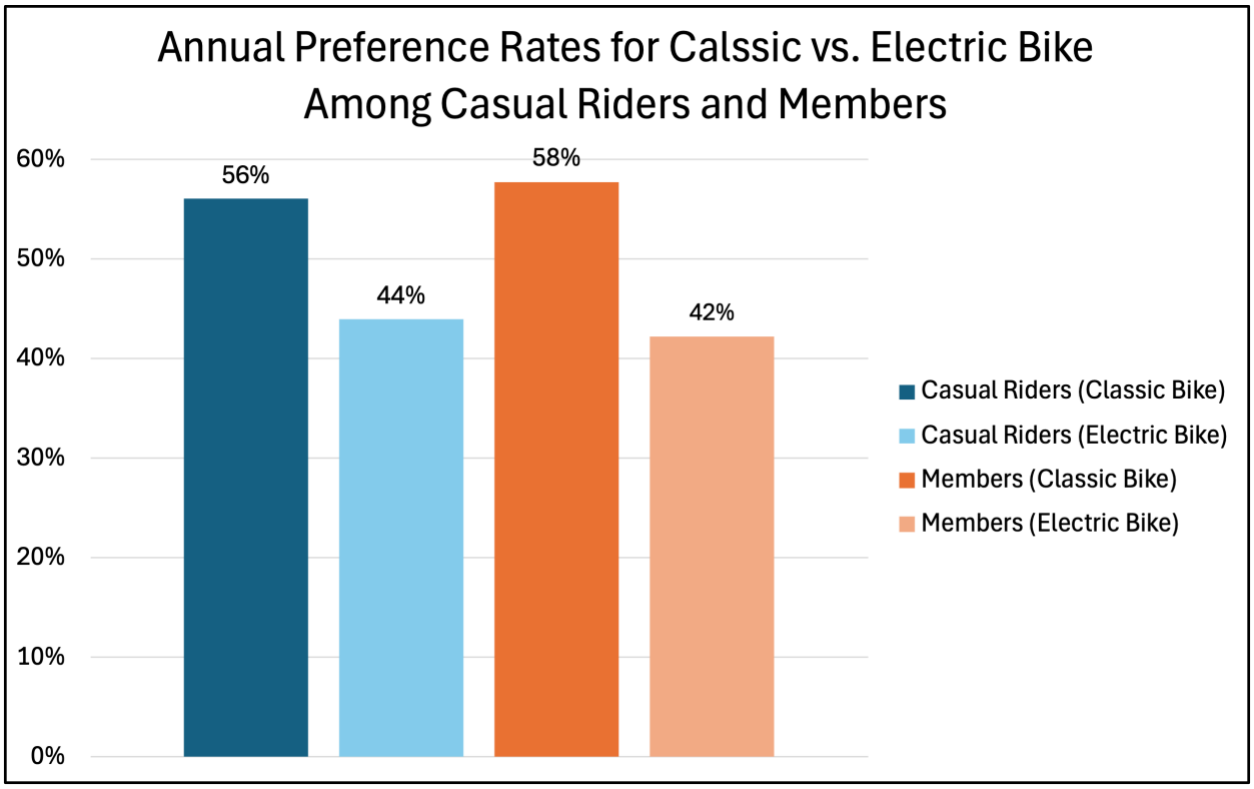


Figure 10 Comparing the preference rates of classic bikes and electric bikes for casual riders (shades of blue) and members (shades of orange).

Location of Starting Stations for Casual Riders and Members

As seen in **Figure 8** and **Figure 9**, most casual riders' starting stations are concentrated around the touristic areas and lakefront, while members tend to start their rides in the business district and transit hubs (most likely for commuting to work). **Figure 11** shows the location of the starting stations for casual riders (blue) and members (orange) relative to the rest of Chicago.

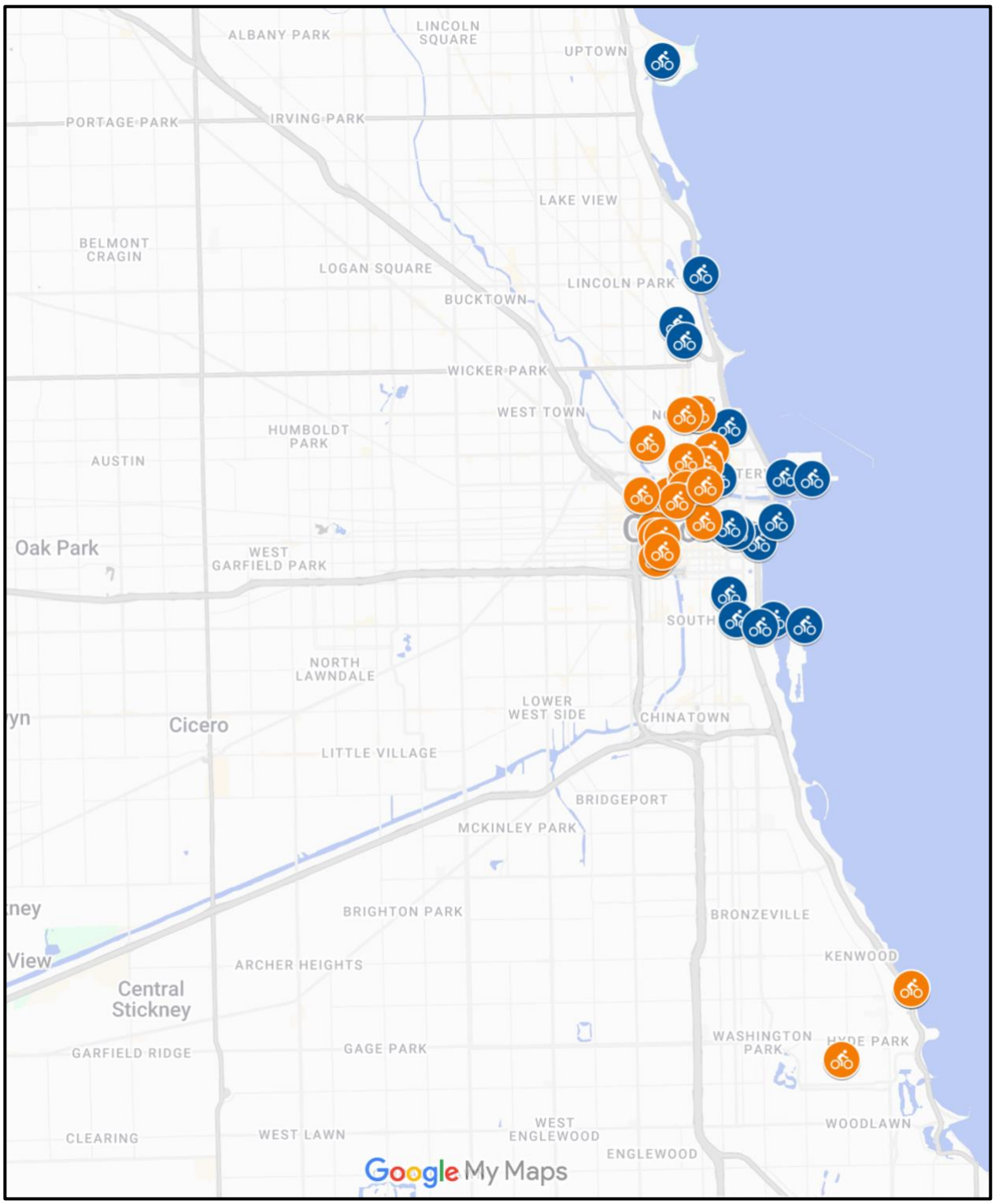


Figure 11 Casual riders' starting points (blue bikes in the map) tend to be concentrated around the lakefront and touristic areas while members (orange bikes in the map) are more likely to start their ride in the business strict and near transit hubs. Image was created with Google My Maps (https://www.google.com/maps/d/).

ACT

Taking into consideration the results of the analysis presented above, which aimed at understanding how casual riders and members use Cyclistic bikes differently, here are top three recommendations for converting casual riders to members to be implemented in the marketing campaign.

1. The marketing campaign should be ready to be launched in February/March as the number of casual riders starts to increase, especially on the weekends.

2. Focus on raising awareness about the variety of rides (classic bikes, electric bikes, etc.) as well as the number and location of available stations.

3. Offer promotions, discounts, or payment flexibility as part of onboarding to encourage purchasing annual memberships by casual riders.

REFERENCES

Divvy. (n.d.). *Divvy trip data*. Retrieved October 1, 2025, from https://divvy-tripdata.s3.amazonaws.com/index.html

Google LLC. (2025). *BigQuery* [Web-based software]. https://cloud.google.com/bigquery

Google LLC. (n.d.). *Map showing the location of starting stations for casual riders and members relative to the rest of Chicago* [Screenshot]. Google Maps. Retrieved October 21, 2025, from https://www.google.com/maps/d/

Microsoft Corporation. (2025). *Microsoft Excel (Version 16.99.2)* [Computer software]. https://www.microsoft.com/