Cyclistic Data Analysis Report

James McBurnie

Table of Contents

1. Introduction

1.1 About the Company

Cyclistic is a bike-share company launched in 2016 that has a fleet of 5,824 geotracked bicycles and a network of 692 stations across Chicago. Bikes can be unlocked from one station and returned to any other station in the system anytime.

While Cyclistic's flexible pricing plans have powered its expansion, the company's financial analysts have concluded that annual members are more profitable than casual riders (customers who purchase single- or full-day passes).

1.2 Business Task

This analysis seeks to gain key insights into the differences in bike usage patterns between annual and casual riders of Cyclistic's bike sharing service in order to assist the design of data-driven marketing strategies aimed at converting casual riders into annual members.

1.3 Data Sources

Data is sourced from the open Divvy dataset (used by Google for this fictional business).
Data Source: https://divvy-tripdata.s3.amazonaws.com/index.html

1.4 ROCCC

- **Reliable:** The data comes directly from Cyclistic itself, and is thus a reliable dataset.
- **Original:** This data is Cyclistic's own proprietary trip data, and thus is original.
- **Comprehensive:** With thirteen columns in each dataset respectively, and a wide range of data and data types, this data is comprehensive.
- **Current:** The data is from the period September 2024 to September 2025, and is thus current while also covering a reasonable time period.
- **Cited:** This data is properly documented as official data from Cyclistic.

1.5 Data Integrity

To verify the data's integrity, checks for missing values, consistent data types, temporal consistency (end time > start time), and checks for alignment between similar variables in the two datasets were performed.

1.6 Privacy, Security and Licensing

This data lacks any personal identifiers, ensuring compliance with privacy standards. Given it was provided by Cyclistic, licensing for internal analysis was implicitly granted. The data has been stored in a local environment and accessed only for analytical purposes.

2. Data

2.1 Data Composition

Row Name	Row Description	Data Type
ride_id	Unique identifier for the trip	String
rideable_type	Type of bike (classic_bike, electric_bike)	String
started_at	Timestamp for when the trip started	Date Time
ended_at	Timestamp for when the trip ended	Date Time
start_station_name	The name for the station from which the bike was hired	String
start_station_id	Unique identifier for the station from which the bike was hired	String
end_station_name	The name for the station at which the bike was returned	String
end_station_id	Unique identifier for the station at which the bike was returned	String
start_lat	Latitude of the starting station	Float
start_lng	Longitude of the starting station	Float
end_lat	Latitude of the ending station	Float
end_lng	Longitude of the ending station	Float
member_casual	Identifies whether the user is a member or casual	String

There are some structural issues due to variables with inconsistent naming and the potential need for datatype conversion. This will be addressed in the data cleaning phase.

2.2 Dataset Overview

Size: 6,360,797 rows × 13 columns

Purpose: Preparing Cyclistic Bike Share data for robust analysis

2.3 Data Cleaning Steps

Below are the steps I took to clean the data. Full data cleaning code can be seen in the R Markdown file.

1. Assessed Null Values
Counted missing entries and calculated null percentages per column.

2. Removed Duplicates
Identified and filtered duplicate rows to ensure data integrity.

3. Standardized Date-Time Fields
Converted started_at and ended_at columns to datetime format for accurate time calculations.

4. Filtered Unusable Dates
Retained only records within the period from 2024-09-01 to 2025-09-30.

5. Validated and Sanitized Categories
Checked and harmonized values in rideable_type and member_casual columns.

6. Calculated Trip Durations
Created a duration column; removed trips with impossible or extreme durations (≤60 seconds or >24 hours).

7. Dropped Incomplete Data
Removed rows and columns with excessive missing values to improve data quality.

8. Enhanced Features
Added a day_of_the_week column to support trend analysis.

9. Outlier Detection
Flagged and reviewed anomalous trip durations for further investigation.

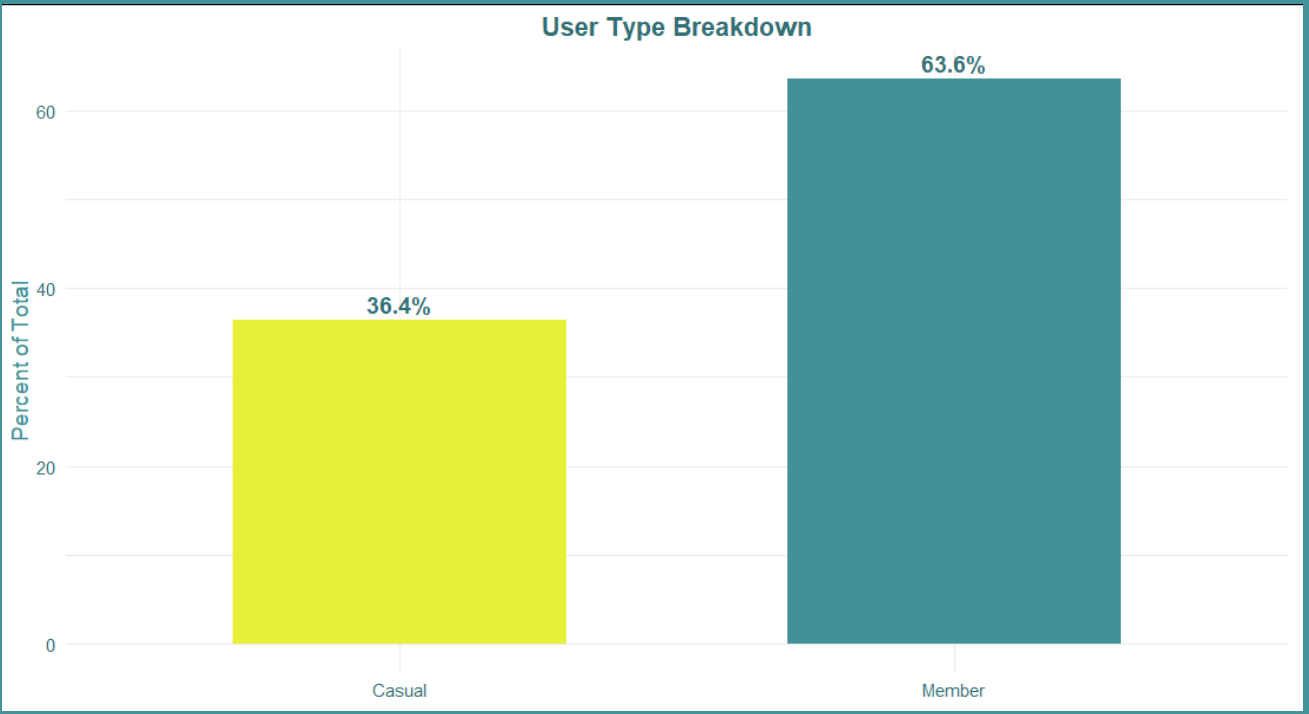
10. Validated Geographic Accuracy
Filtered out records with coordinates outside expected city limits.

11. Add Month and Season Columns
Added new columns based on the started_at variable to track monthly usage and seasonality.

3. Analysis

3.1 User Type

The chart below shows that about two-thirds of trips are made by members rather than casual users. Although our dataset does not provide counts of individual users by membership status, it would be valuable to assess the proportion of members in the overall user base. This would allow us to compare utilisation rates and investigate the average cost per trip for members compared to casual users.



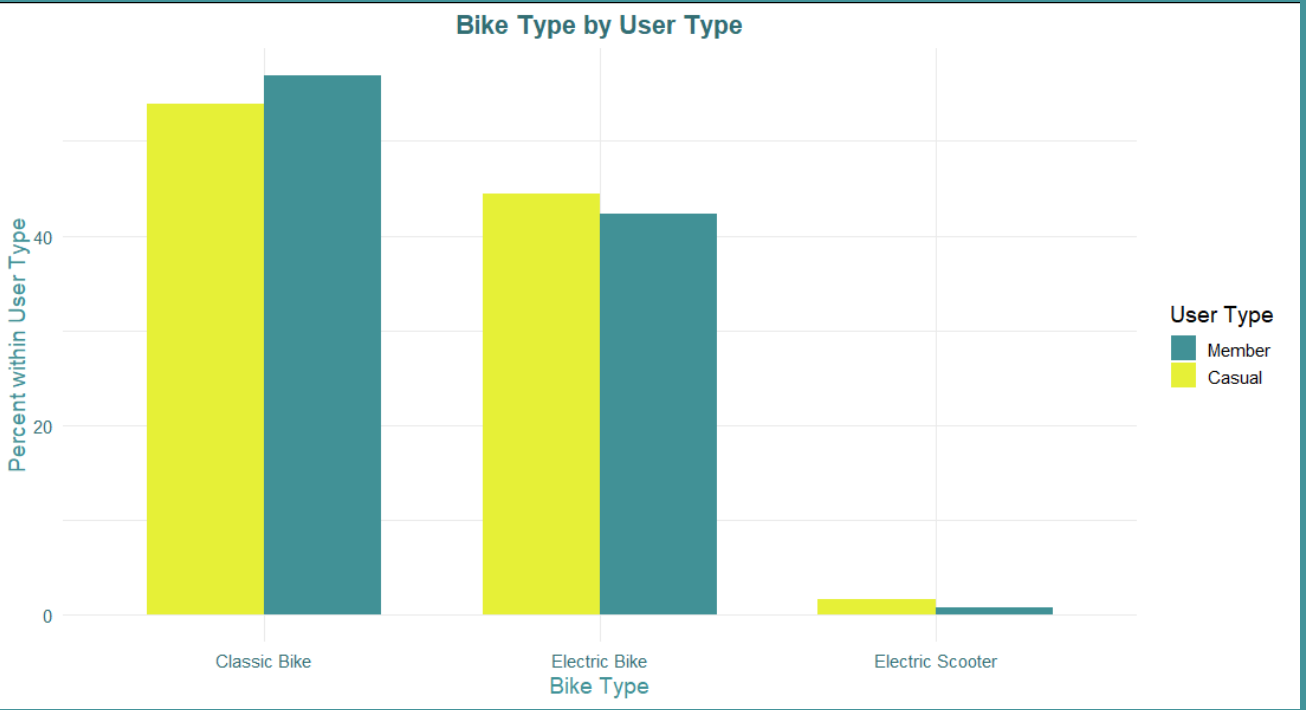
User Type	Number of Rides	Percent of Total
Casual	1549650	36.40%
Member	2707738	63.60%

The data also shows that members generally take shorter trips, averaging 12.21 minutes, while casual users average 24.31 minutes per ride. This suggests that members use the bikes primarily for regular transportation, such as commuting or short errands. Casual riders, who take longer trips, might be using the bikes for leisure or sightseeing.

User Type	Min (mins)	Q1 (mins)	Median (mins)	Mean (mins)	Q3 (mins)	Max (mins)	SD (mins)	IQR (mins)
Casual	1	7.36	12.96	22.59	24.31	1439.76	45.91	16.95
Member	1	5.22	8.75	12.21	14.76	1438.66	19.37	9.53

3.2 Bike Type

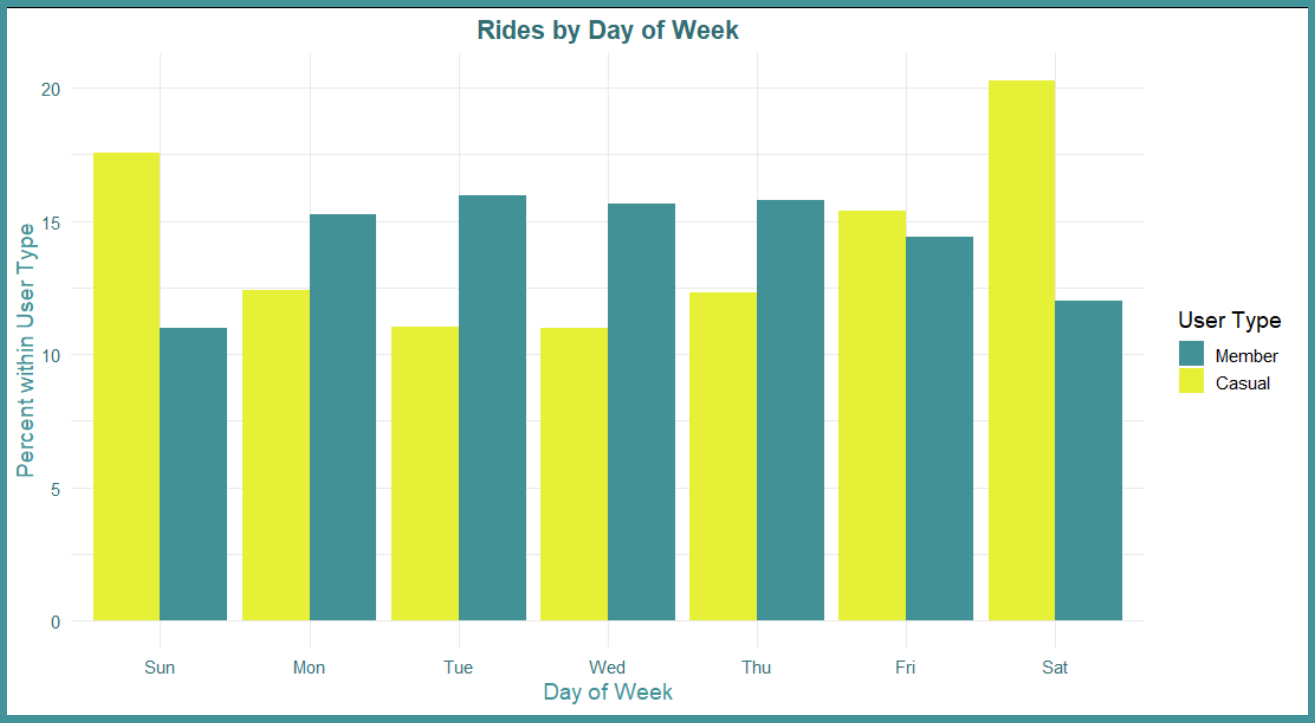
Classic bikes remain the most popular choice for both groups, accounting for more than half of trips regardless of membership. Electric bikes constitute a substantial share as well, between 42 and 44 percent of rides. Electric scooters are used much less frequently and represent only about one percent of total trips. It is important to note, however, that we do not have details on how many bikes of each type are actually available in the fleet.



Bike Type	User Type	Number of Rides	Percent within Group
Classic Bike	Casual	835302	53.90%
Classic Bike	Member	1540900	56.91%
Electric Bike	Casual	689086	44.47%
Electric Bike	Member	1145170	42.29%
Electric Scooter	Casual	25262	1.63%
Electric Scooter	Member	21668	0.80%

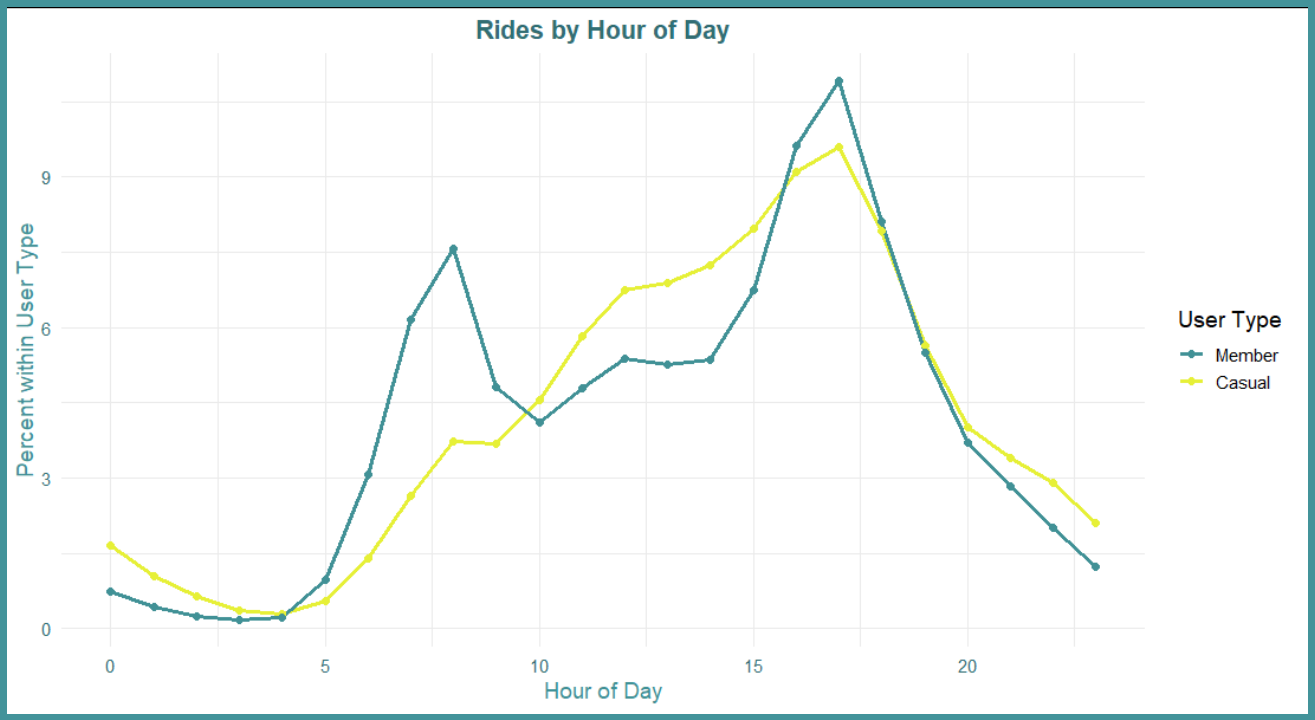
3.3 Rides by Day of Week

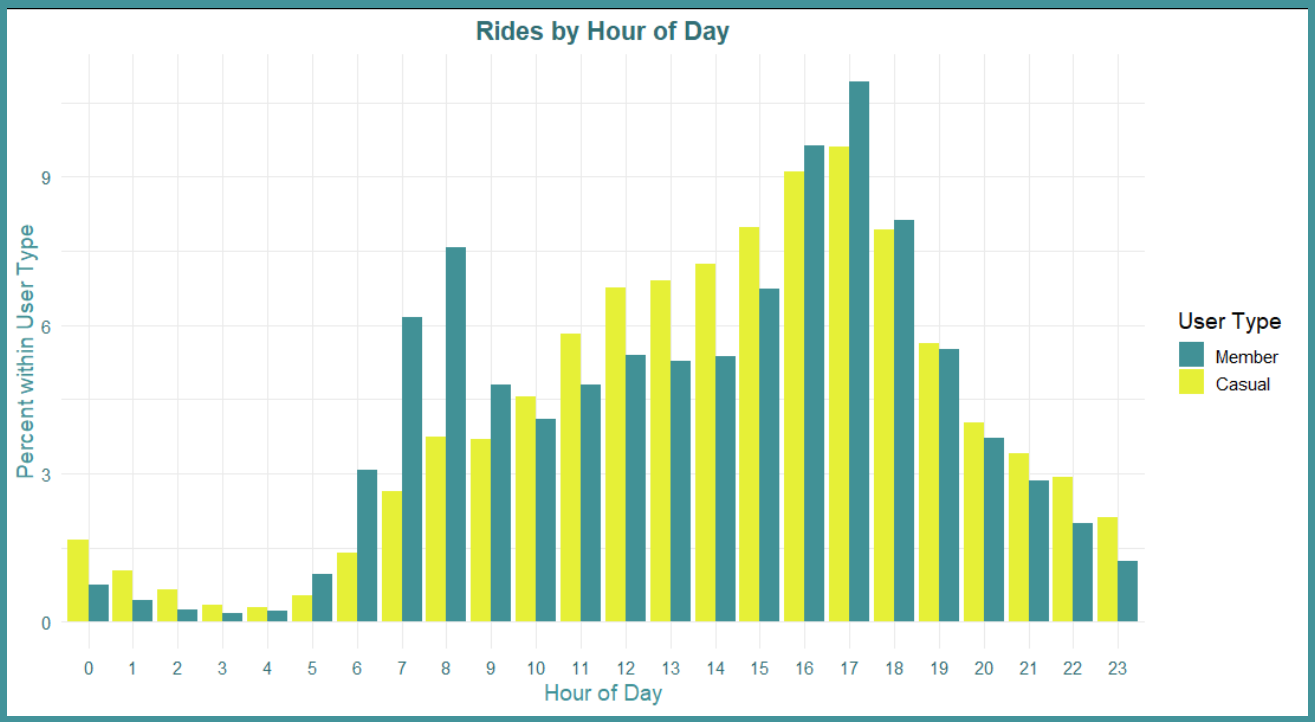
Analysis by day of week reveals that members are more active during the work week, with higher trip counts from Monday through Friday. Casual users show their highest activity from Friday to Sunday, more typical of recreational patterns.



3.4 Rides by Hour of Day

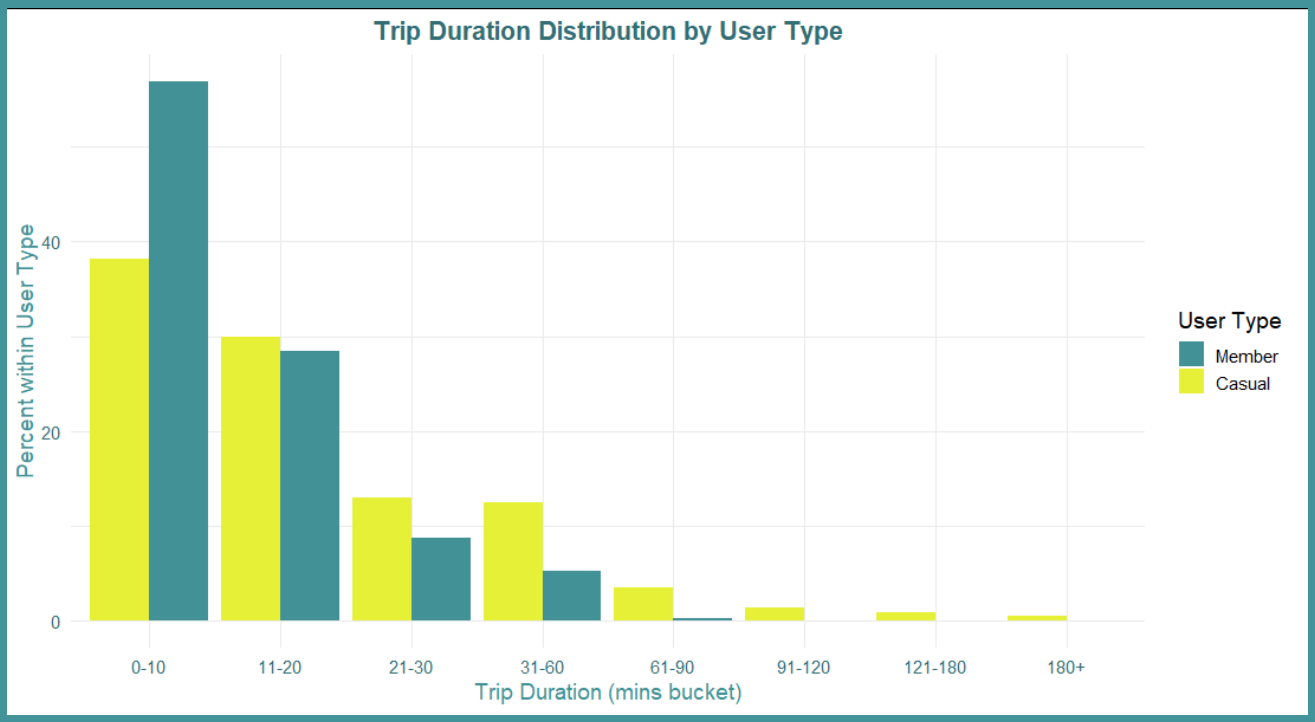
Looking at usage by hour, member activity peaks at the start and end of standard workdays, which indicates the system is well suited to commuters. Casual users are most active in the afternoon, which may reflect weekend outings or social activities. There is no corresponding early morning or evening peak for casual users that might suggest regular round-trip commuting.





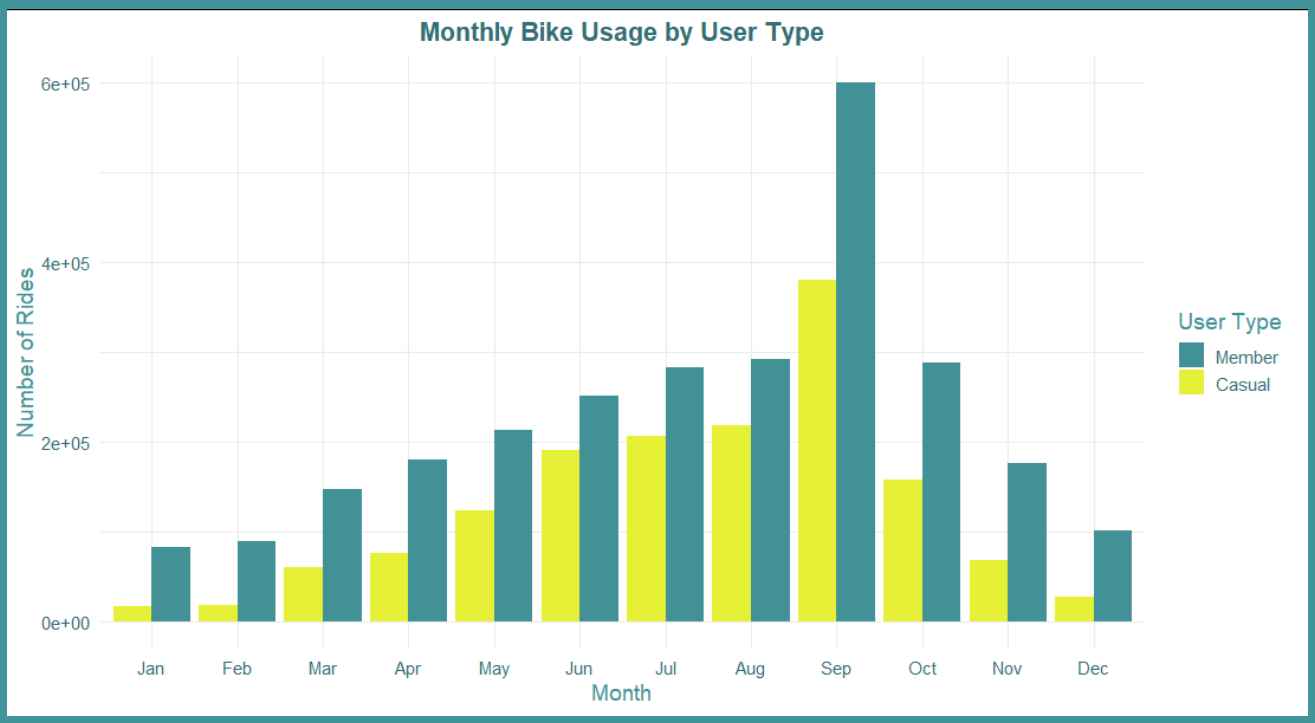
3.5 Rides by Duration

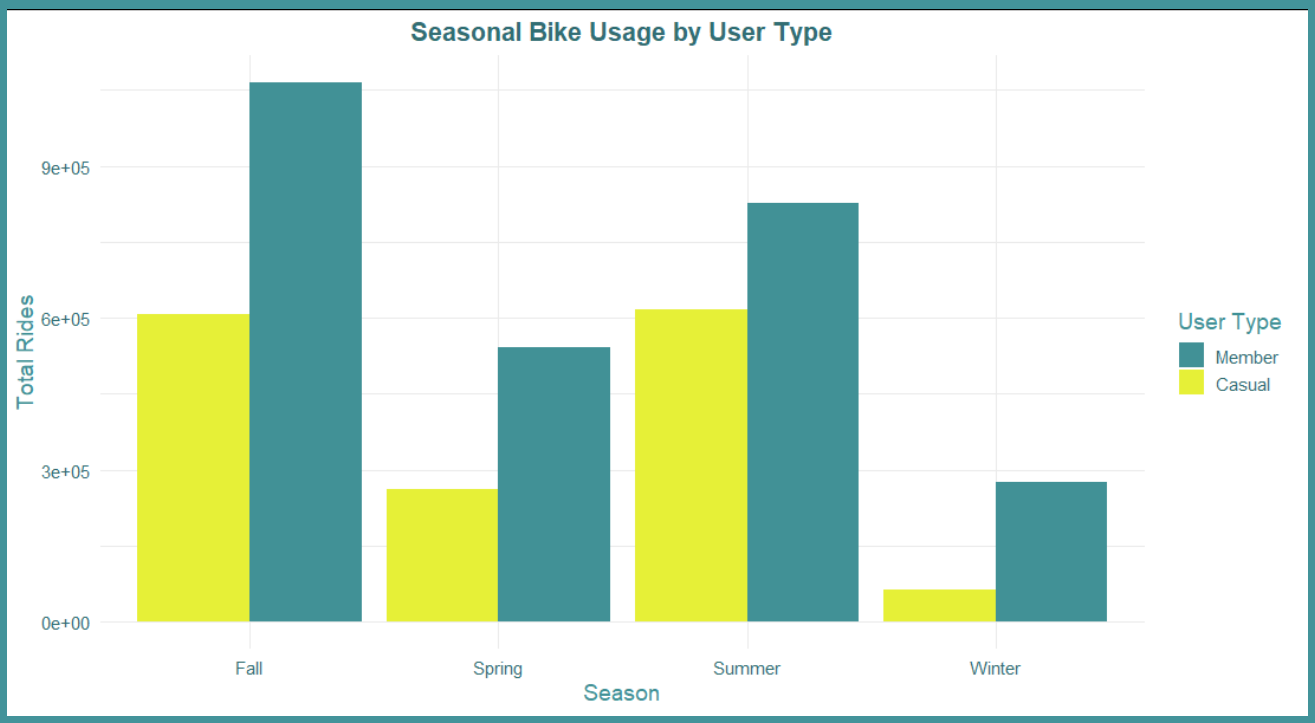
The distribution of trip durations also indicates that members tend to take significantly more short rides, with 56.9 percent of member trips lasting between zero and ten minutes, compared to 38.2 percent for casual users. Members may feel more comfortable using the bikes for quick trips because their annual membership is not linked to pay-as-you-go usage. Casual riders, in contrast, appear to favour longer journeys that maximize the value of a single rental period.



3.6 Rides by Season

Bike share usage in Chicago shows clear seasonality. Ride numbers are much lower in December, January, and February when cold weather and snow are common, making cycling less practical. As conditions improve in spring, both commuting and recreational trips increase steadily. The peak in bike share usage occurs during autumn and summer, times when Chicago experiences longer days and moderate to warm temperatures. The strong spike observed in September stands out and would be worth investigating further using multiple years of data to determine if this is a regular pattern and to identify possible causes.





4. Recommendations

Converting casual users into members can be challenging; however, this analysis provides clear insight into where marketing efforts should focus.

Members typically use bikes for short trips (0–20 minutes) and at a higher rate of utilisation than casual users. Marketing should consider promoting the value of short trips (highlighting ease of use, affordability, sustainability, and last-mile convenience).

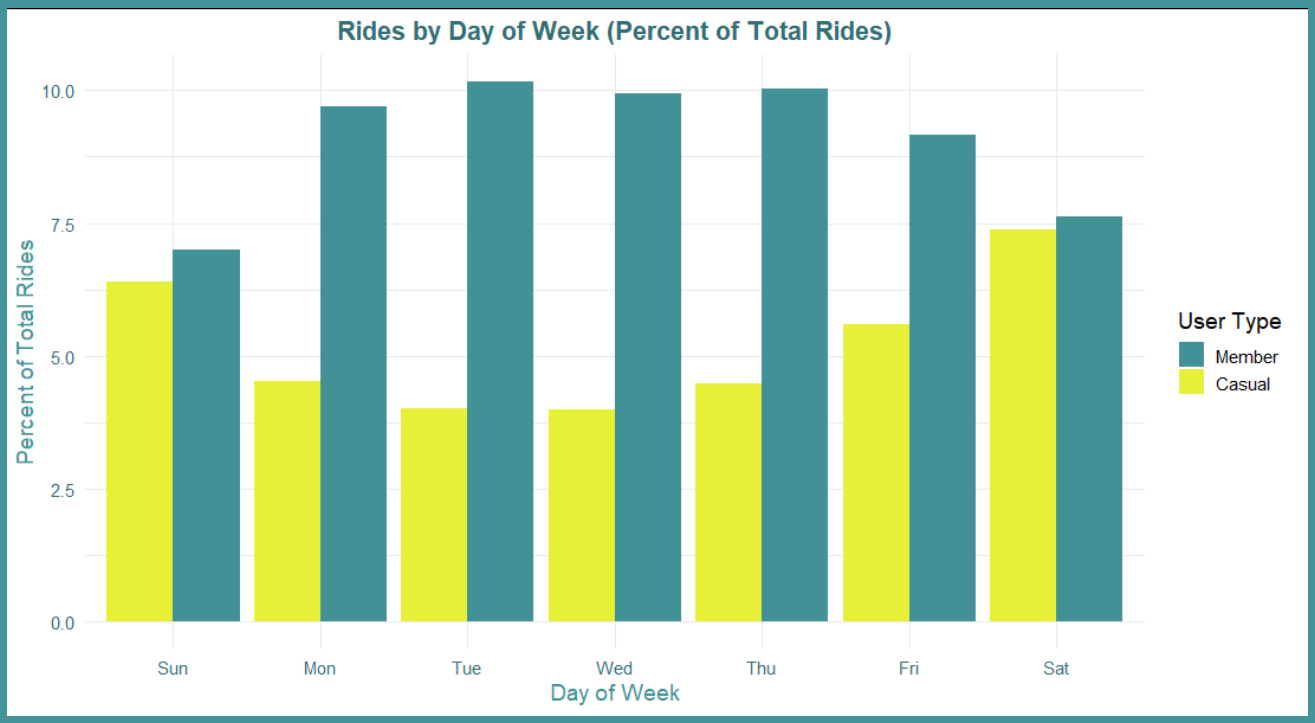
Additionally, Cyclistic should position itself as an integral part of daily routines, whether that be commuting to work, going to the gym, or running errands. A referral program that rewards riders for bringing a friend could further encourage habitual and social use.

Finally, further analysis of member utilisation costs could quantify potential savings for casual users who convert to membership.

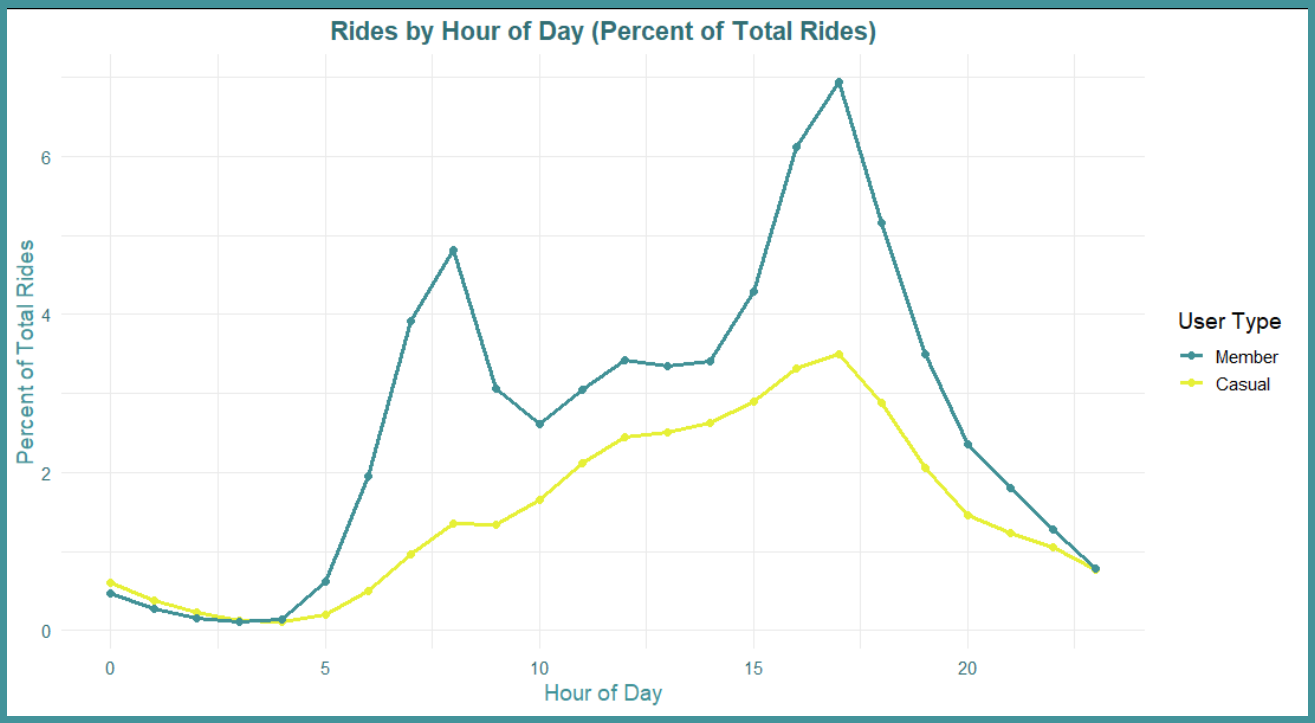
5. Appendix

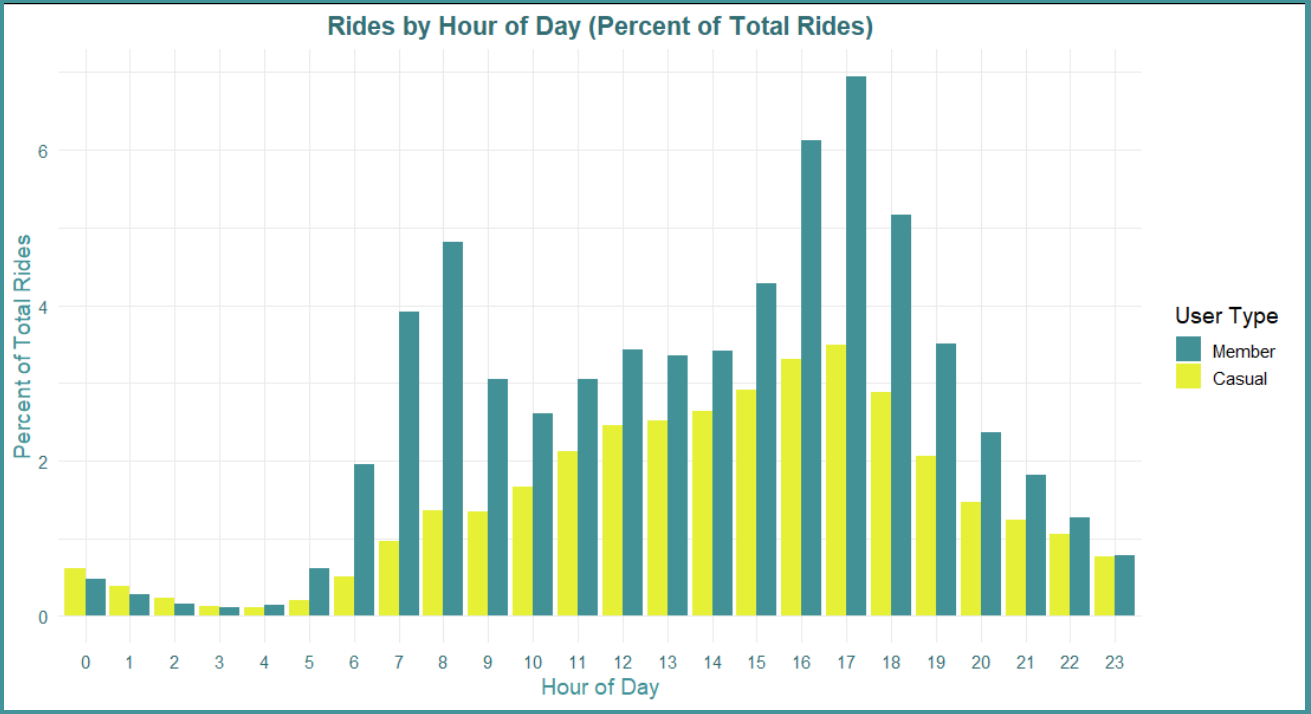
This appendix is populated with charts not included in the above analysis.

5.1 Rides by Day of Week (Percent of Total Rides)



5.2 Rides by Hour of Day (Percent of Total Rides)





5.3 Trip Duration by User Type

User Type	Trip Duration (mins)	Number of Rides	Percent within User Type
Casual	0-10	591907	38.20%
Casual	11-20	464608	29.98%
Casual	21-30	201134	12.98%
Casual	31-60	193881	12.51%
Casual	61-90	54829	3.54%
Casual	91-120	21220	1.37%
Casual	121-180	14476	0.93%
Casual	180+	7595	0.49%
Member	0-10	1540582	56.90%
Member	11-20	771320	28.49%

1–10 of 16 rows

Previous 1 2 Next

User Type	Trip Duration (mins)	Number of Rides	Percent within User Type
Casual	0-10	591907	38.20%
Casual	11-20	464608	29.98%
Casual	21-30	201134	12.98%
Casual	31-60	193881	12.51%
Casual	61-90	54829	3.54%
Casual	91-120	21220	1.37%
Casual	121-180	14476	0.93%
Casual	180+	7595	0.49%
Member	0-10	1540582	56.90%
Member	11-20	771320	28.49%

1–10 of 16 rows

Previous 1 2 Next

5.4 Cumulative Distribution of Trip Duration