Cyclistic: Bike-Share

How Does A Bike-share Navigate Speedy Success?

Rishika Gour

Google Data Analytics Cyclistic Capstone

Introduction

Welcome to the Cyclistic bike-share analysis case study! Throughout this case study, I've tackled various real-world tasks typical of a junior data analyst. I was employed at Cyclistic, a fictional company, and had the opportunity to interact with diverse colleagues and team members. To address the essential business queries, I followed the data analysis process steps, including asking questions, preparing data, processing, analyzing, sharing findings, and taking action.

Statement of the Business Task

Dive into the data to uncover the differences between annual members and casual riders. Try to understand what might motivate casual riders to go for a membership, and think about how digital media could affect our marketing strategies. We're looking to gain valuable insights from this analysis, which will help us shape our marketing plans to encourage casual riders to become annual members.

Key Stakeholders

- Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations.
- Lily Moreno: The director of marketing and my manager.
- Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.
- Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

A description of all data sources used

The data used for our analysis covers the past twelve months, spanning from November 2021 to October 2022. We sourced this data from Divvy Data, and each dataset was initially compressed in .zip file format. The naming convention for the .zip files follows the month and year in which the data was collected. For example, the file containing data from December 2021 was labeled as "202112-divvy-tripdata.zip." This dataset has been generously made available to us by Motivate International Inc., under a data license agreement. To adhere to data privacy regulations, we've ensured that no personally identifiable rider information is used. The data is organized on a monthly basis, with each month's data stored in its respective CSV file. The dataset consists of Cyclistic's own clientele, who are avid bike riders. It's considered ROCCC: Reliable, Original, Comprehensive, Current, and Cited. All the files share consistent columns, and each column contains the appropriate data type. In total, the dataset comprises over 5 million rows and includes the following columns:

- unique ride id
- rideable type (member or casual)
- start and end time
- start and end docking station (name and id)
- start and end latitude/longitude
- bicycle type (classic, docked, electric)

Tools Used

- Utilized Excel, RStudio, and Power BI for data cleaning, analysis, and visualization.
- Employed Excel for data sorting, filtering, and formatting due to its user-friendliness.
- Utilized RStudio for more advanced data organization and analysis.
- Created visualizations using Power BI, known for its ability to generate compelling visuals with ease.

Cleaning or Manipulation of data
Step 1: Collect Data

```
X1_202111 <- read_csv("1-202111-divvy-tripdata.csv")
X2_202112 <- read.csv("2-202112-divvy-tripdata.csv")
X3_202201 <- read.csv("3-202201-divvy-tripdata.csv")
X4_202202 <- read.csv("4-202202-divvy-tripdata.csv")
X5_202203 <- read.csv("5-202203-divvy-tripdata.csv")
X6_202204 <- read.csv("6-202204-divvy-tripdata.csv")
X7_202205 <- read.csv("7-202205-divvy-tripdata.csv")
X8_202206 <- read.csv("8-202206-divvy-tripdata.csv")
X9_202207 <- read.csv("9-202207-divvy-tripdata.csv")
X10_202208 <- read.csv("10-202208-divvy-tripdata.csv")
X11_202209 <- read.csv("11-202209-divvy-tripdata.csv")
X12_202210 <- read.csv("12-202210-divvy-tripdata.csv")
```

Step 2: Wrangle Data And Combine Into A Single File

- Compared column names each of the files.
- While the names don't have to be in the same order, they do need to match perfectly before we can use a command to join them into one file.
 colnames(X1_202111) colnames(X2_202112) colnames(X3_202201) colnames(X4_202202) colnames(X5_202203) colnames(X6_202204) colnames(X7_202205) colnames(X8_202206) colnames(X9_202207) colnames(X10_202208) colnames(X11_202209) colnames(X12_202210)
- Renamed columns to make them consistent with X12_202210 (as this will be the supposed going-forward table design for Divvy)
- Inspected the dataframes and looked for incongruencies.
- Converted ride_id and rideable_type to character so that they can stack correctly.
- Stacked individual month's data frames into one big data frame.
- Removed lat and long.

Step 3: Cleaning Up And Adding Data To Prepare For Analysis

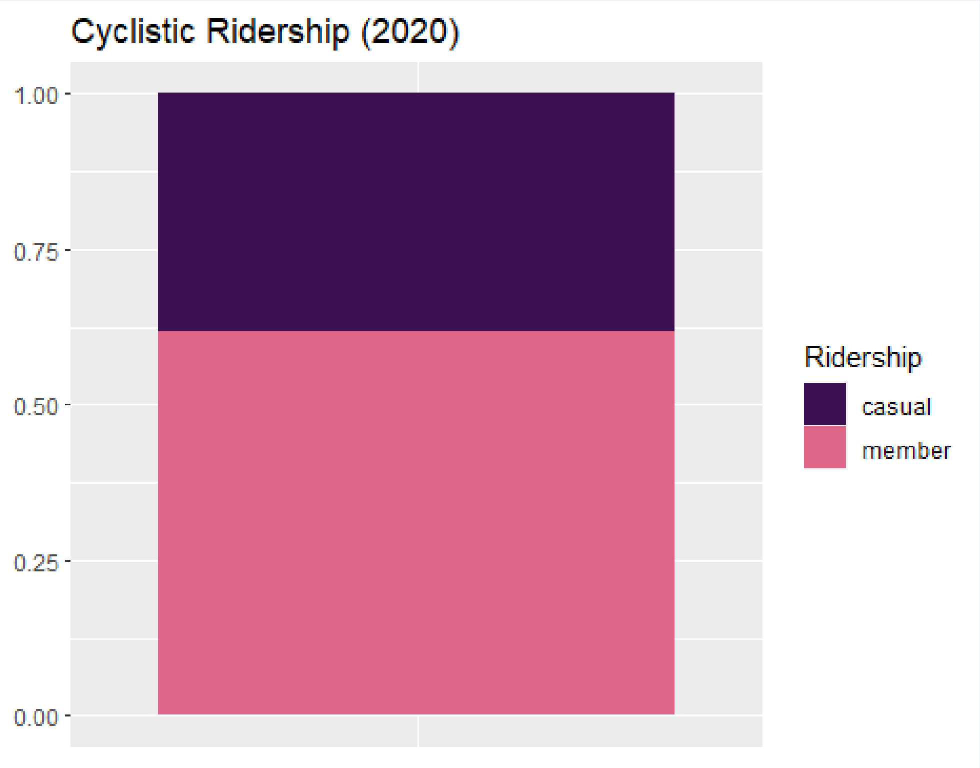
- Inspected the new table that has been created.
  colnames(all_trips) #List of column names
  nrow(all_trips) #How many rows are in data frame?
  dim(all_trips) #Dimensions of the data frame?
  head(all_trips) #See the first 6 rows of data frame. Also tail(all_trips)
  str(all_trips) #See list of columns and data types (numeric, character, etc)
  summary(all_trips) #Statistical summary of data
- Reassigned to the desired values.
- Checked to make sure the proper number of observations were reassigned.
- Added columns that list the date, month, day, and year of each ride.
- Added a "ride_length" calculation to all_trips.
- Inspected the structure of the columns.
- Converted "ride_length" from Factor to numeric so we can run calculations on the data
- Removed "bad" data.
- Created a new version of the dataframe (v2) since data is being removed.

Step 4: Conducting Descriptive Analysis

- Conducted Descriptive analysis on ride_length.
- Compared members and casual users.
- Found the average ride time by each day for members vs casual users.
- Analyzed ridership data by type and weekday.
- Created weekday field using wday().
- Calculated the number of rides and average duration .
- Calculated the average duration.

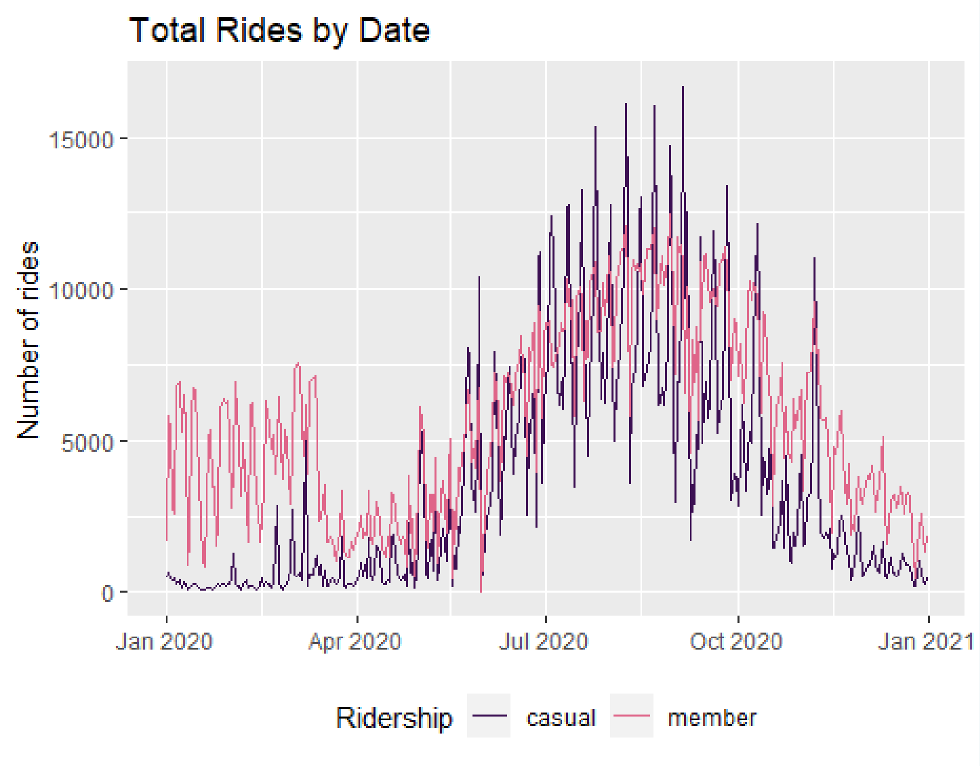
Visualizations and Key Findings

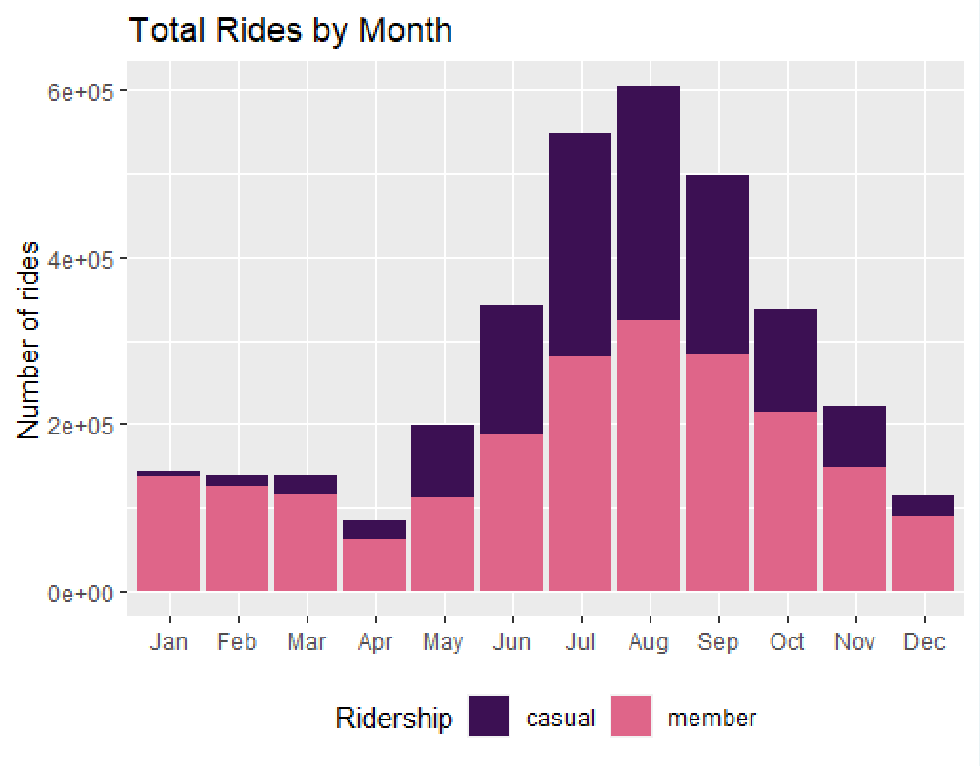
Proportion Of Casual Riders and Member Riders



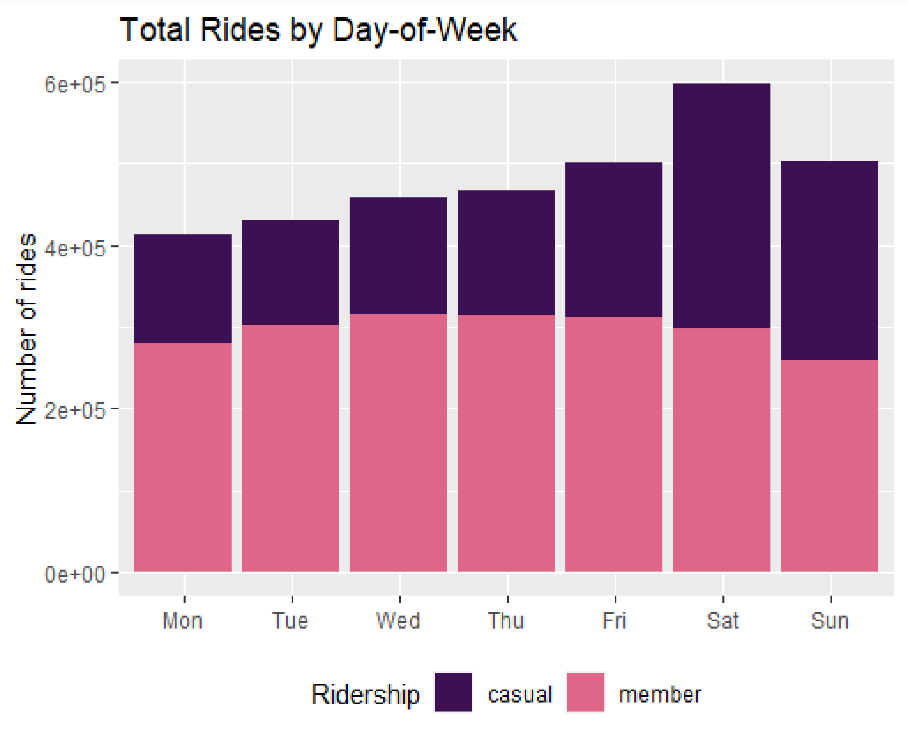
It's evident that over half of Cyclistic users have opted for the membership choice, and members tend to ride more compared to casual riders.

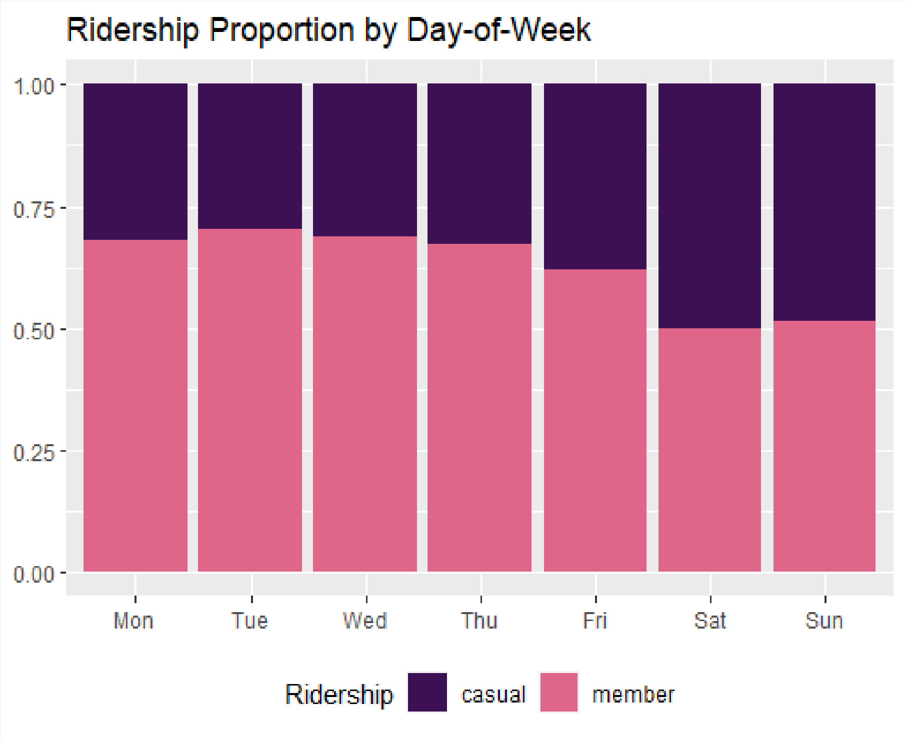
Proportion of Daily Rides and Monthly Rides





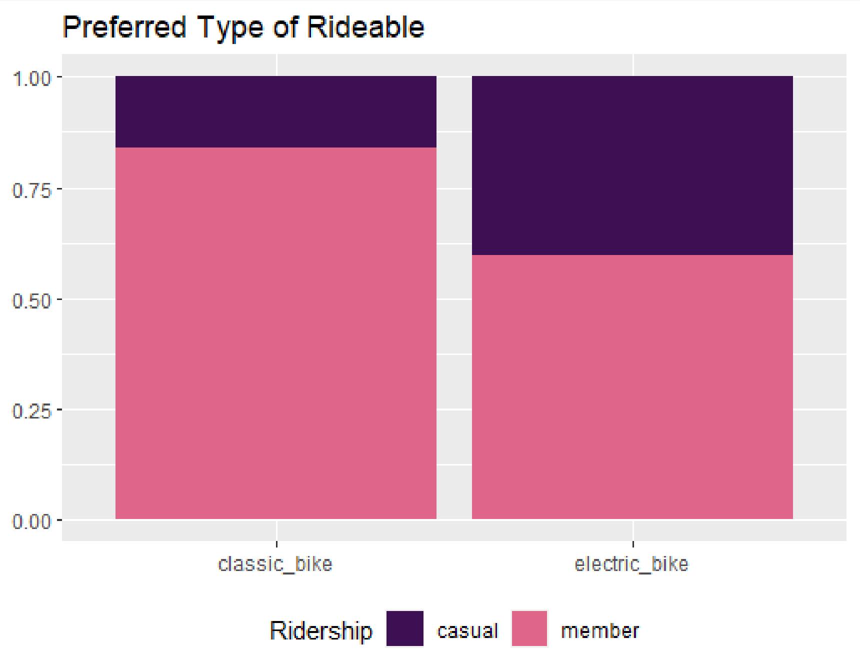
When examining rides by the day of the week, we observe an uptick in total rides during the weekend, notably on Saturdays. Members exhibit more consistent weekday usage of Cyclistic, accounting for 60-70% of the total rides each day. On the weekend, casual riders boost their riding activity by 20%.

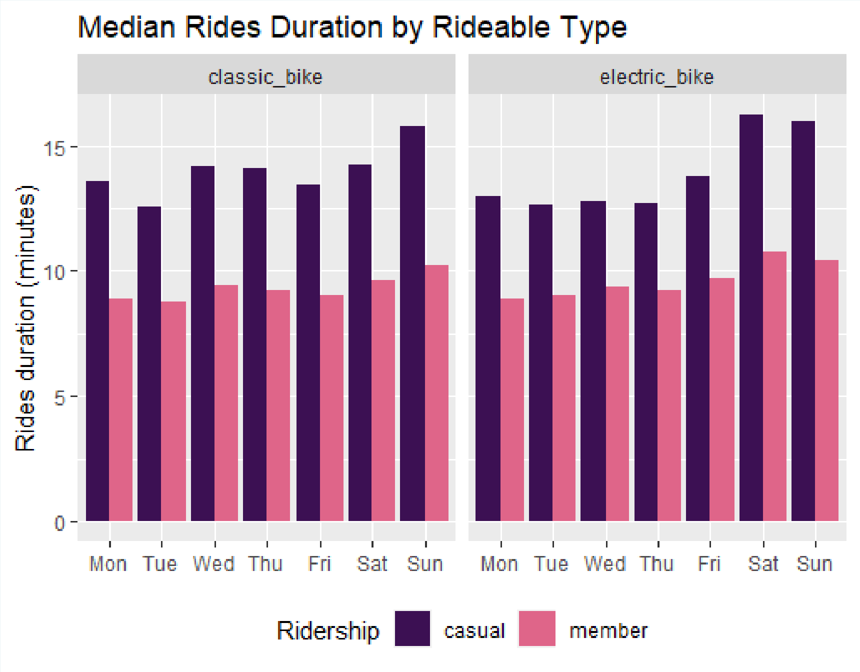




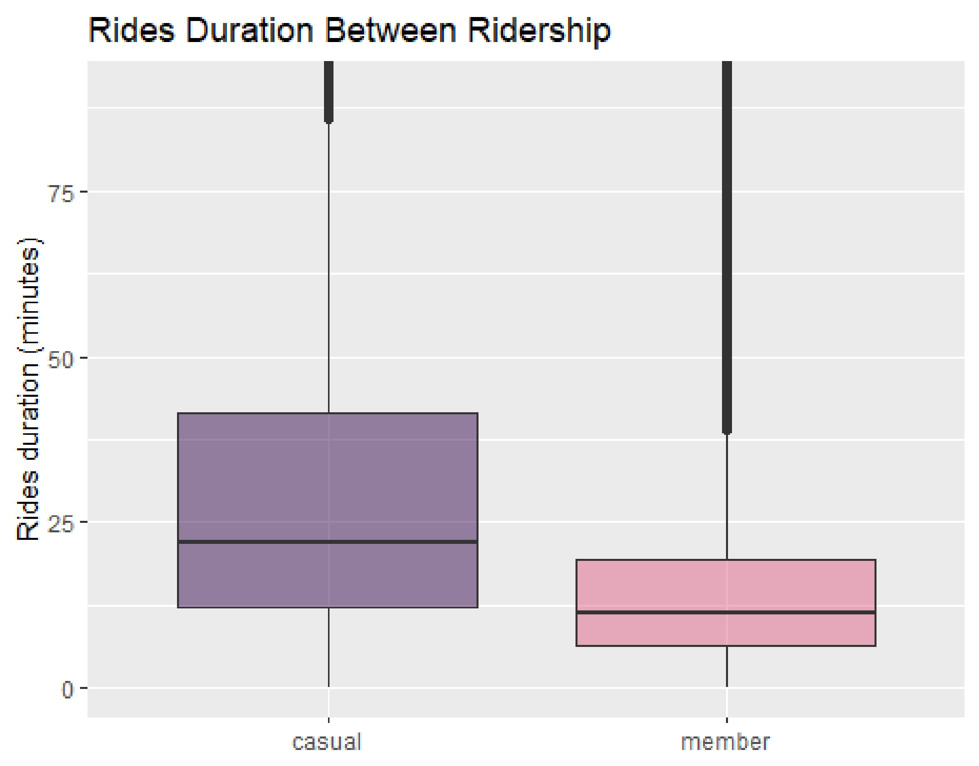
Proportion of Rideable

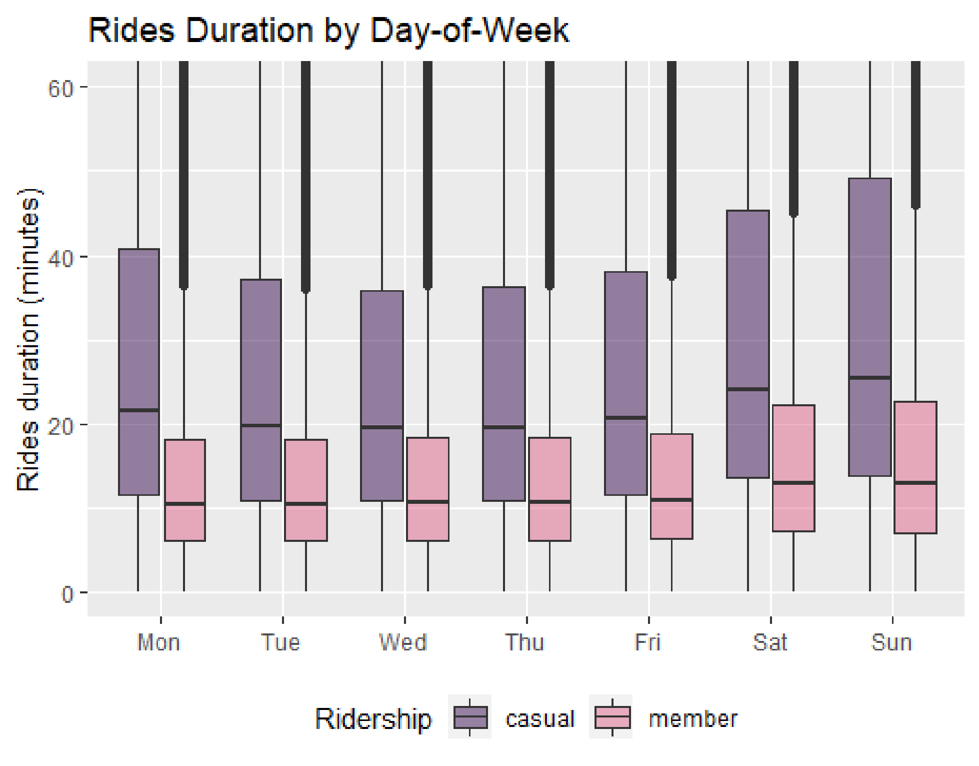
Although there's a distinct inclination among members for classic bikes, the ride duration doesn't appear to be significantly impacted by the type of rideable. In contrast, casual riders have a preference for electric bikes and tend to embark on longer rides, particularly during the weekends.



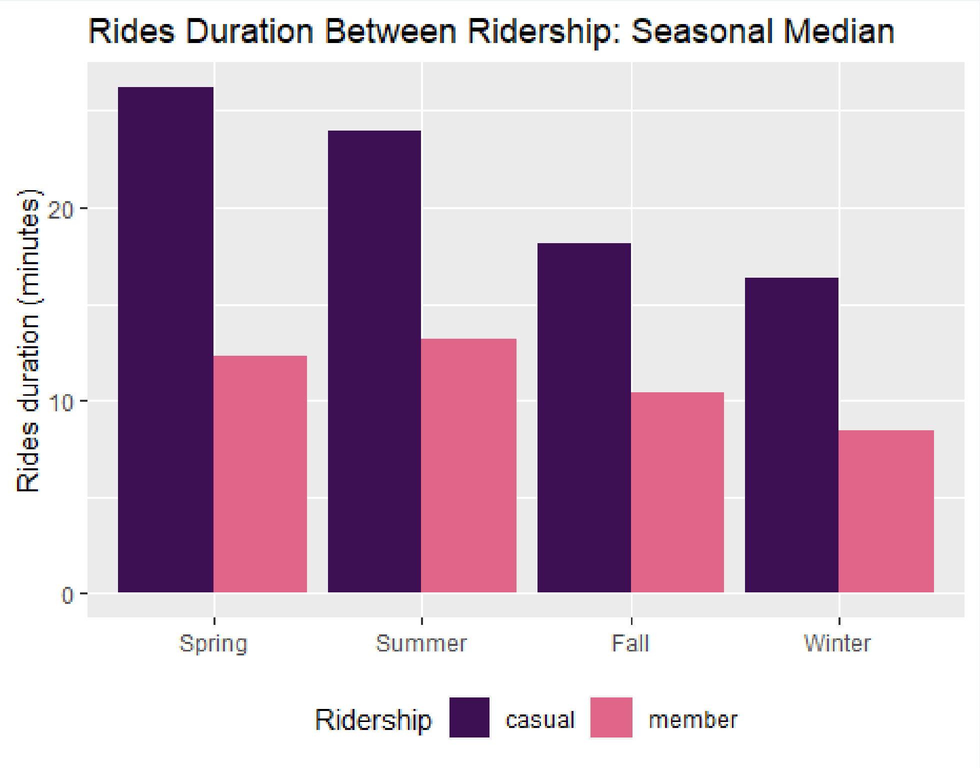


Rides Duration





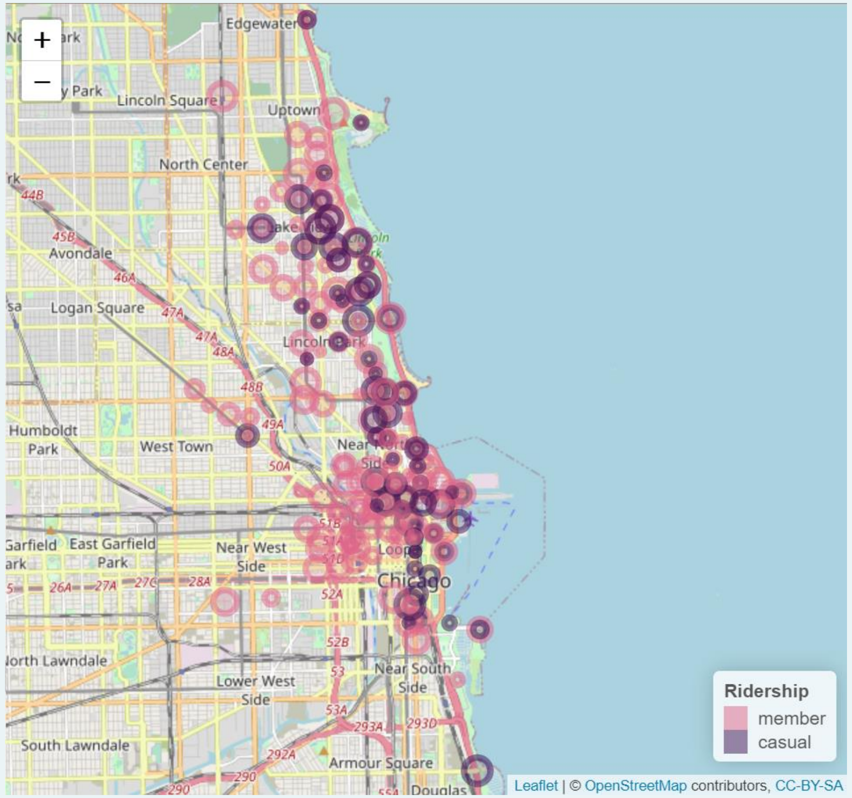
Rides Duration



When we compare the distribution, it becomes apparent that casual riders have longer and more variable ride durations, approximately twice as long as those of members. Specifically, the median ride duration is 22 minutes for casual riders and 11 minutes for members. For casual riders, 75% of their rides are completed in under 42 minutes, while for members, 75% of their rides are under 19 minutes.

Both plots exhibit positive skewness, as evidenced by the mean ride duration being greater than the median. For casual riders, the mean ride length is 42 minutes, and for members, it's 19 minutes. This skewness is primarily driven by the presence of outliers, representing exceptionally long ride durations.

Location Map



In the map provided, you can see 200 stations with the highest ride starts. Casual riders are denoted by purple circles, while members are represented by pink circles. The size of the circles corresponds to the total rides originating from each station.

As illustrated on the map, members tend to cluster further inland, initiating their rides from stations situated within the city. In contrast, casual riders often commence their rides at stations nearer to the coastal area.

Summary of the Analysis

- Casual riders tend to use Cyclistic bikes for extended periods and have a higher weekend and warm-season ridership. They show a preference for electric bikes and docking stations near the lake, indicating a more leisure-focused use of Cyclistic.

- In contrast, annual members appear to have a more utilitarian approach to the bike-sharing service. They ride regularly on weekdays, opting for shorter rides and classic bikes. Their rides usually begin from stations located deeper within the city.

Top Three Recommendations

1. Launch seasonal marketing campaigns, such as special offers for membership enrolment during peak months like July and August, as well as on weekends.

2. Provide discounts for short rides during weekday commuting hours.

3. Examine the potential impact of increasing the number of docking stations within the city on membership registrations.