Data Analysis Course Capstone

Nic Sevic

2025-04-01

Case Study Description: How does a bike-share navigate speedy success?

Scenario:

You are a junior data analyst working on the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members.

Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use the bikes to commute to work each day.

About the company: In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members. Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Goal:

Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics.

Three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

You will produce a report with the following deliverables to answer the first question:

1. A clear statement of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top three recommendations based on your analysis

Case Study Work

Statement of the business task

Cyclistic, a bike-share company in Chicago, wants to increase the number of annual memberships to drive long-term profitability. The primary goal is to understand how casual riders and annual members use Cyclistic bikes differently. By analyzing historical bike trip data, the aim is to identify trends and usage patterns that will inform marketing strategies to convert casual riders into members. The findings will guide targeted marketing campaigns designed to increase membership conversion and secure executive approval for proposed strategies.

Description of all data sources used

The analysis uses Cyclistic's historical bike trip data from Q1 2019 and Q1 2020. These datasets contain information on individual bike trips, including start and end times, trip durations, and user types (casual riders vs. annual members). The data was formatted as CSV files and has been loaded into R for cleaning, processing, and analysis.

Note: Since Cyclistic is an example company, data made publicly available by Motivate International Inc. was used. It has been vetted for privacy, excluding any personally identifiable information.

Documentation of any cleaning or manipulation of data

To prepare the Cyclistic bike-share data for analysis, the following steps were taken:

Data Import and Initial Setup

The quarterly datasets for Q1 2019 and Q1 2020 were imported. Packages such as tidyverse and conflicted were loaded to streamline data manipulation and handle any function name conflicts.

Column Standardization and Merging

Column names in the 2019 dataset were renamed to match the 2020 format to ensure consistency across both datasets. Data types were aligned (e.g., IDs and bike types converted to character), and the two datasets were then combined into a single dataframe using bind_rows().

Dropping Deprecated Fields

Columns no longer used in the 2020 dataset (such as latitude, longitude, birth year, gender, and tripduration) were removed to maintain consistency.

Data Cleaning and Transformation

User Type Consolidation: The member_casual column originally used multiple labels for user types. These were standardized to "member" and "casual" to match the 2020 schema.

Datetime Features: New columns were created to extract the date, month, day, year, and day of the week from the ride start time. This supports more granular temporal analysis.

Ride Duration: A new field ride_length was calculated by subtracting the ride start time from the end time. This value was then converted to numeric to enable statistical analysis.

Filtering Invalid Records: Rides with negative durations or those starting from the internal station "HQ QR" were removed, as they likely represent test or maintenance trips.

Data Aggregation and Summary Statistics

Basic descriptive statistics (mean, median, min, max) were calculated for ride duration across all users and compared between members and casual riders. Aggregations by day of the week were also performed to explore usage patterns over time.

Visualization Preparation

Factor levels for days of the week were reordered to follow the natural calendar sequence (Sunday through Saturday). Summary tables were generated to support visualizations of ridership counts and average ride durations by user type and weekday.

Summary of your analysis

After cleaning and preparing the dataset, the following analysis steps were performed to uncover ridership trends and usage patterns among casual and member riders:

Export of Summary Data

A CSV file was generated containing the average ride duration by user type and day of the week. This export supports external analysis and visualization.

Route Popularity Analysis

A new field was created to represent each ride route by combining the start and end station names. Routes were grouped and summarized to identify the most frequently traveled paths, along with average ride duration and user type distribution (member vs. casual).

Station Usage Analysis

Separate dataframes were created to identify the most common start and end stations across all rides. Each was broken down further by total number of rides, and ride counts for each user type. This helped identify high-traffic stations favored by different user segments.

Top and Bottom Stations by Rider Type

For deeper insight, the top and bottom 10 stations (by ride count) were extracted for both members and casual users—considering both starting and ending stations. These subsets were combined to create a single unified dataset capturing notable patterns in rider behavior at key stations.

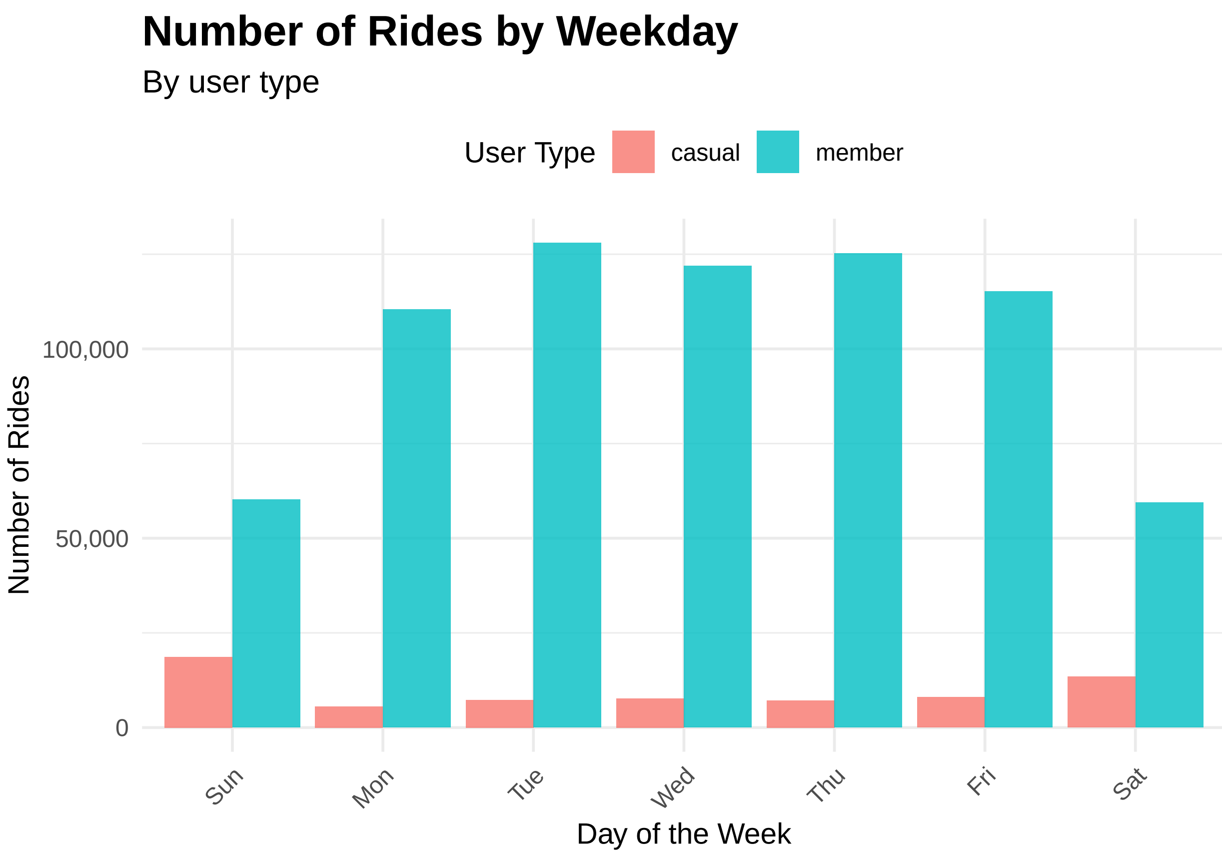
Data Reshaping for Visualization

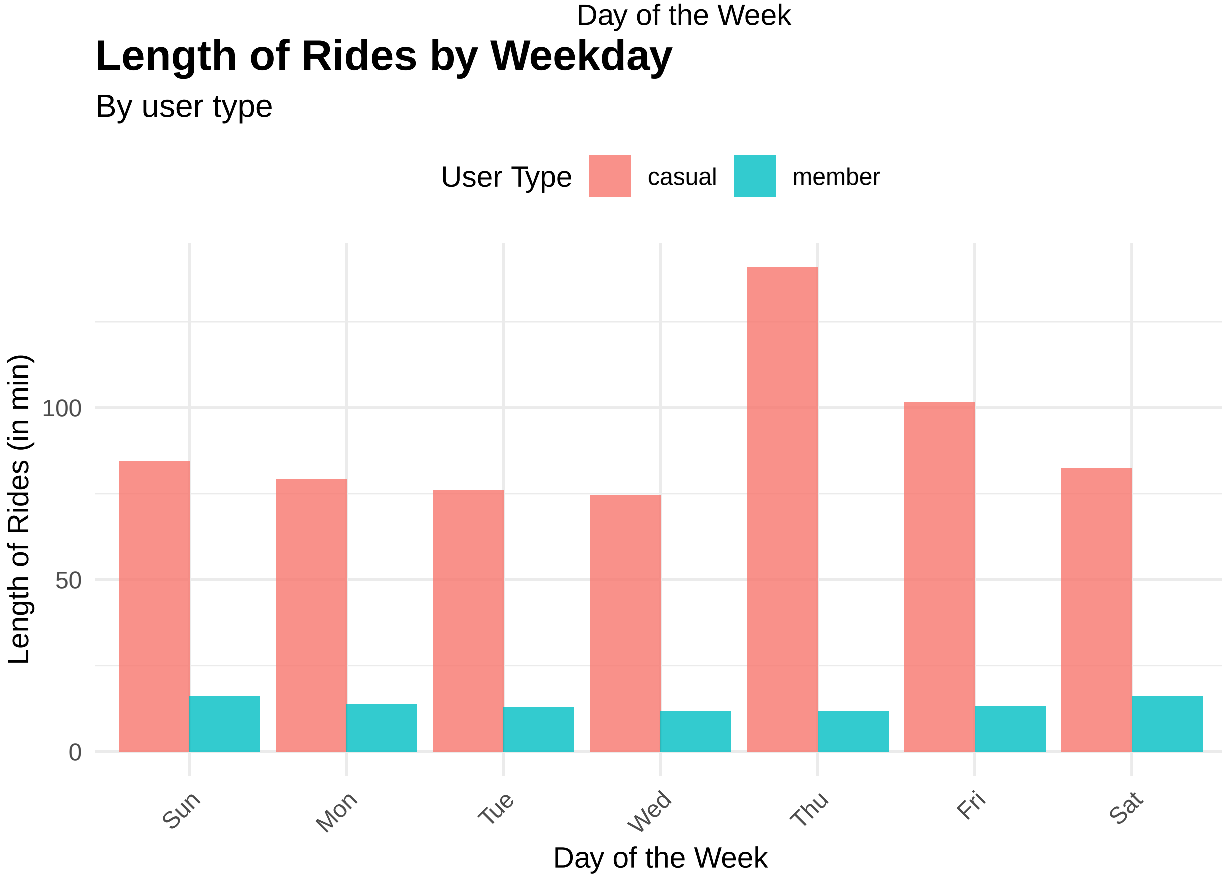
The station summary data was reshaped into a long format to support flexible plotting and analysis. Rider type and station function (start or end) were explicitly labeled, and counts were categorized for comparison between members and casual users. To focus on meaningful trends, only stations with 50 or more rides were retained in the final dataset.

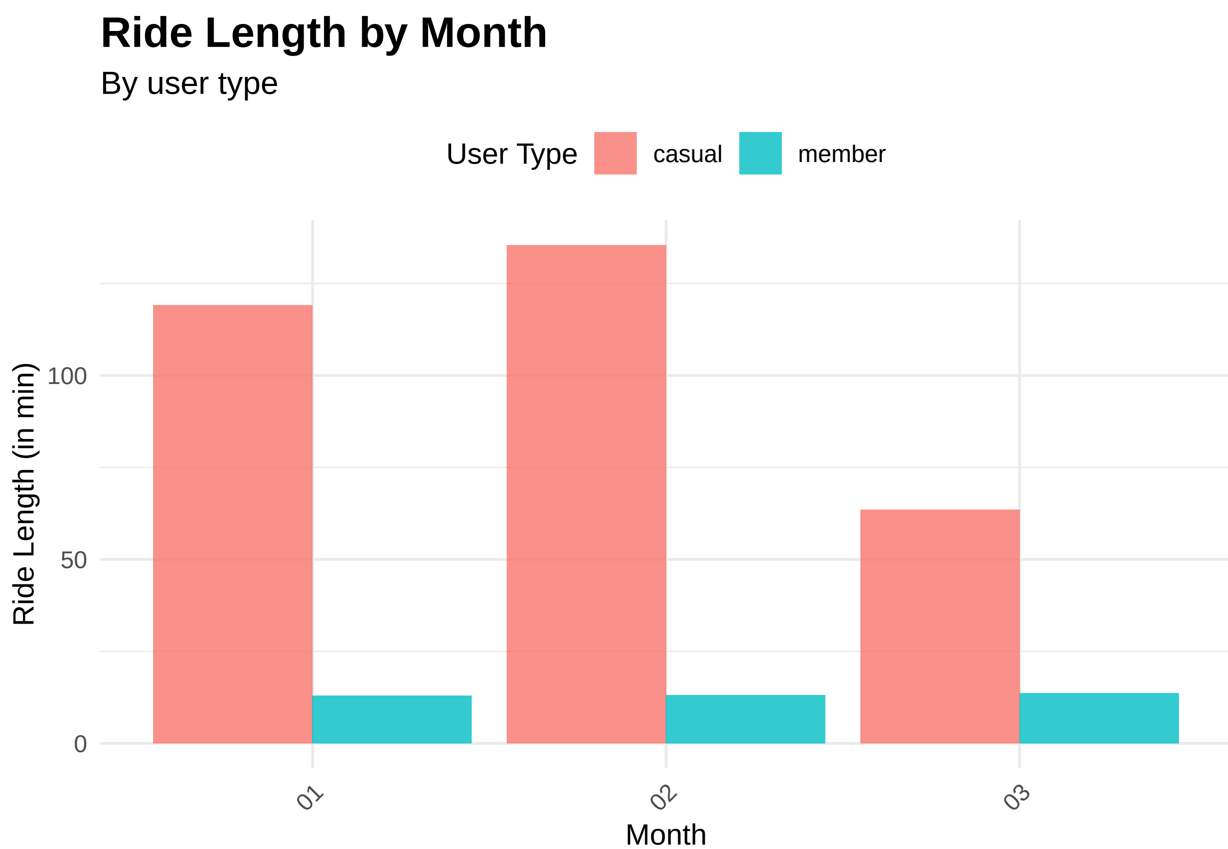
Key Findings

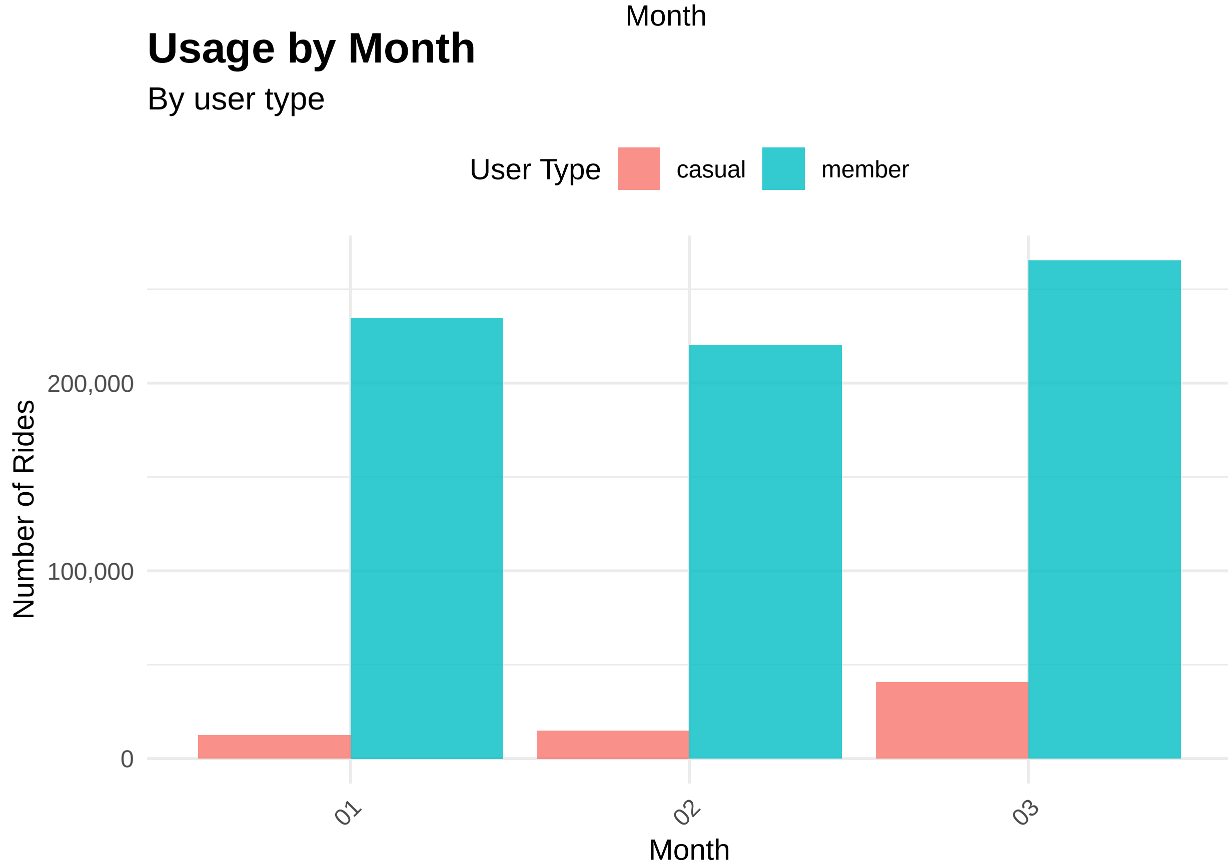
Overall members are taking the bulk of the rides and their rides are shorter but more consistent than casual riders. Casual riders are taking longer rides and are riding more frequently but for shorter periods in March than in January or February. Areas of primary use differ pretty drastically for Casual vs Member riders, offering differing opportunities for recruitment.

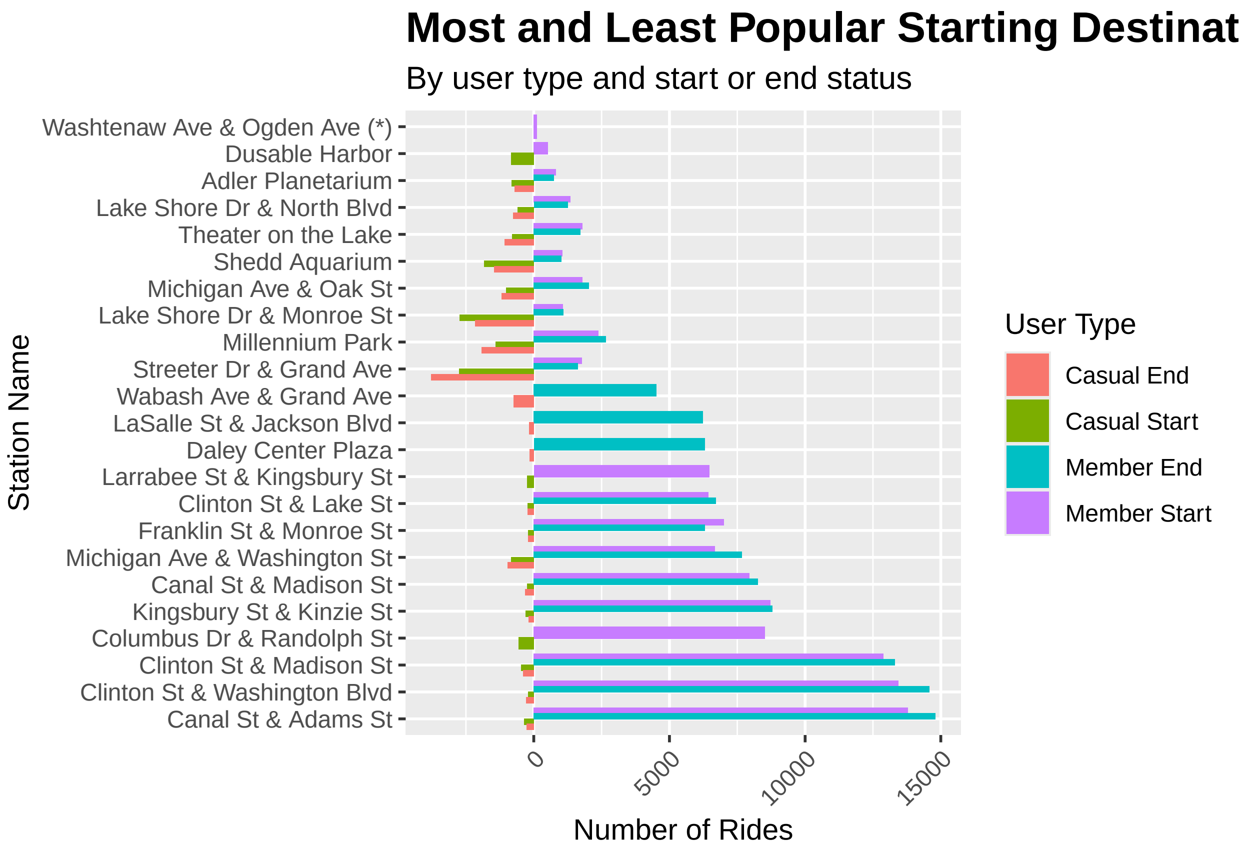
Supporting figures






Ride Length by Month
By user type
User Type
casual
member
Ride Length (in min)
100
50
0
01
02
03
Month


Usage by Month
By user type
User Type
casual
member
Number of Rides
200,000
100,000
0
01
02
03
Month



Strategic Insights

Casual riders skew heavily toward weekend usage, longer ride times, and recreational behavior. Members show consistent weekday usage, with shorter ride durations—indicative of commuting patterns. Spring seems like the best time to try and convert due to the increase in number of rides in March.

Recommendations

Weekend Promotions: Offer casual riders discounted weekend passes to incentivize repeat use.

Convert Commuters: Target casual weekday users with promotions around commuter memberships. Offering to credit the cost of their last ride to membership is one possibility.

Geographic Focus: Deploy ads and signage at high-volume tourist hubs like Navy Pier and Millennium Park. Consider highlighting convenience of routes to/from those points.

Gamify Membership: Launch challenges rewarding consistent weekday use to shift behavior.

Next Steps

Perform geospatial clustering of routes. Run A/B tests on promotional campaigns. Integrate weather and event data to deepen behavioral modeling.