Case study: How does a bike-share navigate speedy success

Karla Leiva Arauz

July 10, 2025

1 Introduction

Cyclistic is a bike-share program that began in 2016 and has since become an essential part of Chicago's transportation network. With a fleet of 5,824 geotracked bicycles and 692 stations throughout the city, Cyclistic offers users the flexibility to unlock a bike at one station and return it to any other. The program also promotes accessibility by providing inclusive bike options for riders with disabilities, including reclining bikes, hand tricycles, and cargo bikes, ensuring a more equitable riding experience for all.

The objective of this study is to provide insights that will support the design of marketing strategies aimed at converting casual riders into annual subscribers. Does customers who purchase single-ride passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members. To guide the analysis, the team has identified three key questions, listed below.

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?

The focus of this report is on answering the first key question.

2 Data Sources

This analysis will use Cyclistic's historical trip data, specifically the Divvy 2019 Q1 and Divvy 2020 Q1 datasets provided by Motivate International Inc. The data is structured in a table format and includes a variety of data types such as character, numeric, double, and time values. Due to data privacy restrictions, personally identifiable information about riders is not available. As a result, it's not possible to link individual pass purchases to credit card numbers, determine if casual riders live within the Cyclistic service area, or identify whether they've made multiple purchases. In the table below you will find the headers of the columns and the data type of each one.

Header Name	Data Type
ride_id	character
started_at	character
ended_at	character
day_of_week	number
ride_length	time
rideable_type	character
start_station_id	number
start_station_name	character
end_station_id	number
end_station_name	character
member_casual	character

Table 1: Data type

Note: The data sets contain more columns, but these ones are gonna be deleted during data preparation on R.

3 Data Preparation

Tools used:

- Excel — For basic data cleaning and visualization.
- R Studio — For advanced data analysis.

This was the process carried out in Excel:

A column named `ride_length` was created and its format was changed to Time (HH:MM:SS). The calculation for the length of each ride was performed by subtracting `started_at` from `ended_at`.

Another column, `day_of_week`, was also created to calculate the day each ride started using the `WEEKDAY` function. For example: `=WEEKDAY(C2,1)` (where Sunday = 1 and Saturday = 7). The format for this column was set to General.

In the dataset named Divvy 2020 Q1, a filter was applied, and it was observed that the `ride_length` column contained some values displayed as "#########". This occurred because the result of the subtraction was negative. Since the end time cannot be earlier than the start time, those rows were removed.

All columns were reviewed for blank cells. In the Divvy 2019 Q1 dataset, blank cells were found in the `gender` and `birthyear` columns. However, those rows were not deleted.

The Divvy 2020 Q1 dataset was also checked for blank cells and none were found.

This was the process carried out in R:

In the analysis carried out in Excel, I identified that some column names differed while containing the same information. For example, start_time (from Divvy 2019 Q1) was equivalent to started_at (from Divvy

2020 Q1). After collecting the data from both datasets, several columns in the Divvy 2019 Q1 file were renamed to ensure consistency with Divvy 2020 Q1.

The data frames were stacked into one big frame call ”all_trips” to continue with the analysis. Before this, the columns `ride_id` and `ride_length` were converted to character to avoid conflicts during the merge and to maintain control over data types.

Additional steps were taken to ensure that the data was clean and ready for analysis. These steps can be reviewed in the linked script below. Moreover, new variables were created to break down the date into month, day, and year components.

In the next step, the mean, median, maximum, and minimum ride durations (ride_length) were calculated for members and casual users using the aggregate() function.

The data was grouped by user type and weekday, and both the number of rides and average duration were computed. The results were ordered using the arrange() function.

For data visualization, bar charts were created using ggplot2 to compare the number of rides by user type and weekday. Another bar chart displays the average ride duration by user type and weekday.

You can check the code at this link for more details: Cyclistic Analysis Script (R).

4 Exploratory Analysis

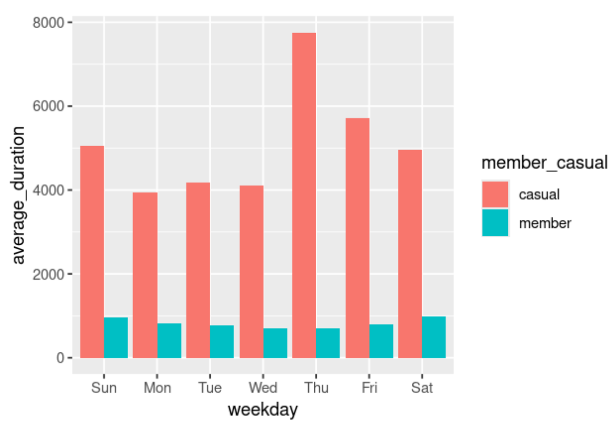

Figure 1: Average duration of trips

Key Insights:

1. Casual riders take significantly longer trips:
   - On all days of the week, casual riders (shown in salmon color) have a much higher average trip duration than members (shown in teal).
   - For example, Thursday shows the largest difference: casual riders average nearly 7800 seconds (over 2 hours), while members stay below 1000 seconds (less than 20 minutes).

2. Weekday vs. weekend usage patterns:

   - Casual riders tend to take longer trips on weekends (Sunday and Saturday) and on Thursday, suggesting more recreational or leisure use.
   - In contrast, annual members show a short and consistent trip duration throughout the week, likely using bikes for commuting or routine activities.

3. Greater variability among casual riders:

   - There is high variation in the average trip duration across days for casual riders.
   - This suggests that casual riders use the bikes in a more spontaneous way and likely for leisure purposes.

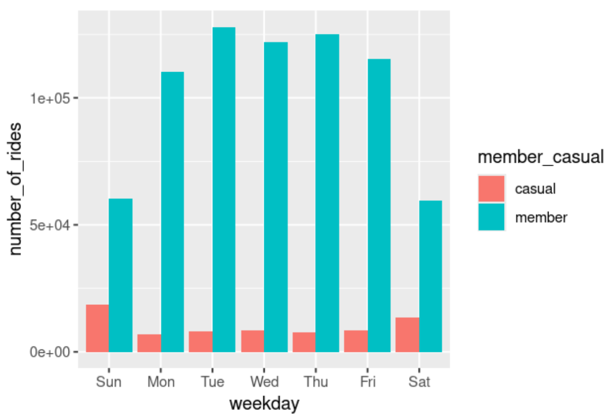

Figure 2: Total Number of Rides

Key Insights:

1. Annual members take significantly more rides than casual riders:

   - Across all weekdays, members (shown in teal) consistently take a much higher number of rides than casual riders (shown in salmon).
   - The difference is especially pronounced from Monday to Friday, suggesting members primarily use bikes for weekday commuting or regular activities.

2. Members show high and stable usage throughout the workweek:

   - The number of rides by members peaks from Tuesday to Friday, all with over 120,000 rides per day, indicating a strong routine usage pattern, likely for commuting or scheduled activities.

3. Casual riders are more active on weekends:

   - Unlike members, casual riders take more rides on weekends (Saturday and Sunday) than on weekdays.
   - This suggests recreational or leisure-focused usage, aligning with more free time or tourism on weekends.
   - Monday through Friday shows consistently lower ride counts for casual users, reinforcing that they are less likely to use the bikes for commuting.

5 Recommendations

- **Convert Casual Riders to Members**
  - Create targeted membership offers for casual riders who frequently ride on weekends (e.g., "Weekend Warrior" discounts).
  - Use in-app pop-ups or post-ride emails to promote trial memberships after multiple long trips.
- **Expand Marketing for Leisure Use**
  - Promote Cyclistic as a leisure activity on weekends through social media and tourism platforms.
  - Partner with parks, museums, and local attractions to offer bike-and-visit promotions.
- **Optimize Bike Availability for Members During Weekdays**
  - Rebalance bikes early in the morning in residential and business areas to match commuter demand.
  - Ensure docking stations near offices and transit hubs are well-stocked during peak weekday hours.
- **Create Targeted Ride Plans Based on Usage Patterns**
  - Offer custom plans such as a "Weekend Pass" for tourists or a "Commuter Pack" for frequent users.
  - Implement flexible pricing based on time of week, ride duration, or usage frequency.
- **Improve User Experience with Trip Planning Tools**
  - Add route suggestion features in the app that highlight scenic or leisure-friendly paths.
  - Provide maps or guides for longer bike routes and nearby points of interest.
- **Use Data to Forecast Demand and Plan Operations**
  - Apply predictive models to anticipate daily demand by user type.
  - Adjust maintenance schedules and staffing, e.g., more support on weekends for casual riders and early weekday support for members.

6 Next Steps

- Detect which stations attract tourists vs. commuters and assess if they are well served.
- Incorporate weather data to assess how rain, temperature, or seasonality affects ridership by type.
- Track how often casual riders return over time. Are there patterns that predict conversion to members?
- Explore if casual and member users prefer different types of bikes (e.g., classic vs. electric) and how that influences duration or frequency.
- Segment users beyond just member/casual (e.g., by frequency, distance, time of day) to create more personalized offers and plans.
- Understand how age, gender, or residence areas relate to usage patterns and tailor services accordingly.