

Bike Share Analysis

Zsolt Oláh

2025-12-21

0. Scenario

I am a junior data analyst on the marketing analytics team at 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, our team wants to understand how casual riders and annual members use the bikes differently. Using these insights, we will design a new marketing strategy to convert casual riders into annual members. Before moving forward, company executives must approve our recommendations, so they need to be supported by compelling data insights and professional data visualizations.

1. Ask

The first step of the analysis is to define the problem we are trying to solve. There is a question that should be answered by the end of the project. As mentioned above, this question is:

How do annual members and casual riders use the bikes differently?

And based on the answer, we should provide three recommendations that address the following question:

What actions can turn casual riders into annual members?

2. Prepare

The dataset is publicly available and can be accessed at this link. The data is provided in several segmented formats, and for this analysis we will download the monthly segmented files.

It is important to perform a **ROCCC analysis** on the dataset:

- **R – Reliable:** The dataset is generated by the automated electronic rental system, so it is considered reliable.
- **O – Original:** The data is collected directly by the bike-share company, making it an original source.
- **C – Comprehensive:** The dataset contains sufficient information to answer our question, such as user gender and age, ride date and time, and whether the rider is a casual user or a subscriber.
- **C – Current:** The analysis is conducted in 2025, the data is relevant because it was collected during the previous 12 month (202412-202511)
- **C – Cited:** The dataset originates from the company's own data collection.

After reviewing the dataset, it becomes clear that we should merge the monthly files into a single dataset so that we can process the data in a simple, uniform manner.

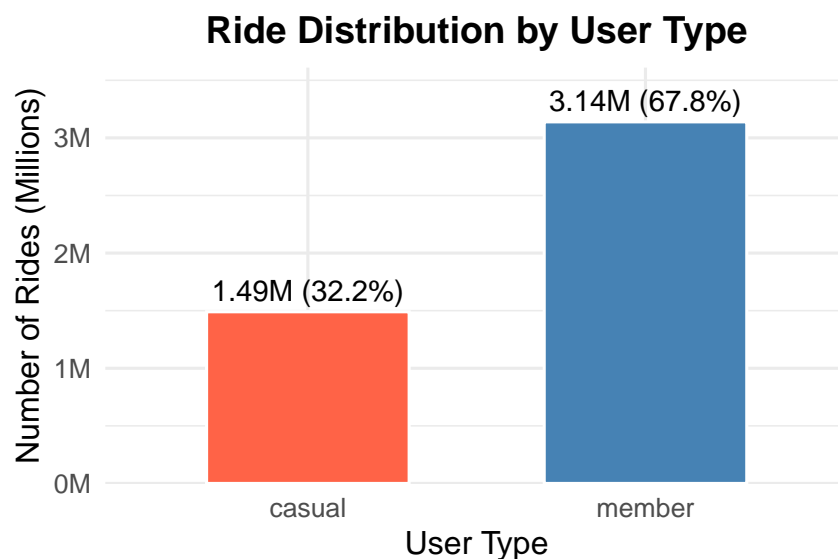
3. Process

- Load the required libraries.
- Import the downloaded dataset. Although all 12 CSV files appear to share the same structure, we enforce consistent formatting by extracting column specifications from the first file and applying them to all subsequent imports to ensure structural integrity before merging.
- Create and save a small sample dataset for external investigation (e.g., using LLMs to explore and get ideas regarding structure, cleaning, and data integrity).
- Generate several new columns derived from existing fields to provide additional analytical perspectives.
- Check for missing values, examine where they occur, and remove rows containing null values in columns essential to the analysis.
- Inspect numeric columns for outliers and remove invalid entries (e.g., negative durations or unrealistic average speeds).
- Review datetime columns for anomalies; no outliers were detected.
- Convert categorical columns to factors and inspect their levels to identify potential misspellings or inconsistencies; no issues were found.
- Check for duplicated rows: none were present.

4. Analysis and Share

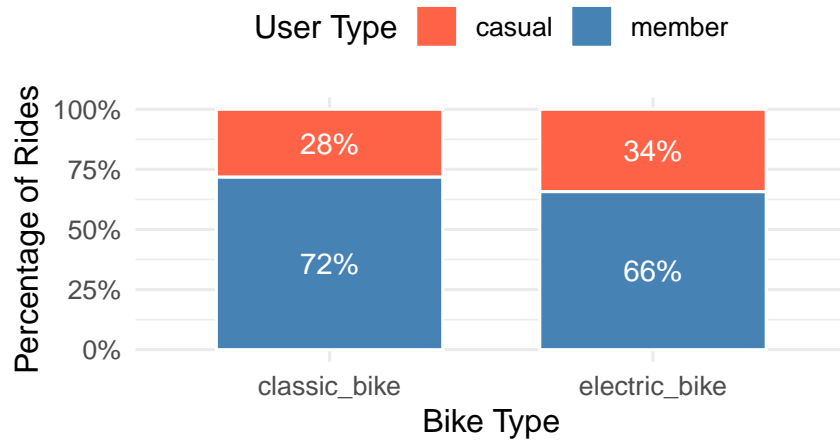
Because the ultimate goal is to convert casual riders into annual members, it is important to first evaluate whether the casual rider segment is large enough to justify targeted marketing efforts. To do this, we begin by examining the proportion of rides coming from each user group.

It can be stated confidently that the target segment is large enough to have potential to get more revenue from turning them to annual members so we proceed the analysis.



Next, we check whether there is any difference in bike types used between the two user groups. As the image shows, there is only a minimal difference: casual users tend to use electric bikes slightly more often than members.

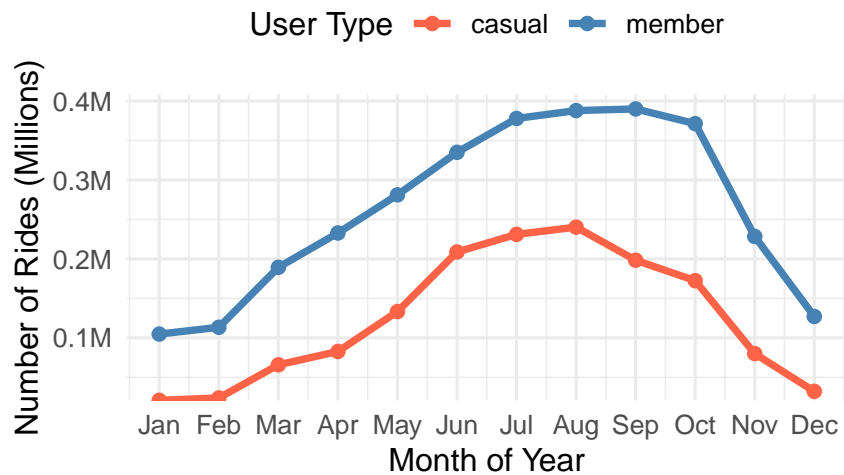
Classic vs. Electric Bike Usage by User Type



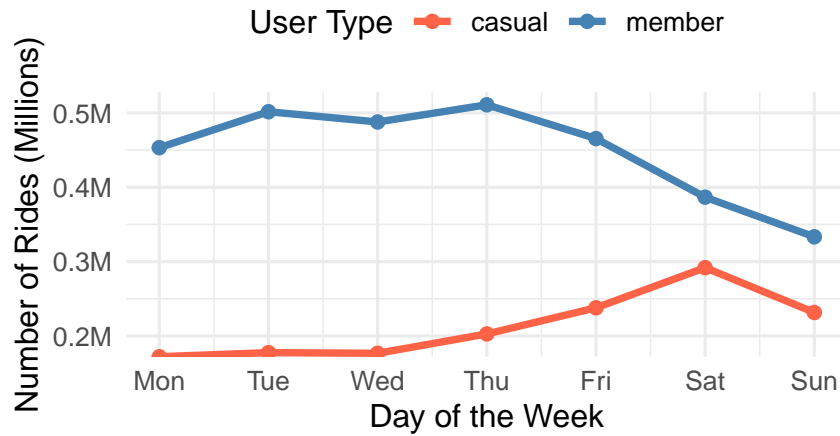
The following two charts show the usage patterns across the year and the week. The yearly pattern is quite similar for the two user groups, but the weekly trends reveal the real difference:

- **member riders** use the service more during the first four days of the week, after which usage declines toward the weekend,
- meanwhile, **casual riders** hardly use the service during the first three days of the week, and their usage increases steadily as the weekend approaches, peaking on Saturday.

Monthly Ride Usage Pattern by User Type



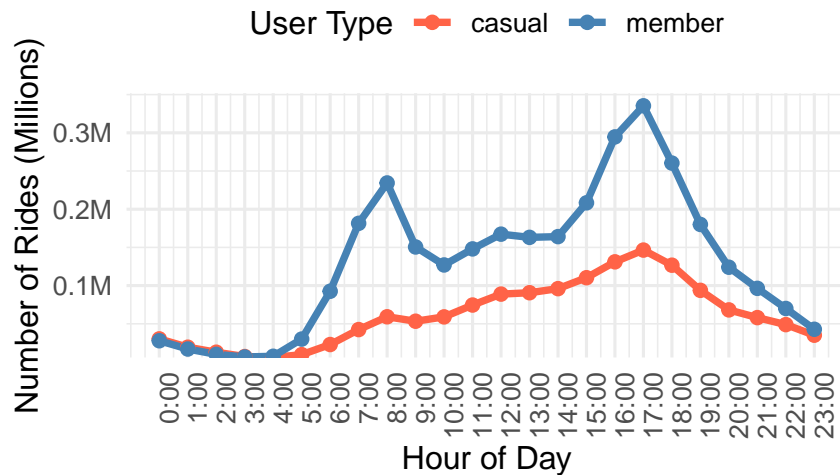
Day of the Week Ride Usage Pattern by User Type



The hourly chart shows another difference between the two user groups:

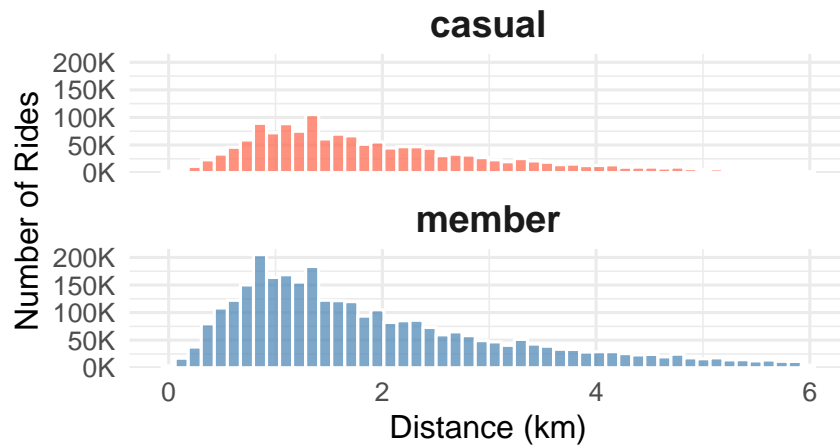
- **the usage pattern of member riders has two peaks:** the first in the morning from 7:30 to 8:30, and the second in the afternoon from 16:30 to 17:30. This clearly indicates that they use the bikes for commuting,
- **although casual riders also show these two peaks, they are far smaller than those of the members.**

Hourly Ride Usage Pattern by User Type

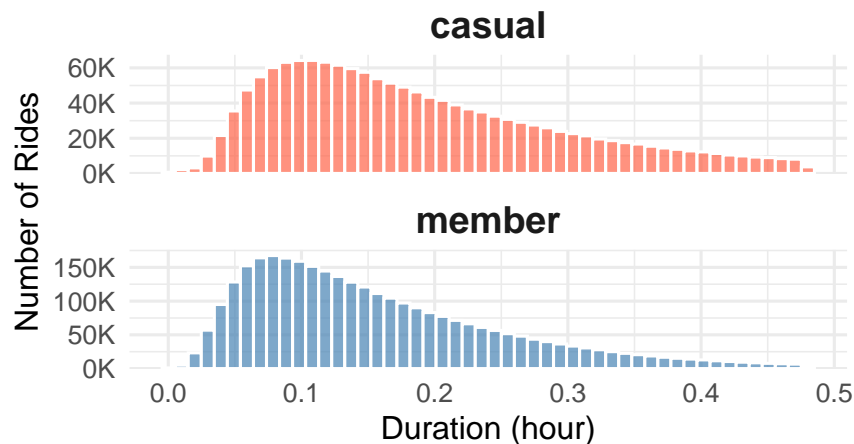


The next two charts show the distribution of ride distances and durations by user group. These visuals clearly indicate that, overall, there is no significant difference between the two patterns: distribution of both user groups are right-skewed and their peaks occur in roughly the same place.

Distribution of Ride Distances by User Type



Distribution of Ride Durations by User Type



5. Act

Based on the analysis above, these are our three recommendations for possible marketing actions to turn casual riders into annual members, focusing on the differences in their usage behavior:

- Since casual riders mainly use our service on weekends, we can develop a new yearly subscription type for weekend users. With this subscription, users would receive discounts on weekend days.
- The analysis also revealed another major usage pattern: users who ride our bikes for commuting on weekdays. For retention purposes, it is worth considering a second special subscription type for them, offering discounts on weekdays.
- We can plan a marketing campaign with a physical presence at docking stations to promote these new subscription types. To increase the efficiency of this campaign, it is worth conducting further analysis to identify whether there are preferred starting or ending docking stations, possibly segmented by the two user groups.