

# Cyclistic Bike-Share Case Study

**Author:** Stergos Kapros

**Date:** January 2026

## Tools & Technologies

For this project, I used Microsoft Excel for data cleaning and preparation, R and RStudio for data analysis and visualization, and Microsoft Word to document the results and insights.

## Ask – Business Task

Cyclistic is a bike-share company that offers both single-ride passes and annual memberships. Riders who use single or daily passes are considered casual riders, while riders with annual plans are members.

The company has identified that annual members are more profitable over time. Because of this, the marketing team wants to understand how casual riders use the bikes compared to members.

The main business question of this analysis is:

### **How do casual riders and annual members use Cyclistic bikes differently?**

By answering this question, the marketing team can design marketing strategies that encourage casual riders to become annual members.

The main stakeholders are:

- ❖ Cyclistic Executive Team
- ❖ Lily Moreno, Director of Marketing
- ❖ Cyclistic Marketing Analytics Team

As a junior data analyst, my role is to clean the data, analyze it, and communicate insights clearly.

## **Prepare – Data Description**

The data used in this case study comes from the Divvy Trip Data provided by Motivate International Inc. and published under the City of Chicago Data License. The dataset is public and anonymized.

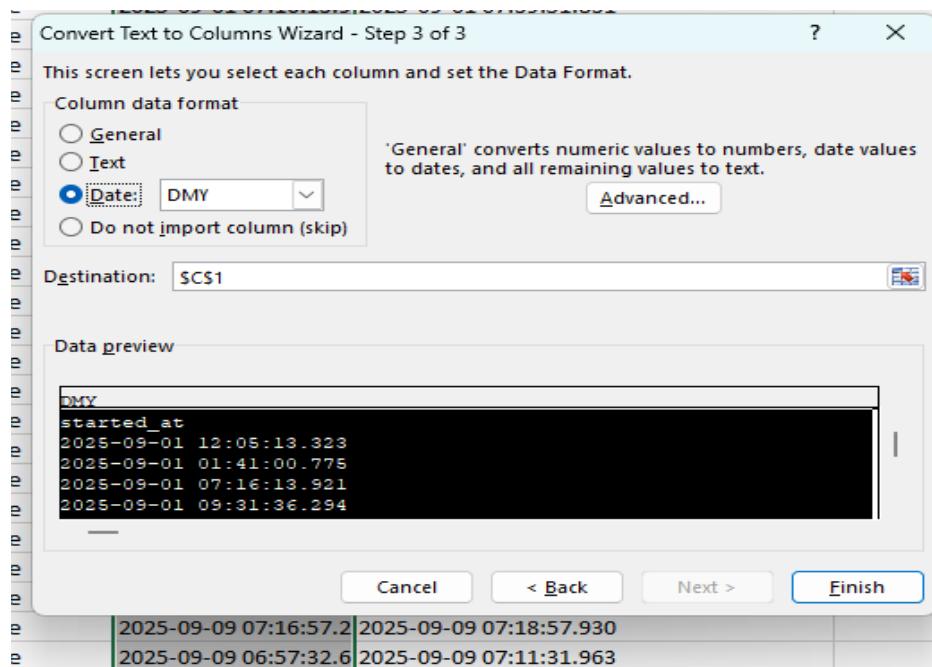
The dataset includes 12 monthly files covering the full year of 2025. Each row represents a single bike trip, with information about the ride time, stations, and rider type.

Although the dataset is public and reliable, it has some limitations. Some station information may be missing, and the data does not include user demographics.

## Process – Data Cleaning (Excel)

The data cleaning process was completed using Microsoft Excel.

First, I checked that all monthly files had the same structure and column names. Then, I converted the start and end time columns from text format to proper date and time format.



This image shows how the `started_at` and `ended_at` columns were converted from text into date and time format so calculations could be performed correctly.

Next, I calculated the duration of each ride by subtracting the start time from the end time and created a new column called `ride_length`. This made it easier to compare ride duration between casual riders and members.

At the same time, I created a `day_of_week` column to identify when each ride started. This allowed me to explore differences between weekday and weekend usage patterns. Before moving to the analysis stage, I checked summary values such as average and maximum ride length, as well as the most common day of the week, to make sure the data was realistic.

ride_length	day_of_week	Average ride length	Max ride length	Mode day_of_week
0:13:58	3	0:11:17	0:59:57	5
0:05:05	7			
0:11:36	5			
0:03:35	5			

This image shows a simple check using basic summary values to confirm the data is correct.

To ensure data quality, I checked the dataset for missing values, formatting issues, and invalid entries, such as negative or unrealistic ride durations. Any problematic records were identified and removed to keep the data accurate and consistent.

A3372D51337	electric_bike	2025-09-26 18:47:26.8	2025-09-26 18:56:43.841
A39D986C2A5	electric_bike	2025-09-22 08:14:40.2	2025-09-22 08:27:40.261
L7B73EB1B9F	electric_bike	2025-09-21 22:00:37.3	2025-09-21 22:03:37.6 Peoria St & Kinzie St CHI02098
34AC78C583C	electric_bike	2025-09-21 21:31:27.1	2025-09-21 21:48:02.386
'485ADE5443	electric_bike		
'21C1BF630D	electric_bike		
iD3FCEE64F4	electric_bike		
ACCE437B2CF	electric_bike		
09DE249710C	electric_bike		
CAE03BEACE9	electric_bike		
351FDF8E0DB	electric_bike		
39DDB5A7DDE	electric_bike		
?6B624332F9	electric_bike		
X0410B471C2	electric_bike		
05EF87942C7	electric_bike	2025-09-15 07:40:43.7	2025-09-15 07:47:45.705
3C2E0D3A4E	electric_bike	2025-09-15 08:28:35.9	2025-09-15 08:35:23.203
B79A9AB5ECA	electric_bike	2025-09-09 04:43:26.4	2025-09-09 04:56:07.668
15C7C9127A3	electric_bike	2025-09-09 07:16:57.3	2025-09-09 07:19:57.920

This image shows how missing or problematic values were identified and handled during cleaning.

Finally, I reviewed the data using sorting and pivot tables to double-check that the numbers looked right and that there were no unexpected values.

Average ride length per user type		Average ride length per day_of_week & user type							Count of rides per day_of_week & user type														
Row Labels	Average of ride_length	Row Labels	Average of ride_length	Column Labels	1	2	3	4	5	6	7	Grand Total	Row Labels	Column Labels	1	2	3	4	5	6	7	Grand Total	
casual	0:16:59	casual	0:24:00	0:15:09	0:14:47	0:14:44	0:16:13	0:16:14	0:20:28	0:16:59			casual		2287	3384	4931	4591	3763	4210	4914	28080	
member	0:11:53	member	0:12:51	0:11:41	0:13:38	0:10:59	0:11:04	0:11:23	0:11:41	0:11:53			member		7103	17572	22622	20960	16216	15684	12297	112454	
Grand Total	0:12:54	Grand Total	0:15:34	0:12:15	0:13:50	0:11:39	0:12:02	0:12:25	0:14:12	0:12:54			Grand Total		9390	20956	27553	25551	19979	19894	17211	140534	

This image shows pivot tables used to verify the accuracy and consistency of the cleaned data.

After cleaning, all files were saved in a separate clean data folder and were ready for analysis.

202501-divvy-tripdata clean	20/1/2026 1:56 μμ	Microsoft Excel Worksheet	24.648 KB
202502-divvy-tripdata clean	20/1/2026 1:56 μμ	Microsoft Excel Worksheet	26.253 KB
202503-divvy-tripdata clean	20/1/2026 3:19 μμ	Microsoft Excel Worksheet	50.361 KB
202504-divvy-tripdata clean	20/1/2026 2:06 μμ	Microsoft Excel Worksheet	65.386 KB
202505-divvy-tripdata clean	20/1/2026 2:11 μμ	Microsoft Excel Worksheet	87.937 KB
202506-divvy-tripdata clean	20/1/2026 2:15 μμ	Microsoft Excel Worksheet	116.579 KB
202507-divvy-tripdata clean	20/1/2026 2:19 μμ	Microsoft Excel Worksheet	131.239 KB
202508-divvy-tripdata clean	20/1/2026 2:24 μμ	Microsoft Excel Worksheet	131.321 KB
202509-divvy-tripdata clean	20/1/2026 3:24 μμ	Microsoft Excel Worksheet	119.503 KB
202510-divvy-tripdata clean	20/1/2026 2:33 μμ	Microsoft Excel Worksheet	108.371 KB
202511-divvy-tripdata clean	20/1/2026 2:36 μμ	Microsoft Excel Worksheet	60.695 KB
202512-divvy-tripdata clean	20/1/2026 2:39 μμ	Microsoft Excel Worksheet	24.617 KB

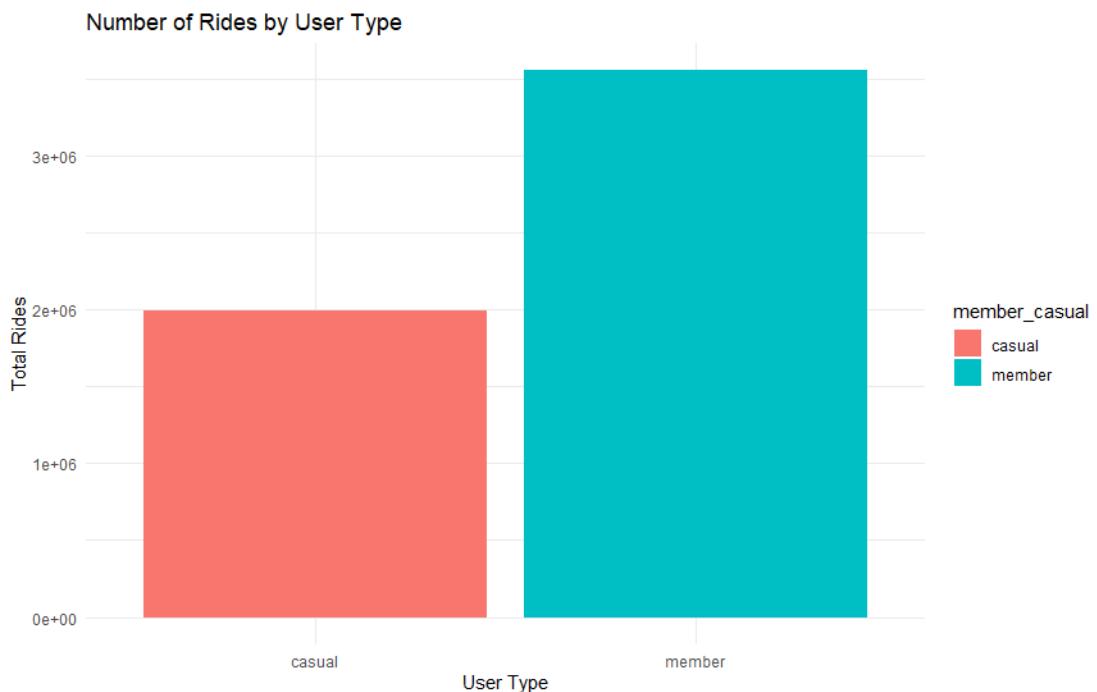
## Data Cleaning Summary

- ❖ Before cleaning, the dataset included all bike trips from 12 monthly files.
- ❖ During cleaning, trips with zero or negative ride length were removed.
- ❖ Trips longer than 24 hours were also removed to avoid unrealistic values.
- ❖ Rows with missing or invalid timestamps were excluded.
- ❖ The final dataset contains only valid bike trips and was used for analysis in R.

## Analyze – Data Analysis (R)

After finishing the cleaning in Excel, I imported all 12 files into R and combined them into one dataset. I added time-based columns, such as month and day of week, to help with the analysis.

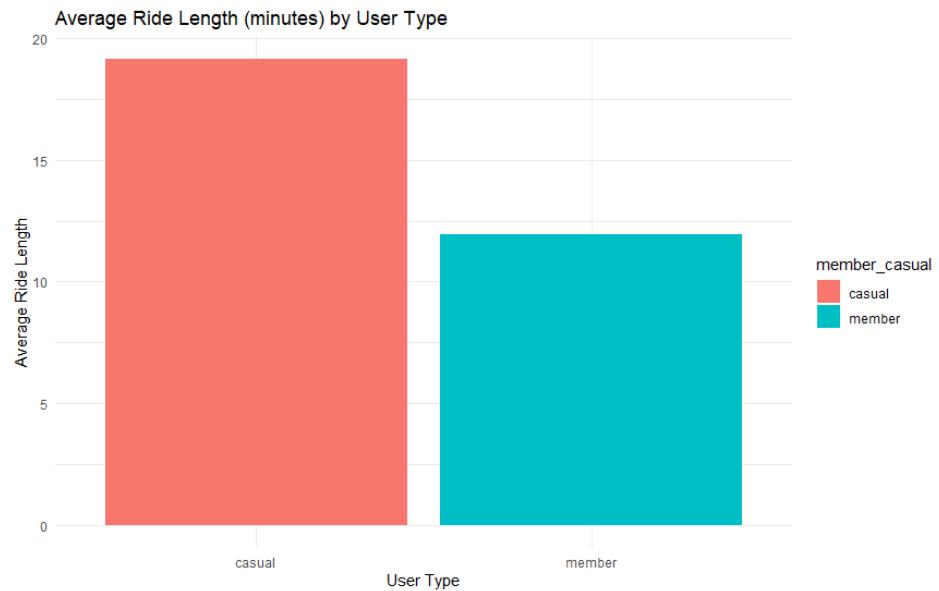
### ❖ Rides by User Type



This chart compares the total number of rides taken by casual riders and annual members.

**Members completed more rides overall, which shows that they use the service more frequently than casual riders.**

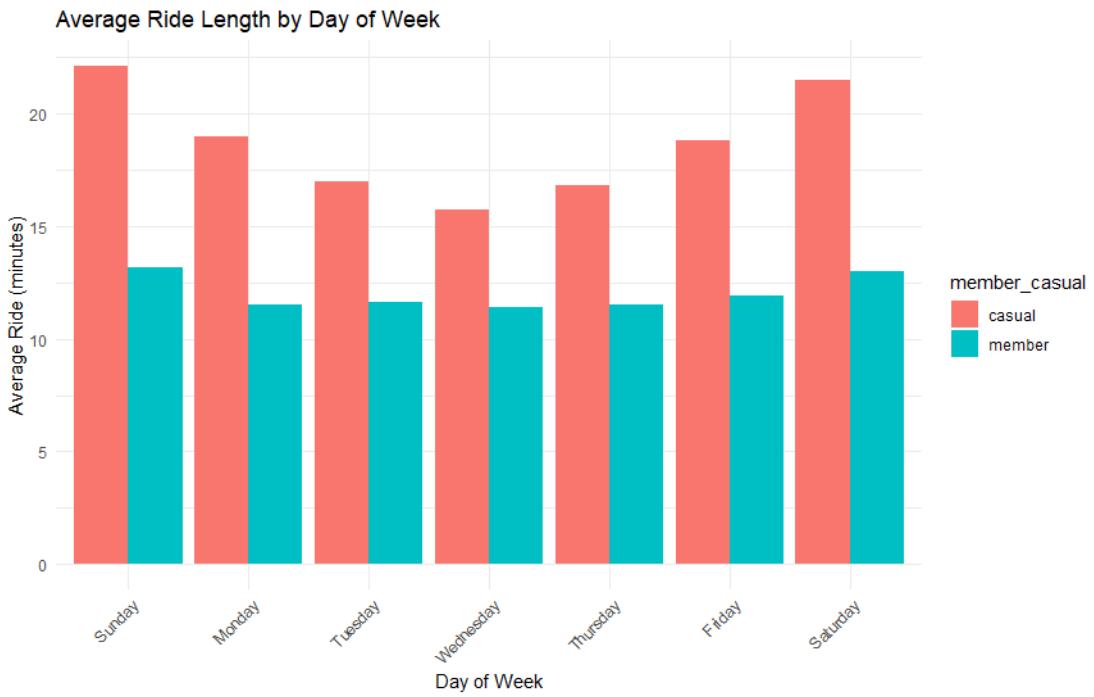
### ❖ Average Ride Length by User Type



*This chart compares the average ride duration between casual riders and members.*

**Casual riders have longer average ride durations, while members tend to take shorter trips. This suggests that casual riders often use bikes for leisure, while members use them more for practical trips.**

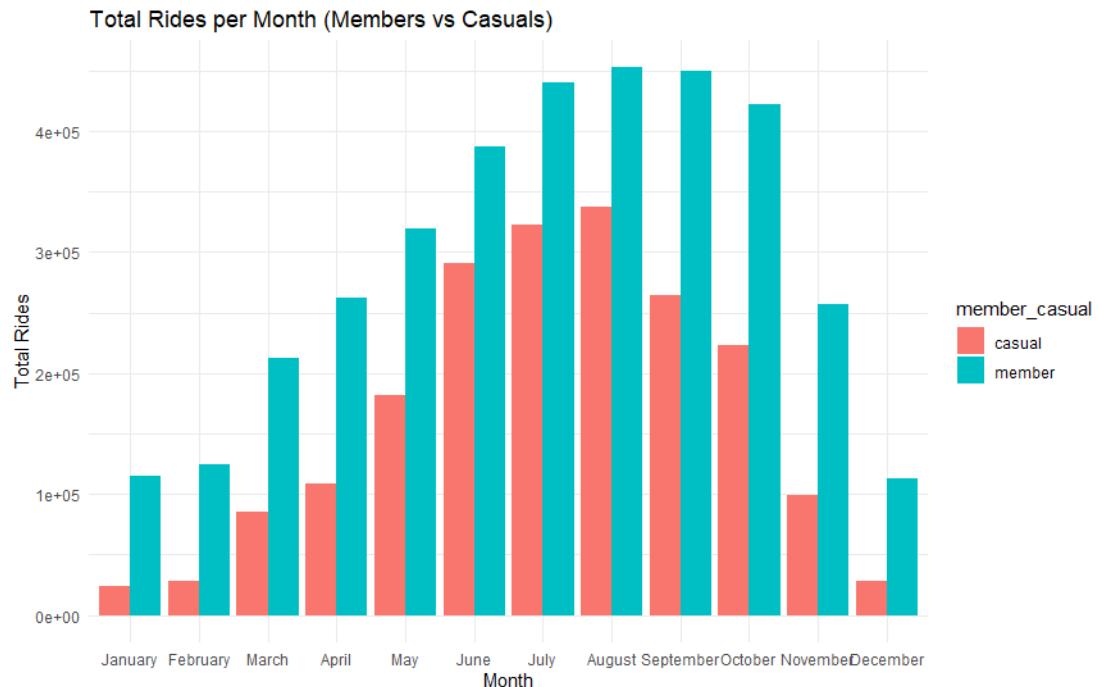
### ❖ Average Ride Length by Day of Week



This chart shows how average ride duration changes throughout the week for each user type.

**Casual riders usually take longer trips on weekends, while members ride for similar lengths throughout the week.**

### ❖ Total Rides by Month



*This chart shows how ride volume changes throughout the year for casual riders and members.*

**Both groups ride more in the summer, but casual riders show much bigger changes between winter and summer compared to members.**

## **Share – Visual Communication**

The visualizations created in R help communicate the results clearly. They make it easy to compare rider behavior between casual riders and members and allow stakeholders to quickly understand key trends without technical knowledge.

## **Act – Recommendations**

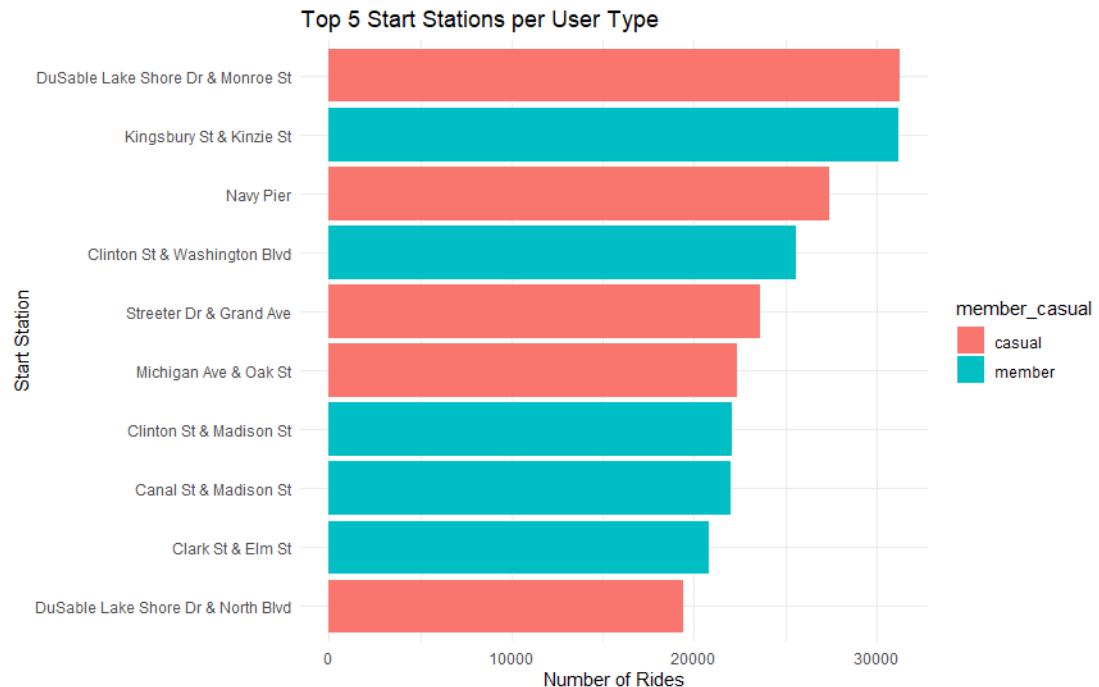
Based on the analysis, the following actions are recommended:

- ❖ Focus on casual riders during busy times, especially weekends and summer.
- ❖ Show clearly why an annual membership is a better option than single or daily passes.
- ❖ Promote memberships while riders are using the bikes, through the app or at stations.

***The impact of these actions could be measured by tracking how many casual riders become annual members.***

## Extra Analysis (Top Start Stations)

This analysis was not part of the original Google Data Analytics case study. I added it to further explore the data.



*This chart shows the most common start stations for casual riders and members.*

Casual riders and members do not use the same stations equally.  
Understanding these differences can help Cyclistic better understand how each user group uses the service.

## Conclusion

This case study shows clear differences between casual riders and annual members. Members ride more frequently but for shorter durations, while casual riders take longer, leisure focused trips.

By understanding these patterns, Cyclistic can design more effective marketing strategies that convert casual riders into annual members and support long-term business growth.