Cyclistic Bike Share Case Study:
How does a bike-share navigate speedy success?

Project Objective: The objective of this project is to analyze how casual riders and annual members use Cyclistic bikes differently. The goal is to provide data-driven insights that support the design of a targeted marketing strategy aimed at converting casual riders into annual members, thereby increasing the company's revenue and long-term customer base.

Ask Phase Summary

Central Problem: Understand how annual members and casual riders differ in their usage of Cyclistic bikes.

Objective: Analyze usage data to identify behavioral differences between the two user groups.

Data Source: Historical trip data provided by Cyclistic.

Key Questions:

- How do annual members use the bikes in terms of volume, time of day, and seasonality?
- How does this differ from casual riders across the same variables?

Prepare Phase Summary

First Actions

- I downloaded all trip data for the previous 12 months (from **June 2024 to May 2025**).
- I used **Excel** to perform the initial data cleaning and exploration steps.
- Each dataset contains **13 columns**, including ride start/end times, start/end stations, and rider type (`member_casual`).

ROCCC Principles Check:

- **Reliable**: Data is provided by **Motivate International Inc.**, operator of the Divvy system.
- **Original**: Primary source data.
- **Comprehensive**: Covers 12 months of trips, enough to capture seasonal and behavioral patterns.
- **Current**: Most recent 12 months of data.
- **Cited**: Public dataset with terms of use available on Divvy's official website.

Bias and Credibility Considerations:

- The data is **anonymized** (no names, addresses, or payment info), which protects privacy but limits behavioral analysis.
- There is **no demographic data** (e.g., age, gender, income), preventing demographic insights.

Data Cleaning Issues:

- The `started_at` and `ended_at` fields contained **milliseconds**, which caused errors when formatting as date/time in Excel.
- I resolved this by removing the last **4 characters** (the dot and 3 digits) using:
  =LEFT(D2, LEN(D2) - 4)
- I then replaced the original columns with cleaned values.

Ride Duration Validation:

- I created a new column, `trip_duration`, by subtracting `started_at` from `ended_at` to check for invalid rides.
- Most files were clean, but in **November 2024**, I found **43 rides with negative durations**.
- I paused to review the project guidelines for how to handle these anomalies.

Process Phase Summary

- During this phase, I identified the need to calculate ride duration using a column named `ride_length`.
- I renamed the previously created `trip_duration` column to `ride_length` across all datasets.
- I also created a new column called `day_of_week` to indicate the weekday of each ride.
  I used the following Excel formula:
  =WEEKDAY(C2, "dddd")
- I continued using **Excel** as my main tool for this phase due to its power, documentation, and my familiarity with it.
- A key part of the processing phase is checking for errors. Since I had previously identified issues, I handled them here:
  *I **removed 43 rows with negative ride durations*** from the November 2024 dataset.
- With the data cleaned and key columns prepared, I was confident to proceed to the analysis phase.

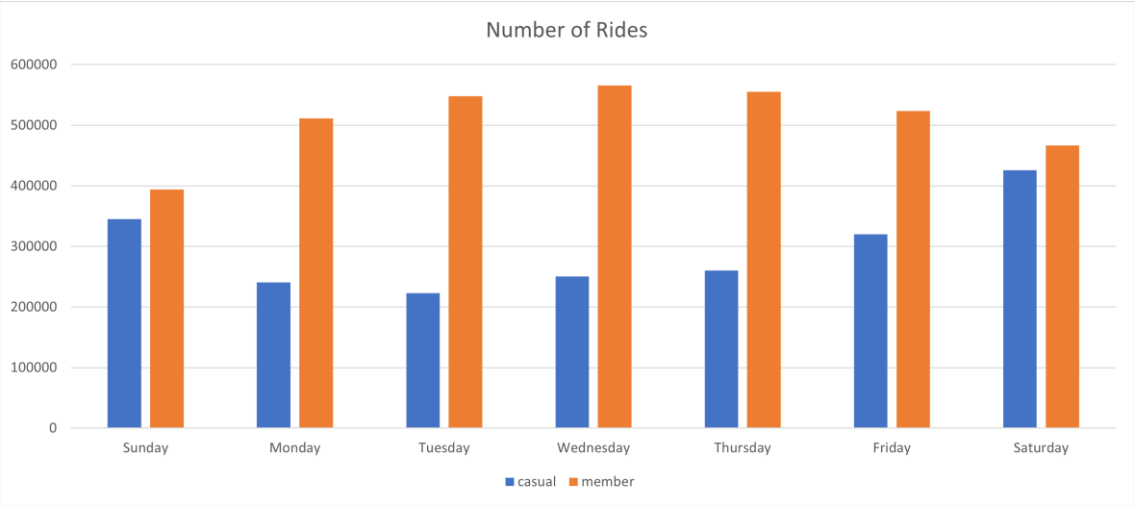
Analyze Phase Summary

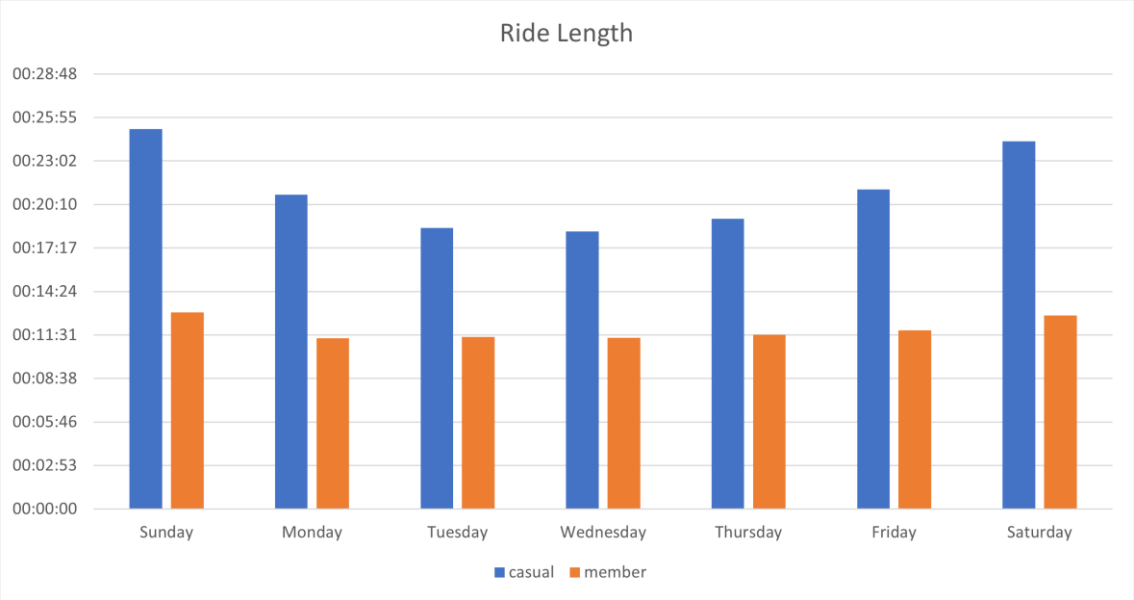
- During the analysis phase, I was asked to perform key calculations. A major task involved aggregating all data for deeper insights. However, due to Excel's row limitations (around 1,048,000 rows), I was unable to combine the full dataset in a single table using Excel alone.
- To work around this, I calculated the required values separately in each of the 12 monthly worksheets using **Pivot Tables**, and then **manually aggregated** the results into a summary sheet.
- Initially, the `day_of_week` column displayed numerical values (1–7), which made interpretation less intuitive. To improve readability, I used Excel's **Find and Replace** tool to convert numbers into actual weekday names (e.g., Monday, Tuesday).
- I then created a new worksheet called **"Dynamics"**, where I inserted **two Pivot Tables per month**:
  - The **first** Pivot Table calculated the **average ride length**, with `member_casual` as rows and `day_of_week` as columns.
  - The **second** Pivot Table showed the **total number of rides**, with the same structure.
- This setup allowed me to analyze and compare user behavior across both rider types and days of the week more effectively.

Share Phase Summary

Preparing Visualizations

- To generate visualizations, I first recreated summary tables based on the results from the Pivot Tables.
  I used Excel formulas such as `AVERAGE()` and `SUM()` to aggregate values across the 12 monthly sheets, for example:
  =AVERAGE(B4;B12;B20;...)
  =SUM(L4;L12;L20;...)
- These tables allowed me to calculate the overall average ride length and total number of rides for each day of the week, by rider type.
- Using these summary tables, I created **bar charts** to visually represent the two metrics:
  - One chart for **average ride duration**
  - One chart for **number of rides**





Key insights observed:

- **Annual members** show consistent usage across the weekdays, with higher ride volumes during workdays.
  This suggests that members primarily use the service for **commuting** (e.g., work or school).
- **Casual riders** have noticeably higher usage on **weekends**, and their ride length and frequency both increase starting on **Fridays**.
  This indicates more **leisure-oriented usage**.

Act Phase Summary

Key conclusions based on the analysis:

1. **Annual members take more rides overall**, including on weekends.
   However, **casual riders tend to have longer ride durations**, which could inform pricing strategies — such as offering time-based discounts to encourage more frequent use among members.
2. Since annual members primarily use bikes on **weekdays**, this pattern suggests they rely on Cyclistic for **commuting** (e.g., to work or school).
   This insight could shape **marketing campaigns** targeting casual riders by highlighting the benefits of using bikes for daily transportation.
3. Casual rider activity is lowest between **Monday and Thursday**, both in **ride count** and **ride duration** (especially Tuesday through Thursday).
   These days could be leveraged for **special promotions or member incentives**, supported by targeted marketing efforts to increase engagement during off-peak periods.

Wrap-Up Summary

Completing this first case study was a valuable and challenging experience. Although the project provided a general structure and instructions, I encountered some technical difficulties along the way.

Initially, I planned to use **Google Sheets** as my main platform, but due to the large volume of data — likely larger than originally expected when the case study was designed — it became impractical to work within that environment.

As a result, I switched to **Microsoft Excel** and used it throughout all phases of the project. For my first projects, I intentionally chose to work only with spreadsheets to deepen my understanding of formulas, functions, and spreadsheet-based workflows.

However, as mentioned earlier, the dataset size exceeded Excel's limits, especially when trying to combine all 12 months into a single table. To overcome this, I decided to calculate and summarize the data separately using **Pivot Tables**, and then manually aggregate the final results.
While this approach allowed me to complete the analysis, it also limited deeper exploration and cross-month insights that would have enriched the study.

In the end, I believe the best approach for future projects of this scale would be to **combine tools** — using **Excel** for initial data exploration and cleaning, and **R or SQL** for aggregation, analysis, and data visualization at scale.