Business Task

For this project I will be working with Divvy Bikes user data to locate trends between the consumer data. I will be focusing on the differences between how members and casual riders use Cyclistic bikes. The end goal of this analysis is to have three data driven business recommendations aimed to influence casual riders to become members.

Data Sources Used

I used the Divvy trip data from 2023 to conduct my analysis. These files are organized by month and stored as wide data. They contain information on specific bike trips that took place in the year of 2023. This information includes start and end times, dates, and locations.

Cleaning and Manipulating the Data

Before we complete our analysis we need to clean up the data to make analysis possible and ensure confidence in the results. This includes: importing the data into R, removing blank spaces, checking the data types, and removing unnecessary data.

I have decided to use R to process and analyze the data.

Loading Libraries

First, we install and load libraries that will allow us to work with the data.

Package	Description
janitor	For cleaning
lubridate	To manipulate time/date data
skimr	For viewing data
tidyverse	For data transformation

Cleaning Steps

Setting the Working Directory

After loading the datasets into R, we begin by setting the working directory to ensure that file paths are referenced correctly.

Initial Data Exploration

At this point we are prepared to clean the data. To start, we take a look at our data using skim_without_charts which gives us some general information about our data including data types, min/max/average for numeric values, and how many entries there are for each variable.

Checking Column Compatibility for Joining

Since the data is split across 12 months, we need to ensure that the datasets are compatible for merging. We use the compare_df_cols_same() function to verify that all columns across datasets have consistent names and data types. In this case, the function returned TRUE, confirming that the datasets can be safely combined

Combining the Data

Now that we have ensured compatibility between all 12 datasets, we use bind_rows() to merge them into a single dataset titled bike_data. This gives us a unified dataset to begin working with that includes all trip information for the year.

Verifying Dataset

After combining the data, we run the summary() function to review the combined dataset. This function helps us check that the merging was successful by displaying summary statistics (like count, min, max, and mean) for each column.

Checking for Duplicates

To ensure data accuracy, we use the get_dupes() function to identify any duplicate entries in the ride_id column. Since each ride should be unique, any duplicates could skew the analysis. Removing or addressing duplicates is critical at this stage.

Creating New Columns for Analysis

Next, we add several new columns that will be useful for our analysis. First, we add columns for day, month, year, and day of week, which we extract from the start time. Then we use a simple function to calculate the length of the rides in seconds, using the ended_at and started_at columns. Now we remove unnecessary columns by transforming the dataset into one that only contains the type of bike, start time, rider type (member or casual rider,) year, month, day, day of week, and ride length.

Analysis Summary

The analysis was conducted using R for data manipulation and calculation, along with the ggplot2 package for visualization. The steps taken helped identify key differences between how members and casual riders use Cyclistic bikes. I created three key visualizations, which will be discussed later in this report.

The ggplot2 package was especially useful for creating custom visualizations, as it allowed me to incorporate various elements and customize the charts to fit my needs. While I initially struggled with ggplot2, this project provided valuable learning experiences, and I now have a stronger understanding of how to create effective visualizations using the package.

Key Calculations

The following calculations were conducted in R to compare the behavior of members and casual riders.

	Members	**Members and Casual Riders**	**Casual Riders**
Mean Ride Length	750.7899	1090.159	1693.472

- Overall Mean Ride Length: 1090.159 seconds (approximately 18 minutes)
- Mean Ride Length for Members: 750.7899 seconds (approximately 12.5 minutes)
- Mean Ride Length for Casual Riders: 1693.472 seconds (approximately 28 minutes)

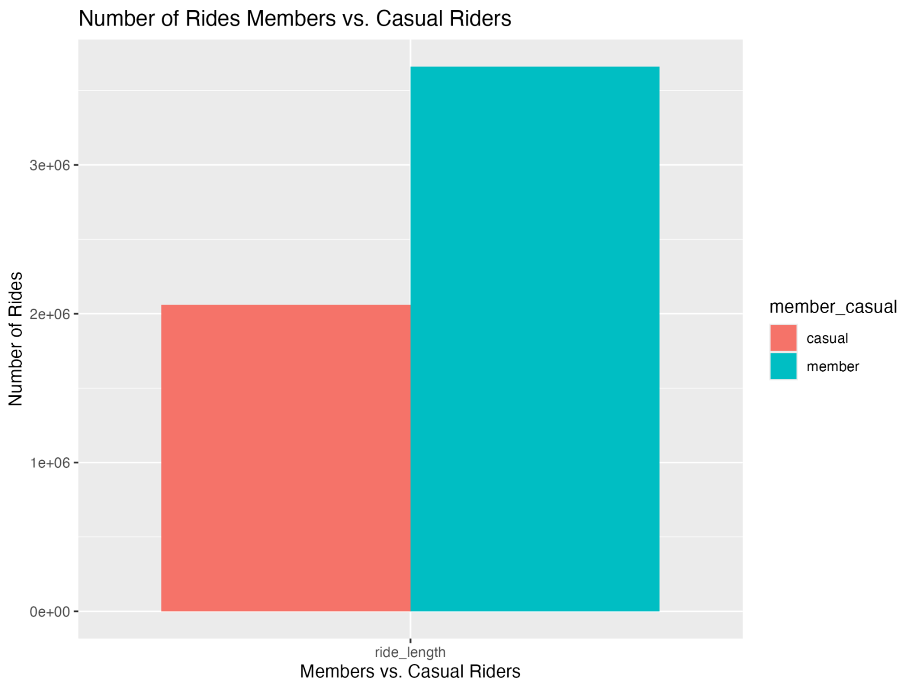
These results show that casual riders take significantly longer trips than members, which aligns with the hypothesis that casual riders are more likely to use the bikes for leisure, whereas members may use them for more frequent but shorter trips, possibly for commuting.

Transition to Tableau

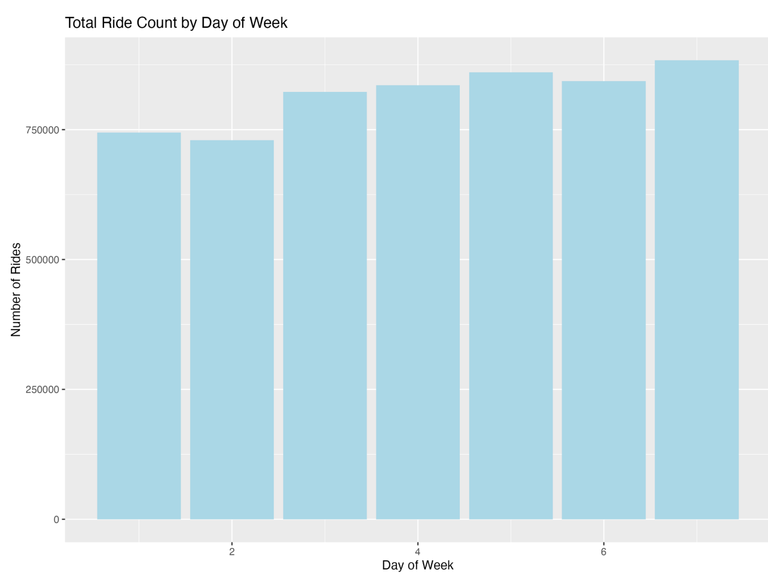
After completing the initial calculations and creating visualizations in R, I exported the cleaned dataset and imported it into Tableau Public. This allowed me to create additional visualizations to explore and communicate trends in the data more effectively. The visualizations created in Tableau helped me to confirm existing trends and uncover new trends.

Visualizations and Key Findings

I did three visualizations in R, and several more in Tableau Public. Let’s first discuss the visualizations that I did in R using ggplot2.

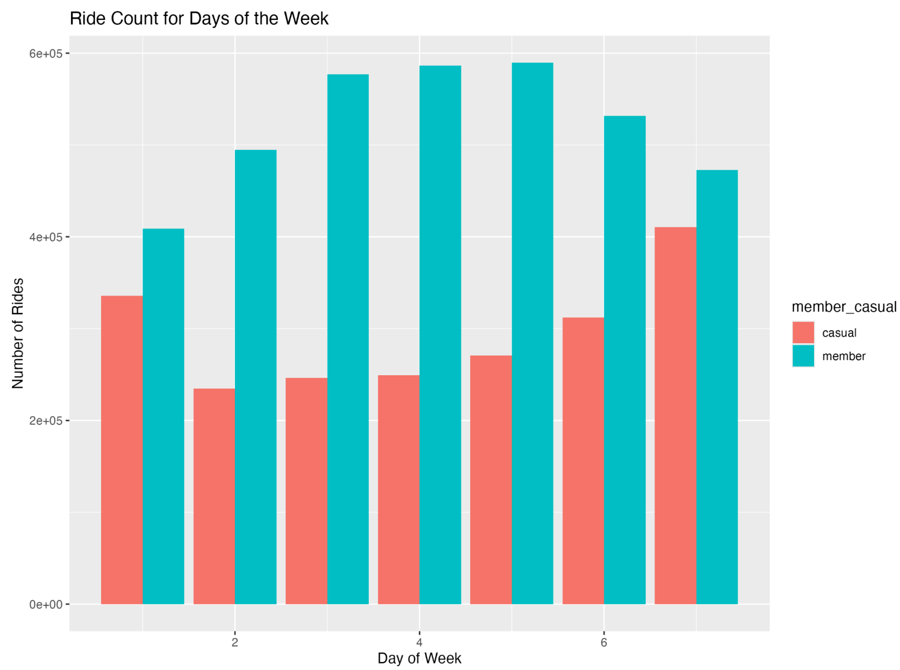


Here is the first visual I created, comparing the number of rides done by members and casual riders to compare how many trips are being made by each rider type. As you can see, the members did almost twice as many rides as casual riders in 2023. The fact that there are so many rides done by members suggests that we need to target those that don't ride as often and influence them to become members. We can do this by advertising the pricing benefits that come with being a member.

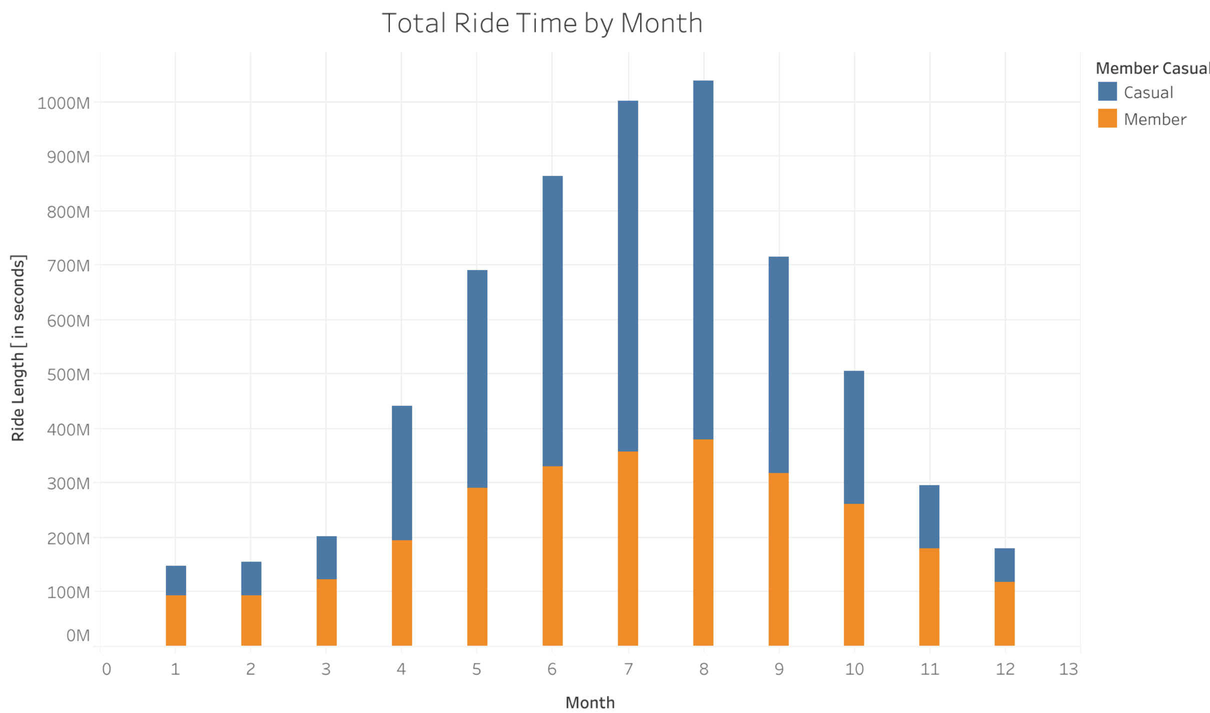


For my second visual, I decided to check the total number of rides sorted by days of the week to check for any trends that may have happened as the weeks went on. The way that R sorts days of the week is by starting on Sunday, so this graph starts on Sunday and ends on Saturday.

There isn't much insight to be gained from this other than to say, each day of the week has similar numbers.

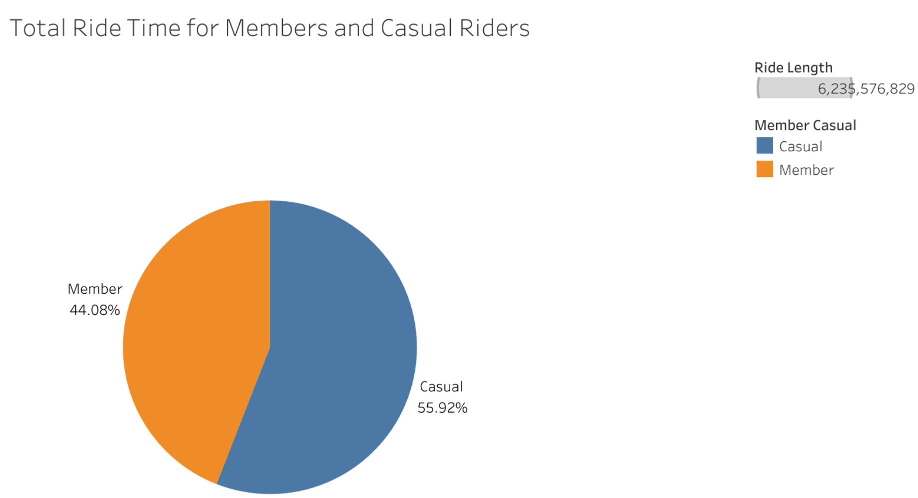


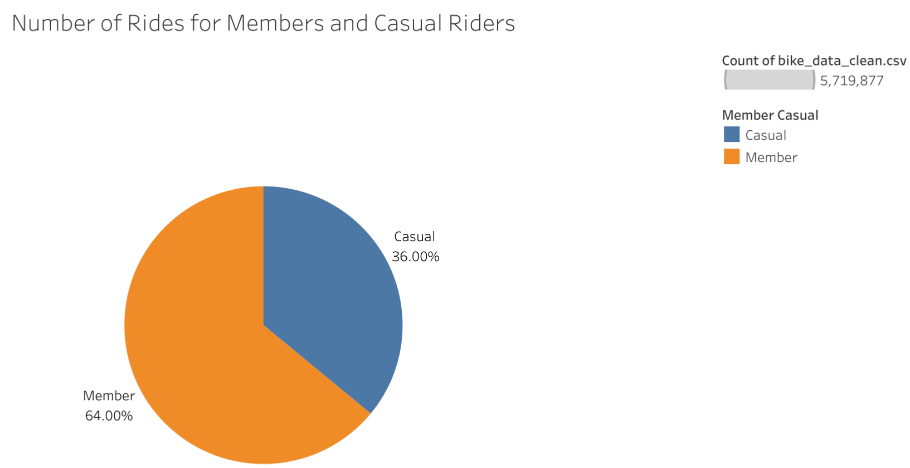
For my final visualization done in R, I decided to again compare the number of rides to days of the week, but this time I filtered the results between members and casual riders. I did this to see if there were any differences in the way each rider types rides throughout the week. The graph shows us that casual riders have the most rides recorded on Saturday and Sunday, with the numbers slowly increasing as the work week goes on. While the casual riders have their busiest days on the weekend, members use the service most during the workweek. The fact that these graphs are almost inverse of each other proves that casual riders and members are using the bikes differently than each other. If we can get more casual riders using the bikes like members are, we could successfully convert more riders into members since customers save money with a membership if they have to use the service a lot.



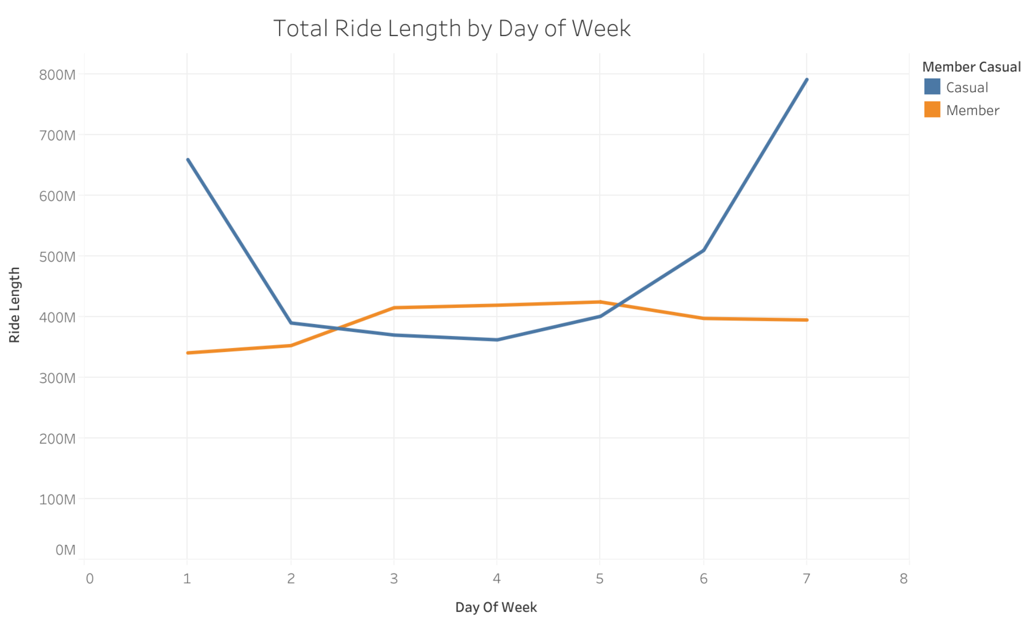
Moving on to my Tableau Public visualizations, I started by looking at the total ride time by each month, sorted by members and casual riders. I did this with the intent of comparing it to the number of rides per month. I discovered that in the cold months of October through March, members have slightly more ride time than casual riders. However, from April through September, casual riders have more ride time than members, with the numbers peaking in August. The increased number of rides during the summer months make them best for advertising to casual riders, as we can see that this is when they are using the service most.

Next I made a very similar graph, except this time I was comparing the count of bike trips by each month sorted by members and casual riders to see how it compared to the graph of total ride time. This is telling us that for every single month, members go on more rides than casual riders, with there being especially more members than casual riders riding during the coldest months. This shows us that the riders that take many trips are members. This presents an opportunity to compel casual riders that use the service a lot to become members. We can target these groups by running ads that explain the savings benefits available to the consumer.

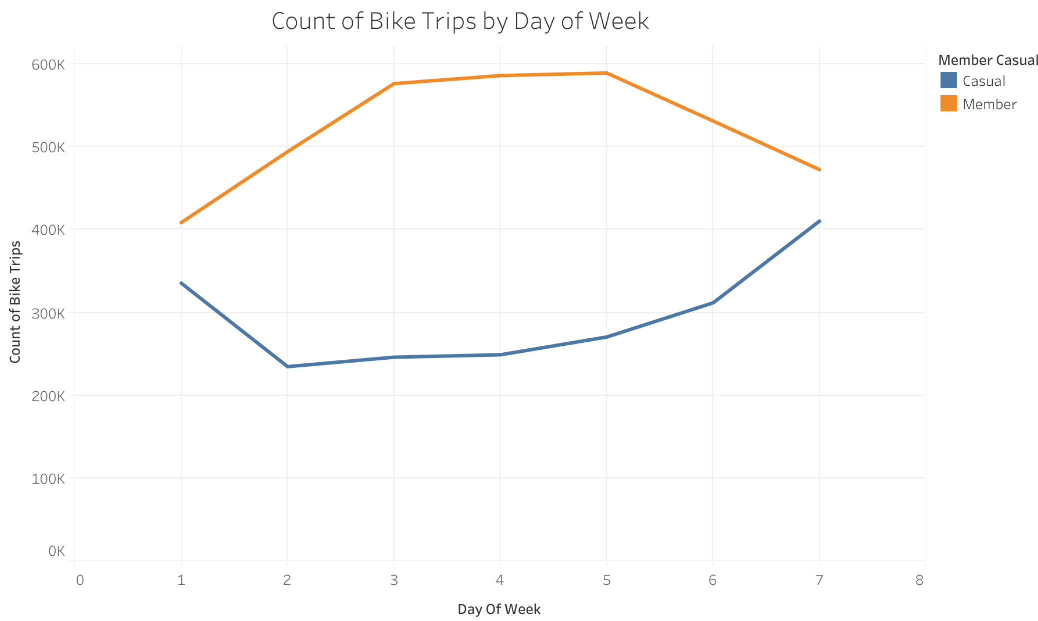




Moving on from bar graphs, I made two pie charts to look at the behavioral differences between rider types. The first pie chart compares the total ride time of members and casual riders. The second chart compares the number of rides between members and casual riders. These charts showed that casual riders have more ride time, but members have a greater number of rides. The greater ride length for casual riders makes it clear that casual riders are taking longer trips on average than members. We can use the pricing benefits of becoming a member to persuade casual riders to become members.



The final two charts that I made were line charts. I made them to gain insight about how each member type is using the service throughout the week. Above, I compared the total ride length of members and casual riders based on the days of the week. Similarly to R, Tableau organizes days of week starting on Sunday. This chart shows opposite trends for members and casual riders. Casual riders have the most ride time on the weekends, and the members have more ride time on the weekdays. The high numbers of total ride time on the weekend suggest that running digital adverts near bike check-in locations on the weekends would be the most effective.



My final chart compared the count of bike trips, sorted by day of week just like the last chart. I made this visualization to ensure that it looked similar to the last one. Here we can see again that the members have significantly more rides than casual riders. The high frequency of trips made by members suggests an opportunity to convert casual riders to members by promoting the use of Cyclistic bikes for commuting.

Overall, these visualizations highlight a clear distinction between casual and member riders: members tend to ride more frequently and for shorter periods, likely for commuting, while casual riders favor longer, leisurely rides on weekends. These insights form the basis for targeted strategies to convert casual riders into members.

Top Three Data-Driven Recommendations

1. Members primarily ride during the weekdays.
• Insight: Members likely use the bikes for commuting to work or school.

• Action: To increase membership, target casual riders by promoting the convenience and cost savings of using Cyclistic bikes for their daily commutes. Highlight the time and money saved compared to other commuting options.

2. Casual riders mostly ride on weekends.
• Insight: Casual riders likely use the bikes for leisure activities on weekends when they have more free time. This is supported by the longer ride durations compared to members.
• Action: Convert casual riders to members by offering pricing plans that appeal to their leisure habits. For example, run targeted ads at bike check-in locations that emphasize the benefits of unlimited 45-minute rides and discounted rates for e-bikes and scooters for members.

3. Casual riders take longer rides than members.
• Insight: Casual riders may appreciate the convenience and ease of e-bikes for longer trips.
• Action: Encourage casual riders to become members by informing them that an annual membership reduces e-bike rates to $0.18/min (matching the cost of casual classic bike rentals) and includes free unlocks. Use in-app notifications or ads at bike stations to highlight this cost-saving opportunity.