Table of Contents

Chicago Cyclistic Case Study for Google Analytics Professional Certificate

Project Content: This document, titled *"Analysis Description,"* a PowerPoint presentation, two R Markdown files (for cleaning & analysis, runnable on Kaggle [link]), their HTML versions for easy reference, an Excel file with its PDF, and more are all available via the [Project Content] link on GitHub.

Introduction

As a junior aspiring Data Analyst, I have completed a Capstone project for the Google Analytics Professional Certificate. I performed extensive data analysis for a fictional bike-share company, Cyclistic, to help them attract more customers. In this scenario, I was part of Cyclistic's marketing analytics team and worked to help my company achieve its business goals.

Actual data from the City of Chicago's Divvy bicycle-sharing service has been provided by Lyft Bikes and Scooters, LLC (Bikeshare) to use for educational purposes. According to Wikipedia

> "Divvy is the bicycle sharing system in the Chicago metropolitan area, currently serving the cities of Chicago and Evanston. The Chicago Department of Transportation owns the system, and Lyft has operated it since 2019. As of Sept 2021, Divvy operated 16,500 bicycles and over 800 stations, covering 190 square miles."

I used Cyclistic's historical trip data (from Divvy) to analyze trends from 2023. Because Cyclistic is a fictional company, I was restricted from authentic communication with the executive team. As a result, I made several assumptions within this analysis that were not included in the data analysis scenario provided by the Certificate course. These assumptions are noted within the report.

The Divvy bike-sharing company actively provided services throughout the project, so I checked their website and prices and even listened to bloggers' comments for real-time information.

This document presents six phases of data analysis: Ask, Prepare, Process, Analyze, Share, and Act. It links to supporting documents from R Markdown, Excel, and PowerPoint.

About Cyclistic from the course material

Cyclistic is a successful bike-share company based in Chicago. Users can unlock bikes from one station and lock them back into any other station within the system. Cyclistic users are more likely to ride for leisure, but about 30% use bikes to commute to work each day. The majority of riders opt for traditional bikes; about 8% of riders use assistive options.

Customers who purchase single-ride or full-day passes are referred to as Casual riders. Customers who buy annual memberships are Cyclistic Members.

Cyclistic's finance analysts have concluded that annual members are significantly more profitable than casual riders, making the company's future success dependent on maximizing the number of annual memberships. The marketing director, Lily Moreno, believes a solid opportunity exists to convert Casual riders into Members.

Data Discrepancies

Some differences were present between the scenario description provided by the Certificate course and the Divvy data. According to the 2023 dataset for analysis, their website, and my phone call to customer service, Divvy's bicycle-sharing service does not offer reclining bikes, hand tricycles, and cargo bikes, most likely considered assistive options, as stated in the course material. Additionally, the description from the course material states Cyclistic has 692 docking stations. However, Divvy data shows 964 unique docking station names in January 2023 and 1100 in December 2023. Possibly, the number of unique station names can be more significant due to some typos in their names. However, the course material information was likely from an older 2016 data set. Overall, these numbers are good evidence of a well-growing business.

Cyclistic's data included three types of bikes: classic, electric, and docked bikes. The "docked bike" type was a mystery because the Divvy website and customer service stated that they do not have "docked bikes." The analysis phase provides an in-depth look into the "docked" bike type.

Ask Phase

Business Goal and Task

This analysis aimed to find marketing strategies for increasing the number of annual memberships by convincing Casual riders to become Cyclistic Members.

This task was completed by finding differences and trends in bike use between Members and Casual riders regarding the number of trips, trip duration, and price. Additionally, an investigation was conducted into how digital media can help marketing tactics.

Stakeholders

Lily Moreno: The director of marketing and my manager.

Cyclistic's Executive Team: The executive team will decide whether to approve the recommended marketing program.

Question To Be Answered:

1. What distinguishes annual Members' and Casual rider's bike trip behaviors?
2. What marketing strategy will convince Casual riders to buy an annual membership?
3. Discover how digital media can help in marketing tactics.

Prepare Phase

Data Origin & Credibility

Provided by course material: [Link to Divvy datasets]

The course material allows access to many datasets, such as CSV files from 2013 to the previous month from the current date. I started this project in April 2024 and found that the last available dataset was March 2024. New data is continuously being added. I chose the data from January to December 2023 because it is easier to see trends by looking at the year's data.

A note from the course material:

> "The datasets have a different name because Cyclistic is a fictional company. For this case study, the datasets are appropriate and will enable you to answer the business questions. Motivate International Inc. has made the data available under this license (see link below). You can use this public data to explore how different customer types use Cyclistic's bikes."
>
> [License]

After observing and checking the Data License Agreement, I can state that the data appeared reliable, original, comprehensive, current, and appropriately cited. It does not have any personal or credit card information.

Data Limitations

The data doesn't include any demographic information or type of bike ride (Single Ride or Day Pass), nor does it give the actual number of annual Members or Casual riders. Therefore, it is impossible to check, for example, how often, on average, one rider takes a bike trip in a week and compare this between all kinds of bike users or to find the exact proportion between permanent and occasional Casual users. This data could help filter the data more accurately to make better conclusions and clarify how to reach business goals and tasks in the future.

Process Phase

The data cleaning performed in R Markdown can be found [here] (GitHub link), making viewing the R codes in the HTML file easy. You can access the R file through the [Project Content] link and check how it runs on Kaggle [link].

The Process Phase section doesn't include explanations or purposes for each R code. Instead, it describes the data cleaning process, from downloading 12 months of raw data from an online source to saving the combined and cleaned dataset as an—RData (R data) file.

The tools used for data cleaning & analysis were RStudio Desktop and Microsoft Excel.

Note: To differentiate charts and tables in the Process Phase from those in the Analysis Phase, I have named them differently: **Tibbles** for tables and **Plots** for charts in the Process Phase. In the Analysis Phase, they are referred to as **Tables** and **Figures**.

Appropriate file-naming conventions were used to organize and name all project folders and subfolders to ensure consistency and ease of navigation. Specifically, files used for the GitHub repository were combined in the same folder, while other files were grouped in the "Bike Project Start Files" folder. These conventions applied to various file types, including:

- .docx (Microsoft Word Document) and .xlsx (Excel spreadsheets)
- .Rmd (R Markdown documents) and .RData (R data files),
- .pptx (PowerPoint presentations) and .pdf (Portable Document Format files)

Using functions for data cleanliness observation and evaluation points.

I employed a structured approach to observing the data cleanliness and drawing conclusions. This involved systematically examining the dataset using specific functions as needed for different data observations. Specifically, I utilized the following functions and provided corresponding findings about the extent of data cleanliness in R Markdown after executing them:

1. readr::read_csv(): Reads the CSV file into a data frame, ensuring accurate data import.
   - Verified the initial structure and content of the dataset.
   - Confirmed correct data import with appropriate column names and data types.

2. head(): Displayed the first few rows of the data frame.
   - Checked for any anomalies or inconsistencies in the initial rows.
   - Observed that the data appeared consistent and well-structured in the sample preview.

3. glimpse(): Used `glimpse()` to get a compact, transposed view of the data frame.
   - Examined the overall structure, including column names, data types, and sample values.
   - The `glimpse()` function provided a quick overview, reaffirming the data's integrity.

4. skim_without_charts(): Generates summary statistics without visual charts.
   - Assessed descriptive statistics for each column's mean, median, and missing values.
   - The detailed summary statistics helped identify potential data quality issues, ensuring the dataset was ready for further analysis.

The streamlined approach to handling multiple .zip files directly from online sources.

Specifically, it:

1. Defines the base URL and file names for 12 monthly datasets:

```r
base_url <- "https://divvy-tripdata.s3.amazonaws.com/"
file_names <- paste0(202301:202312, "-divvy-tripdata.zip")
```

2. Initializes structures to manage data dynamically.
3. Implements a loop to:
   a. Download .zip files from the provided URL into memory;
   b. Extract .csv files without saving intermediate files locally;
   c. Read and store the extracted data frames in the R environment for data cleaning and analysis.
4. After the initial data observation from the column specifications returned by the readr::read_csv() function, I found that each month's data is structured similarly with 13 columns, and the column types are correct for each month.
5. The 12 months of datasets were combined into one data frame using the bind_rows() function.

Beginning of the Modification and Data Cleaning

I will not analyze based on bike station locations, so I did not select the four columns with longitude and latitude.
Note: Modification and Data Cleaning steps numbering format will remain consistent as 1), 2), 3), 4), and so on, even with long explanations between steps.

1) Created an initial data frame with nine columns relevant to data cleaning.
2) A column 'ride_length' was added to represent the trip duration in minutes. It is calculated from the differences between the 'ended_at' and 'started_at' columns and rounded to two decimal places.
3) Added the ‘day_of_week’ column from ‘started_at.’
4) Checked the detailed statistics, including missing values, values distribution, and overall dataset cleanliness, by using the glimpse() and skim_without_charts() functions:

Positive Observations:

- Number of rows = n_unique for ride_id column - no duplicates;
- Same number of characters (16) in ride_id column;
- Whitespace and empty cells: 0 in each column;
- Column names: Correct;
- Character length variation in column ‘rideable_type’ and four station-related columns (names and IDs) is normal for this data type;
- The ‘ride_length’ column is in the “difftime” (<dttm> - duration) format, which is correct.

Data Issues to Address:

1. Completeness: missing values in four columns with station names and their IDs.
2. The ‘ride_length’ column has outliers in p0 (minimum: -16656.52 min) and p100 (maximum: 98,489.07 min) trip durations.

Explanation and Approach to Fixing the Two Issues:

The missing values in the four station-related columns have two reasons: specific problems with classic bikes and because electric bikes can be parked in two different ways. Specifically:

a. The problems with classic bikes: According to [globalaffairs.org](), over 200 bikes are stolen in Chicago every month. Next, based on Divvy customer reviews, I hypothesized that some long-duration rides are due to improper parking. Some customers stated that riders are overcharged if a bike is parked improperly because it is treated like a bike that was not returned. On the other hand, extremely short trips are due to problems with bike unlocking, canceling the rides, or because some bikes can be broken.
b. Electric bike parking rules: According to the Divvy Bikes website, “you can either dock or lock your e-bike (not both at once). Dock at any Divvy station or use the cable to lock at any e-station or the 500+ Divvy-approved public bike racks for no additional cost”. So, it is typical for electric bikes that were not docked to have missing values in the “start & end_stations_name” columns.

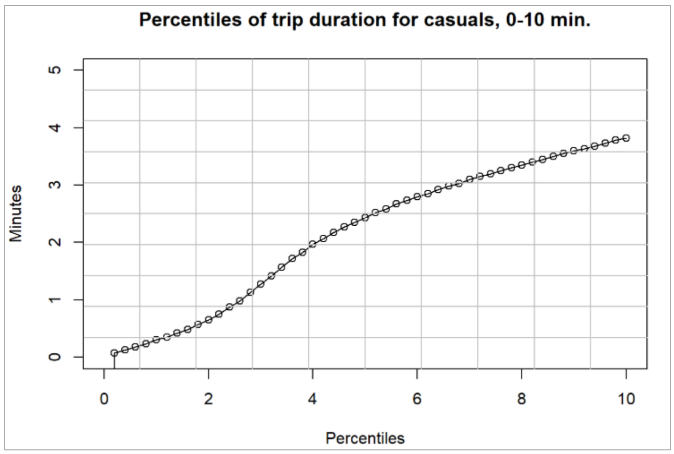
Two steps for fixing the two data cleanliness issues:

1. Delete all rows with missing values in four columns with station names and their IDs, except if ‘rideable_type == electric_bike,’ and check the number of rows removed.
2. Identify and apply a reasonable trip duration trim range for further data Analysis Phase.

Modification and further Investigation for a Reasonable Trip Duration Range.

5) Using the R function `filter()`, I deleted all rows with missing values in four columns with station names and their IDs, except if ‘rideable_type = electric_bike.’ The number of rows removed was 7314.

6) Created a data frame with six key columns to serve as the foundation for all subsequent data cleaning and analysis processes.
7) Evaluated shorter ride lengths for trip durations by analyzing the low-end quantiles of the distribution percentiles after completing the following steps:
   a. Created a chart with trip durations of 0-10 minutes in percentiles for Casuals for general visualization.



Plot 1. *Trip duration 0-10 minutes in percentiles for Casuals*

   b. Created 'Tibble 1' for the trip duration in the 2nd to 5th percentile range and 'Tibble 2' in the 97th to 100th percentile range (both for Casuals'):

##	Percentile	Minutes
## 2%	0.020	0.65
## 2.5%	0.025	0.93
## 3%	0.030	1.27
## 3.5%	0.035	1.65
## 4%	0.040	1.97
## 4.5%	0.045	2.22
## 5%	0.050	2.43

Tibble 1 *in the 2nd to 5th percentile range*

##	Percentile	Minutes
## 97%	0.970	80.73
## 97.5%	0.975	88.17
## 98%	0.980	97.85
## 98.5%	0.985	111.38
## 99%	0.990	132.92
## 99.5%	0.995	176.01
## 100%	1.000	12136.30

Tibble 2 *in the 97th to 100th percentile range*

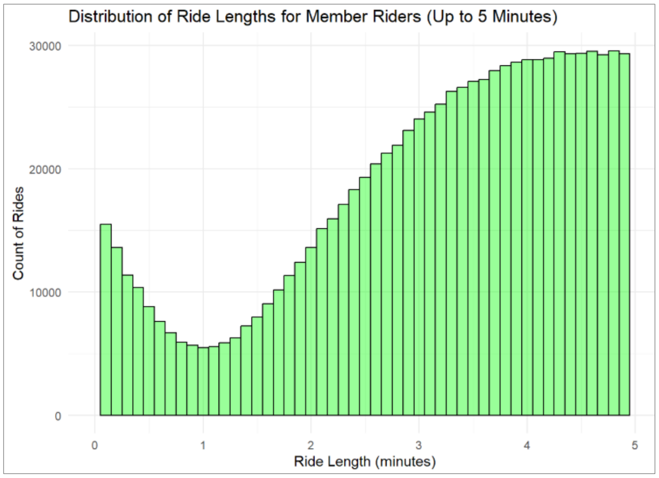
1. Insights from the Percentile Plot 1 and Tibbles 1 and 2.

- **Steady Cumulative Distribution (Linear Relationship):**
  The nearly straight line on the chart indicates that trips under 2 minutes form a consistent proportion of the dataset. Trimming this range removes a predictable portion without drastically changing the distribution.

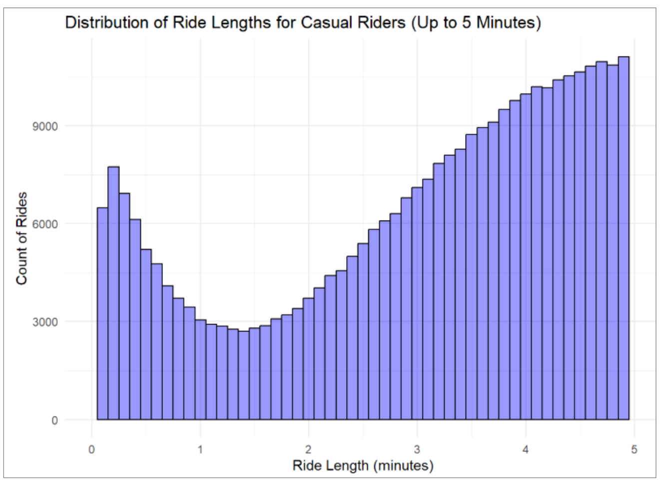
- **Focus on Meaningful Percentiles**:
  Tibble 1: Short trips, such as those up to the 4th percentile, often represent non-representative behaviors (e.g., accidental unlocks or test rides). Removing these ensures the analysis centers on meaningful durations reflecting genuine usage patterns.

- **Tibble 2:** 99.5% of trips are under 180 minutes (3 hours), while 100% include clear outliers like 12,136.3 minutes. Removing trips exceeding 12 hours ensures the analysis focuses on practical durations that accurately represent typical usage patterns.

8) Created histograms for Members' and Casuals' ride durations separately for trim duration range analysis. The x-axis represents Ride Length (in minutes), the y-axis shows the Count of Rides, and the bin width is set to 0.1 minutes.



Plot 2: *Histogram showing Members' ride durations with 0.1-minute bins, covering 0 to 5 minutes.*



Plot 3: *Histogram showing Casuals' ride durations with 0.1-minute bins, covering 0 to 5 minutes.*

2. Insights from the Histograms

Members and Casuals Histograms show similar "cap" patterns, indicating that accidental unlocks, system errors, or improperly started trips are common issues for both groups. Therefore, the following insights focus on Casuals' Histogram:

- **Unusually High First Bins (Casuals Outliers):**
  The drop from ~7,700 trips in the 2nd bin (centered at 0.2 minutes) to ~2,900 trips in the 14th bin

(centered at 1.4 minutes) highlights anomalies like accidental unlocks or system errors. These records don't reflect genuine trips and could skew metrics such as average trip duration.

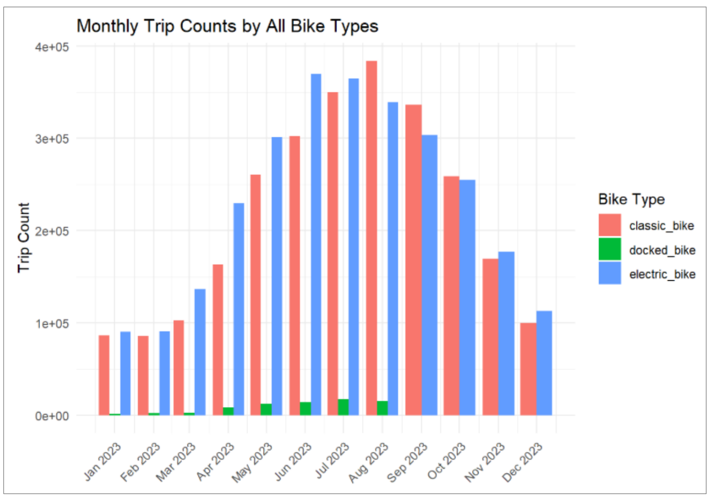
- **"Deep Wave" Suggests Unusual Behavior**:
  The dip around the 1.4-minute mark indicates that very short trips are inconsistent, suggesting they may result from atypical behavior (e.g., trial rides or test usage). The 2-minute cutoff effectively captures the diminishing relevance of shorter trips while filtering out the noise from the initial spike.

3. Combining insights from both charts:

- This discrepancy between the smooth trend of the cumulative chart and the "wave" pattern of the histogram highlights how aggregated data can mask detailed variability, underscoring the importance of analyzing data from multiple perspectives.

Conclusion: The duration trim range of 2 minutes to 12 hours (720 minutes) is considered reasonable.

9) Filtered the dataset to retain rows where ride_length is between 2 minutes and 12 hours.
   - Number of rows removed: 266067.
10) Created Plot 4 for visualizing trip counts to identify other potential outliers.
11) Created Tibble 3: 'Missing Bike Types in Monthly Data.'



Plot 4: *Visualizing Trip Counts to Identify Other Potential Outliers*

```
## # A tibble: 4 × 2
##   month      rideable_type
##   <date>     <chr>
## 1 2023-09-01 docked_bike
## 2 2023-10-01 docked_bike
## 3 2023-11-01 docked_bike
## 4 2023-12-01 docked_bike
```

Tibble 3: *Missing Bike Types in Monthly Data*

a. Plot 4 reveals that the trip count for docked bikes is negligible compared to other bike types and looks like it is missing from September to December 2023
b. Checked the presence of each rideable_type in each month and printed Tibble 3, which lists the Missing Bike Types in Monthly Data.

Conclusion: In four months, the 'docked_bike' is missing from September to December 2023.

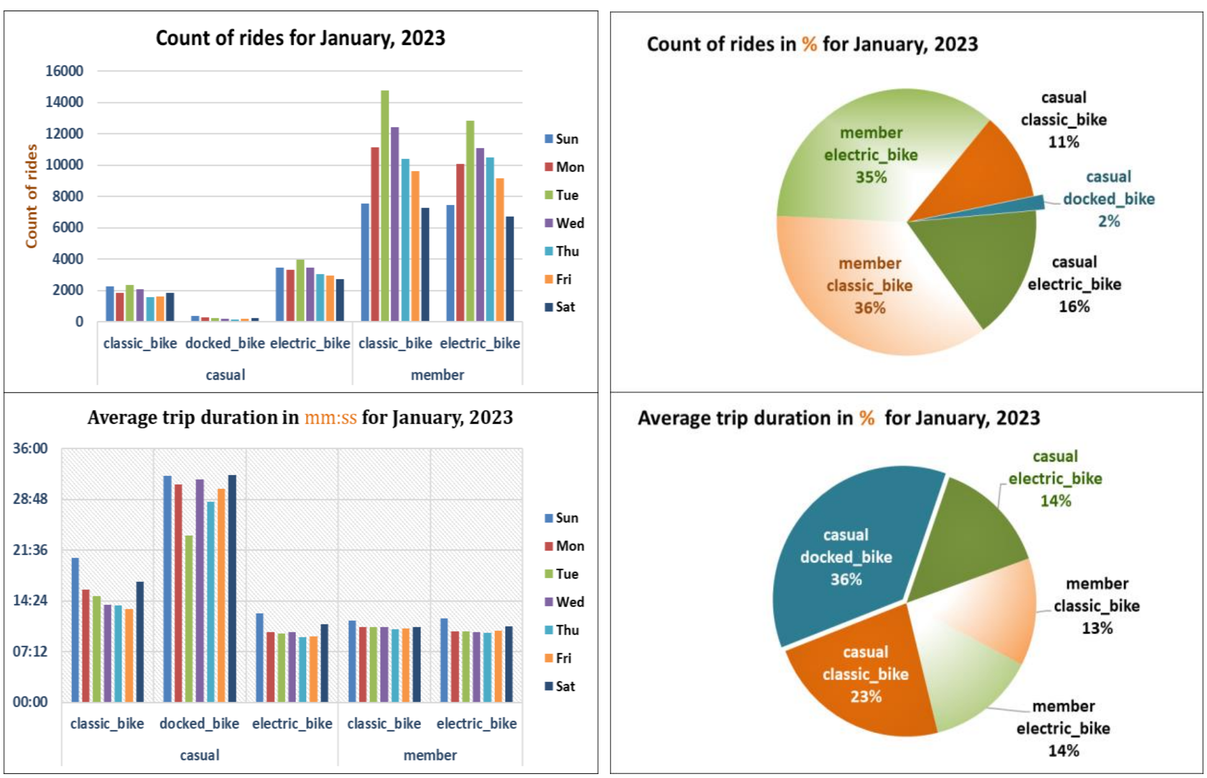
Investigation of Docked Bike Behavior Exclusively in Microsoft Excel:

- The January, April, and August historical datasets were each uploaded into separate Excel workbooks.
- Data cleaning was performed after evaluations of the data cleanliness in this R Markdown.
- Deleted all rows with missing values in four columns with station names and their IDs, except if 'rideable_type == electric_bike.'
- Created a columns 'ride_length' and 'day_of_week'.
- Trimmed trip duration outliers to within the range of 2 minutes to 12 hours.

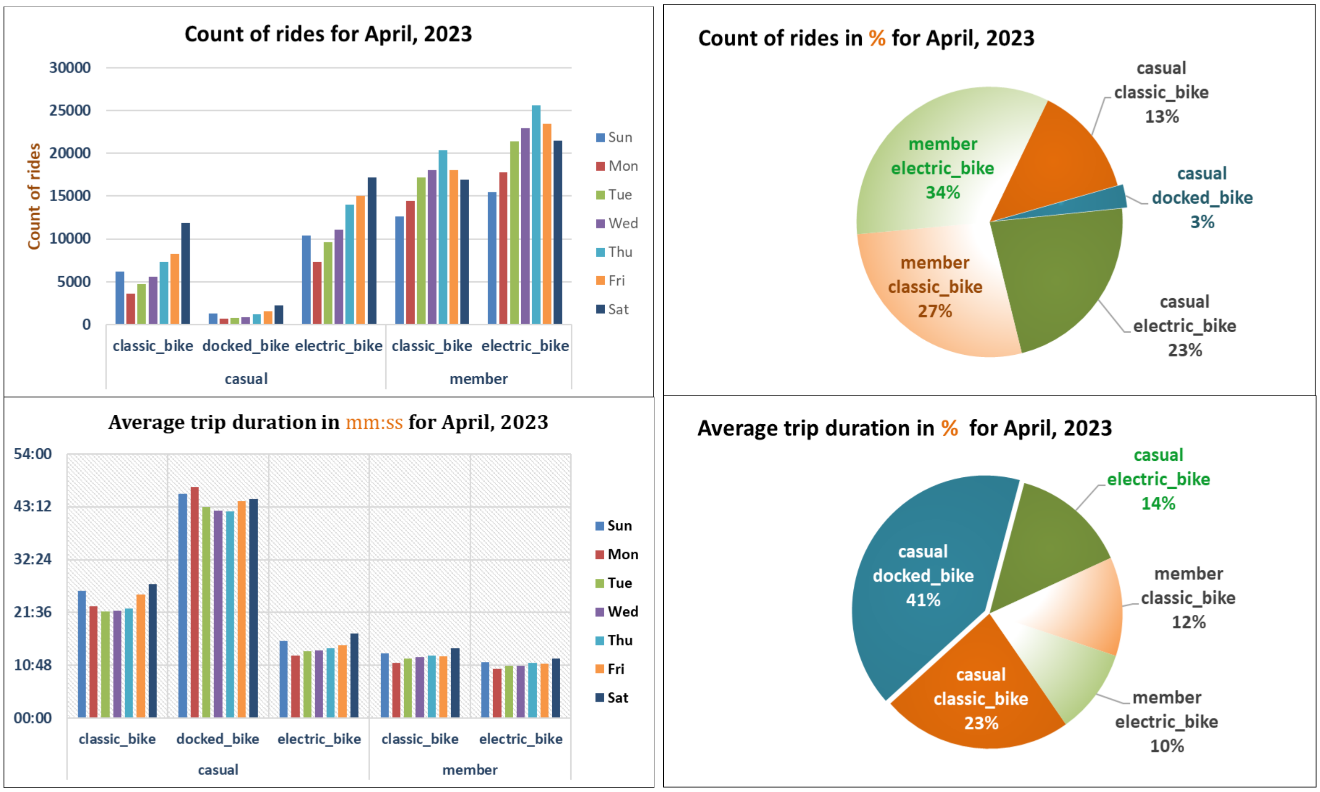
12) Analysis of docked bike behavior using Excel Pivot Columns Charts and Pie Charts is below:

Docked Bike Data Discrepancy

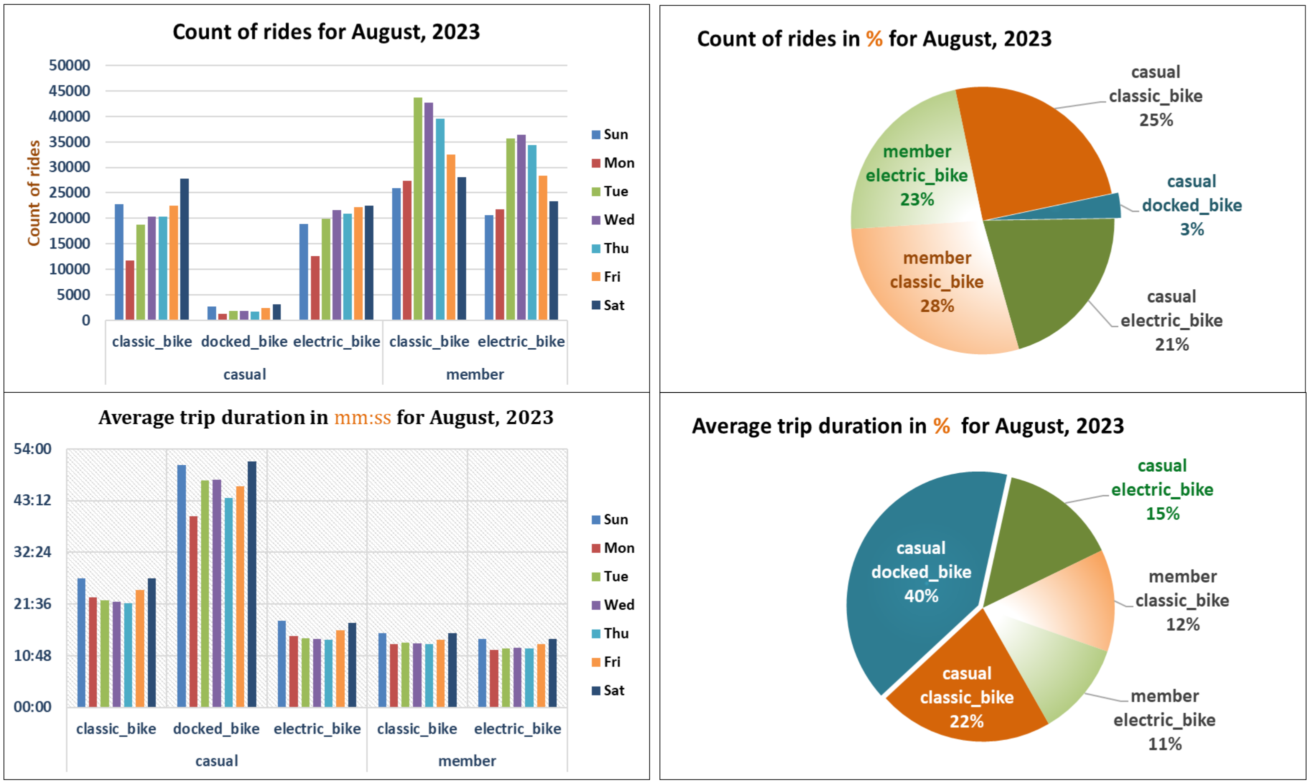
The "docked bike" was present in datasets from January to August 2023. Since September 2023, docked bikes have been removed from monthly datasets and, apparently, from service entirely. I could not access internal company information to know conclusively what a docked bike is. I chose three months to investigate: January, April, and August 2023. I created a bar and corresponding monthly pie charts, as shown below in **Excel Chart Compilation 1-3**. Excel Pivot Tables were used to make the initial data observation and analyze the "docked bikes" behavior or its origin. My hypotheses and subsequent analysis are below.



__Excel Chart Compilation 1:__ Docked Bike Data Investigation for January 2023



__Excel Chart Compilation 2:__ Docked Bike Data Investigation for April 2023



Excel Chart Compilation 3: *Docked Bike Data Investigation for August 2023*

<u>Hypothesis 1</u>

The first hypothesis was that docked bikes are another name for classic bikes. However, docked bikes showed very different behavior in the charts than other bike ride types. The pie charts illustrate that docked bikes have the most minor total count of rides (2-3%) and the highest average trip duration compared to other bike types. This suggests that docked bikes are unique and different from classic bikes.

Finding 1: Based on the chart illustrations, I concluded that docked bikes are not another name for classic bikes as initially hypothesized.

<u>Hypothesis 2</u>

The second hypothesis investigated was that docked bikes are reclining bikes, hand tricycles, and/or cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bicycle. According to the Cyclistic company description provided by the course material:

> “Cyclistic sets itself apart by offering reclining bikes, hand tricycles, and cargo bikes. It makes bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike. Most riders opt for traditional bikes; about 8% use the assistive options.”

Finding 2: If this is true, it explains a tiny number of trips and longer trip duration. Fewer of these types of individuals ride at a slower pace, which can be reflected in longer trip duration. However, the data set and my investigations into the Divvy program did not provide any evidence to conclude this.

<u>Hypothesis 3</u>

Divvy offers significant discounts for people in households receiving SNAP, WIC, LIHEAP, FAFSA, or public housing assistance. With lower prices, this smaller group of users can take a longer duration and a smaller number of trips. This is less likely for “docked bike” users, but they can be classified as a separate group. However, this deeper investigation is beyond the scope of my assignment.

- Excel Pivot Columns and Pie Charts revealed that “docked bikes” exhibited very different behavior—a tiny number of rides and a significantly longer average trip duration—indicative of potential outliers.

Conclusion: Since the “docked bikes” data are outliers and no longer in service, I removed them from the dataset as irrelevant to our analysis. These charts can help Cyclistic make a better business decision if it wants to offer the “docked bikes” service again.

13) Data Cleaning after Analysis in Excel: Removed rows where rideable_type = “docked_bike.”
14) As the final step before saving, a quick data observation using the glimpse() function was performed to ensure data readiness.
15) Modified and cleaned dataset saved as an .RData object, ready for the Analysis Phase.

Analysis Phase

The data analysis performed in R Markdown can be found [here] (GitHub link), making viewing the R codes in the HTML file easy. You can access the R file through the [Project Content] link and check how it runs on Kaggle [link].

Note: To differentiate charts and tables in the Process Phase from those in the Analysis Phase, I have named them differently: **Tibbles** for tables and **Plots** for charts in the Process Phase. In the Analysis Phase, they are referred to as **Tables** and **Figures**.

I analyzed the data from least to most important so that at the end of this analysis, we could see a clear picture and concentrate on the most critical differences in bike usage between annual Members (Members) and Casual riders (Casuals). This approach made it easier to conclude the new marketing strategy and achieve the business goal.

Member vs. Casual Rider Breakdown

Figure 1 indicates that the ratio of the total number of trips by Casual users to Members is approximately 1:2. This unweighted total yearly average (35.2%/64.8% = 0.54) represents the rider type trip count divided by the total trip count. Later, I will compare it with the weighted yearly average based on monthly and weekly averages.

The data doesn't include a personal ID column but a trip ID ("ride_id" column). As a result, I could not directly compare the actual number of customers and their proportion between categories.

Assumption: The trip count proportion roughly mirrors the customer ratio among rider types; further validation will follow with additional data.

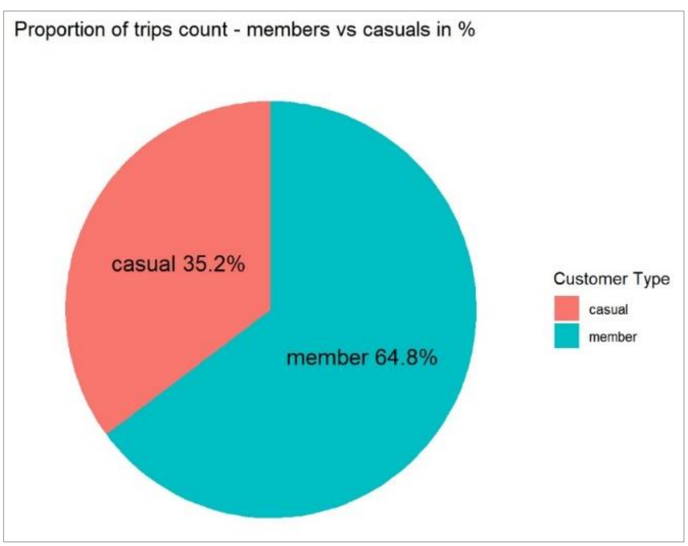


Figure 1: *Pie chart for the trip number of members vs. casuals in %.*

Figure 2 shows a further breakdown of the typical rider. For classic bikes, the ratio of Casual users to Members in the trip count is approximately 1:2, while for electric bikes, this ratio is about 2:3.

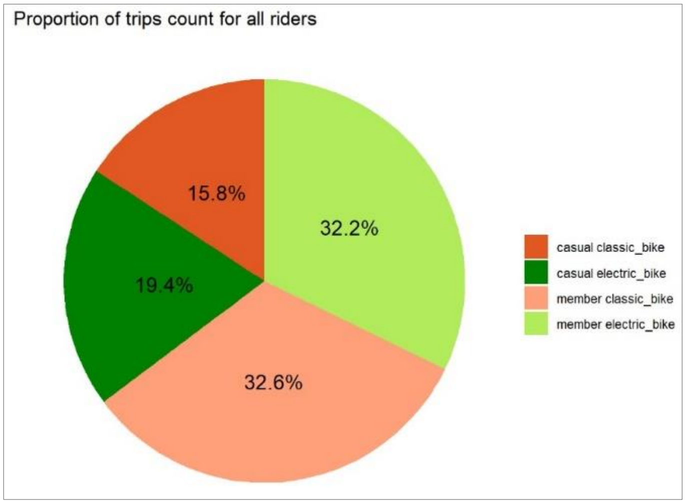


Figure 2: *Pie chart for the trip number of all riders in %.*

Later in this analysis, we will explore other comparisons of typical riders. However, these ratios significantly impact most numbers and chart appearances.

Trip Duration Percentiles Investigation

Let's delve into other aspects of pricing and bike ride time duration. Because a trip's price depends on how long a trip is, it is essential to know how long most riders use their bikes.

R programming codes can show the percentage of trip duration within a specific range.

Table 1: Members' Durations

	Percentile	Minutes
97%	0.97	37.5
97.5%	0.975	39.6
98%	0.98	42.1
98.5%	0.985	45.5
99%	0.99	52.1
99.5%	0.995	70.0
100%	1	718.1

Table 2: Casuals' Durations

	Percentile	Minutes
97%	0.97	75.2
97.5%	0.975	81.4
98%	0.98	89.7
98.5%	0.985	101.2
99%	0.99	118.6
99.5%	0.995	153.2
100%	1	719.4

Tables 1 and 2: *Trip Duration Percentiles for Members and Casual Riders in 2023*

Insight 1: Most users rarely exceed their free ride time: Members have a 45-minute free ride limit, and Casual customers have a 180-minute free ride limit. 98% of Members use up to 42 minutes, and 99.5% of Casuals use up to 153.2 minutes.

Insight 2: Although Casuals have 180 minutes of free rides, 98% of them take trips lasting less than 90 minutes. For our marketing challenge, we should focus on targeting free ride time. I'll revisit these figures later.

Typical Week Analysis

I want to answer whether we can develop a better marketing strategy based on the variation in trip counts across different days of the week. Figures 7 and 8 show the breakdown of the number and duration of trips for Causals and Members throughout the week.

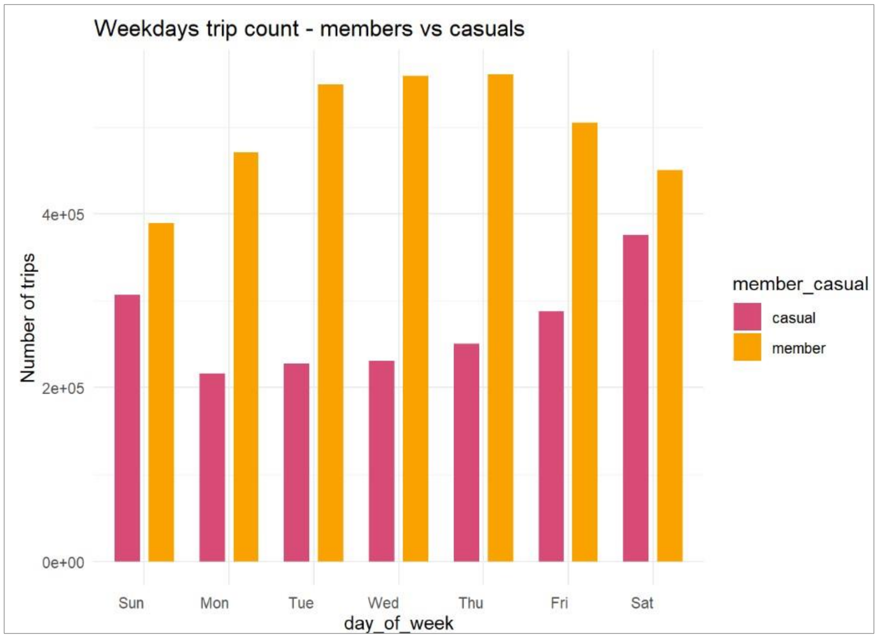


Figure 3: *Number of Trips Throughout the Week For 2023*

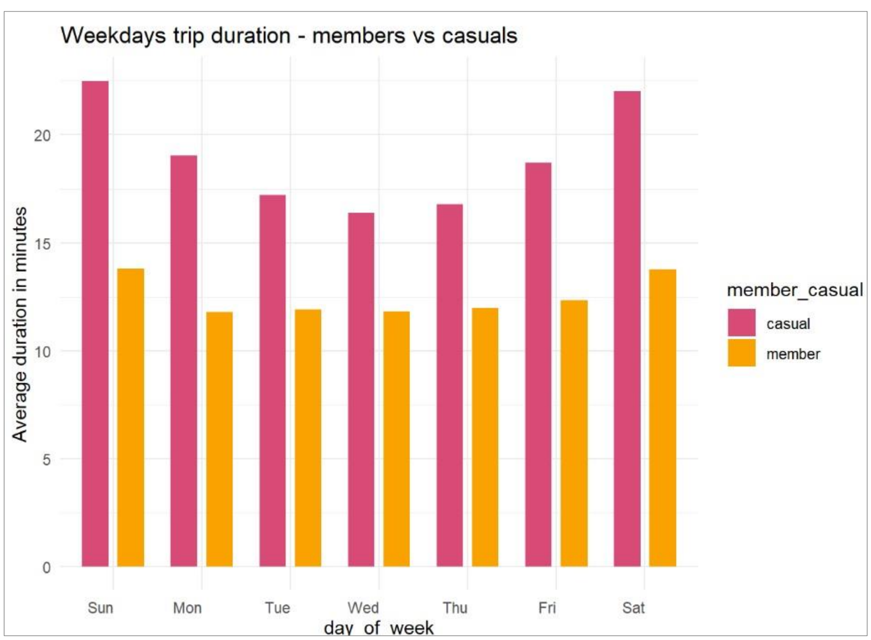


Figure 4: *Duration of Trips Throughout the Weeks of 2023*

Data Interpretation Approach: Given the distinct differences in chart patterns of bike usage between Members and Casual riders during weekdays, it is unnecessary to seek specific numerical comparisons. These chart patterns provide sufficient insights to inform and develop an effective marketing strategy. I will apply the same approach to all other visualizations below, except for the analysis related to weighted average ratios of the trip counts and pricing.

Insight 1: As shown in Figure 3, from Monday to Friday, Members have roughly double the number of trips compared to Casuals. The bar pattern for Members reflects that most are likely commuters to work or school during the week. Casual bike riders likely use bikes for leisure. Casual trips increase on the weekends and Fridays.

Insight 2: Figure 4 shows that Casuals' average trip duration on weekdays is more significant, which is appropriate for this type of user. Casuals are usually not hurrying to work and ride longer than Members, increasing the number of minutes on weekends as expected.

Insight 3: These two charts prove that both categories of riders still look very active on any day of the week. The consistency of these charts' patterns suggests that Casual riders use bikes as a lifestyle, which is a good opportunity to convert them to annual Members. This is also supported by the monthly column charts in Excel, as shown previously in Excel Chart Compilation 1-3, where each chart displays weekday data.

Suggestion: An option to increase the profitability of Cyclistic may be to raise prices on certain days based on demand. However, changing prices during the week may discourage Casual riders, as people want the freedom to choose when they ride. We face risks such as losing customers, failing to convert Casual riders to Members, and disappointing current riders, which requires deeper analysis.

Also, an hourly analysis could determine the best pricing schedule. But changing prices at certain hours might make leisurely riders feel rushed, risking customer loss. We must seek win-win marketing strategies that encourage annual memberships without turning current customers away.

Monthly Analysis

Reviewing monthly trends will likely give us insights into typical ridership patterns. The following two charts show that the average trip duration for each month of the year was much longer for Casual riders than for those who bought an annual membership. Figure 5B presents an aggregated view of all trip duration charts to see how they appear when all users are separate. I have not accounted for the differences based on trip count at this moment. Instead of conducting a deep analysis based on these charts, I compared the same type of users for more straightforward observation and analysis, showing us a better picture of typical ridership.

Reflection: Later, after examining differences in bike ride usage patterns, I grouped the months into two categories: “Cold Season” and “Bike Season.” Bike season includes seven months from April to October, typically featuring longer trip durations and a more significant number of trips. The cold season consists of the remaining five months, from November through March. This division will provide better substantiation later and help focus the information, enabling more insightful analysis from the data visualization charts.

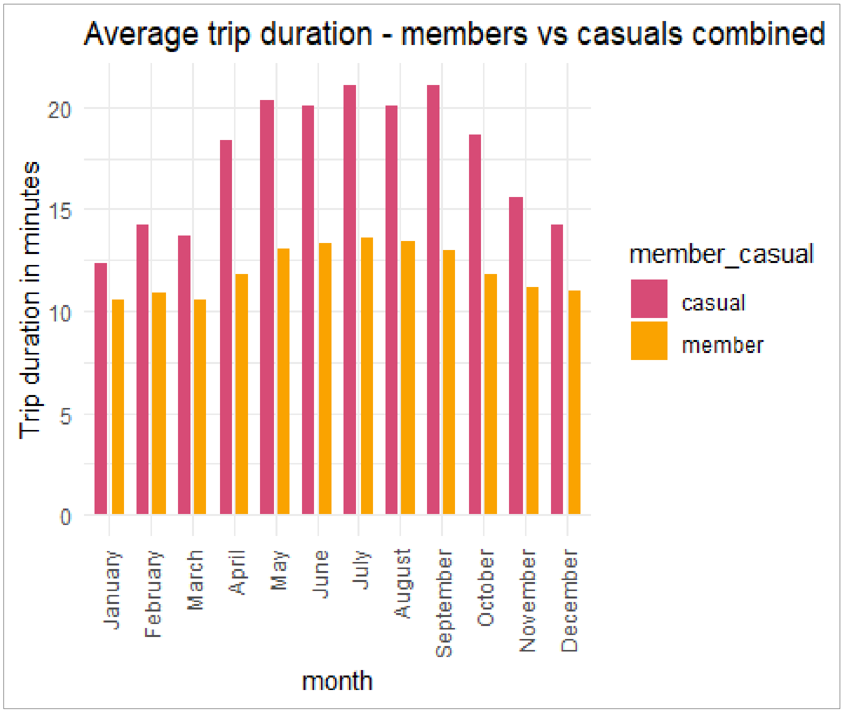


Figure 5A: Month by Month Breakdown of Trip Duration for Casuals and Members

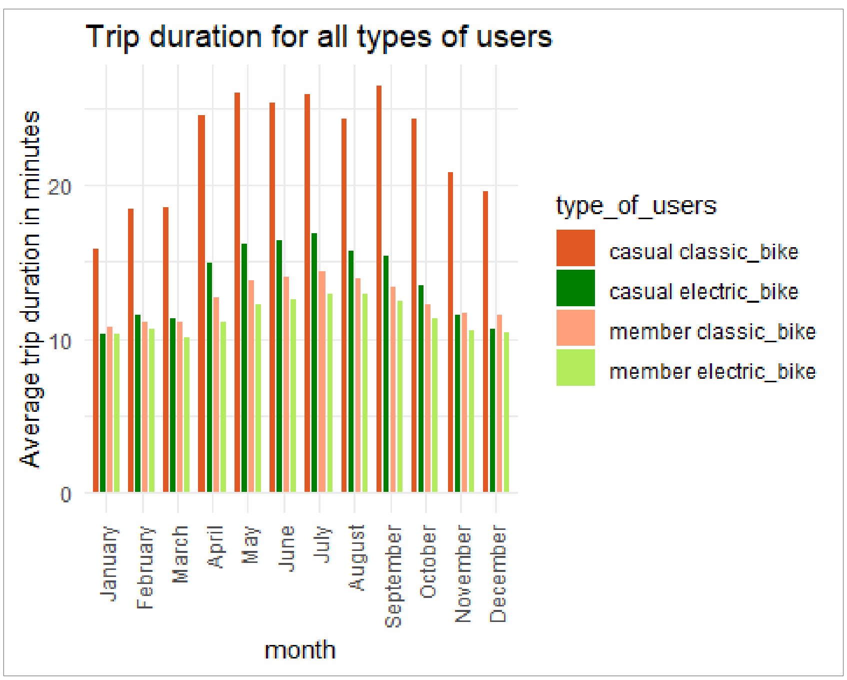


Figure 5B: Month-by-Month Breakdown of All Rider Types

But now, let us compare the same types of users…

Classic Bike Users

Insight 1: When analyzing classic bike trips separately, we can see a normal distribution of the average trip duration and trip count between cold and warm seasons: colder weather results in fewer trips.

The number of trips for both types of users peaks in August, followed by July and September.

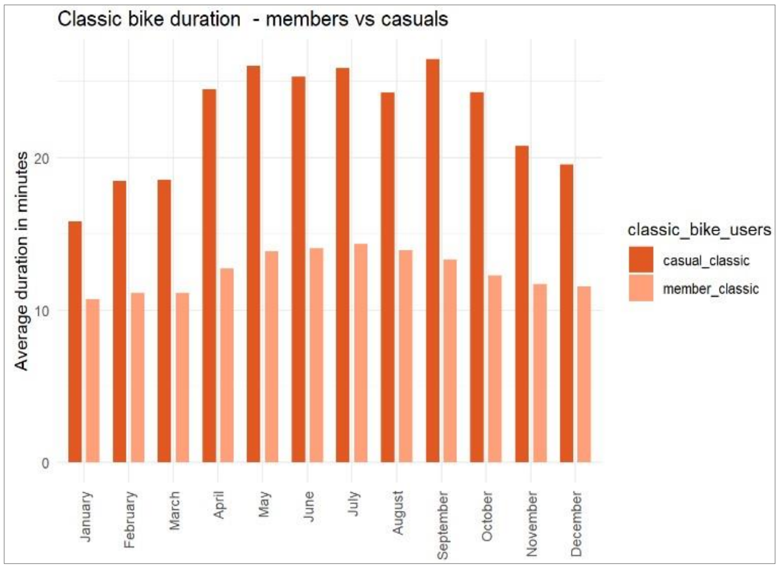


Figure 6: *Average Trip Duration for Classic Bike Riders*

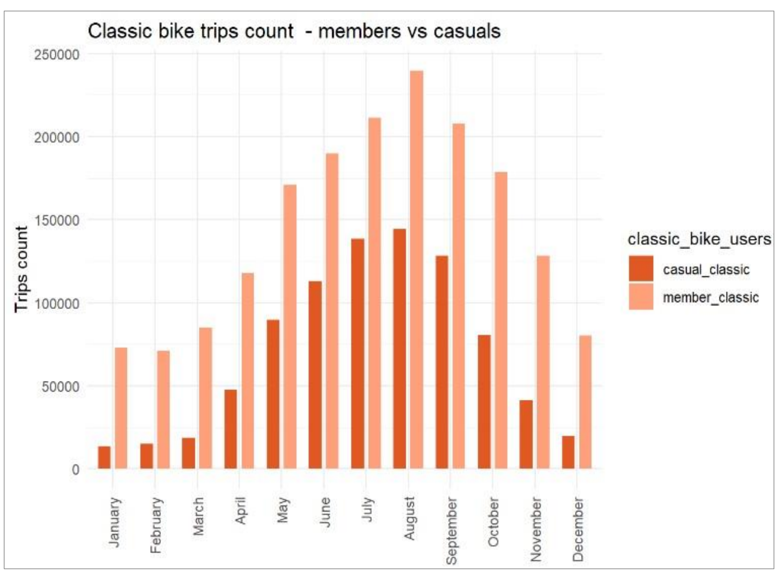


Figure 7: *Monthly Trip Count for Classic Bike Riders*

Insight 2: Observing the increased number of trips and knowing that Members are usually commuters, it is clear that this surge in trip numbers is not due to more job openings or an influx of students during the summer. Instead, it indicates that Members are also using bikes for leisure and maintaining a healthy lifestyle.

Electric Bike Users

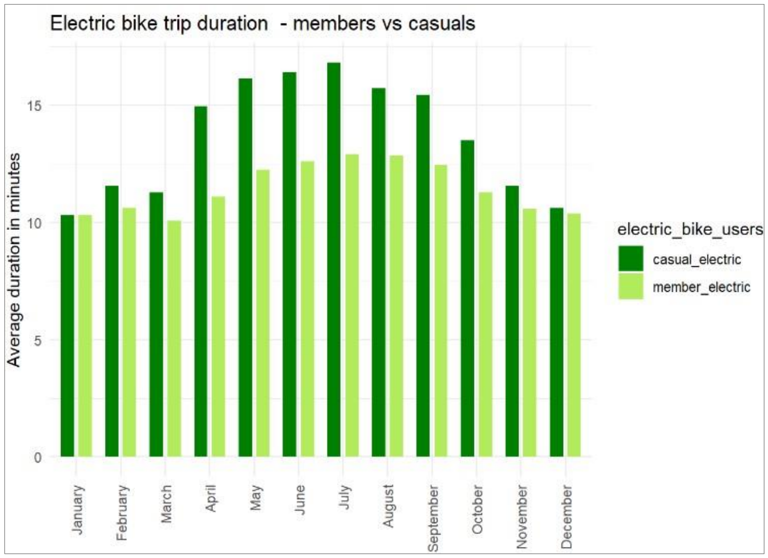


Figure 8: *Average Trip Duration for Electric Bike Riders*

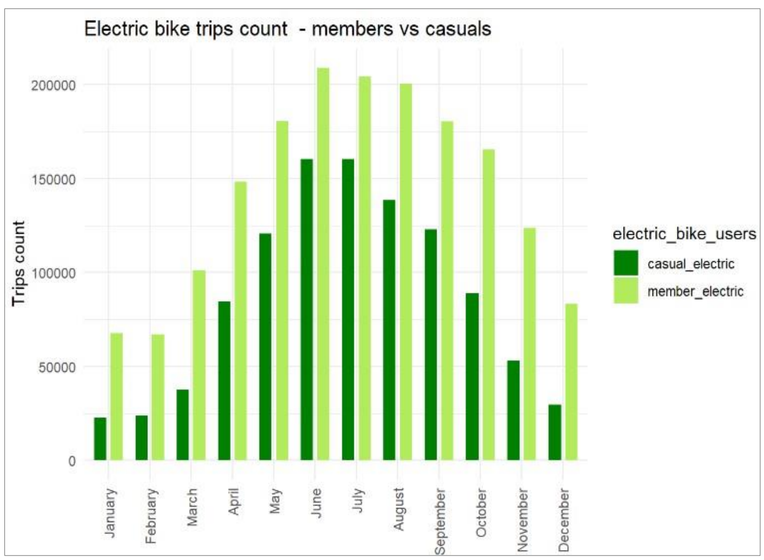


Figure 9: *Monthly Trip Count for Electric Bike Riders*

Figures 8 and 9 show patterns that are close to being similar for electric bike users and classic bike users, so we can gain the same insights as mentioned above.

Insight 3: The difference in trip duration for the same type of bike between Members and Casuals:

If we compare Casual trip duration patterns from January to December in Figure 8 to Figure 6, then according to Figure 8, Casual electric bike users have an average trip duration range of approximately 11 to 18 minutes, whereas Casual classic bike users (Figure 6) range between 16 and 27 minutes. The difference in trip duration between Casual classic and electric bike users is likely influenced by bike speed and price. Prices are discussed in more detail later in this report.

Cold And Bike Season Analysis

As mentioned before, I divided the year into Cold and Bike Season.

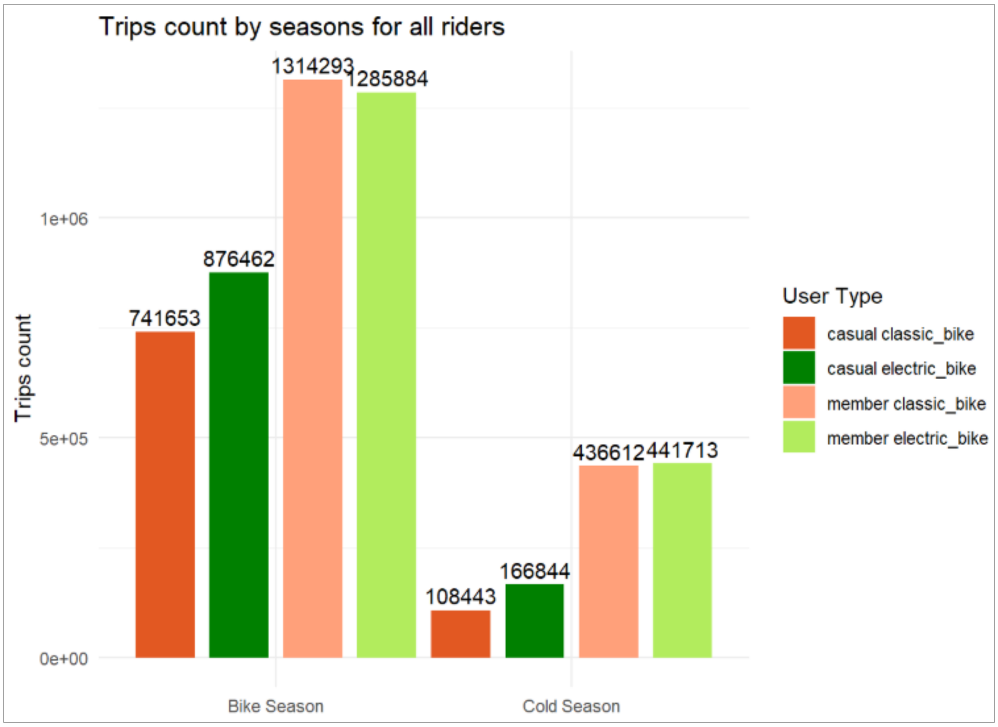


Figure 10: *Bike Season and Cold Season Comparison of Trip Counts*

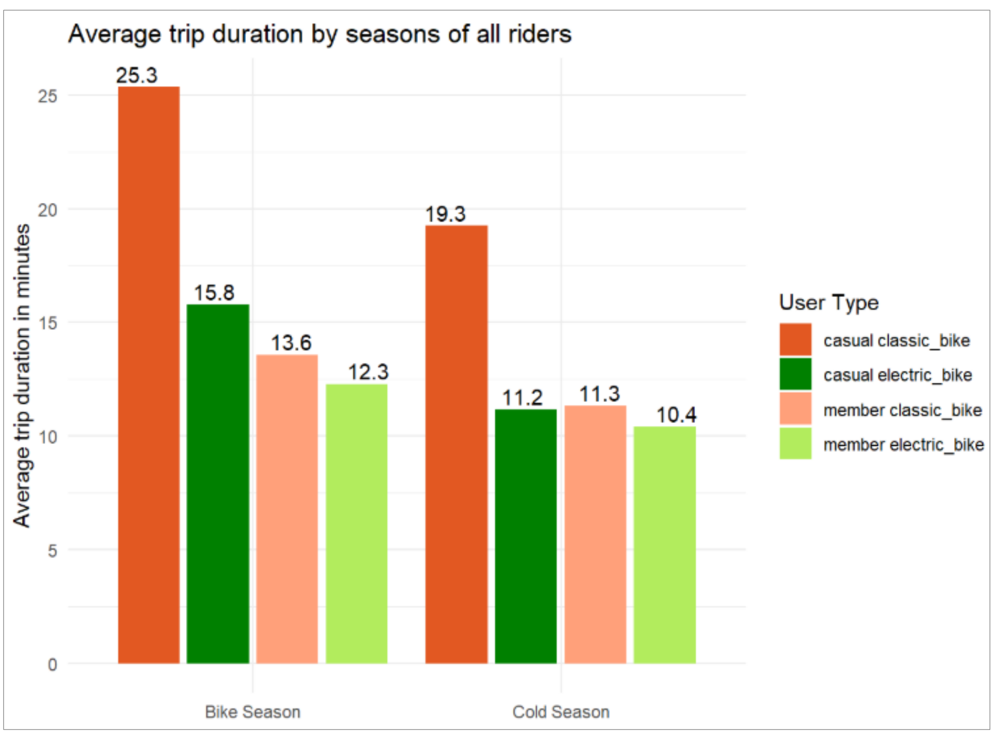


Figure 11: *Bike Season and Cold Season Comparison of Average Trip Durations*

As a reminder, the cold season is from November to March, and the bike season is from April to October. Figures 12 and 13 show two compact graphs above representing the number of riders' activities in these two seasons.

Insight 1: Given the small number of trips during the cold season, casual riders will doubtfully take advantage of any promotions. However, this offering can convert Casuals into Members in April and May as the riding season picks up.

Insight 2: Examining Members' trip counts and average trip durations in Figures 12 and 13 reveals that their data bars are similar between each bike type in each season, likely because most of them are commuters.

Insight 3: Casual riders use electric bikes more frequently, resulting in a higher trip count than classic bikes. Therefore, targeting Casual electric bike riders with promotions and advertisements for membership is a strategic approach. Promotions will be discussed in the Share Phase.

Table 3: Casuals' Bike Season

Percentile	Minutes
60%	15.7
65%	17.6
70%	19.9
75%	22.8
80%	26.6
85%	32.2
90%	41.5
95%	61.0
100%	719.4

Table 3 reflects the percentiles of the trip durations shown in Figure 11. Specifically, the 15.8- to 25.3-minute range for classic bikes represents 60% to ~80% of Casual riders' trips during the Bike season.

Insight 4: Casual customers who are very active, with 60% to 80% taking longer trips, have a strong potential for conversion to Members. Additionally, the large number of Casual users throughout the year and across all days of the week suggests that many are frequent or repeat riders, making them more likely to become Members.

Casuals-to-Members Trip Count Weighted Average Ratio Analysis

Figure 12 Description:

a) Figure 12 below shows the weighted average ratio method, accurately reflecting the yearly trend. Considering the higher ridership in summer, this method leads to a more realistic ratio.
b) The weighted average ratio reflects the overall contribution of Casual and Member riders to the entire dataset, highlighting how trip counts vary from week to week and month to month.
c) Figure 12 shows how the ratios of monthly averages smooth out the fluctuations of weekly ratios, making it necessary to differentiate the year into two parts. This differentiation is because ratios vary from 0.24 in January to 0.72 in July, according to the chart scale, resulting in an average ratio of 0.61 for Bike Season and 0.31 for Cold Season.
d) The dot markers on the light-blue line represent 52 weeks, while the dot markers on the orange line represent the 12 months of the year.
e) In Figure 12, the Yearly Unweighted Average line is not shown.

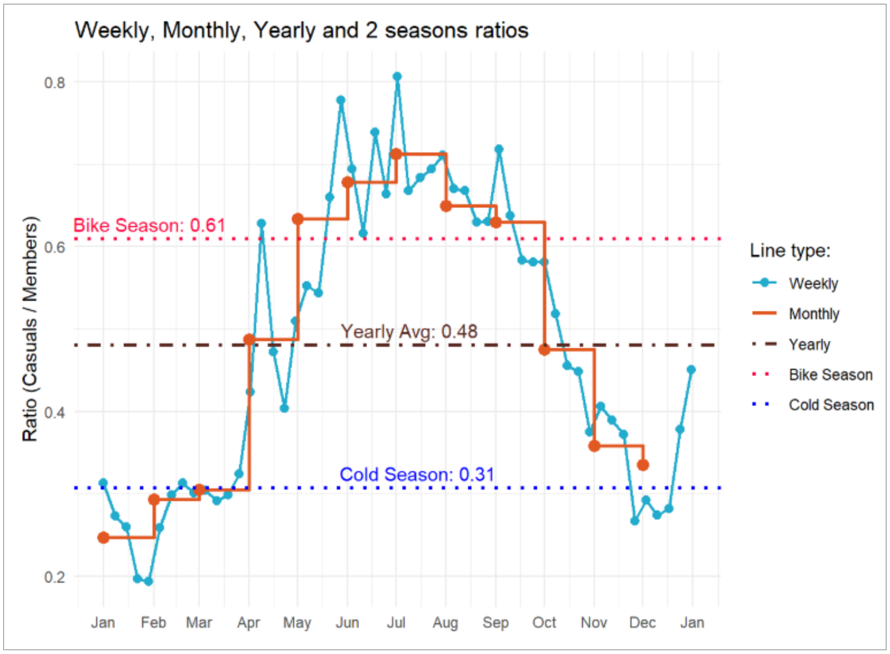


Figure 12: Weighted Averages for Weekly, Monthly, Yearly and Seasonal Ratios

Assumption: The weekly ratio line on the chart spikes from April to September. This trend is likely influenced by varying weather conditions during these months, such as stormy days, which can significantly impact bike usage patterns.

Insight 1: The weighted Annual average ratio of 0.48 is close to the April and October ratios. The months between them have much higher numbers, resulting in a Bike Season ratio of 0.61.

Type of Season (for Casuals only)	**Trip Count**
Bike Season Trips	1,618,115
Cold Season Trips	275,287
Ratio (Bike/Cold)	5.88
Difference (Bike - Cold)	1,342,828

Table 4: *(Auxiliary Table for Figure 10) Total number of both seasons' trips, their ratio, and difference*

Note: The values in Table 4 (above) were derived from R code aggregations for Figure 10 and Table 5 (below) from codes for Figures 1 and 12. These values were calculated and compiled into a data frame and table at the end of the R Markdown script.

Evaluation 1: Let's define the Cold and Bike Seasons ratios of 0.31 and 0.61, in combination with Table 4 data, which was created from the aggregation of Figure 10 trip count values.
The 30% difference (0.61 - 0.31 = 0.30) in the combined trip count ratio between Bike Season and Cold Season reflects the impact of an additional 1,343,773 Casual rider trips in Bike Season, which is 5.88 times more than in Cold Season.

Insight 2: Based on the comparison of the numbers above, remarkably, there were 1,343,773 additional Casual rider trips in Bike Season compared to Cold Season. Most of the marketing strategy should focus on the Bike Season period.

Insight 3: Even though the marketing strategy should focus on the Bike-Season period, the actual number of Casual trips during the Cold Season remains decent—275,607 (Table 4 data).

Evaluation 2: The unweighted yearly ratio of 0.54, previously calculated for data in Figure 1, gives more importance to summer months, which have a higher ratio. This skews the result to 6% higher than if we accounted for seasonality (0.54 – 0.48 = 0.06, which is 6%).

Category	Average
Bike Season (April-Oct Avg)	0.6098
Cold Season (Jan-Mar & Nov-Dec Avg)	0.3085
Weighted Yearly (Weeks)	0.4812
Weighted Yearly (Months)	0.4842
Unweighted Yearly Ratio	0.5443

***Table 5:** Unweighted and Weighted Averages for Yearly and Seasonal Ratios*

Evaluation 3: Table 5 shows a negligible difference between yearly by week and yearly by month weighted average ratios: 0.4812 and 0.4842. Figure 12 shows how monthly data smooths out intra-week fluctuations but reflects the overall trend.

Insight 4:

a) The aggregated averages remain close since the ratios are similar (0.481 vs. 0.484). That is evidence that data appears evenly distributed over the year (no extreme outliers or seasonal spikes), as shown by the holistic view of consistent chart patterns across all other charts.
b) However, seasonal analysis reveals distinct patterns: during the bike season, there is a higher Casual-to-Member ratio (0.61) due to increased Casual ridership, whereas the cold season sees a lower ratio (0.31) as Casual trips decline more sharply, as shown in Figure 10. This highlights how aggregation smooths seasonal variations while capturing key behavioral trends in rider activity.

Conclusion: Our assumption that trip counts roughly mirror the customer ratio among rider types is correct for Bike Season only because:

a) It has a better ratio (0.61) than the Unweighted (0.54) and Weighted (0.48) Average yearly ratios.
b) As concluded before, most users rarely exceed their free ride time.
c) The average trip durations in Figure 11 fall within the 15.8 to 25.3-minute range for electric and classic bikes, representing 60% to 80% of Casual riders' trips during the Bike Season. This indicates that few riders take more than one trip per day.

Analysis of Bike Ride Costs

At this point, we need to know the real benefits and prices for Members and Casuals. Without this, I could only give superficial suggestions for new marketing strategies. The benefits and 2024 prices can be found at divvybikes.com, which is open to the public.

In a real-world scenario, I would have access to both current and historical data on usage, prices, and other relevant metrics. For this case study, although the prices from 2024 differ from those in 2023, I will apply the 2024 prices to the 2023 data for further analysis. This approach allows me to simulate a close-to-real business project.

For Casuals:	**Single Ride**	**Day Pass**
	$1 + $0.18/min	**$18.10/day**
Classic prices	$1 unlock + $0.18/min	**3 hours free**, >$0.18/min
E-bike prices	$1 unlock + $0.44/min	Free unlocks + $0.44/min

For Members:	**$143.90/year**
Classic prices	**45 min free**, > $0.18/min
E-bike prices	Free unlocks + $0.18/min

Above are April 2024 Divvy prices as shown on divvybikes.com/pricing. The complete pricing table and graphs from Excel are available in PDF copy for easy viewing here. I also linked the Excel worksheet with price tables, charts, and formulas to manipulate prices and visualize changes across all users here: link.

Pricing Structure: Users can purchase a membership, a Day Pass, or a Single Ride.

"Free single rides": Members and Casuals have an unlimited number of "free single rides" during the day for classic bikes. Members have a "free single ride" up to 45 minutes long and will be charged per minute beyond that time if the bike is not locked. The "Day Pass" purchased by Casuals also includes unlimited "single free rides," each up to 3 hours. Beyond 3 hours, the Casual user will be charged per minute. None of the electric bike users have a "free single rides" benefit.

"Free unlock": The Day Pass, priced at $18.10, includes a "free unlock" for classic and electric bikes. Unlocking a bike without the Day Pass cost $1.

Price Point: For electric bikes, the Day Pass does not include a "single free ride." Therefore, it does not usually benefit the customer because they are charged $18.10 and an additional $0.44 every minute they ride. To make purchasing the Day Pass more cost-effective, customers must unlock the bike at least 19 times during the day. I even called Divvy customer service to clarify this, and they confirmed it.

Benefits Consideration: The tables and charts I created do not consider all the benefits of being a Member or using a Day Pass for Casual riders. One of the benefits is that a rider can lock and unlock their bike an unlimited number of times during the day and use the "free single rides" benefit as long as they do not exceed the free ride time included in the Day Pass or membership purchase price.

Explanation of the structure of the following two charts (Figures 13 & 14):

Figure 13: *Overall Graph of Cost Per Hour of All Ride Types*

Figure 13: Showing the price of each type of user up to 12 hours of a single continuous trip. The prices at the end of the chart's lines show the cost for a 12-hour ride for reference—which is impractical for most riders due to the high cost, but it is helpful to see the differences over a long period. As a reminder, all data is trimmed to the 2 – 720-minute range. The classic Single Ride line is not shown here.

Note: I assumed an average Member rides 5 days a week because most are commuters. Based on my calculations (see PDF), an annual membership costs $0.55 daily. I used this $0.55 rate with the price per minute for further calculations. If Members ride 7 days a week, the cost drops to less than $0.40 a day. This difference is not significant and is for reference only.

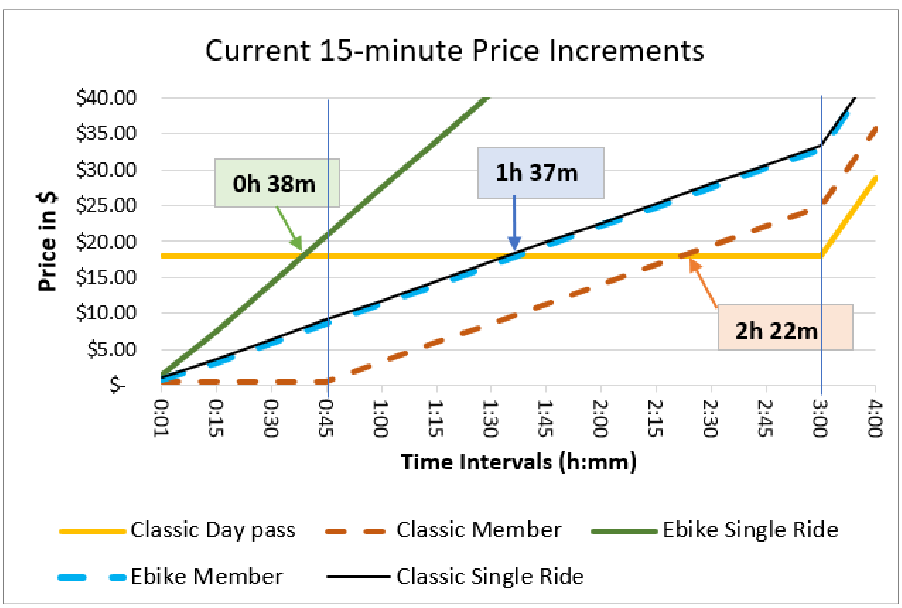


Figure 14: *First 4 Hours Price Comparison of All Ride Types*

Figure 14: The chart focuses on the first four hours of the trip, covering most of all riders. It has a 15-minute increment scale of up to 3 hours. From 3 to 4 hours, there is a 1-hour increment due to space constraints on the x-axis, causing an abnormal bend at the 3-hour mark. Refer to Figure 13 for a more straightforward comparison, as the chart has a consistent 1-hour increment scale.

The horizontal yellow line shows 3 hours of "free ride" for a Day Pass before it slopes up. The thin vertical blue lines at 0:45 minutes and 3:00 hours on the x-axis serve as reference points for free ride durations.

Analysis based on Figure 14

Breakdown 1: The thin black line for Casuals' Classic Single Ride overlaps the blue dashed line for Members' E-bike, indicating similar prices for these bikes and users. Members pay an extra $0.55 daily, while Classic Single Ride users pay $1 to unlock the bike—a nearly invisible $0.45 difference. Though being a Member is slightly cheaper daily, the upfront cost of $143.90 is high compared to the $1 unlock fee, with the total per-ride difference being just $0.45.

Insight 1: There is too little benefit to being a Member if the prices for Casuals' Classic Single Ride and Members' E-bike are the same, aside from the ease of riding an e-bike.

Breakdown 2: The text boxes with time stamps in Figure 14 show where each line crosses the yellow Classic Day Pass line. Up to these points, the crossing trip type is cheaper than a Day Pass with 3 hours of free ride. After these points, the Classic Day Pass becomes more inexpensive than the Members' E-bike (after 1h 37m) and Classic bike (after 2h 22m).

Insight 2: This is unfair to Members, as their benefits should surpass Casual customers in every respect.

Insight 3: The Cyclistic (Chicago Divvy) website does not sufficiently highlight the benefits of purchasing an annual membership. For instance, the membership price of $143.90 is equivalent to the cost of eight Day Passes. Such advantages should be prominently featured as the primary advertisement on their website and phone app.

Suggestion about "free single ride" time: To make the membership more attractive, consider reducing the Day Pass free ride duration from 3 to 2 hours. Since 98% of Casual riders' trips are under 90 minutes, this change would not significantly impact them. Additionally, increasing the "free single ride" time for Members' classic bikes from 45 minutes to 1 hour could further enhance the membership benefits.

Price Change Recommendations

Table 6 and Figure 15 below show price change suggestions to make an annual membership more attractive. Based on adjusting some prices per minute and new single free ride times for Members and Casuals, these suggestions are for reference only and maintain the annual membership price at $143.90, the same as before.

Type of ride	*Current*	*Suggested*
Day Pass	$ 18.10	$ 20.00
Classic Casual	$ 0.18	$ 0.20
Classic Member	$ 0.18	$ 0.16
Ebike Single Ride	$ 0.44	$ 0.44
Ebike Member	$ 0.18	$ 0.16
Single free ride		
Casual	3 hours	2 hours
Member	45 min.	1 hour

Table 6: *Current and Suggested Prices with Single Free Ride Time*

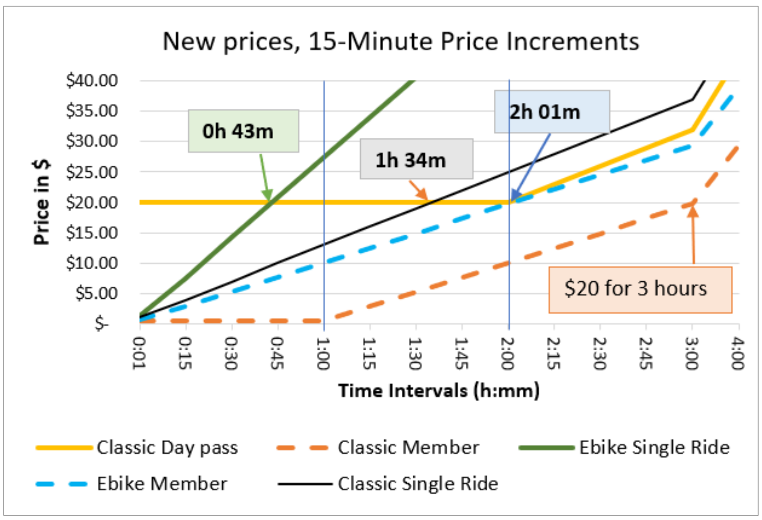


Figure 15: *First 4 Hours Price Comparison of All Ride Types with Suggested Prices*

Insight 1: Being a Member is now always more beneficial than purchasing a Day Pass or other options available to Casual riders. The blue and dark orange dashed lines for Members, indicating a lower price, are now lower than those for Casual riders.

Insight 2: Table 6 shows the new price list, making distinguishing between Members and Casual riders easier. Slightly different prices for Casuals and Members have a psychological effect, similar to $0.99 versus $1.00. This encourages casual riders to evaluate the freedom of everyday bike rides, which is essential. Otherwise, these benefits are obscured by the identical price ($0.18 in this case).

Conclusion: The need to commute naturally drives us to find convenient and cheaper options. Members choose bikes over personal and public transport for convenience and cost.

To motivate Casual users to purchase an annual membership, offering longer "free single rides" for Members (beyond the 45-minute limit for classic bike users) and more competitive per-minute rates could attract those interested in leisure, cardio, entertainment, and a healthy lifestyle. This approach aligns with Casual users' nature and could attract more riders, potentially increasing the number of Members without affecting the company's profit. The analysis shows that Casual riders are frequent and active users. The primary strategy to convert them to Members is to offer more benefits through pricing adjustments.

Share Phase

Link to PowerPoint Presentation of Share and Act phase

Introduction: Based on the analysis, there are 1,343,773 additional Casual rider trips in Bike Season (from April to October) compared to Cold Season (from November to March). The casuals-to-members trip count ratio is also better in the Bike season (0.61) than in the Cold season (0.31). Therefore, most marketing strategies should focus on the Bike Season period.

Furthermore, the current analysis reveals that Casual riders are very active, especially on weekends, and use electric bikes more often than classic ones. The consistency of chart patterns shows that Casual riders tend to ride much longer than Members, primarily for leisure, cardio, and a healthy lifestyle. This presents a significant opportunity to convert them to annual Members.

Converting Casual Riders vs New Customers

After data cleaning and trimming, this 12-month dataset contained 5,380,725 rows representing the number of trips taken in 2023. While this does not reflect the exact number of customers, it evidences the popularity of Cyclistic's bike-sharing service. The analysis suggests that converting Casual riders into Members will be more profitable than targeting entirely new customers with a marketing campaign.

Pricing

Day Pass for electric bikes: The analysis reveals that it is not cost-effective for customers, as it doesn't include a "single free ride." Instead, customers are charged $18.10 plus $0.44 per minute. To make the Day Pass more economical, a customer must unlock the bike at least 19 times on the same day. Divvy customer service confirmed these details, highlighting the unreasonably high cost of the Day Pass for electric bikes.

The exact price per minute of use applies for Classic Single Ride and E-bike Members, which, in some respects, makes a membership not as beneficial even though these are different types of bikes.

Set up the correct leverage of prices: Review the current prices and compare them with those suggested after detailed observation of the related charts from the analysis above. It is recommended that the annual membership price be kept the same at $143.90 unless other market analysis suggests otherwise.

Type of ride	*Current*	*Suggested*
Day Pass	$ 18.10	$ 20.00
Classic Casual	$ 0.18	$ 0.20
Classic Member	$ 0.18	$ 0.16
Ebike Single Ride	$ 0.44	$ 0.44
Ebike Member	$ 0.18	$ 0.16
Single free ride		
Casual	3 hours	2 hours
Member	45 min.	1 hour

Table 6: *Current and Suggested Prices with Single Free Ride Time*

Price Perception Effect: Slightly different prices for Casuals and Members have a psychological effect, similar to the impact of $0.99 versus $1.00 pricing. This encourages casual riders to evaluate the freedom of everyday bike rides, which is essential. Otherwise, these benefits are obscured by the identical price ($0.18 in this case).

Raising prices on weekends and/or at some hours: To increase profitability, consider raising prices on certain days based on demand. However, changing prices during the week may discourage Casual riders who value flexibility. An hourly analysis could identify optimal pricing, but varying prices by the hour might make leisurely riders feel rushed, risking customer loss. We need win-win strategies that encourage annual membership without coercing customers to buy it.

Length of “Free Single Rides”

Suggested “free single ride” time for a Day Pass to 2 hours instead of the current 3 hours because most (99%) of casual riders only ride much below that amount of time. Again, for 99% of Casuals, it will be not overly coercive but a gentle prompt to become a Member.

Figure 15: *First 4 Hours Price Comparison of All Ride Types with Suggested Prices*

With the new prices, being a Member is always more beneficial than purchasing a Day Pass or other options available to Casual riders. Members' blue and dark orange dashed lines are now lower than those for Casual riders.

The linked Excel worksheet allows us to adjust prices and visualize their distribution across all users.

If Members use bikes for leisure, exercise, and a healthy lifestyle as well as Casual riders, increasing the "free single ride" time to 1 hour for classic bikes and offering better per-minute prices for both types of bikes will attract Casual riders, who are using electric bikes more often than classic bikes. This strategy aligns with their nature and should not harm Cyclistic's profit but encourage memberships, creating a win-win situation.

Complimentary Month

Offering a complimentary free month as part of the membership can effectively convert Casual riders into Members, especially in April and May, when people are more likely to ride bikes and notice promotions. Casual trips drop significantly in colder months, so offering this benefit in winter or March will unlikely attract many Casual riders.

Free Ride Days and Other

Since Casual riders use electric bikes more frequently than classic bikes, offering free single rides for electric bikes on holidays or weekends to Members could attract Casual riders to purchase an annual membership. Additionally, hosting customer appreciation days, competitions, and engaging activities for all Members will make bike rides enjoyable and encourage new sign-ups and renewals.

Advertisements

Digital media can help in marketing tactics. The Cyclistic's website and phone app should include advertisement pop-ups, interstitials, and banner ads. The following advertisement messages may be successful in drawing Casuals towards a membership:

- Annual membership costs are the same as only 8-Day or 7-Day Passes at the new prices.
- With the new prices, Members can have a single ride of classic bikes for 3 hours at the same price that Day Pass users pay for just 2 hours. (See Figure 14)
- New pricing adjustments are showing off advantages for Members.
- Details of the benefits for Members, such as the health benefits of riding bikes, spending time outdoors, and staying healthy.
- Posters at the docking stations encouraging membership.

Act Phase

The following actions are recommended based on this analysis:

1. Review and adjust prices for annual Members and Casual riders so that buying an annual membership is more beneficial and attractive.
2. Review and adjust the "free single ride" time for annual Members and Casual riders for the same reason mentioned above.
3. Provide Member appreciation days, competitions, and activities that make bike rides fun.
4. Offer Casuals complimentary months of membership in April or May.
5. Offer free single rides for electric bikes on some holidays or weekends for Members in such a way that attracts Casuals to buy an annual membership.
6. Redesign the company's website and phone app to better advertise new Cyclistic membership prices and benefits. Place ads on Instagram, Facebook, or similar platforms.
7. Conduct further analysis of Members and causals to determine more marketing strategies. Suggested areas for further exploration are:
    a. Add an encrypted customer ID to datasets, allowing a more profound analysis without access to personal information.
    b. Provide analysts with more information about reclining bikes, hand tricycles, and cargo bikes to find a way to continue this service.