Cyclistic Bike Share: Converting Casual Users to Members

Jessica Acker

September 3, 2022

Executive Summary

One of Cyclistic's main goals is to increase profits by converting casual riders into subscribed members. This analysis examines the last twelve months of rider trip data to determine key differences between how casual riders and members utilize Cyclistic's services to provide recommendations that can help Cyclistic reach its business goals. Casual riders are more likely to utilize Cyclistic's services for pleasure over commuting, thus Cyclistic should diversify membership options, such as shorter term subscriptions and tourist passes, to appeal to a wider audience. Marketing strategies should prioritize the seasonal summer peak and be targeted at popular stations among their casual users to increase odds of subscription conversion.

Introduction

Bike-share programs have been introduced in cities around the world in efforts to reduce traffic congestion and air pollution, as well as solve the "last mile" problem of public transport, allowing passengers a transport method to get to and from the end point stations in their travels. Additionally, bike-share programs appeal to residents and visitors alike as an environmentally friendly and healthy recreational activity. Operating a bike-share company can therefore be both profitable and add value to the community.

Cyclistic, a Chicago based bike-sharing company, would like to increase profits by converting casual riders into subscribed members. To do this, the company first needs insight into how these member types differ. Casual riders are defined as customers who purchase either single-ride or full-day passes while members are those who currently have an active annual membership. The purpose of this report is to analyze both user types, determine key differences between them, and design some marketing strategies to recommend to Cyclistic's executive team. The analysis that follows revealed that casual riders most often use Cyclistic's services for pleasure while members more often use their membership for commuting purposes. This suggests Cyclistic should focus on more diversified membership offers to appeal to a wider audience, which is why members are referred to as subscribed members vs annual members throughout this report.

This analysis serves as the capstone project of the Google Data Analytics Certification and is a simulated exercise. Cyclistic is a fictional company, but all data gathered and analyzed come from the real-world company Divvy, a bike-sharing company owned by Lyft Bikes and Scooters, LLC that makes its usage data available online. Google provided fictional

background information about Cyclistic and its business goals as part of the course corresponding to this capstone project. The conclusions of this report apply to the fictional company Cyclistic only and do not apply to or reflect on Divvy in any way.

Methodology

Prior to this analysis, the past twelve months of cycling trip data were acquired from [Divvy's data repository](#) under Lyft Bikes and Scooters, LLC's [licensing agreement](#). All cleaning, transformation, and analysis steps utilized the R statistical programming language and the RStudio IDE. The datasets were loaded into RStudio and combined into a single dataset spanning one year of time from August 2021 to July 2022. The combined dataset contained 5,901,463 rows of observations.

The data cleaning process entailed first removing rows with missing values. The combined dataset was filtered to indicate how many rows contained NA values in any column. There were 1,272,233 rows with missing values. All NAs were dropped, leaving 4,629,230 rows intact. As part of the quality assurance process, the smallest, original dataset (January) was manually reviewed to find the *ride_id* of an observation that had an NA value in one or more columns. An additional *ride_id* with no missing values was also noted down as a control. Then, the combined dataset was filtered for those two *ride_id* observations to verify that only the observation with the missing values was successfully dropped from dataset.

Then, the dataset was checked for duplicates by comparing the number of rows versus the unique number of values of the *ride_id* variable. No duplicates were present.

Next, to help answer the question of how member type affected bike use, the variable *ride_duration* was added to the dataset. This variable calculates the number of minutes each unique ride, based on *ride_id*, lasted. To account for errors in data entry such as inverted *started_at* and *ended_at* variables, false starts, or docking errors, any ride durations less than or equal to zero were filtered out. This removed an additional 273 rows of data, leaving 4,628,957 rows of observations.

To ensure the data cleaning process did not adversely affect the validity of this report's findings, additional checks were performed. The *ride_duration* variables for both the raw dataset, after negative ride durations were removed, and the cleaned dataset, with NAs also removed, were compared to determine if removing the NAs significantly affected the data. The mean ride times were 19.89 and 18.51 minutes, respectively. Compared to the 1.2 million rows of data removed, a difference in average *ride_duration* of 1.38 minutes suggests that the cleaning process did not significantly alter the integrity of the cleaned dataset.

Finally, to answer several smaller questions such as "What month has the most riders?" and "Which day of the week is busiest?" that guided parts of this analysis, additional transformations were made to the dataset. Three new variables, *weekday*, *month*, and *quarter*, were added to the dataset to help answer these questions.

Findings

Cyclistic's goal is to maximize membership by converting current casual riders into subscribed members. Thus, an important first step was to determine the breakdown of current riders by their membership status. As shown in Figure 1 below, 58% of users over the past year were annual members, compared to 42% of casual riders who purchased either single-ride or full-day passes.

Figure 1: Proportion of Riders by Membership Status
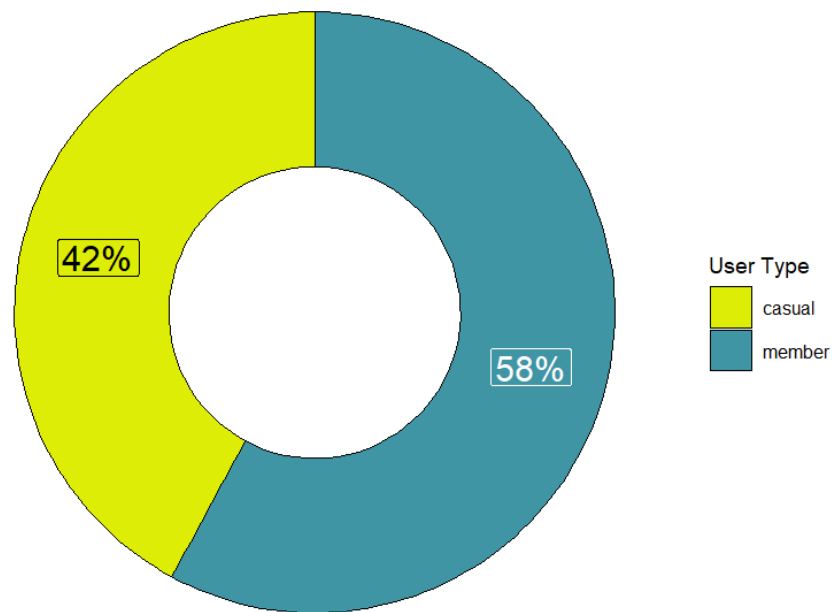
With this basic ratio calculated, the next step was to determine if any significant differences existed between the type of bicycle used and membership status that could suggest patterns in customer preference. Table 1 below contains an important observation—only casual members interacted with the *docked_bike* rideable type. Given the steps taken during the data cleaning process, it is unlikely that this observation came about through error in the data. However, without access to more data or the team responsible for gathering the data, there is no way to determine the significance of this stark disparity. For example, it is unknown if docked bikes were only made available to casual members. Therefore, no additional analysis was performed on this particular subset of data due to the imbalances caused in the data from unknown factors.

Table 1: Proportion of Bike Type Usage by Membership Status

Bike Type	Casual Riders (in Percent)	Members (in Percent)
Classic	37	63
Docked	100	0
Electric	44	56

Next, the number of rides each month was broken down by membership status. Cyclistic operates in Chicago, a city with freezing, snowy winters. Due to this, less rides in winter months were expected across both user types. Figure 2 below corroborates this, showing a sharp decline in total rides starting in November and slowly building up again in March.

In the warmest months, June to September, the average ratio among casual riders to members is 48% to 52%, respectively. In the coldest four months of the year, November to February, the contrast between user types is more pronounced, with only 21.5% of all rides in these months coming from casual riders. Of all months, only August saw more casual riders than members, with 51% of all rides attributed to casual riders as shown in Table 2 below.
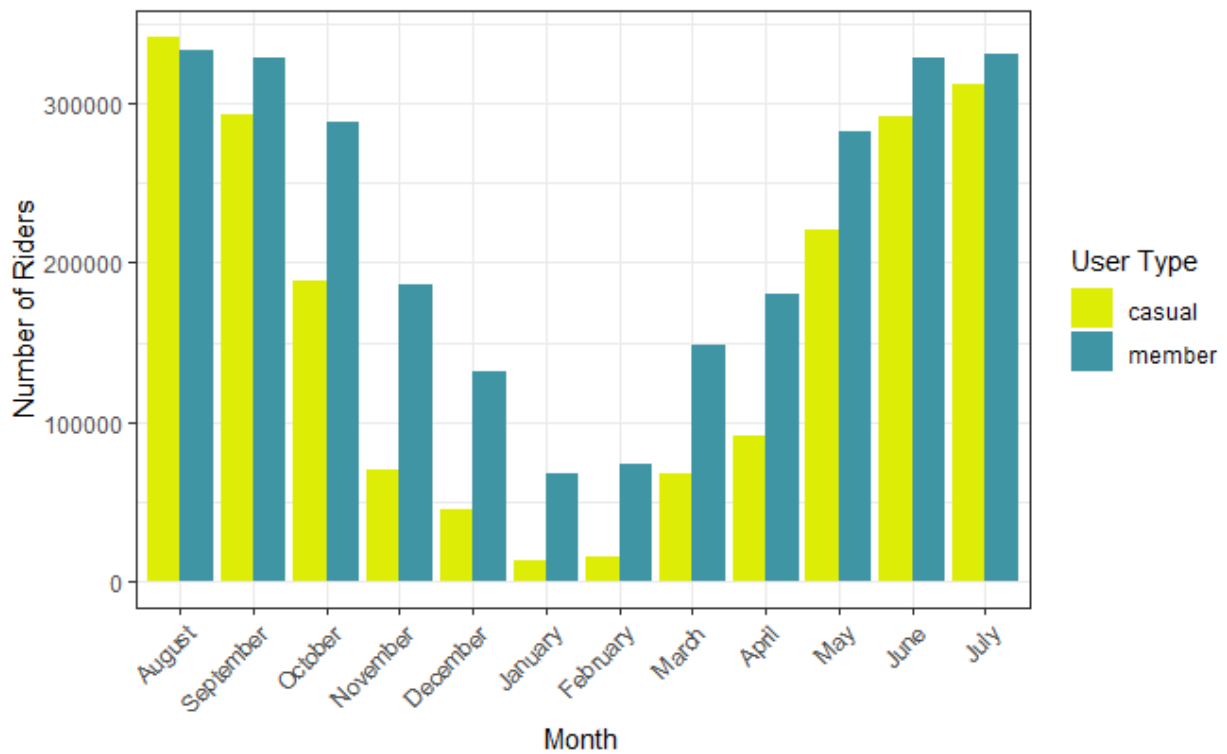
Figure 2: Total Riders each Month by Membership Status

Table 2: Proportion of Riders by Membership Status per Month

Month	Casual Riders (in Percent)	Members (in Percent)
August 2021	51	49
September	47	53
October	40	60
November	27	73
December	26	74
January 2022	16	84
February	17	83
March	31	69
April	34	66
May	44	56
June	47	53
July	48	52

These results suggest that casual riders are utilizing Cyclistic's bikes for leisure purposes and have less interest in riding during less than favorable weather conditions. Additionally, according to U.S. News & World Report, peak tourism season in Chicago is from June to August, the months with the most casual riders. Therefore, it is reasonable to conclude that large portions of casual riders during the summer months are tourists who would see less value in purchasing an annual membership. Other casual riders could be college students who have returned home from university for the summer and do not have access to a car. Shorter-term memberships or flash sales on annual memberships could potentially entice these casual riders to purchase a membership.

With this in mind, it is recommended that Cyclistic diversify membership plans. With additional plans offering three and six month membership terms, Cyclistic would be more likely to convert local, casual riders to seasonal members. These plans could be marketed in spring to attract casual riders when they start to head outside more as the weather begins to warm. Additionally, a one-month plan could be offered during peak tourism season for non-local users wishing to make the most of Cyclistic's services.

Additional analysis broke down ride totals by weekday, separated by user status. Members utilize Cyclistic more often during the week, accounting for 64% of all rides on weekdays, suggesting members use Cyclistic bicycles to commute to and from work. By contrast, casual riders are more abundant on weekends totaling 54% of all rides. Saturday is the most popular day for rides both overall and among casual riders, and Tuesday sees the most rides by members, as shown in Figure 3.

Figure 4 below breaks down the average ride time by user status for each weekday. Regardless of the day of week, casual riders spend, on average, approximately twice as long riding than members. This further supports the conclusion that casual riders more often utilize Cyclistic for leisure purposes.

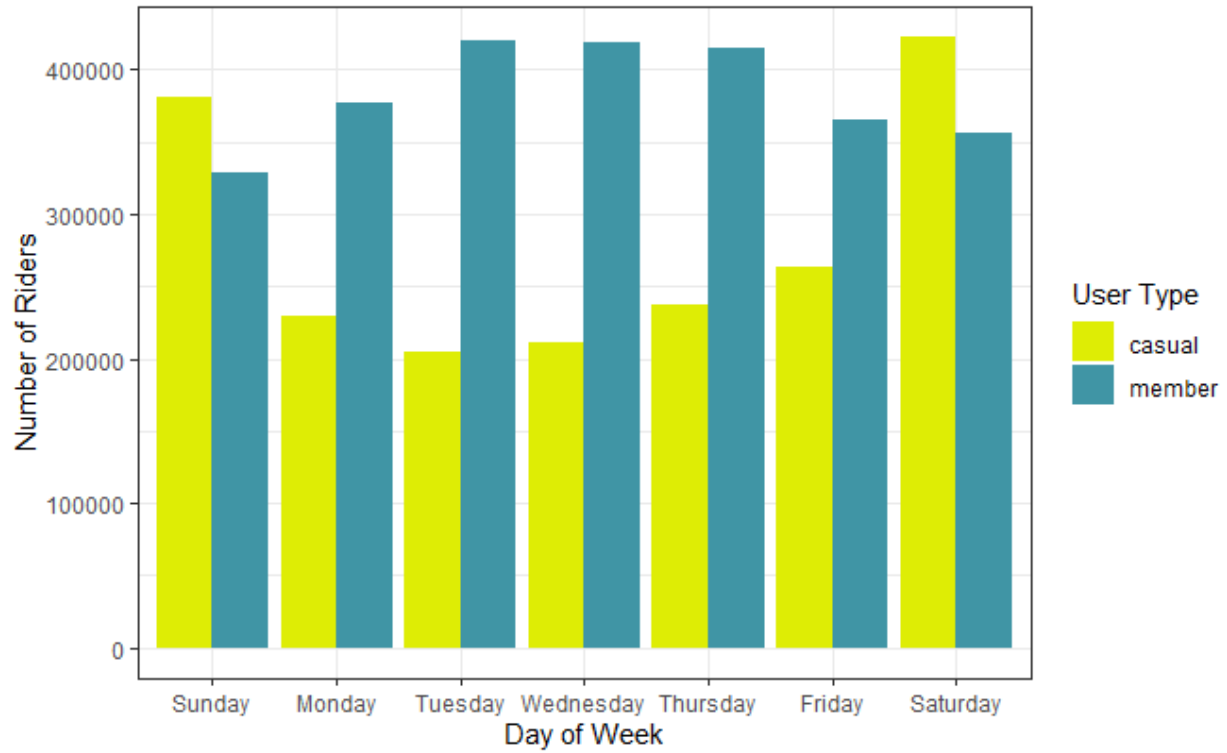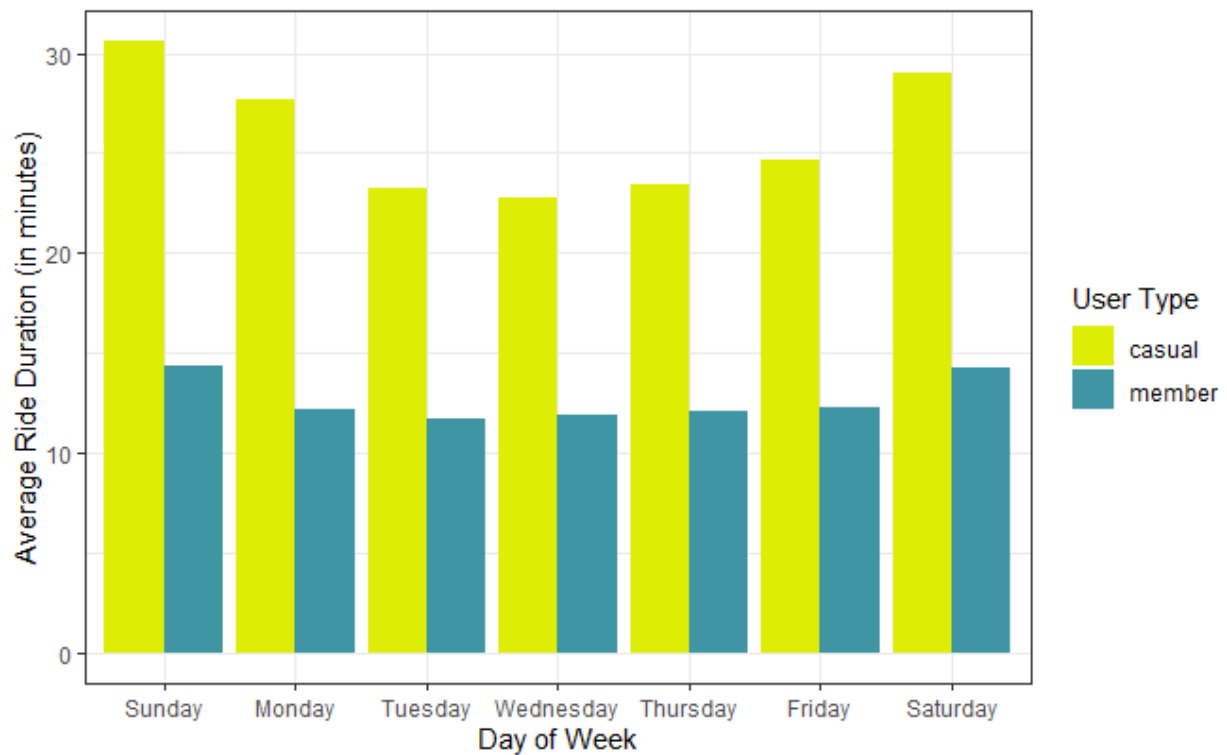*Figure 3: Total Rides per Weekday by Membership Status*

Figure 4: Ride Duration Averages per Weekday by Member Status

Additionally, a look at popular start times shed more light on the varying usage among casual riders and members, as shown in Figure 5. The hours between 4-6pm are, by far, the most popular start times for both casual riders and members but members also see a peak in start times between 7-8am. Not only does this reinforce the previous assertion that casual riders utilize Cyclistic bikes for leisure purposes, it also further supports the conclusion that members are more likely to use Cyclistic for their daily commute than recreationally.

Figure 5: Start Times per Hour by Member Status
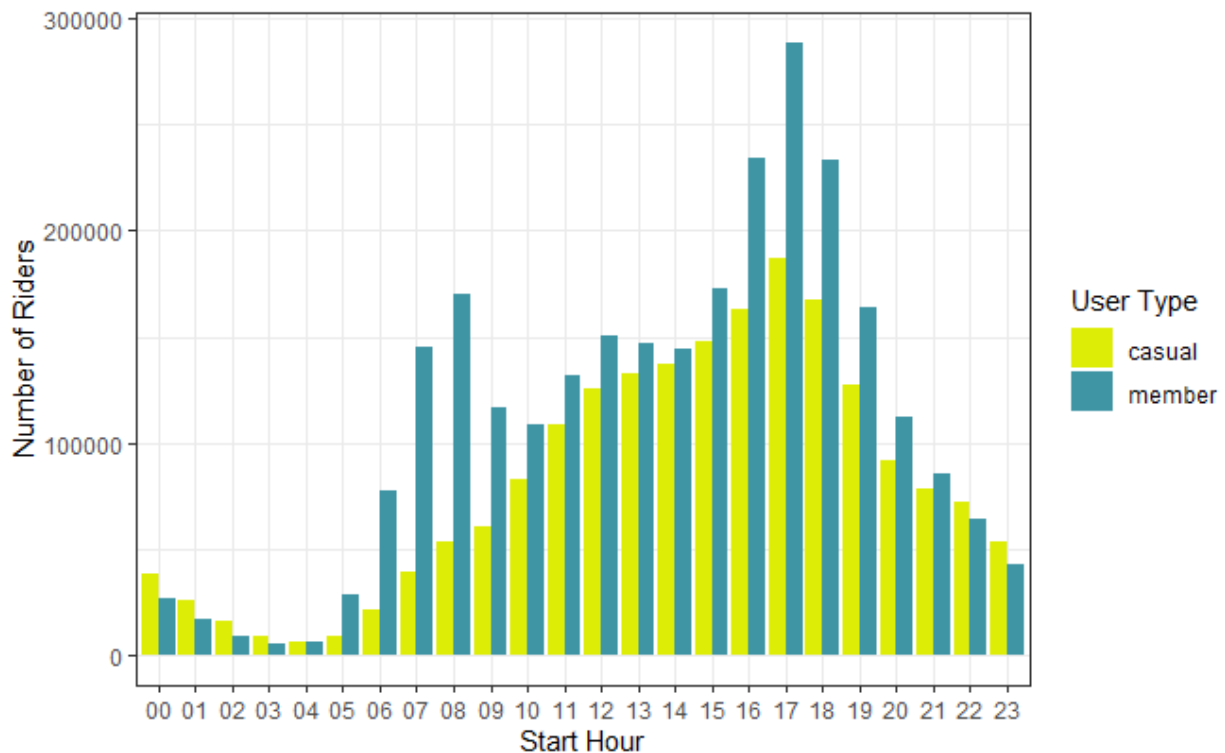
Knowing that members are more likely to use Cyclistic for commuting and that casual riders spend more minutes per ride, the next step is to determine popular start stations to identify any potential pain points in bike availability. Customers utilized a total of 1,306 start stations from August 2021 to July 2022. Start stations were filtered to include only those that had over 10,000 members or 10,000 casual riders over that period. These lists were then combined to determine the most popular stations between both user types, producing twenty-six stations that met these criteria. Table 3 below shows the ten most popular stations of these twenty-six. While not the most popular with members, the Streeter Drive & Grand Avenue station was, by far, the most popular with casual riders, making it the most heavily trafficked station over the year, with more than 77,000 riders in that time.

With this in mind, Cyclistic could introduce member-only benefits aimed at popular stations to ensure members are more likely to have bike availability during peak hours.

This may in turn convince some repeat casual riders to invest in a membership when they otherwise would have relied on single passes only.

Table 3: Popular Shared Start Stations among Casual Riders and Members

Ride Start Station	Casual Riders	Members	Total Riders
Streeter Dr & Grand Ave	60,338	16,683	77,021
DuSable Lake Shore Dr & North Blvd	26,008	17,388	43,396
Michigan Ave & Oak St	26,834	14,277	41,111
Wells St & Concord Ln	17,722	22,252	39,974
Clark St & Elm St	13,945	22,437	36,382
Theater on the Lake	19,606	15,119	34,725
Wells St & Elm St	14,476	19,686	34,162
Clark St & Armitage Ave	14,435	15,831	30,266
Clark St & Lincoln Ave	14,583	14,654	29,237
Wabash Ave & Grand Ave	12,731	16,468	29,199

Finally, Table 4 combines popular start and end stations used by casual riders to identify the ten best stations marketing should target for membership advertisements. Five of these stations saw over 50,000 casual riders each in the last year. Of all the stations that Cyclistic operates, these ten stations should be the focus of an initial round of marketing campaigns for membership conversion. Another twenty-two stations serviced over 20,000 casual riders each in that same period. While these stations do not appear in the table, they could also be the focus of subsequent marketing campaigns.

Table 4: Most Popular Stations among Casual Riders

Station Name	Number of Casual Riders
Streeter Dr & Grand Ave	123,535
DuSable Lake Shore Dr & Monroe St	61,485
Millennium Park	56,866
DuSable Lake Shore Dr & North Blvd	56,236
Michigan Ave & Oak St	55,411
Shedd Aquarium	40,933
Theater on the Lake	40,592
Wells St & Concord Ln	34,893
Clark St & Lincoln Ave	29,343
Clark St & Armitage Ave	29,097

Key Takeaways & Recommendations

Based on the findings above, the following are the key takeaways of this analysis:

- Summer (June - August) is the busiest season. August had the most overall rides, and was the only month with more casual riders than members.
- While weekends gross more rides overall, weekdays see greater member usage.
- Casual riders spend approximately twice as long per ride compared to members.
- Peak start times are between 4-6pm, with an additional peak between 7-8am specific to members.
- These findings suggest casual riders utilize Cyclistic for recreational purposes while members access the service for commuting and day-to-day activities.

These findings point to several recommendations that Cyclistic could implement to increase conversion from casual rider to member status:

- Diversify membership plans. Add three and six month plans for frequent casual riders who typically avoid riding in less-than-favorable weather.
- Add a 30-day pass or a month-to-month membership for non-local riders during peak tourism season and summer break.
- Include additional member-only benefits, such as exclusive membership bikes at popular start stations to ensure members receive priority for their commuting needs.
- Finally, focus marketing campaigns on the most popular stations among casual members.

Limitations

Below, some limitations with the analysis are noted alongside thoughts on how additional data could be utilized to produce further insights into Cyclistic's growth opportunities.

- With the addition of price data for the annual membership, single-ride, and full-day passes, further analysis could determine whether lowering the membership price by a certain percentage would convert more casuals while still increasing profits. This could also provide additional recommendations, such as the introduction of tiered pricing based on membership length, with annual membership providing the most savings for customers.

- The *docked_bike* rideable type should be further researched to determine exactly what it is and why only casual riders interact with it.

- The data were sorted into quarters; however, quarterly findings may have been skewed due to the nonconsecutive calendar year (start end of Q3 2021 and run to beginning of Q3 2022), so they were were not utilized in this analysis. Further data could provide greater context to the seasonal variability of each quarter's data.