Cyclistic Analytics Report

Data Analyst: Ruby Smith
Created On: 9/24/2024
Last Updated: 10/10/2024

Purpose: *Compare & Contrast Annual Members vs. Casual Riders*
The Cyclistic Team is looking to increase annual memberships. Marketing team wants me to compare annual memberships to casual riders over the last year and come up with ideas to convert their casual rides to annual memberships. My goal is to analyze the data and compare casual rides and annual memberships over the last year to help with marketing strategies.

Questions:
Why do people choose casual rides over annual memberships?
Are there any trends when comparing casual rides to annual memberships?
Are there more casual rides on specific days of the week?
Are there people who are casual riders but use Cyclistic all the time?
Are there special geographic locations used more than others for casual riders?
What marketing strategies could be used to convert casual riders to annual memberships?
Rewards, Sign-up bonuses, Referral bonus, Discounts, Customer Feedback and Community.

PREPARE
Key Tasks For Case Study:

- Locate and Analyze data of casual riders and annual memberships to understand trends.
- Create visualizations and show insights into analyzed data about riders and communicate it to all stakeholders (Cyclistic, executive team, Marketing manager, and marketing team).
- Help come up with ideas to maximize the number of annual memberships.

Data:
Data collected by Motivate International Inc.: Index of bucket "divvy-tripdata"
https://divvy-tripdata.s3.amazonaws.com/index.html
Data Useable by License: Data License Agreement | Divvys Bikes
https://divvybikes.com/data-license-agreement
The data itself cannot be shared stand-alone, but the data can be shared with my analysis and report.

Data Organized: 12 .csv files named by Year, Month, and file name (ex. 202403-divvy-tripdata). The .csv files contain data on each bike trip taken during the specified month in the name of the file. Each file contains 13 columns of data in wide format with several hundred thousands of rows. The Primary Key for each dataset is the 'ride_id' column.

Column Names: ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, and member_casual.

Bias: I feel that there is little to no bias in the data. The datasets contain all rides, it is reliable, current, and from the company itself. The only issue is some locations for the starting_station name and id and also the ending_station name and id, there is nothing and is marked as 'null.' Integrity: It looks like all data imported correctly. All .csv files imported contain the same number of columns and are all named the same and have matching data types. Datasets are kept private and only I can have access to them.

Prepare Phase Conclusion and Insights: The datasets contain information that will help me find trends between casual riders and memberships. I have data that shows types of bikes used, how long the rides took, and starting and ending locations.

There are some areas I have noticed that do not have names or id's for starting locations or ending locations and some latitude and longitude is not entered in as well. I will explore the datasets more thoroughly to understand. Possible hypothesis: error reading some locations and possibly not turning in bike at end of route.

I want to combine all data into one place so it can be easily accessed and updated when necessary. I want to get trip time and compare memberships and filter the dataset with other categories. I want to separate the date and time and look for trends in the days of the week. I want to compare latitude and longitude on Tableau for any geographical trends. I also plan to separate categories by bike type as well for any trends.

PROCESS

Tools used for Analysis:
Google Cloud - imported .csv to cloud because of file size
BigQuery - imported .csv from cloud to BigQuery to analyze data with SQL
GitHub - to store queries used in SQL
Google Sheets - to store tables created from SQL queries to use for visualizations
Google Search - to help with query organization when stuck
Notes - from Coursera data analytics course

Changes Made to Original Data:

- Combine September 2023 - August 2024 into one dataset
  rs-project-01-429415.cyclistic_bike_data.annual_202309_202408_tripdata
- Remove latitude and longitude NULLs

- Check for duplicates with the ride_id column, found 171. Remove duplicates and create an updated dataset.
  rs-project-01-429415.cyclistic_bike_data.annual_202309_202408_tripdata2
- Created new column 'trip_time_seconds'
- Separated date from started_time and ended_time
- Created 2 new columns 'start_date' and 'end_date'.
- Create column 'day_of_ride' for weekday names.

Updated Column Names: ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, member_casual, trip_time_seconds, start_date, end_date, and day_of_ride.

Process Phase Conclusion and Insights: I decided to remove the rows that do not have latitude and longitude values. I want to compare locations geographically and without that information, I would be unable to do so. There were 7526 entries out of 5.6 million (<0.2%). I felt that with what I would like to analyze, taking out less than 0.2% of incomplete data would be ok and it would all still be in original data stored. I have removed 171 duplicated as well and restored integrity to the data. I have completed the tasks I wanted to do from the Prepare Phase and in doing so created 4 new columns.

ANALYSIS

GitHub Link For All SQL queries: Case_Study_Cyclistic
RubyRene90/Case_Study_Cyclistic: Scenario: non-membership vs membership (github.com)

Part 1: 1_Understanding_DataSet
Part 2: 2_annual_tripdata_cleaning
Part 3: 3_annual_tripdata_analyzing
Part 4: 4_annual_tripdata_analyzing2
Part 5: 5_annual_tripdata_analyzing3

Google Sheets Document: *'202309_202408_CyclisticAnalyzedData'*
https://docs.google.com/spreadsheets/d/1DlsFo9a9v79fZ12HbLIfPqG-iBwvY2XONY1JVraHPBs/edit?usp=sharing

Sheet 1: Annual
Sheet 2: Monthly
Sheet 3: Weekday
Sheet 4: Quarterly
Sheet 5: Time_Intervals
Sheet 6: Popular_Locations
Sheet 7: Q2_Popular_Locations

Beginning analysis phase by running **annual** aggregate functions. When running MIN on the 'trip_time_seconds', the answer that came back was negative. Also, when running MAX, the answer was over 24 hours. I decided since both results were small when researched, I pulled all

data for any time that was negative (344 results) and over 24 hours (389 rows) and removed them from the cleaned dataset. I created a Table on Sheet 1: Annually with 3 rows and 8 columns.

Column Names: rider_status, rider_count, percent_of_riders, annual_avg_minutes, annual_min, annual_max_minutes, classic_count, electric_count

Continued with running **monthly** aggregate functions. I created a Table for casuals on Sheet 2: Monthly with 24 rows and 9 columns.

Column Names: month, rider_status, rider_count, percent_of_rider_type, monthly_avg_minutes, min, max_minutes, classic_count, electric_count

Continued with running **weekly** aggregate functions. I created a Table on Sheet 3: Weekday with 14 rows and 10 columns.

Column Names: weekday, total_weekday_count, rider_status, rider_count, percent_of_total, weekday_avg_minutes, min, max_minutes, classic_count, electric_count

Continued with running Count and Average for **quarterly** aggregate functions. I created a Table on Sheet 4: Quarterly with 56 rows and 5 columns.

Column Names: rider_status, weekday, rider_count, quarterly_avg_minutes, quarter

I checked different trip time intervals to see any differences between memberships and casuals. I created a Table on Sheet 4: Time_Intervals with 8 rows and 4 columns.

Column Names: time_interval, membership, casual, all_riders

I queried for the top 20 starting and ending stations. I was able to get the starting and ending latitude and longitudes for each of the top results. I created a Table on Sheet 5: Popular Locations with 80 rows and 6 columns.

Column Names: rider_status, location, station_name, rider_count, latitude, longitude

I also queried the top 20 starting and ending stations for the 2nd quarter (December - February). I created a Table on Sheet 6: Q2_Popular_Locations with 80 rows and 6 columns.

Column Names: rider_status, location, station_name, rider_count, latitude, longitude

Analysis Phase Conclusion and Insights: After running queries in SQL and creating more manageable tables in google sheets, I can see trends in the data between September 2023 - August 2024. Membership rides currently count for 65% of all rides annually. Casual riders (non-membership riders) have a higher annual average in ride time by a 9 minute difference

compared to membership riders. Also, there is no significant difference between all_riders, membership, or casual riders when it comes to using classic or electric bikes.

When separating the data on a monthly basis, I noticed both membership and casual ride count drop in the winter months between December 2023 and February 2024. The casual ride count drops more significantly than membership ride count. Also, the casual average ride time also drops to more closely resemble membership ride time average.

After separating and analyzing my data on a weekly basis, I noticed membership rider count was higher on weekdays than weekends while casual rider count was the complete opposite, higher on the weekends and lower on the weekdays. Both membership riders and casual riders have an increased trip time average on the weekends.

I also separated my data into quarterly aggregates to see if there was any difference between seasons (quarters and seasons matched for the most part). I noticed that rider_count was higher for both membership riders and casual riders in June - August. Both counts dropped in December - February but casual rider count dropped significantly everyday of the week. The average trip time stays about the same throughout the year with the exception of December - February, there is a small decrease in average trip time between both rider types.

I decided to separate the ride count into different time intervals. The majority of membership ride time is between 1-5 min and 5-10 min whereas the majority of casual ride time is between 5-10 min and 15-30 min. There is also a larger number of casual riders who have taken bike rides over 1 hr than member riders.

I analyzed the top 20 station locations for starting and ending locations. I noticed that a lot of membership trips and casual trips did not share the same station locations. I was able to pull the latitude and longitude for each station location to hopefully use to visualize data geographically.

After noticing a complete drop in casual rider count for 2nd quarter (December-February), I decided to also analyze the top 20 station locations for 2nd quarter to see if there were any differences. The results were more similar to overall membership station locations than overall casual station locations.

SHARE

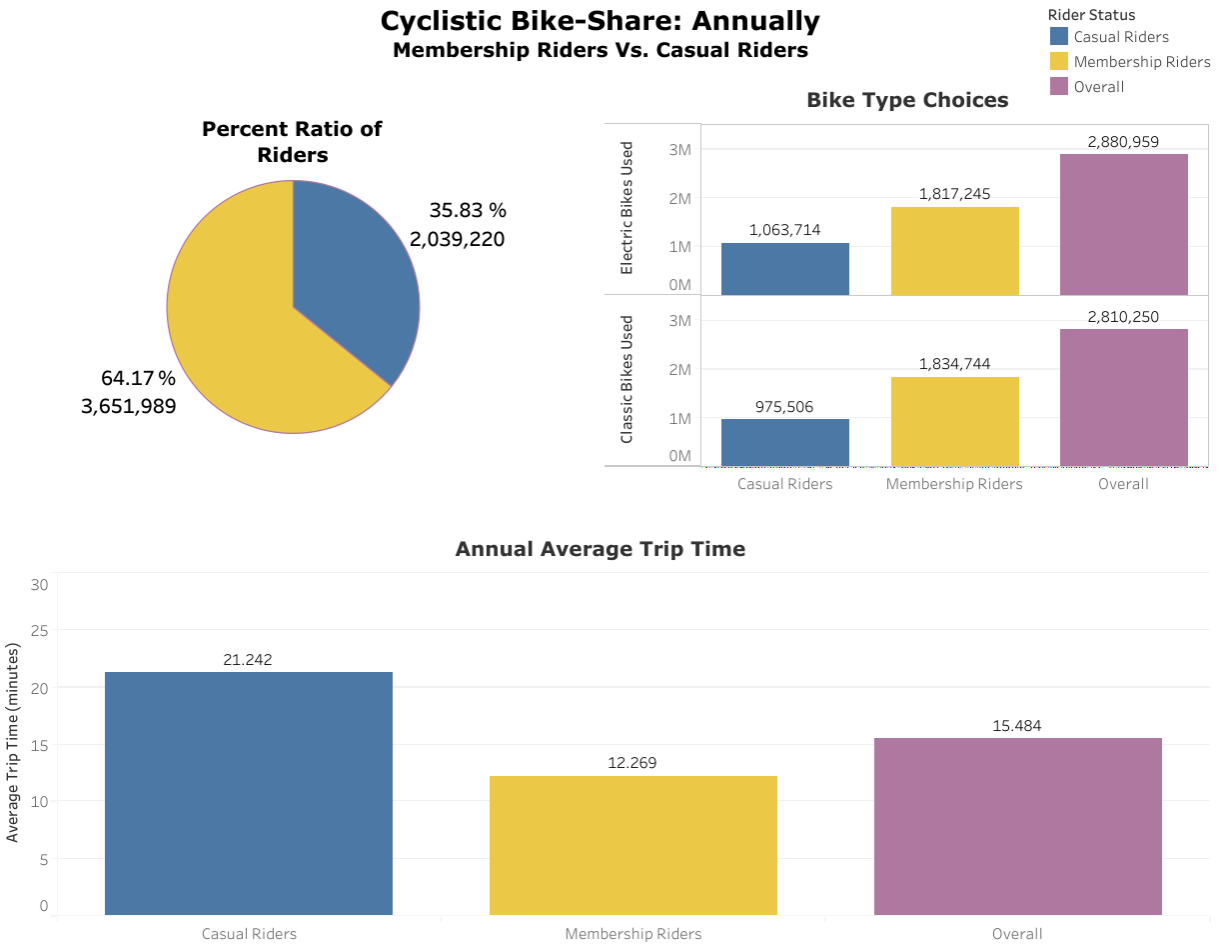
After organizing and formatting data, I imported my google sheets document, *'202309_202408_CyclisticAnalyzedData'* **https://docs.google.com/spreadsheets/d/1DlsFo9a9v79fZ12HbLlfPqG-iBwvY2XONY1JVraHPBs/edit?usp=sharing**,

to Tableau Public to create visuals for the analyzed data.
https://public.tableau.com/app/profile/ruby.smith/vizzes

4 Vizzes available:
Cyclistic-Location Popularity
Cyclistic Annually, Monthly, and Weekday Dashboards -Tabs
Cyclistic Time Intervals
Cyclistic Quarterly - Tabs

I saved the vizzes as .jpeg files and used them to create a Case Study Slide Show.
Google Slides *'2024_Cyclistic_Case_StudyRS'*
https://docs.google.com/presentation/d/193xSKJ1-WgIpOcSgydBVMSnMISDIsP1sYIa5yYSAQ9o/edit?usp=sharing

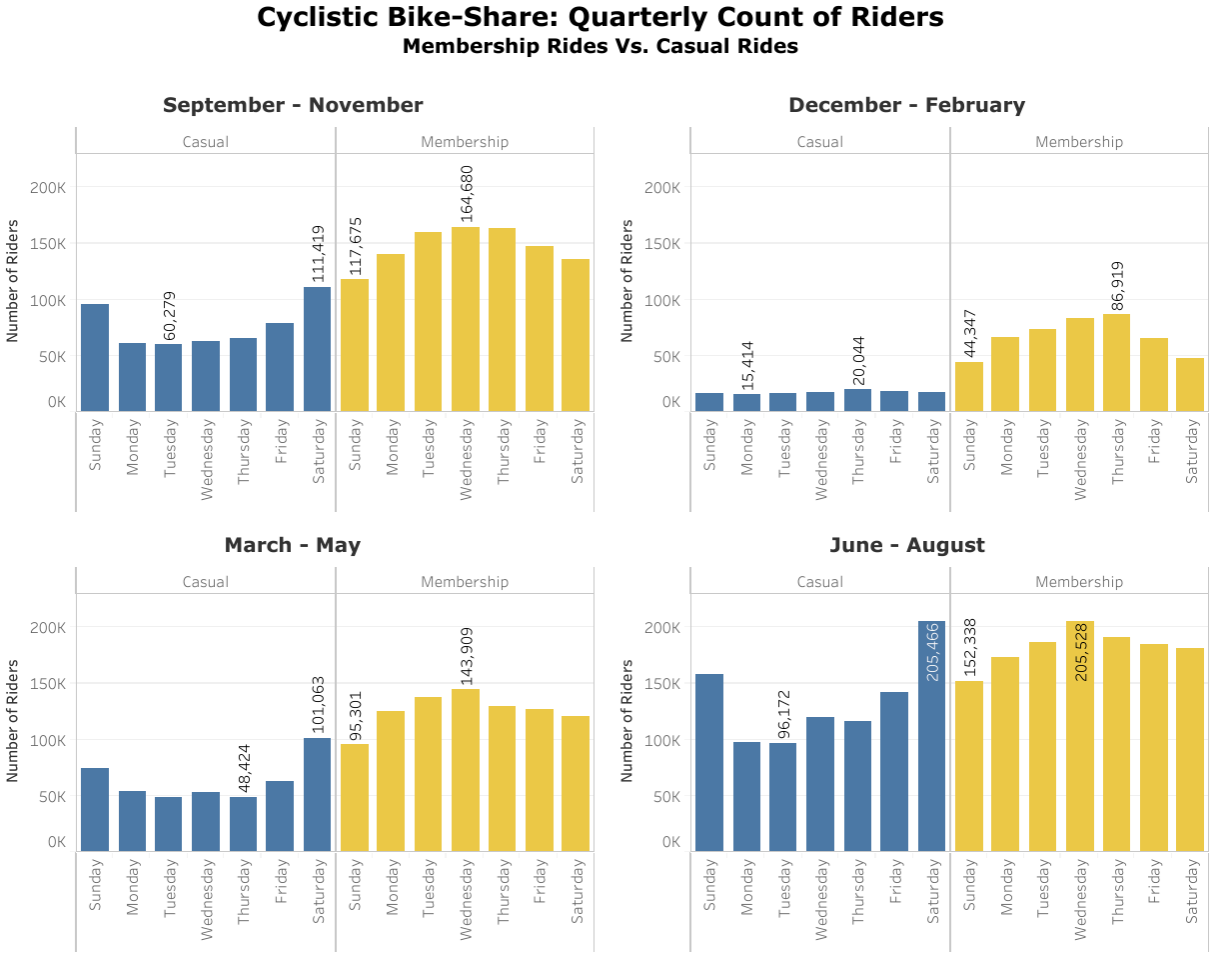
Included are the visuals I used and the 'speaker notes' I wrote to remember what I wanted to say not including talking about the visuals or breaking down analysis.



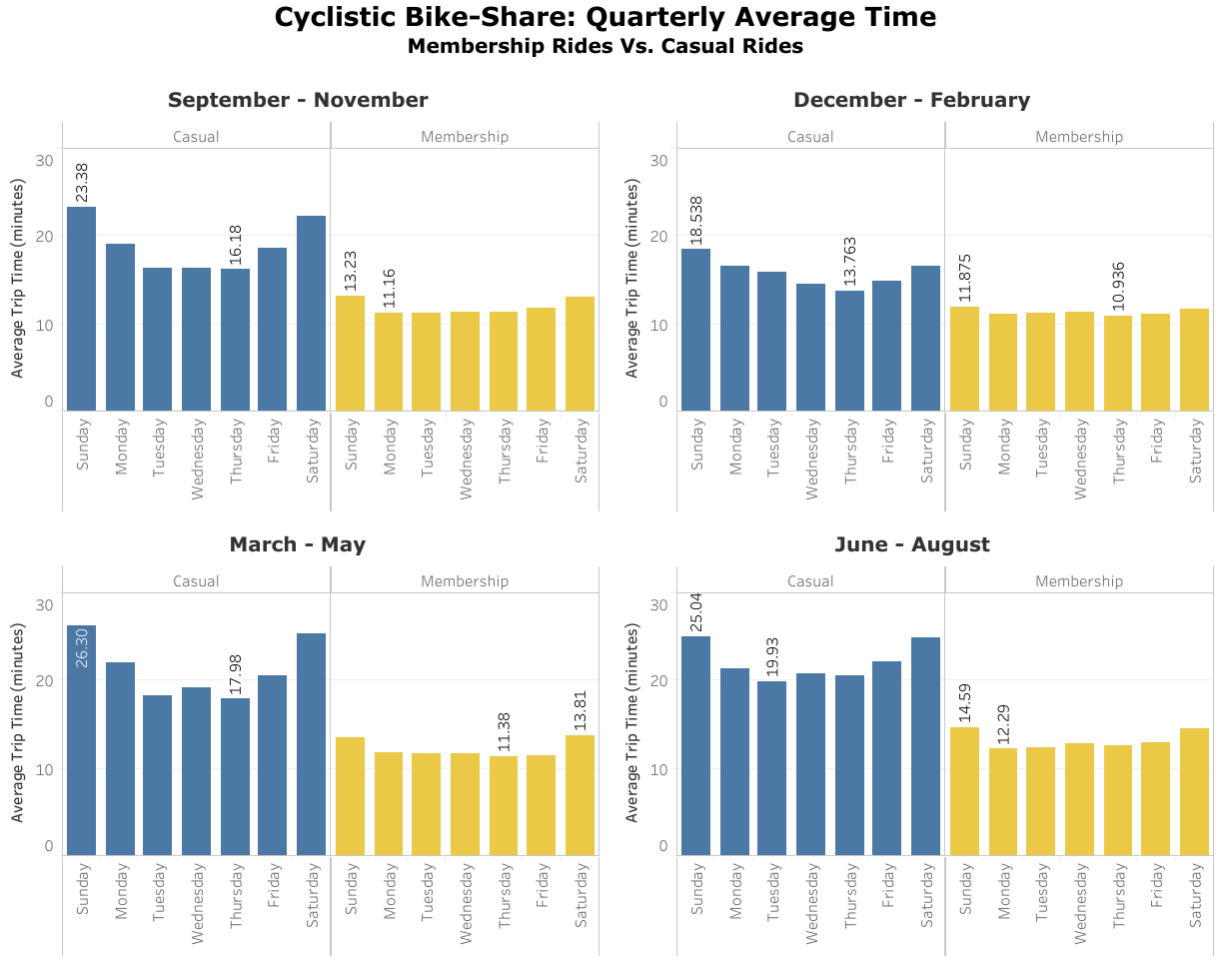
I saw that during Sept - Aug, 65% of Cyclistic riders had an annual membership and about 35% were casual users who rented bikes each time they used them. There are two bike choice options, the classic, manual bike, and the electric bike. There was no significant difference in choice but a slight preference in electric bikes for casual riders. I did break down the bike type more periodically but the results were the same.

I wanted to understand if there were any trends in the length of time of each trip. I looked at the average time of each trip and noticed that on an annual basis casual riders rode their bikes an average of 9 minutes more each trip compared to membership riders. Riders with annual memberships only rode for a little over 10 minutes whereas casual riders tend to ride 20 min at a time on average. (I feel that if members are averaging only a little over 10 minutes, then more

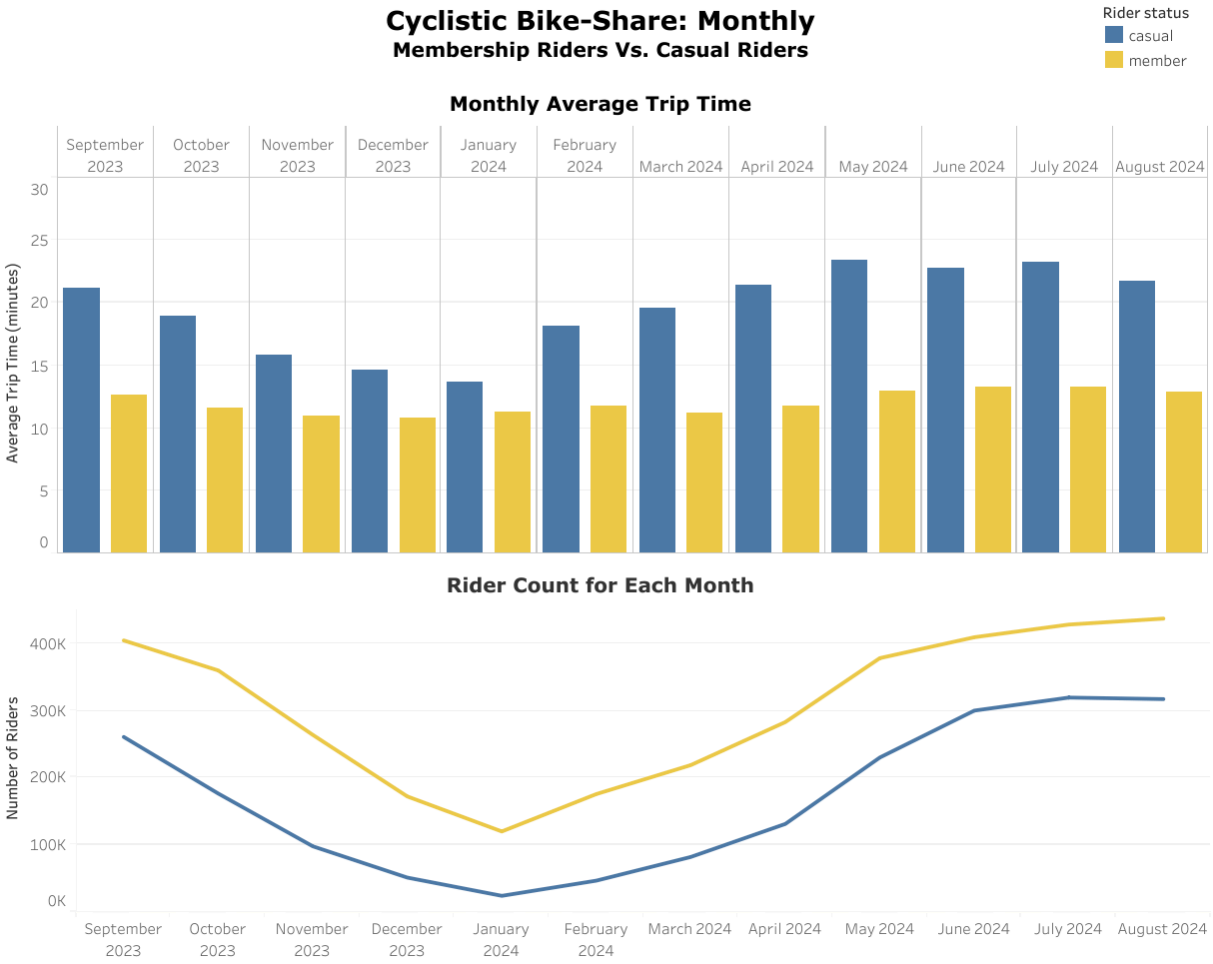
riders are using it for shorter distance destinations. I think of jobs possibly, visiting people, running errands, school, etc. Casual riders are averaging a little over 20 minutes so they take longer bike rides. I think of exercising, site-seeing, school, visiting people, etc.)



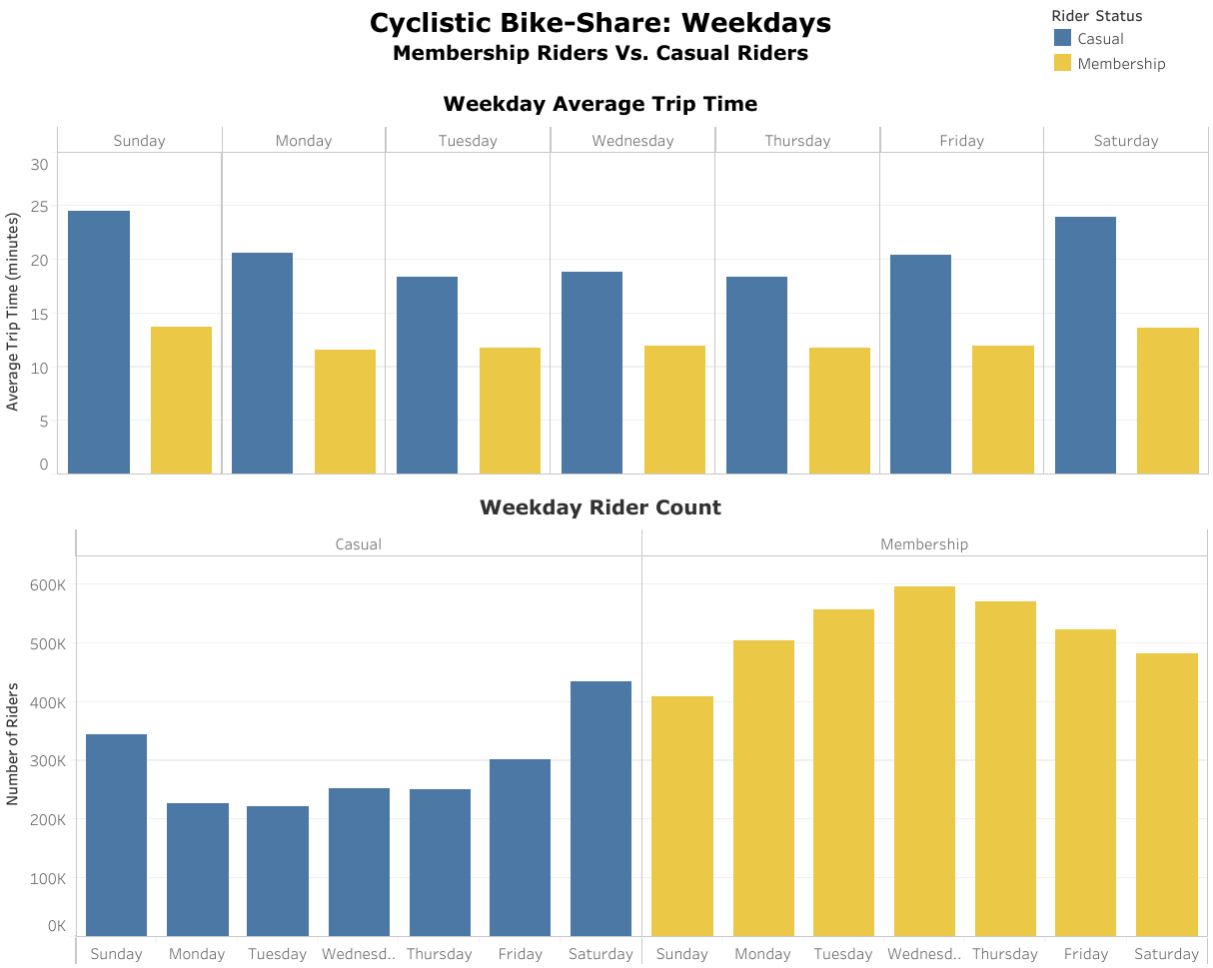
When separating my year into quarters, I noticed each quarter nicely fit into each season for the most part – Fall, Winter, Spring, and Summer. I have organized each quarter into days of the week as well to see if there were any differences. I noticed throughout the year riders with annual memberships use Cyclistic Bike-Share service throughout the week more and casual riders keep the weekends busier. December - February, during the coldest months, is the slowest quarter with the least amount of trips taken. Casual trips taken are very low and there is a decrease in annual membership trips as well. June - August , the warmest months, are when most riders take trips. Casual trips are very high during the weekends and higher during the weekdays. Membership trips taken are at the highest during the summer quarter. (After looking at some geographic data, I wanted to break things down into quarters. I was able to organize by season too and I am glad I did. Fall seems busy, Winter is DEAD, Spring is picking back up, and summer is full blown busy. I wonder why it's dead in the winter for casuals and why it's extra busy on weekends for casuals. It being busier on weekdays also solidifies members using services more for job or school needs. Something more routine. I wonder if less people exercise outdoors in the winter or if its something else that has numbers way down in casuals)



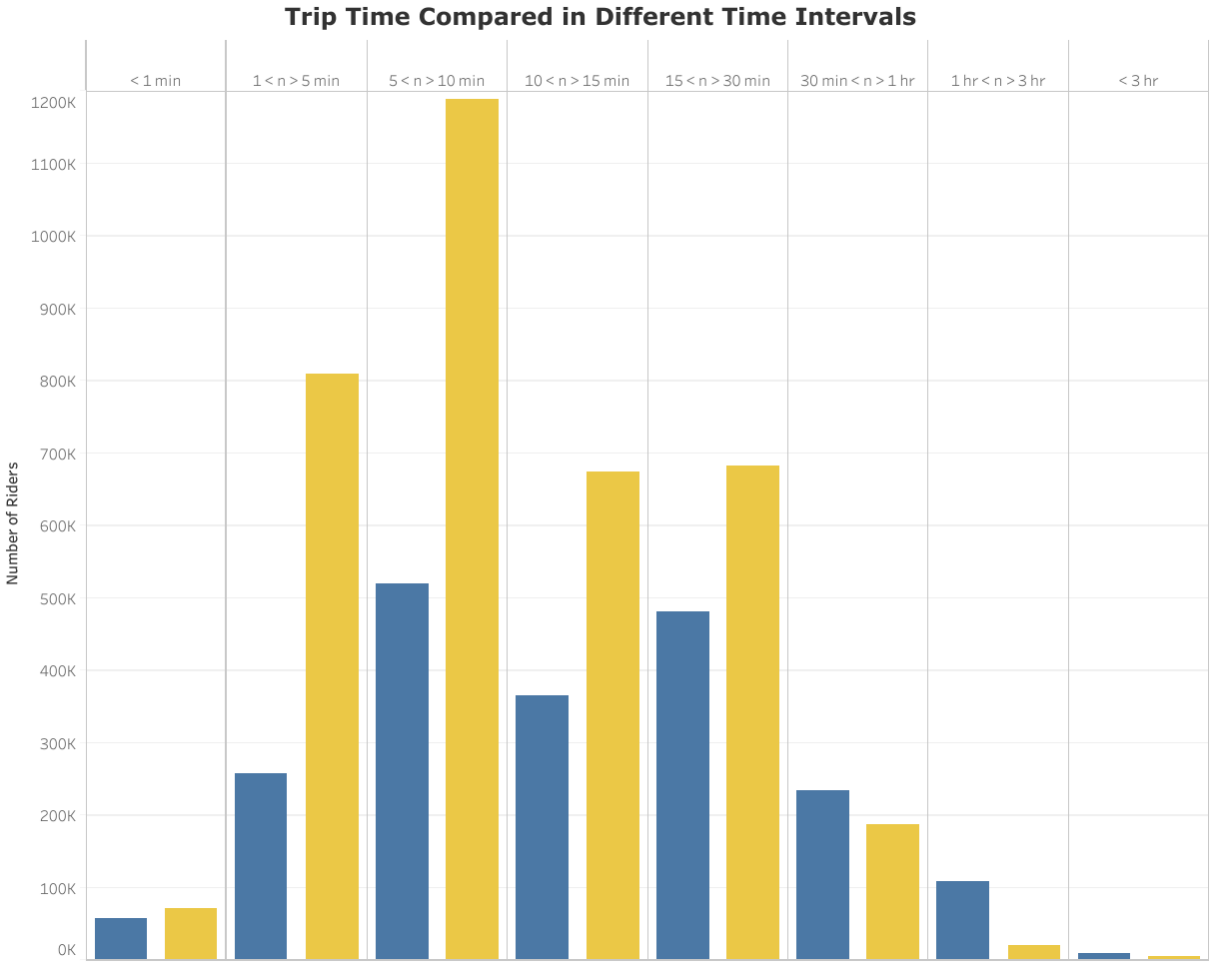
Quarterly, I again looked at the average length of time of each trip. Knowing that Dec - Feb was the least busy quarter and June - August was the most busy quarter, I wanted to see if there were any differences in trip time. Membership riders' average trip time stays about the same throughout the whole year whereas casual riders decrease trip time in the winter months and increase trip time in the summer months. (Seeing quarterly average broken down again solidifies my feel for members using the bikes for more routine things, it remains the same all year long. Casuals basically stays the same with a slight decrease in average in the winter months)



We are looking at each month separately here. As you can see riders with annual memberships on average take trips that are about ~12 min throughout the year, it doesn't fluctuate too much. Casual riders on average take longer bike trips than membership riders. Their average starts around 20 min or above most of the year. It does drop down Nov - Feb, but we did notice that when it was broken down by quarter as well. We also noticed that there were more trips taken in the summer and less in the winter. We can see that the trend follows through smoothly when broken down each month. (I feel bike rentals follow the seasons whether riders have a membership or not. I still feel that people with memberships are using bike rentals more on a routine and going less distance and casuals use bikes less on routine or for exercise and go longer distances.)



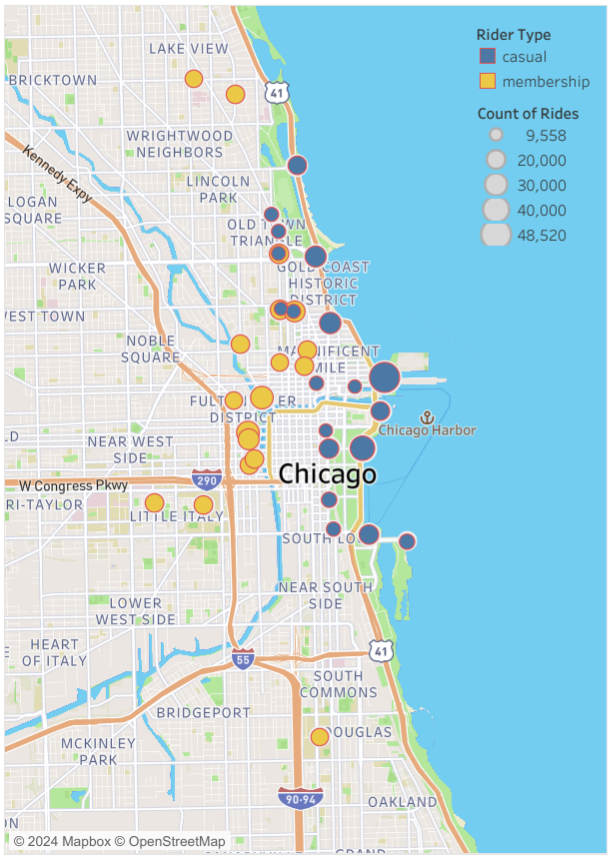
I separated data on a weekly basis and right away noticed that no matter the day, annual membership riders maintain their average trip time. They keep to between 10 - 15 minutes but their are significantly more riders during the weekday than compared to the weekend. Casual riders, on the other hand, have a higher average trip time on the weekend and maintain a higher average trip time every day of the week when compared to members. Casual riders are more prominent on the weekend and less active on weekdays. (When I look at the data over a weekly basis, I feel that membership riders cover the weekdays and the weekends pretty fully. I hypothesize that these membership riders use rentals to get to work or school on weekdays and use it on the weekends for jobs or exercise. I hypothesize that casual riders use rentals more for exercising or doing things that are further distances away.)



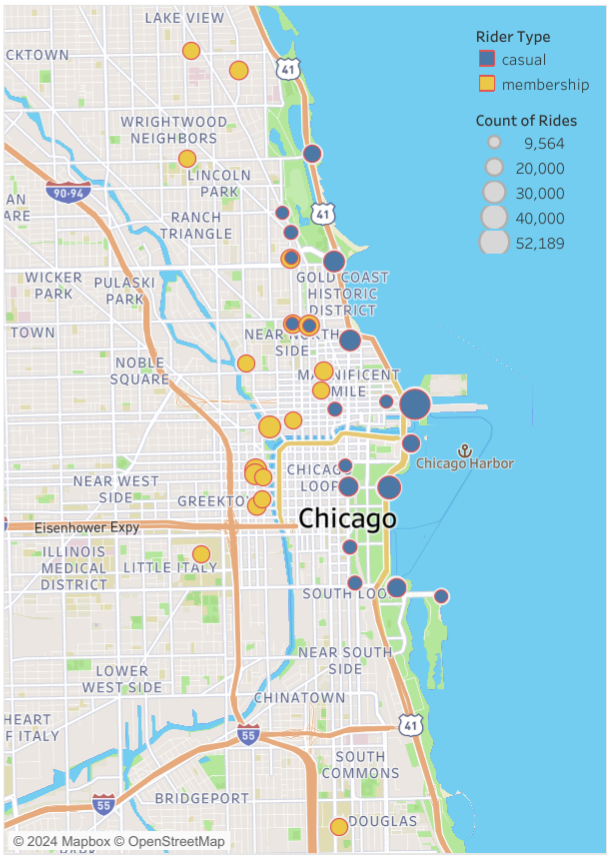
I decided to separate the ride count into different time intervals. The majority of membership ride time is between 1-5 min and 5-10 min whereas the majority of casual ride time is between 5-10 min and 15-30 min. There is also a larger number of casual riders who have taken bike rides over 1 hr than member riders. (When breaking it down like this, you can see that the majority of riders with memberships use it for short periods of time. There are a good amount of casuals that have the same ~10 min average, but why are there more casuals percentage wise having longer trip times?)

Cyclistic Bike-Share: Geographic Comparison of Top 20 Station Locations
Membership Riders Vs. Casual Riders

Starting Location Popularity

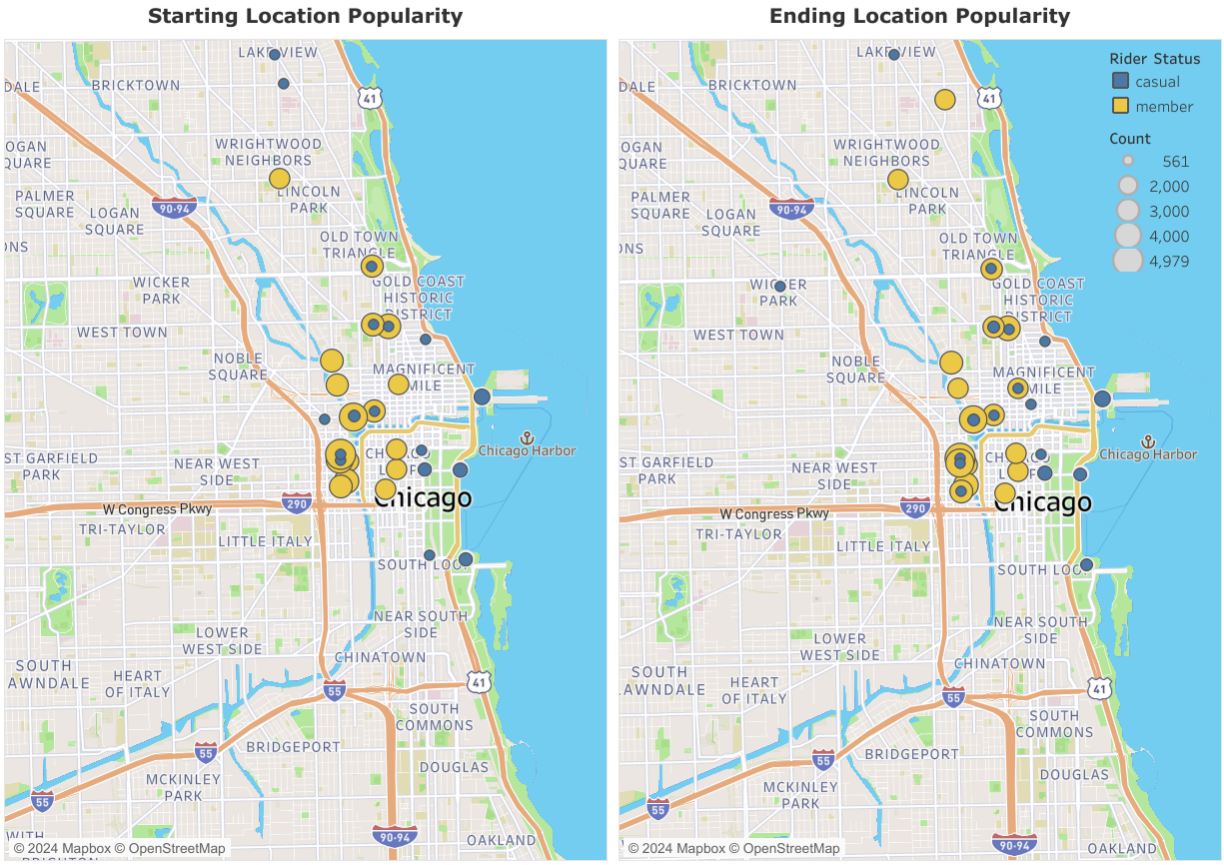


Ending Location Popularity



Lastly, I wanted to compare popular station locations and see if I could map them out. I noticed that riders with annual memberships and casual riders did not share a lot of starting and ending locations. Most membership riders stay more inland. Popular locations are around residential areas, businesses, train station, hospitals, and colleges. Most casual riders stay closer to Lake Michigan. Popular casual stations are around things to do such as, beaches, theaters, museums, Millennium Park, Navy Pier, Soldier Field, etc. (Looking at the data geographically, I can see that casual riders are going to fun places all along the lake. Casual riders are less frequent than membership riders so i assume its for more of an outing or special occasion. I would assume then that most casual riders are tourists who are sight seeing or people who are going out for a special occasion or something.)

Cyclistic Bike-Share: Geographic Comparison of Top 20 Station Locations
Quarter 2: December - February
Membership Riders Vs. Casual Riders



I was able to get the most popular station locations during the 2nd quarter, December - February and I noticed that casual riders have a lot of similar station locations to membership riders during this time period. Casual trip time average was a little higher but more similar to membership riders during quarter 2 as well. (After going back and looking at Q2 geographically separately, I feel it makes more sense for most casual riders to be tourists or people getting on on special outings. During the winter the about the same number of trips occur everyday and it's a small number closer to member average in trip time. Also, geographically, routes are more similar than in the other months. I feel these casual people are more likely to be Chicago residents or people that use Cyclistic Bike-Share more on a normal basis.)

ACT

After analyzing Cyclistic bike riders and looking at differences in how they use the Bike-Share, I have a few possible **recommendations.** By analyzing our data we can see that casual riders frequent leisure areas and are most likely people going out on special occasions or tourists visiting a big city. We can separate those riders from our other casual riders that more closely resemble annual membership riders and market to them. We can see what are the most popular station locations and advertise to casual riders in those areas. The metra train station is also a very popular area. When we looked at the geographic data for Q2, we can see it is popular for casual riders as well. I feel advertising near here will help create more annual memberships and

also casual riders as well who wish to try out Cyclistic Bike-Share. Lastly, my final recommendation to possibly advertise and offer during the slow, winter months the gift of an annual membership. Anyone can pay for a membership ahead of time and gift it to someone to use for the year. This could be helpful for parents looking for a gift to make their kids' life easier in college or a spouse helping their other spouse be able to get to work more efficiently. I feel all of these ideas would lead to an increase in casual riders to become annual membership riders.

Top 3 Recommendations

1. Look into stations that are not located at ‘sight seeing’ or ‘leisure areas’ to market to casual riders.
2. Promote marketing around metra train stations for new casual riders and potential annual memberships.
3. Offer annual memberships as a gift idea during winter holiday months.