Data Analysis Summary: Cyclistic Marketing

Introduction

Taken from overview.pdf

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

Preface

This is a sample case study provided by Google Career Certificates as a part of the Google Data Analytics Professional Certificate.

Due to the nature of this being an independent project with no open communication for inquiries, several assumptions need to be made:

1.) The data is taken from a real-world company Divvy, so we'll use their business model to represent the fictional company of Cyclistic.
2.) As this is a publicly available dataset, private information is not available for internal analysis.
3.) Only data from April 2020 to June 2021 is available, so we can only analyze that time frame.
4.) We'll assume this their population data from that time frame rather than a sample.
5.) Major real-world events taking place during that time frame and likely would have had an impact will be considered.

Ask

Main Issue:

How can Cyclistic's marketing strategy be improved to have casual riders buy more memberships?

Since it was previously established that memberships are more profitable for the company than casual riders, the general plan would be to maximize the number of members by converting casual rides. In addition, since the data given is limited and more of this type cannot be gathered and there is not an expert/specialist to consult, our recommendations will likely be focused solely on what is within this data and any external data without much guesswork.

Beginning Questions

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders buy Cyclistic annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members?

Additional Information

- Cyclistic offers bikes that are more inclusive to those with disabilities and/or cannot use standard two-wheeled bikes
- Many used the bike share for leisure, but around 30% use it for commuting

Prepare

Summary

The overall data provides various details about Cyclistic's individual bike shares from April 2020 to June 2021 in Chicago, USA. Various information such as the place the ride started, membership type, and time started are recorded. However, some information is not readily available, such as personally identifiable information that could be helpful for internal analysis.

Lastly, data can provide more data, such as finding out the time spent on a bike share using the starting and ending time, meaning that several variables need to be planned and will be listed below.

Files Provided

- overview.pdf
- 202004-divvy-tripdata.csv
- 202005-divvy-tripdata.csv
- 202006-divvy-tripdata.csv
- 202007-divvy-tripdata.csv
- 202008-divvy-tripdata.csv
- 202009-divvy-tripdata.csv
- 202010-divvy-tripdata.csv
- 202011-divvy-tripdata.csv
- 202012-divvy-tripdata.csv
- 202101-divvy-tripdata.csv
- 202102-divvy-tripdata.csv
- 202103-divvy-tripdata.csv
- 202104-divvy-tripdata.csv
- 202105-divvy-tripdata.csv
- 202106-divvy-tripdata.csv

Programs Used

- Microsoft Excel
- RStudio
- Tableau

Variables Within the Data

Variable	Data Type	Description
ride_id	Identifier	Unique ID given to each bike share session
rideable_type	Categorical Nominal	The type of bike share ride. All sessions use docked bikes.
started_at	Date Time	When the customer started their bike share session
ended_at	Date Time	When the customer ended their bike share session
start_station_name	Categorical Nominal	The name of the station where the bike share session started
start_station_id	Categorical Nominal	The internal ID of the station where the bike share session started
end_station_name	Categorical Nominal	The name of the station where the bike share session ended
end_station_id	Categorical Nominal	The internal ID of the station where the bike share session ended
start_lat	Continuous Numerical	The latitude of the station where the bike share session started
start_lng	Continuous Numerical	The longitude of the station where the bike share session ended
end_lat	Continuous Numerical	The latitude of the station where the bike share session ended
end_lng	Continuous Numerical	The longitude of the station where the bike share session ended
member_casual	Categorical Nominal	Whether or not the customer is an annual member with Cyclistic or not at time of the bike share session

Variables Created from Data

Variable	Data Type	Description
date_started_at	Date	The date when the customer started their bike share session
time_started_at	Time	The time when the customer started their bike share session
day_started_at	Ordinal Nominal	The day of the week when the customer started their bike share session
date_ended_at	Date	The date when the customer ended their bike share session
date_ended_at	Time	The time when the customer ended their bike share session
date_ended_at	Ordinal Nominal	The day of the week when the customer ended their bike share session

These variables will be later added in the Process section

Process

Data Cleaning Overview:

1. Merge Several Data Sources
2. Create New Columns
    a. Split dates and times
    b. Columns based on day of the week
    c. Column based on time spent on bike share
3. Clean the Combined Data
    a. Remove NA Values
    b. Remove Duplicate Values
    c. Remove Outliers (Geolocation)
    d. Remove Outliers (Date/Time)

Note:

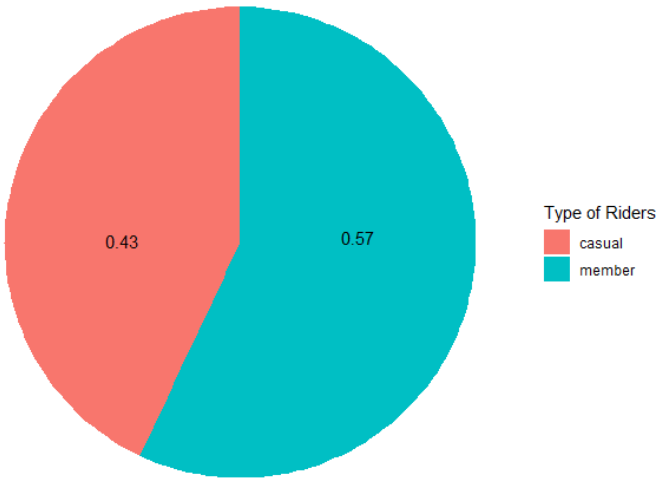
All the data cleaning was done in RStudios. For a more detailed overview of the steps taken, please refer to the R file *cleaning.R* in the R folder.

The cleaned .csv file of the data is named *td_cleaned.csv* in the Data folder.

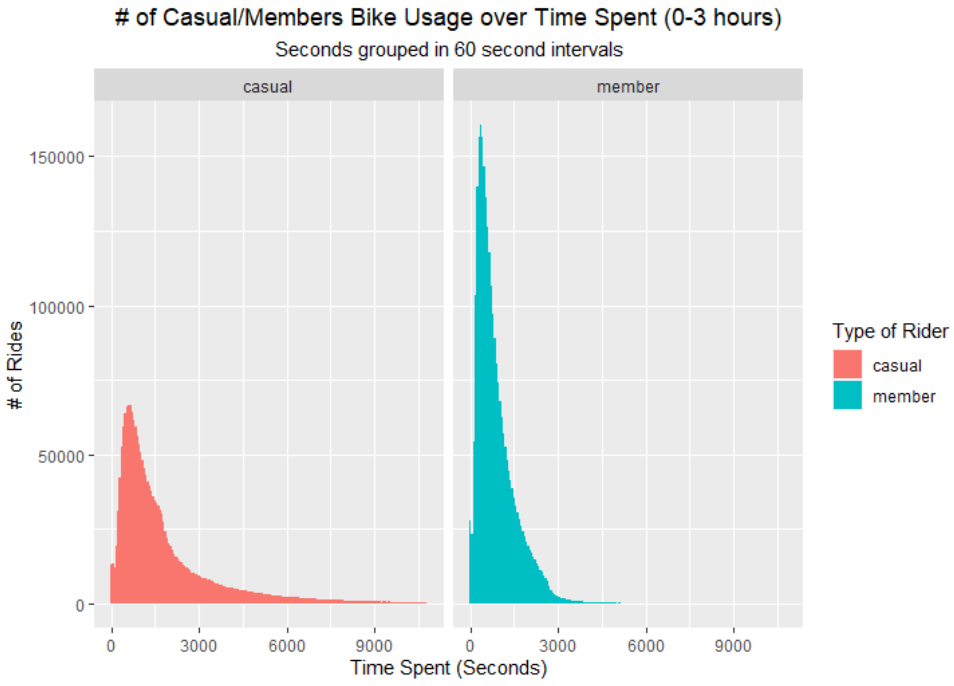
Analyze

For specific code, please refer to the R file *analysis.R* in the R Folder, *stations.twb* in the tableau folder, and *station_frequency.xlsx* in the data folder.

Percentage of Total Rides

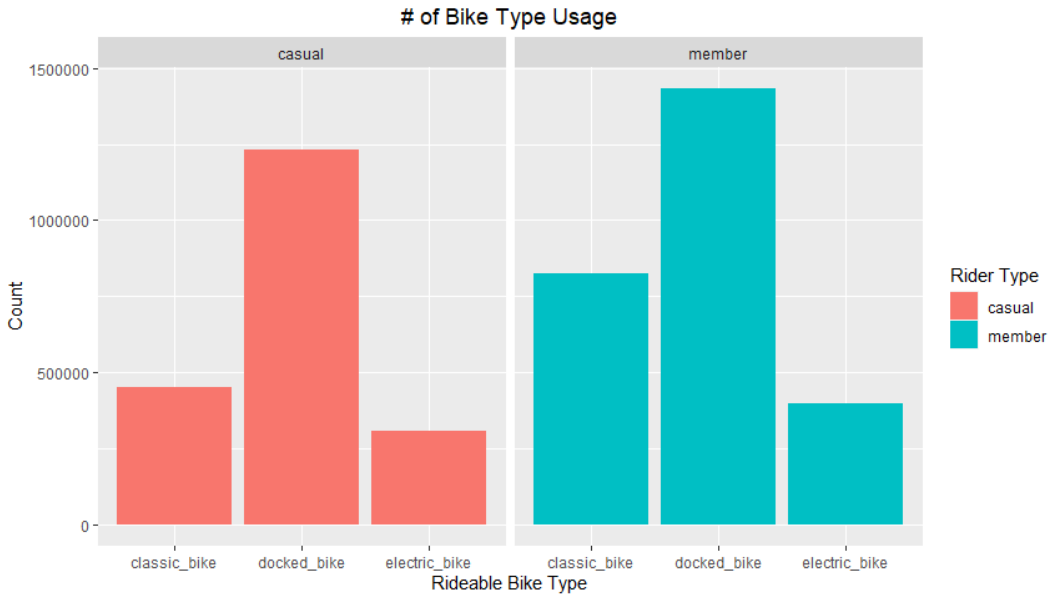


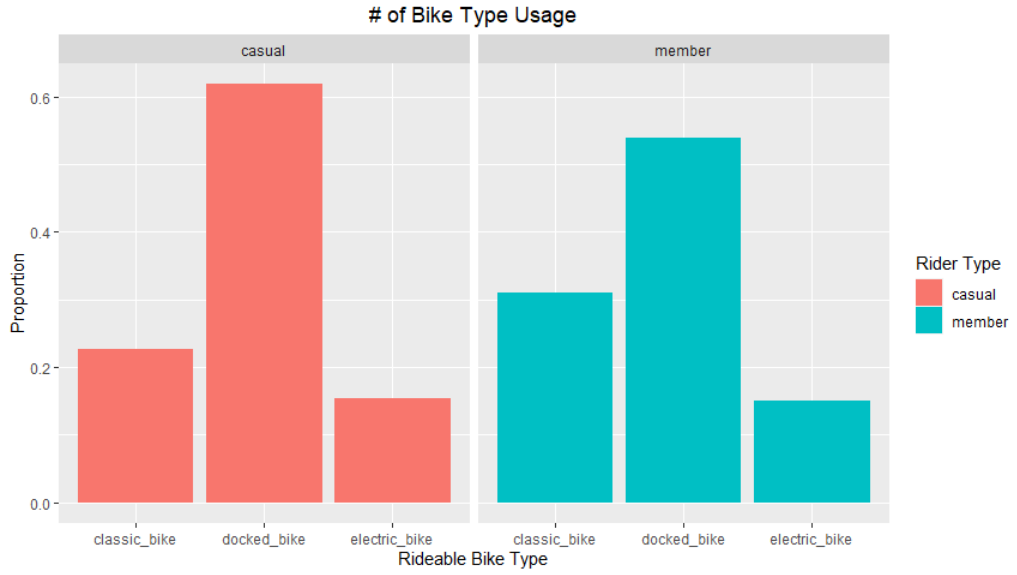
Time on Bike Share



Type of Bike Used

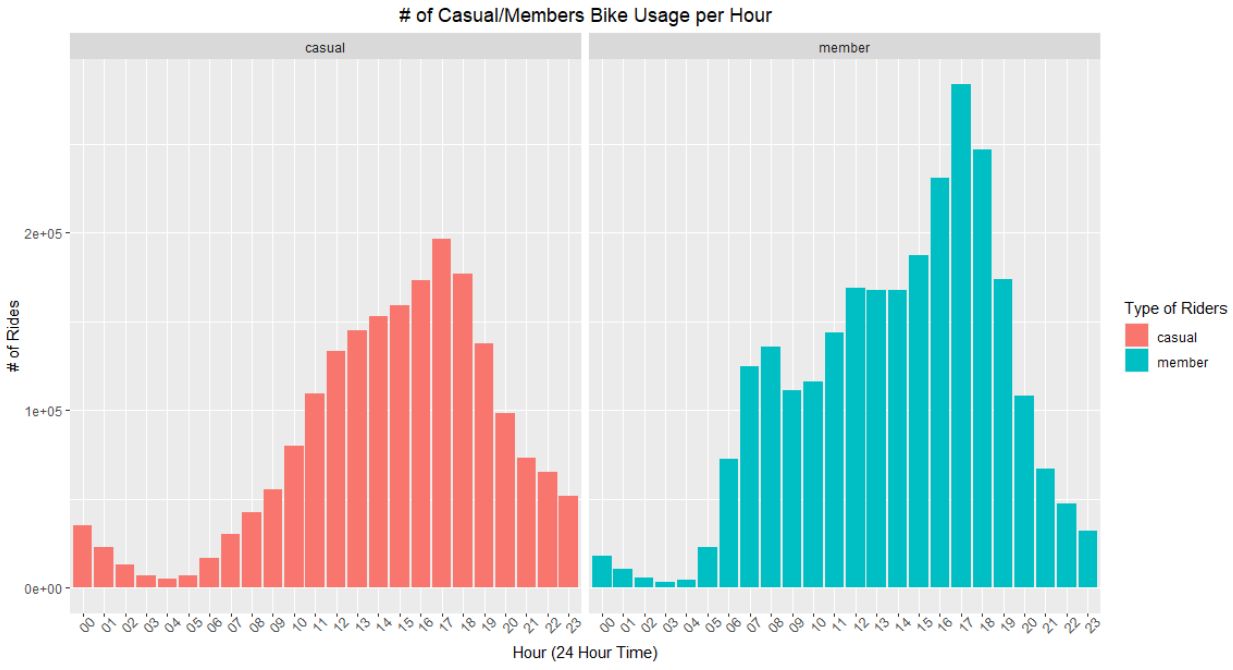
	Casual Riders	Members
Classic Bike	451,816 (22.71%)	823,007 (31.03%)
Docked Bike	1,231,138 (61.89%)	1,432,357 (54.00%)
Electric Bike	306,379 (15.40%)	397,190 (14.97%)
Grand Total	**1,989,333 (100.00%)**	**2,652,554 (100.00%)**





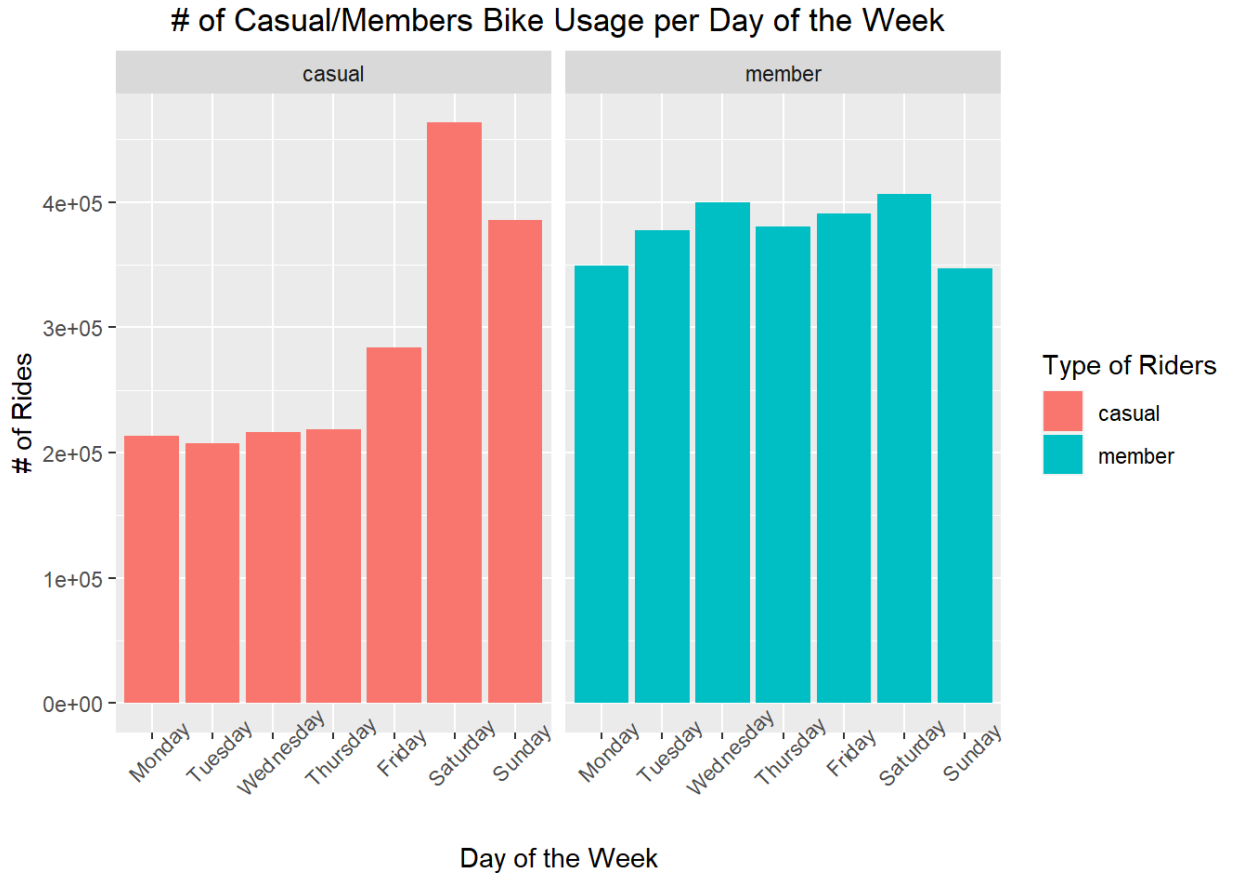
Daily Usage

	Casual Riders	Members
00	35,114	18,170
01	23,002	10,628
02	13,269	5,585
03	6,824	3,324
04	5,034	4,872
05	7,218	23,263
06	17,010	72,893
07	30,535	124,551
08	42,832	135,640
09	55,405	111,284
10	80,157	116,450
11	109,294	143,867
12	133,520	169,080
13	144,807	167,826
14	152,815	167,463
15	159,376	187,527
16	173,162	230,676
17	196,542	283,760
18	177,172	246,793
19	137,624	173,890
20	98,164	107,992
21	73,365	66957
22	65,263	47,692
23	51,829	32,371
Grand Total	**1,989,333**	**2,652,554**



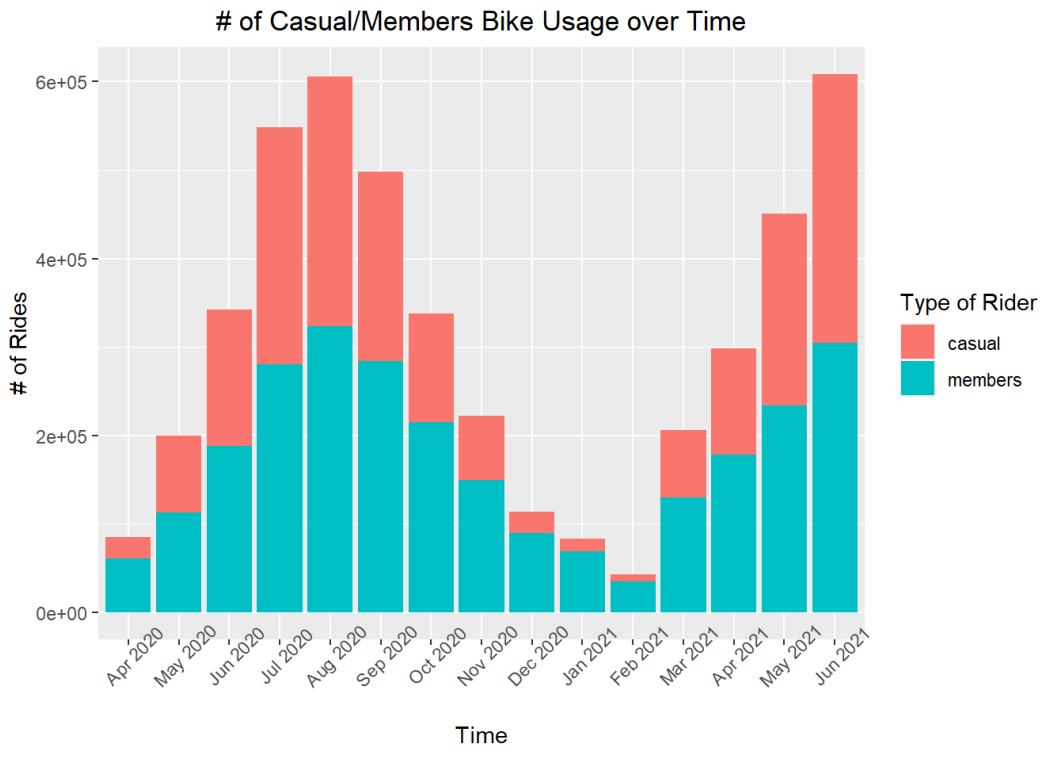
Weekly Usage

	Casual Riders	Members
Monday	213,199 (10.72%)	349,573 (13.18%)
Tuesday	207,784 (10.44%)	377,592 (14.24%)
Wednesday	216,731 (10.89%)	400,161 (15.09%)
Thursday	218,283 (10.97%)	380,750 (14.35%)
Friday	284,275 (14.29%)	390,810 (14.73%)
Saturday	463,502 (02.33%)	406,640 (15.33%)
Sunday	385,559 (19.38%)	347,028 (13.08%)
Grand Total	**1,989,333 (100.00%)**	**2,652,554 (100.00%)**



Season Usage

	Casual Riders	Members
April 2020	23,566	61,054
May 2020	86,693	113,079
June 2020	154,329	187,711
July 2020	268,103	280,514
August 2020	281,945	323,707
September 2020	214,672	283,556
October 2020	122,317	215,058
November 2020	72,847	149,069
December 2020	24,492	89,046
January 2021	14,690	68,818
February 2021	8,613	34,381
March 2021	75,641	130,046
April 2021	120,418	177,781
May 2021	216,823	234,155
June 2021	304,184	304,579
Grand Total	**1,989,333**	**2,652,554**

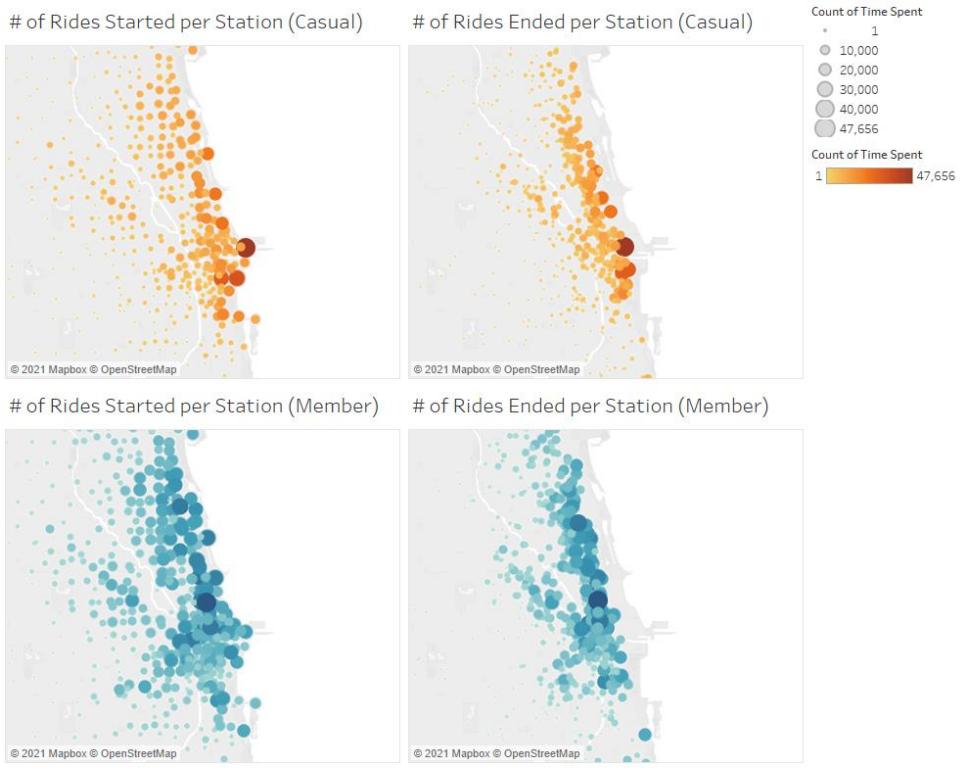


Location

	Start Station	End Station
Casual Riders	**1. Streeter Dr & Grand Ave** **2. Lake Shore Dr & Monroe St** **3. Millennium Park** 4. Michigan Ave & Oak St **5. Lake Shore Dr & North Blvd** **6. Theater on the Lake** 7. Indiana Ave & Roosevelt Rd 8. Shedd Aquarium 9. Michigan Ave & Lake St 10. Clark St & Elm St	**1. Streeter Dr & Grand Ave** **2. Lake Shore Dr & Monroe St** **3. Millennium Park** **4. Lake Shore Dr & North Blvd** **5. Theater on the Lake** 6. Michigan Ave & Oak St 7. Indiana Ave & Roosevelt Rd 8. Michigan Ave & Lake St 9. Clark St & Elm St 10. Wells St & Concord Ln
Members	**1. Clark St & Elm St** **2. Wells St & Concord Ln** **3. Dearborn St & Erie St** **4. Broadway & Barry Ave** **5. Kingsbury St & Kinzie St** **6. St. Clair St & Erie St** **7. Wells St & Elm St** **8. Theater on the Lake** **9. Lake Shore Dr & North Blvd** **10. Wells St & Huron St**	**1. Clark St & Elm St** **2. Wells St & Concord Ln** **3. St. Clair St & Erie St** **4. Dearborn St & Erie St** **5. Broadway & Barry Ave** **6. Kingsbury St & Kinzie St** **7. Lake Shore Dr & North Blvd** **8. Wells St & Elm St** **9. Theater on the Lake** **10. Wells St & Huron St**

Bolded station names are considered hotspots for either casual riders or members

For a complete list, please refer the Excel file *station_frequency.xlsx*



Share

Summary of Analysis

	Casual Riders	Members
Percentage of Total Rides	43%	57%
Time on Bike Share	Used often for longer times	Used often for shorter times
Type of Bikes Used	A higher proportion use docked bikes	A higher proportion use classic bikes (docked bikes are still the most popular choice)
Daily Usage	Somewhat normally distributed around 5pm	Higher total usage from 7am – 9pm with the busiest hours being at 4pm - 6pm
Weekly Usage	Heavily skewed towards Friday, Saturday, Sunday	Relatively evenly distributed throughout the week
Seasonal Usage	Like the members' distribution, but with less bike share usage in off-season months (winter and fall)	Normally distributed bike share usage with it being centered on July - August
Location	Starting locations are mainly centered around the busier districts of Chicago, but the ending ones are even more centered around them	

Looking at the summarized information gathered, casual riders seem to be those who use bike shares recreationally mainly on the weekends or at around rush hour while members use them more regularly for longer times, likely for commuting and daily activities. In addition, while there is a definitively larger number of member bike shares compared to casual riders', Cyclistic still has a massive portion of casual riders it has not fully tapped into yet as 43% of rides are by casual riders. The types of bikes used, while slightly different, are not massive enough to likely be dominate difference between members and casual riders. Lastly, while the types of stations that casual riders and members use tend to be somewhat different, they are all localized around busy districts such as the Chicago Loop.

Act

Detailed Recommendations

Date-based

In terms of date, **usage of bike shares tends to increase after the colder months end (after February)**. Rather than changes due to company growth, the increase from 2020 to 2021 bike shares was likely due to a decrease in COVID-19 lockdown regulations and an increase in vaccinations, making people feel more comfortable leaving home more. Therefore, **the trend in 2021 so far is likely to be closer to normal trends than 2020**. Using this information, it would be advisable to **begin advertising around April or May through summer to capture the greatest number of casual riders.**

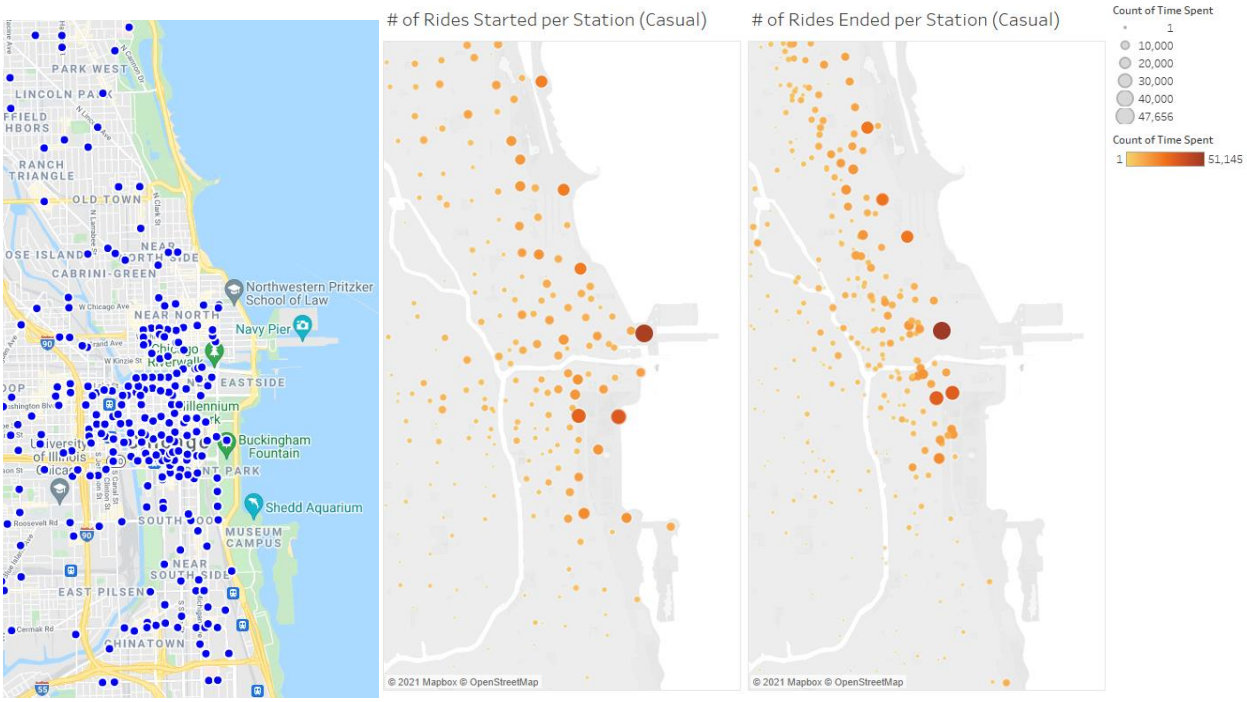
Digital Marketing

Having special deals and promotions on social media can be an effective tool as well to reach a wider audience. With more data, the best type of digital marketing can be ascertained, such as using **search engine optimization, social media, affiliates**, etc. In conjunction with the date-based recommendation, **focusing these efforts during the spring and summer months** would be the best way to maximize the input-output ratio.

Location-based

Considering that a large amount of the ride shares take place in busy districts of Chicago, **showing the appeal of getting around the city whilst avoiding common city traffic can be a good selling point**. The Chicago Loop area is an example of this, where there are many casual riders but also frequency traffic according to the City of Chicago website. For marketing, **putting advertisements on billboards and the bike stations in the Chicago Loop area would be effective in having both frustrated drivers and casual members buy memberships**.

In addition, several stations listed and bolded in Location in Analysis were listed that could also be potentially excellent spots to post advertisements on. Specifically, the **stations at Streeter Dr & Grand Ave, Lake Shore Dr & Monroe St, Millennium Park, Lake Shore Dr & North Blvd, and Theater on the Lake would be the top five spots to use**, as they are hotspots for casual riders starting and ending bike shares.



Shorter Recommendations

Advertising on locations where casual riders have longer bike shares

Increasing advertising around 4-6 pm

Increasing advertising during the weekend