2020 MARKETING ANALYTICS REPORT

Company: Cyclist

GOOGLE DATA ANALYTICS COURSE

Data Analytics Capstone Project, Educational and Demo Purposes

Shervin John Taheri

Data Analytics | Case Study

Contents

Introduction

This project is to show the usage of the data analytics and the data analysis cycle through a business case study, provided by Google's data analytics certification course capstone project.

The capstone project is targeted in explaining an ad hoc descriptive analysis for a fictional company called Cyclistic, however, the analysis is based on real-world historical data provided by the Divvy bicycle services in Chicago, Illinois.

In the coming sections below, a background of the project will be provided. A brief synopsis will be provided on the company and what it provides to the marketplace, the data, and a breakdown of the project's data analysis cycle.

Background

Cyclistic is a bike-share company founded in Chicago, in which individuals can rent a variety of bicycles from stations found throughout the city.

There is a total of 5,824 bicycles available for riders, and a total of 692 docking stations. A ratio of 8.42 bicycles to each docking station.

Cyclistic offers two categories of bikes, assistive and traditional. The rider base is broken down by 92 percent primarily using traditional bicycles, and the remaining eight percent using assistive bicycles. In terms of rider types, 70 percent of its rider base is categorized as leisure riders, while the remaining 30 percent are listed as commuters.

It offers three pricing plans:

1. Single-Ride Passes
2. Full-Day Passes
3. Annual Memberships

The following sections will go through each step of the data analysis process to solve the stakeholder's requested business objective.

Phase: Ask

Stakeholders

1. Cyclistic Executive Team
2. Lily Moreno, Director of Marketing
3. Cyclistic Marketing Analytics Team

Stakeholders Request

Cyclistic's financial analysts have assessed that annual members generate more revenue than casual riders. As such, Lily Moreno, Director of Marketing has initialized the plan to spur future growth by targeting the conversion rate KPI. The goal is to design a marketing strategy aimed at converting casual riders to annual member holders. Both Moreno and the marketing analytics team are the prime stakeholders, in which all relevant information will be passed to the executive team.

Business Objective

This report will be focused on supplying insights in increasing and maximizing the conversion rates of casual users of Cyclist's bike services into annual membership subscribers. As annual membership subscribers are a key profitable segment for the company. The following three guiding factors will aid in increasing conversion rates. They are understanding the key differences between annual membership subscribers to casual riders, factors influence purchasing behavior for annual members, and marketing methods to influence prospect causal riders into annual members. The key goal at the end is to supply data-driven evidence and insights for key stakeholders in their objective decision-making at Cyclist.

Three Primary Guiding Factors

1. Usage differences between casual cyclists and annual members.
2. Purchasing Factors
3. Marketing Influence Factors

Target Group, KPI, and Method:

- **Target Group**: Casual Riders
- **Target KPI:** Conversion Rates
- **Target Method:** Digital Media Marketing

Phase: Prepare

Data

The original data provided by Google is stored on the following link, Divvy-Trip Data. The Divvy trip data is organized in .zip files, which are indexed by month and year. Example, 202004-divvy-tripdata.zip is for the entire month of April 2020.

Given the scope of the data is vast from 2020 to 2022, including quarterly data that ranges from Q1 2014 to Q1 2020, the focus was set on only year 2020. The analysis is focused on the data from April to December 2020. The months from January to March 2020 were not available in the source location.

The data was extracted from the original source location and set to the following location online, Google Drive: Cyclistic Data Source. There you will find all relevant monthly data sets, indexed by their respective month. Given the size of August and analysis reasons from the code, it is split into two separate files, which are respectively called Part I and Part II.

There are no alterations to their original data files, other than being split into two files for technical reasons.

To ensure baseline integrity for the fields, a cross-check is performed across all individual monthly CSV files to ensure accuracy. The cross-check ensures that all tables have the same fields in further analysis. To view the cross-check results, a link to the spreadsheet can be found here, Google Drive: Table Field Check.

Category/Fields

Field	Data Type	Description
ride_id	Character/String	User's Rider ID
rideable_type	Character/String	Type of Bicycle
started_at	POSIXlt	Start Date and Time
ended_at	POSIXlt	End Date and Time
start_station_name	Character/String	Start Station Name
start_station_id	Character/String	Start Station ID
end_station_name	Character/String	End Station Name
end_station_id	Character/String	End Station ID
start_lat	Numeric	Start Station Coordinate Latitude
start_lng	Numeric	Start Station Coordinate Longitude
end_lat	Numeric	End Station Coordinate Latitude
end_lng	Numeric	End Station Coordinate Longitude
member_casual	Character/String	Type of Rider (member, casual)

Credibility

Given the data has been offered through Google's data analytics certification course, there is an assumption that the Divvy trip data across all of the years is credible in its data collection methods.

Further stages will inspect any issues that are in regard to the data structure or its cleanliness, but that is not necessarily questioning the sources trust worthiness.

Phase: Process

The process phase for this project is focused on ensuring that the downloaded data is structured and cleaned to ensure data integrity and accuracy for the analyze phase. The phase has been broken down into the following sub phases, process, clean, merge.

Process

Focus: To extract all relevant CSV files from the Google Drive location and store it into the R environment.

- **Step 1:** Obtain each month's CSV file Google Drive link.
- **Step 2:** Set each month's file encoding to ***'UTF-8-BOM'*** and download each CSV file to a data frame.
- **Step 3:** All blank cells are set to **'N/A'.**
- **Step 4:** Data integrity check is performed by cross-referencing each month's data frame columns against each other. This will ensure that all fields are correct, when later merged.

The format for each month's data frame is, "**tbl\[month_number]\2020_divvy**."
Example: *tbl_04_2020_divvy*

Clean

Focus: To format the data for merging to later use in the analysis phase.

- **Step 5:** The individual month files have date/time stamp format issues for fields, ***started_at*** and ***ended_at***, thus the respective fields are reformatted.
- **Step 6:** Each date/time field format was structured to the following, format = "%Y-%m-%d %H:%M:%S", tz = "UTC".

Merge

Focus: To omit 'N/A' data points and merge all data frames into a single data frame.

- **Step 7:** Merge all individual data frames for each month into a single table/data frame.
- **Step 8:** After merging all "N/A" data points from the single table/data frame are omitted.
- **Step 9:** After merging, all individual monthly data frames are dropped from the R environment.

Table Information, Total N/A's Per Month by Variable

	default_names	tbl_04_NA	tbl_05_NA	tbl_06_NA	tbl_07_NA	tbl_08_NA	tbl_09_NA	tbl_10_NA	tbl_11_NA	tbl_12_NA	Total_NA
1	ride_id	0	0	0	0	0	0	0	0	0	0
2	rideable_type	0	0	0	0	0	0	0	0	0	0
3	started_at	0	0	0	0	0	0	0	0	0	0
4	ended_at	0	0	0	0	0	0	0	0	0	0
5	trip_duration_min	0	0	0	0	0	0	0	0	0	0
6	start_station_name	0	0	0	149	7595	19691	31198	24324	11699	94656
7	start_station_id	0	0	0	152	7691	19901	31405	24434	11699	95282
8	end_station_name	99	321	468	967	10035	23373	35631	26749	13237	110880
9	end_station_id	99	321	468	969	10110	23524	35787	26826	13237	111341
10	start_lat	0	0	0	0	0	0	0	0	0	0
11	start_lng	0	0	0	0	0	0	0	0	0	0
12	end_lat	99	321	468	770	938	789	474	284	111	4254
13	end_lng	99	321	468	770	938	789	474	284	111	4254
14	member_casual	0	0	0	0	0	0	0	0	0	0
15	Total N/A	396	1284	1872	3777	37307	88067	134969	102901	50094	420667

To access and inspect the code, please click the links below:

- [Processing: File Directory 01]
- [Cleaning: File Directory 02]
- [Cleaning: File Directory 03]
- [Merging: File Directory 04]

Phase: Analyze

Exploratory Data Analysis

The (core) table used for the analysis phase is the, **tbl_2020_divvy**.

An added column is created that is called ***trip_duration_min,*** which is a calculation of a trip duration in minutes by solving the difference between the columns, ***ended_at*** – ***started_at.*** This column is used later in the analysis phase and plotting phase.

Structure

```
> str(tbl_2020_divvy)
'data.frame':   3114796 obs. of  14 variables:
 $ ride_id           : chr  "A847FADBBC638E45" "5405B80E996FF60D"
 $ rideable_type     : chr  "docked_bike" "docked_bike" "docked_b
 $ started_at        : POSIXlt, format: "2020-04-26 17:45:14" "20
 $ ended_at          : POSIXlt, format: "2020-04-26 18:12:03" "20
 $ trip_duration_min : int  26 8 14 12 52 5 5 75 5 17 ...
 $ start_station_name: chr  "Eckhart Park" "Drake Ave & Fullerton
 $ start_station_id  : chr  "86" "503" "142" "216" ...
 $ end_station_name  : chr  "Lincoln Ave & Diversey Pkwy" "Kosciu
 $ end_station_id    : chr  "152" "499" "255" "657" ...
 $ start_lat         : num  41.9 41.9 41.9 41.9 41.9 ...
 $ start_lng         : num  -87.7 -87.7 -87.6 -87.7 -87.6 ...
 $ end_lat           : num  41.9 41.9 41.9 41.9 42 ...
 $ end_lng           : num  -87.7 -87.7 -87.6 -87.7 -87.7 ...
 $ member_casual     : chr  "member" "member" "member" "member"
> nrow(tbl_2020_divvy)
[1] 3114796
> ncol(tbl_2020_divvy)
[1] 14
```

Table Field Summary Statistics

Field: ride_id

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Field: rideable_type

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Field: started_at

- Min. :2020-04-01 00:00:30
- 1st Qu.:2020-07-09 22:22:03
- Median :2020-08-20 14:27:00
- Mean :2020-08-20 15:52:00
- 3rd Qu.:2020-10-01 07:49:33
- Max. :2020-12-31 23:59:59

Field: ended_at

- Min. :2020-04-01 00:10:45
- 1st Qu.:2020-07-09 23:07:34
- Median :2020-08-20 14:54:00
- Mean :2020-08-20 16:24:09
- 3rd Qu.:2020-10-01 08:02:49
- Max. :2021-01-03 08:54:11

Field: trip_duration_min

- Min.: 0.00
- 1st Qu.: 8.00
- Median: 15.00
- Mean: 31.76
- 3rd Qu.: 27.00
- Max. :58720.00

Field: start_station_name

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Field: start_station_id

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Field: end_station_name

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Field: end_station_id

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Field: start_lat

- Min. :41.64
- 1st Qu.:41.88
- Median :41.90
- Mean :41.90
- 3rd Qu.:41.93
- Max. :42.08

Field: start_lng

- Min.: -87.87
- 1st Qu.: -87.66
- Median: -87.64
- Mean: -87.64
- 3rd Qu.: -87.63
- Max.: -87.52

Field: end_lat

- Min. :41.54
- 1st Qu.:41.88
- Median :41.90
- Mean :41.91
- 3rd Qu.:41.93
- Max. :42.16

Field: end_lng

- Min.: -87.89
- 1st Qu.: -87.66
- Median: -87.64
- Mean: -87.64
- 3rd Qu.: -87.63
- Max.: -87.44

Field: member_casual

- **Length:** 3114796
- **Class:** Character
- **Mode:** Character

Table Analysis – Summary Statistics

Table: ***tbl_2020_summary:*** Is an analysis of the ***tbl_2020_divvy,*** which analyzes data for each of the categories of bike and member types for each month.

```
> str(tbl_2020_summary)
'data.frame':	9 obs. of  20 variables:
 $ month_name                  : chr  "April" "May" "June" "July" ...
 $ month_num                   : int  4 5 6 7 8 9 10 11 12
 $ sum_rides                   : int  84776 200274 343005 551480 622361 532958 388653 260094 131195
 $ sum_minutes                 : int  2997597 6589351 11338854 20832595 18540521 13207457 7592680 15800235 2030714
 $ avg_minutes_per_ride        : num  35.4 32.9 33.1 37.8 29.8 ...
 $ tot_casual                  : int  23628 86909 154718 269296 289661 230692 145012 88178 30001
 $ tot_members                 : int  61148 113365 188287 282184 332700 302266 243641 171916 101194
 $ tot_classic_bikes_used      : int  0 0 0 0 0 0 0 0 70616
 $ tot_docked_bikes_used       : int  84776 200274 343005 549545 556166 404606 236477 151791 12795
 $ tot_electric_bikes_used     : int  0 0 0 1935 66195 128352 152176 108303 47784
 $ tot_member_dock_bike_used   : int  61148 113365 188287 281089 294965 232832 156594 105753 7852
 $ tot_member_classic_bike_used : int  0 0 0 0 0 0 0 0 59297
 $ tot_member_electric_bike_used: int  0 0 0 1095 37735 69434 87047 66163 34045
 $ tot_casual_dock_bike_used    : int  23628 86909 154718 268456 261201 171774 79883 46038 4943
 $ tot_casual_classic_bike_used : int  0 0 0 0 0 0 0 0 11319
 $ tot_casual_electric_bike_used: int  0 0 0 840 28460 58918 65129 42140 13739
 $ sum_members_minutes         : int  1282727 2183922 3429510 4856338 5571412 4526923 3289648 10789296 1240031
 $ sum_casual_minutes          : int  1714870 4405429 7909344 15976257 12969109 8680534 4303032 5010939 790683
 $ avg_minutes_per_ride_casual  : num  72.6 50.7 51.1 59.3 44.8 ...
 $ avg_minutes_per_ride_members : num  21 19.3 18.2 17.2 16.8 ...
```

Table: ***tbl_2020_rides_weekday_summary:*** A summary statistics data frame that uses ***tbl_2020_divvy,*** to analyze the data about weekday usage by bike and member types.

```
> str(tbl_2020_rides_weekday_summary)
'data.frame':	42 obs. of  7 variables:
 $ weekday       : chr  "Sunday" "Sunday" "Sunday" "Sunday" ...
 $ weekday_num   : int  1 1 1 1 1 1 2 2 2 2 ...
 $ rider_type    : chr  "casual" "casual" "casual" "member" ...
 $ bike_type     : chr  "classic_bike" "electric_bike" "docked_bike" "classic_bike" ...
 $ num_rides     : int  2143 32579 206540 7770 36287 191051 1377 23094 112326 9246 ...
 $ total_minutes  : int  66668 803158 11488063 111141 523285 3609857 31968 465179 5655835 110513 ...
 $ average_minutes: num  31.1 24.6 55.6 14.3 14.4 ...
```

Table: ***tbl_2020_day_trips:*** A summary statistics data frame that analyzes the data from ***tbl_2020_divvy*** to categorize data on minutes ridden by the month and day, and bike and member type. It's conditioned on only analyzing trips less than or equal to 24 hours (≤ 1,440 minutes).

```
> str(tbl_2020_day_trips)
'data.frame':	3111852 obs. of  6 variables:
 $ month_name  : chr  "April" "April" "April" "April" ...
 $ month       : int  4 4 4 4 4 4 4 4 4 4 ...
 $ day_num     : int  1 1 1 1 1 1 1 1 1 1 ...
 $ minutes_ride: int  14 36 11 6 8 8 20 20 13 6 ...
 $ member_type : chr  "member" "member" "member" "member" ...
 $ bike_type   : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
```

Table: ***tbl_2020_day_trips:*** A summary statistics data frame that analyzes the data from ***tbl_2020_divvy*** to categorize data on minutes ridden by the month and day, and bike and member type. It's conditioned on only analyzing trips greater than 24 hours (> 1,440 minutes).

```
> str(tbl_2020_day_plus_trips)
'data.frame':	2944 obs. of  7 variables:
 $ month_name    : chr  "April" "April" "April" "April" ...
 $ month         : int  4 4 4 4 4 4 4 4 4 4 ...
 $ day_num       : int  1 2 2 2 2 2 2 2 2 2 ...
 $ minutes_ride  : int  2917 1499 1760 2077 1499 1499 1808 23433 2785 36156 ...
 $ member_type   : chr  "casual" "casual" "casual" "casual" ...
 $ bike_type     : chr  "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
 $ days_check_out: num  2.03 1.04 1.22 1.44 1.04 ...
```

To access and inspect the exploratory analysis on monthly and weekly data, please click below:

- Exploratory Data Analysis: File Directory 05
- Exploratory Data Analysis: File Directory 06

Phase: Share

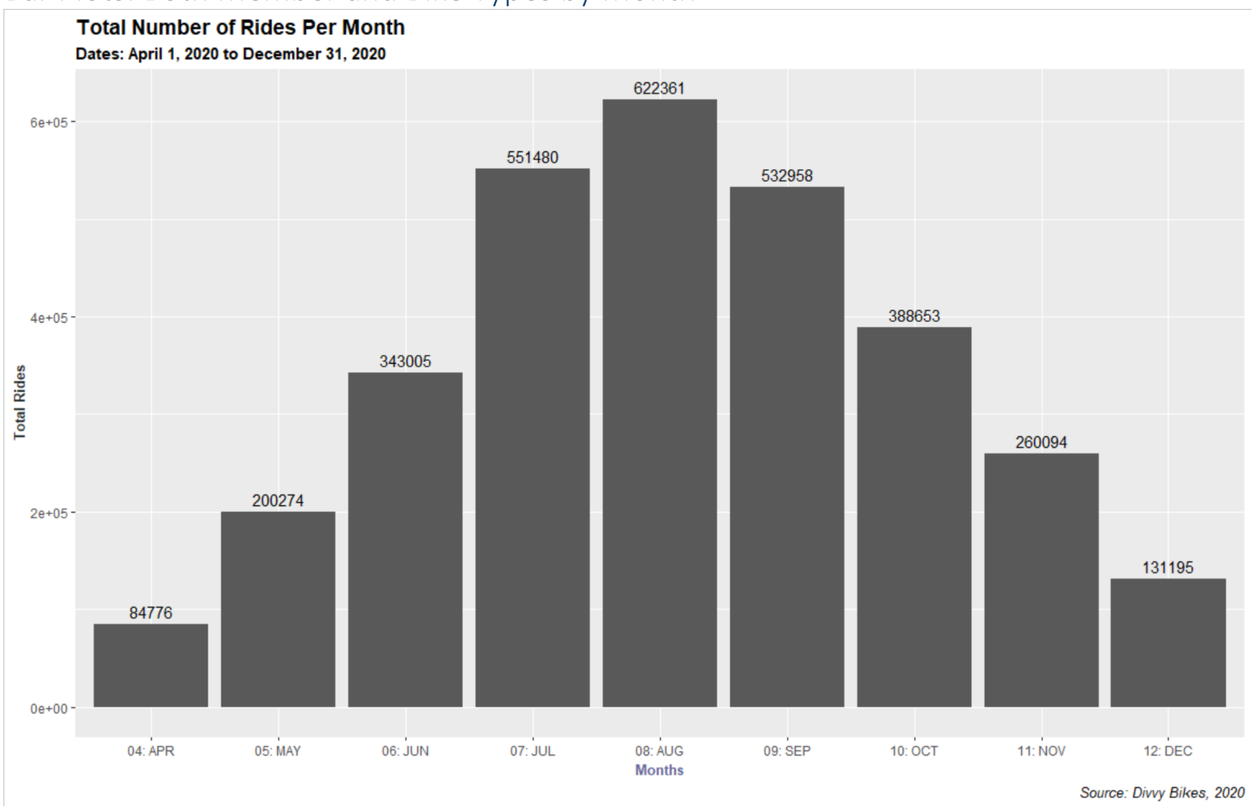
Technical Details:

Programming Language: R
Package Used: ggplot2

The following tables are used for plotting:

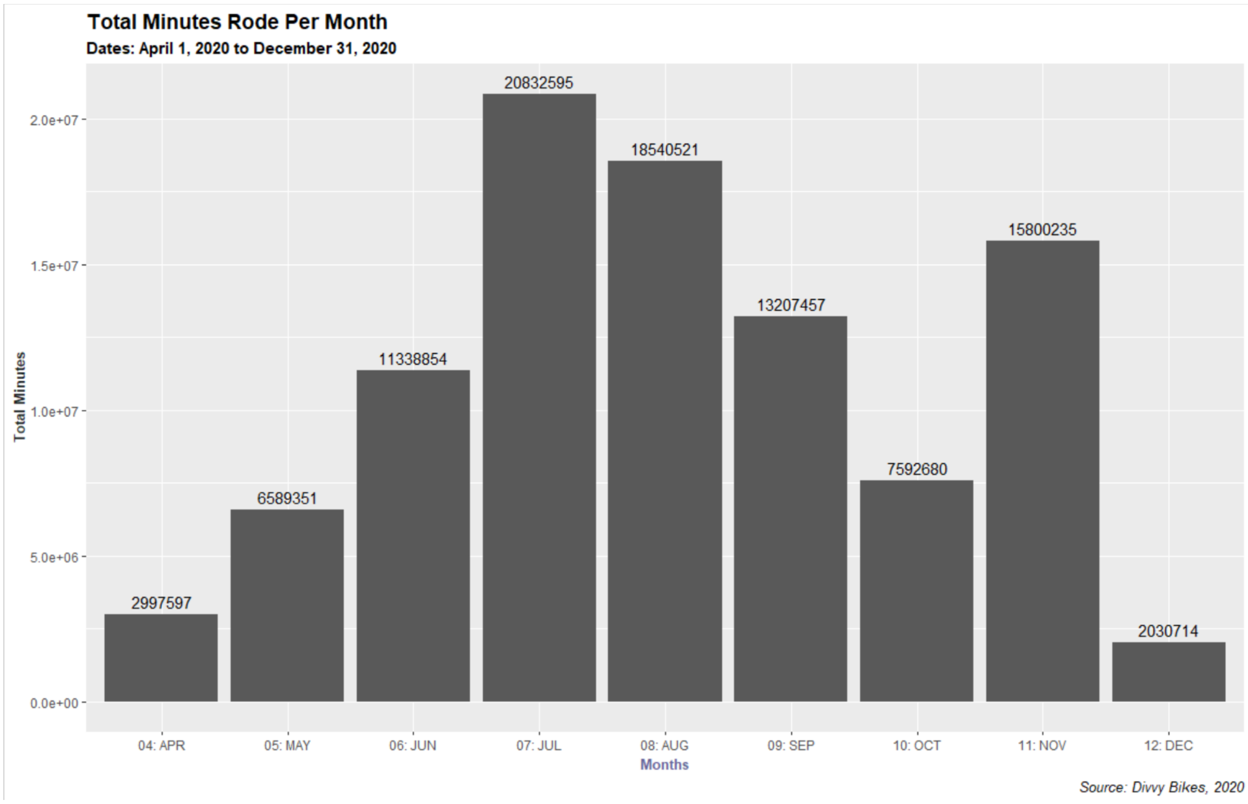
- tbl_2020_summary
- tbl_2020_rides_weekday_summary
- tbl_2020_day_trips

Bar Plots: Both Member and Bike Types by Month



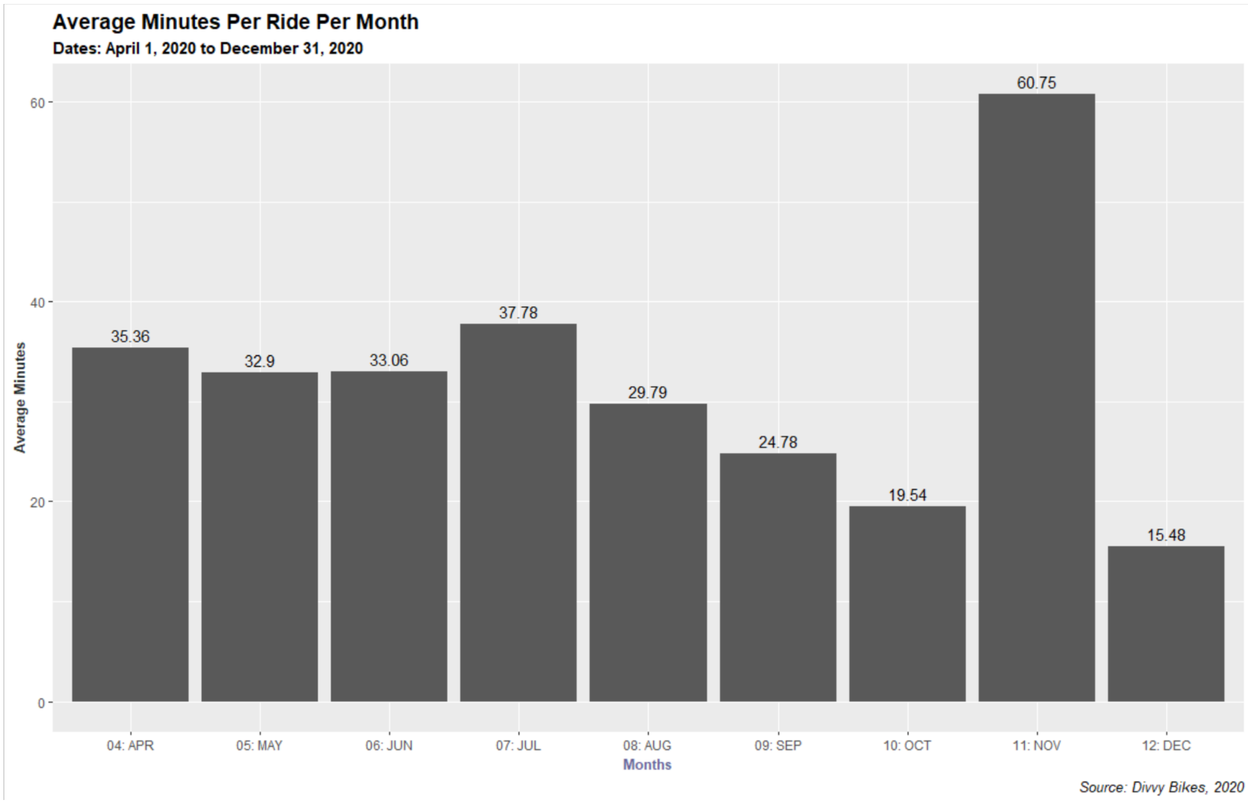
(Figure 1.0)

The total number of rides per month from April to December 2020, follows a normal distribution. The plot seems to illustrate a seasonal usage pattern. In the warmer months (summer, June-Sept) shows a higher number of rides, with August (average warmest month of summer) having the highest count. The colder months as guessed illustrate a lowering of rides as it approaches December.



(Figure 1.1)

The total minutes rode per month, it would also seem that it would a normal distribution, however, in November the total minutes rode spikes before dropping back to its all-time low in December. This could be an anomaly spike, however, further investigation on the reasoning why November spikes could prove to be insightful. Additionally, this can be used as a comparison statistic for earlier and future years for usage patterns. Again, the pattern seems to follow such that as the months approach summer the total minutes rode increases before decreasing as it cools down in fall and winter. Implying that maybe a relationship between seasons and ridership usage are correlated. A further investigation could prove to be insightful.

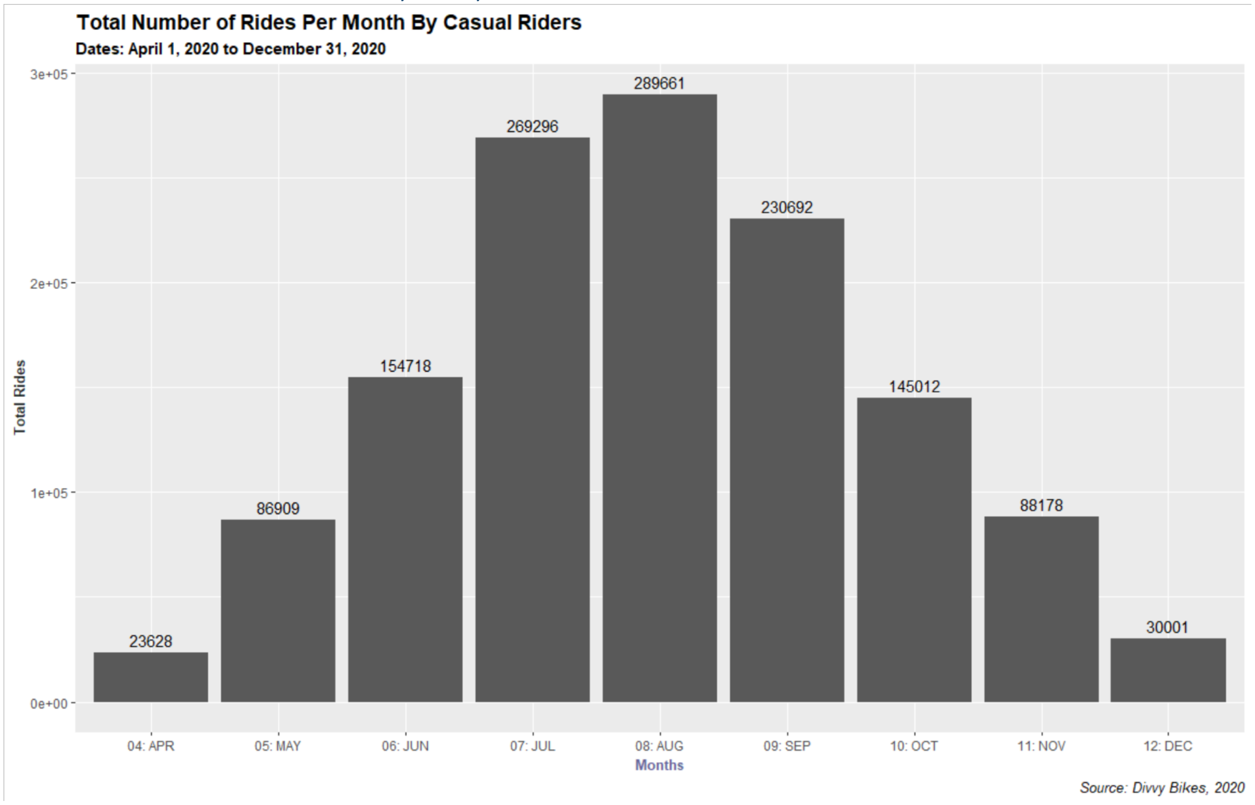


(Figure 1.2)

The average minutes per ride seems to follow a fairly uniform distribution from April to August, before a drop occurs as fall approaches. Again, a spike in November occurs, but this is due to the fact that although total number of rides in November were low, the total minutes rode in that month were high, and thus raises the average minutes per ride.

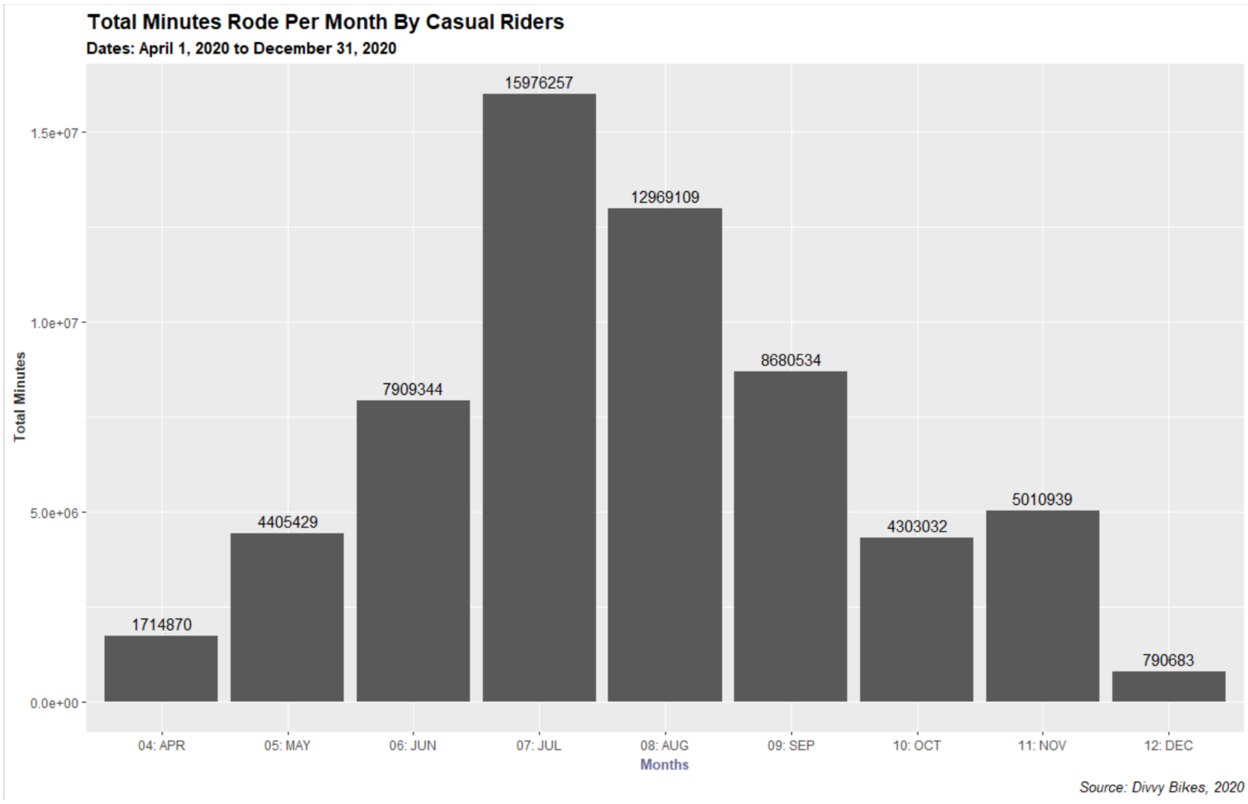
The following three graphs above have both member types (casual and annual) and all bike types (classic, docked, electric).

Bar Plots: Casual Riders Analysis by Month



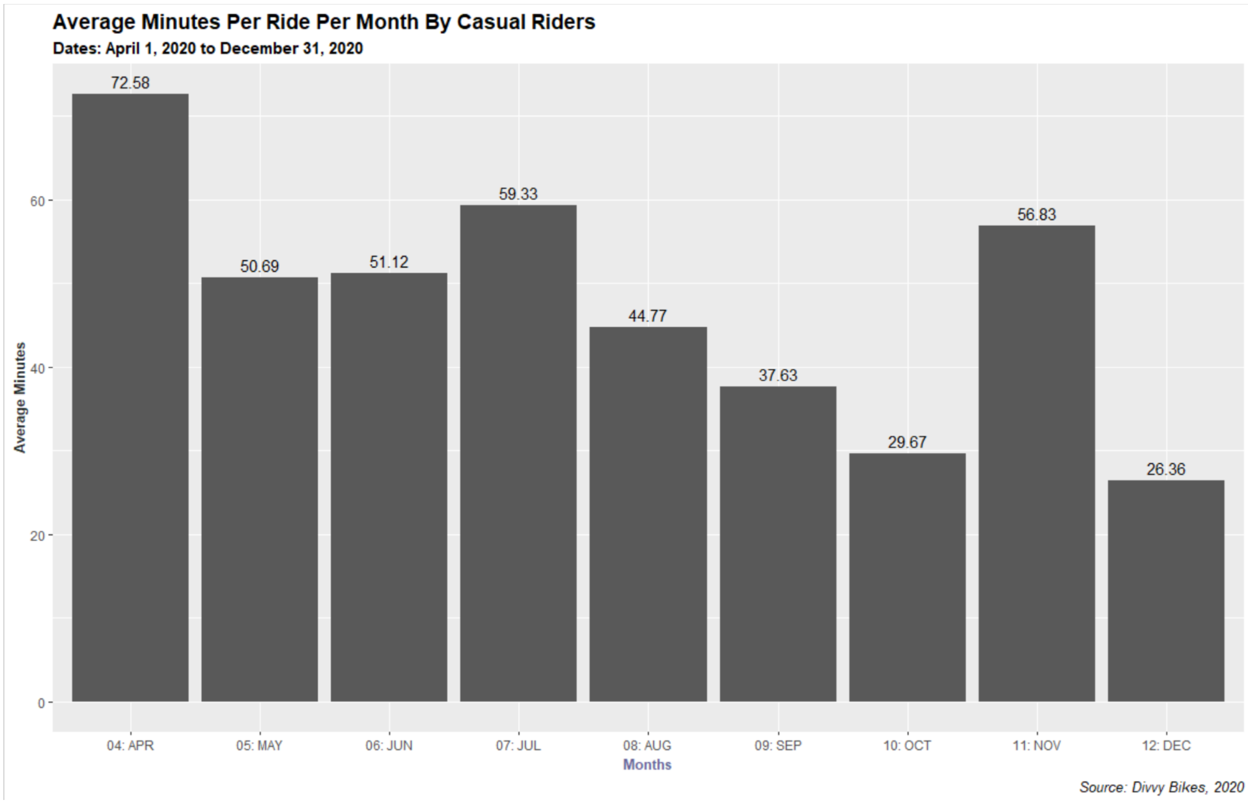
(Figure 1.3)

The total number of rides per month by casual riders shows a normal distribution. With similar trends with the colder months having lesser number of rides and increases as the months approach the warmer seasons (summer). August being the peak of total number of rides in the eight-month period.



(Figure 1.4)

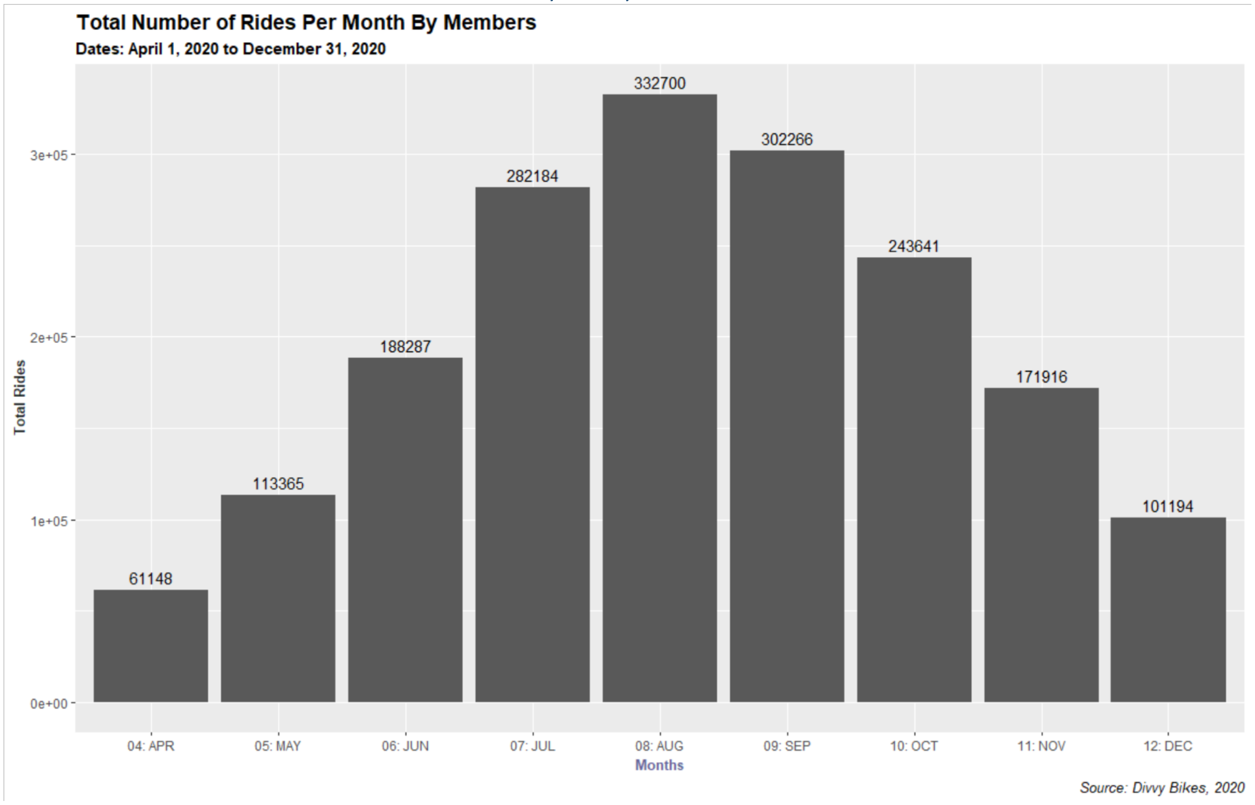
The total minutes rode per month by casual riders here shows a relative normal distribution as well, with the exception that in November there is a small jump in minutes before dropping to its all-time low in December.



(Figure 1.4)

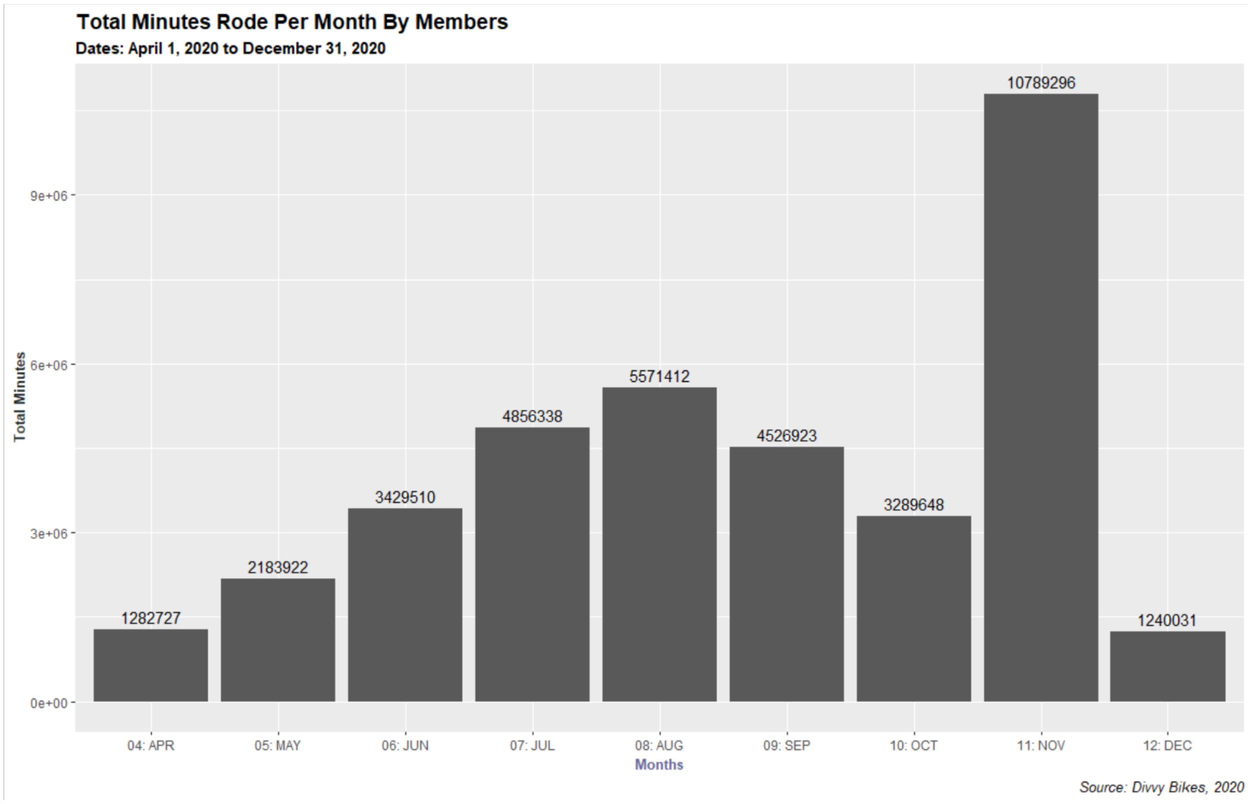
Average minutes per ride for casual riders, shows a fairly marginally decreasing rate over time. With April having the highest average minutes until it reaches its lowest point in December. July is the highest average per month for the summer. The trend follows a similar pattern of decreasing until November. Where again it is most likely due to the spike that occurred in the total minutes rode per month.

Bar Plots: Annual Members Riders Analysis by Month



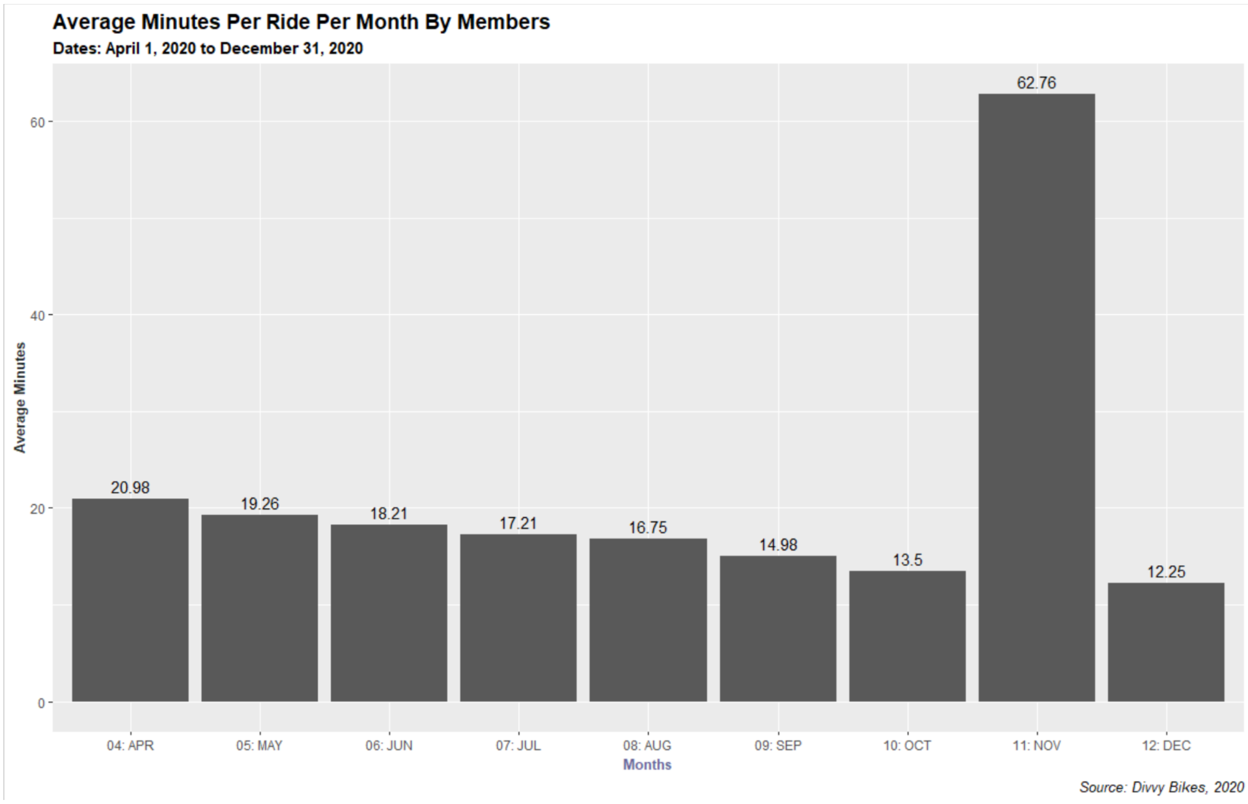
(Figure 1.5)

The total number of rides per month for annual members shows similar trends compared to the other two plots. It forms a relative normal distribution, with the warmer months in summer seeing increase ridership while in the coming cooler months a decrease. August is the peak month for total number of rides per month.



(Figure 1.6)

Total minutes rode per month for annual members is also following a similar trend with normal distribution, however, again in November there is a massive spike in comparison to the previous months. A further investigation on this month's data trend could be insightful to understand the behavioral usage of both annual and casual riders in November. December, one of the coldest months, is the all-time low in the eight-month period.

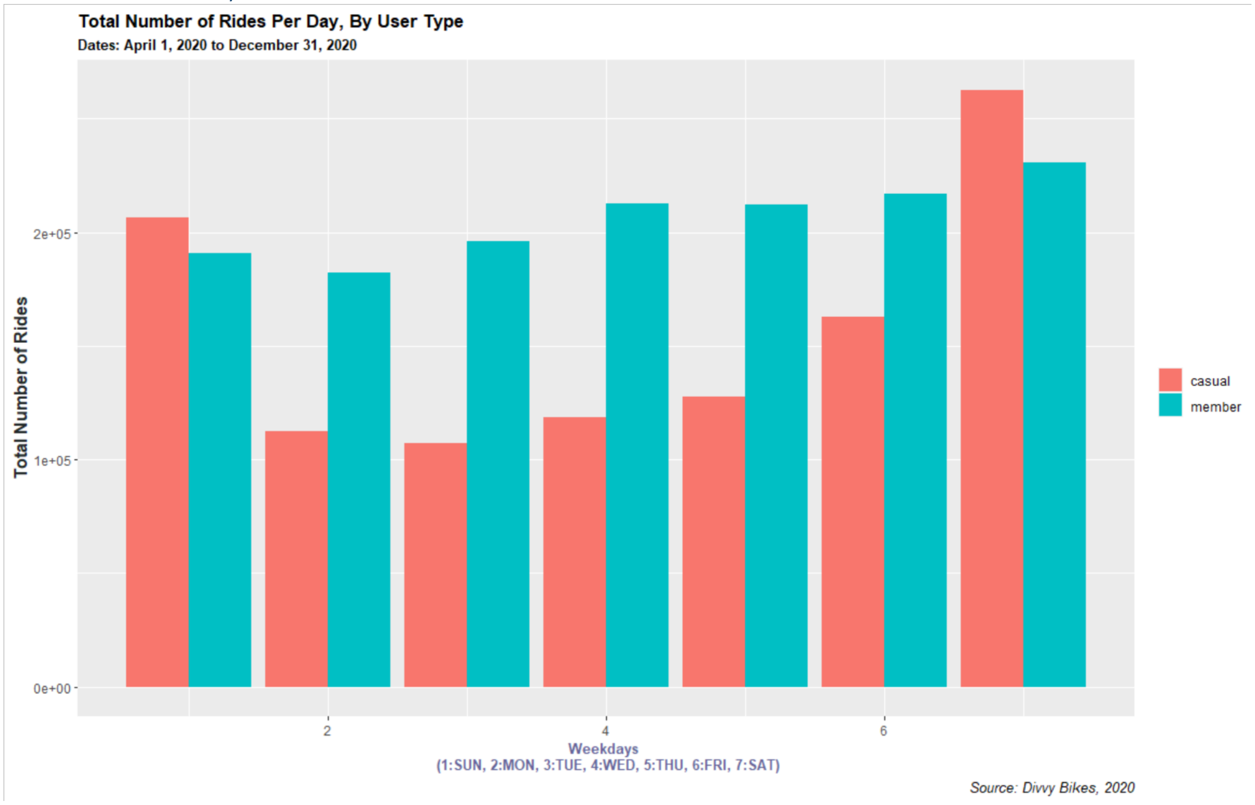


(Figure 1.7)

Average minutes per ride per month show's a relative decrease starting from April all the way to October, before a spike in November. This is most likely due to the spike experienced in total minutes rode in November, causing it to relatively much larger than the other months. Further investigation required for November to understand this behavioral phenomenon.

Weekday Summary Data - By User Type

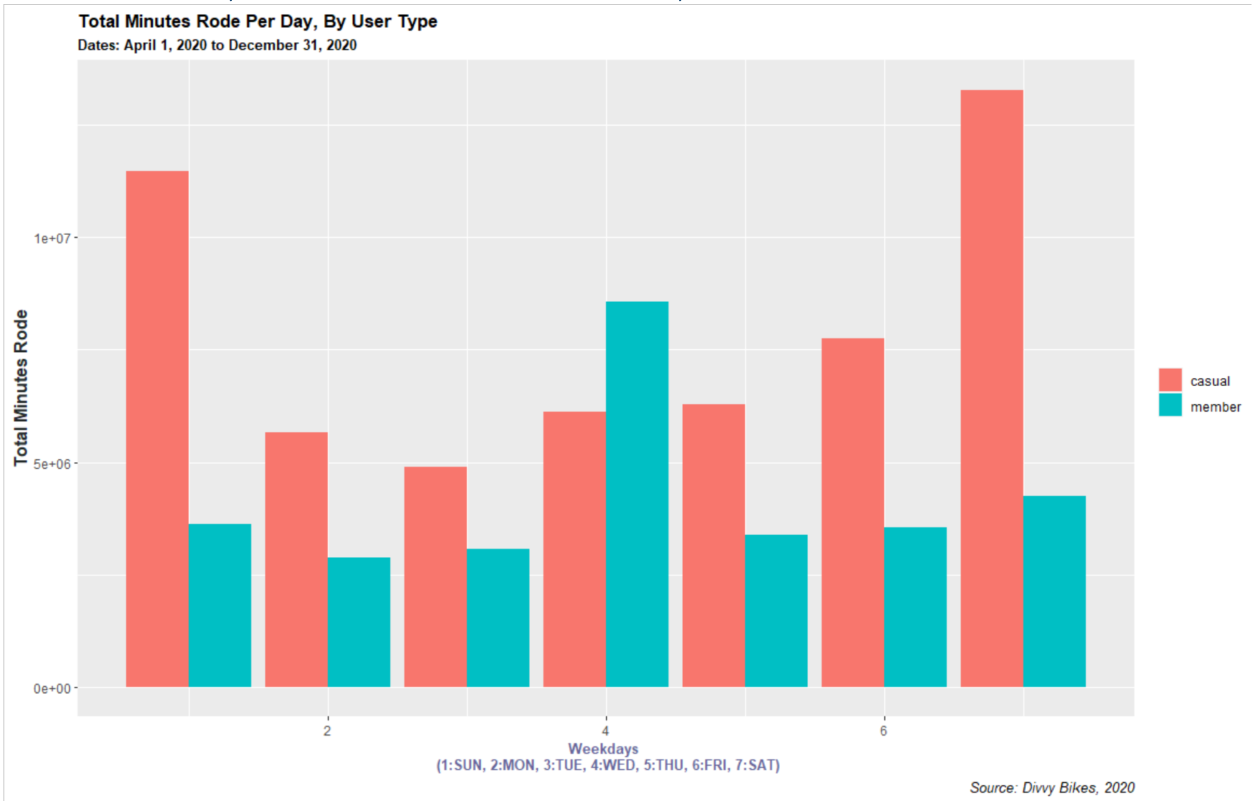
Bar Plot: Weekday Data - Total Number of Rides



(Figure 2.0)

In visualizing the total number of rides throughout the weekday we can assess if a potential pattern is seen in usage between the user types. For annual members throughout the week, their total number of rides is fairly uniform in distribution with it slightly increasing as it approaches the weekend and Wednesday being the highest within the five-day work week. For casual riders, it seems the pattern that the weekend being the highest total number of rides relative the rest of the days. This could be for a potential reason that casual riders are only using the service during their leisure time, while annual members are using it during for commuting. An investigation on correlating behavioral patterns and weekdays could prove insightful.

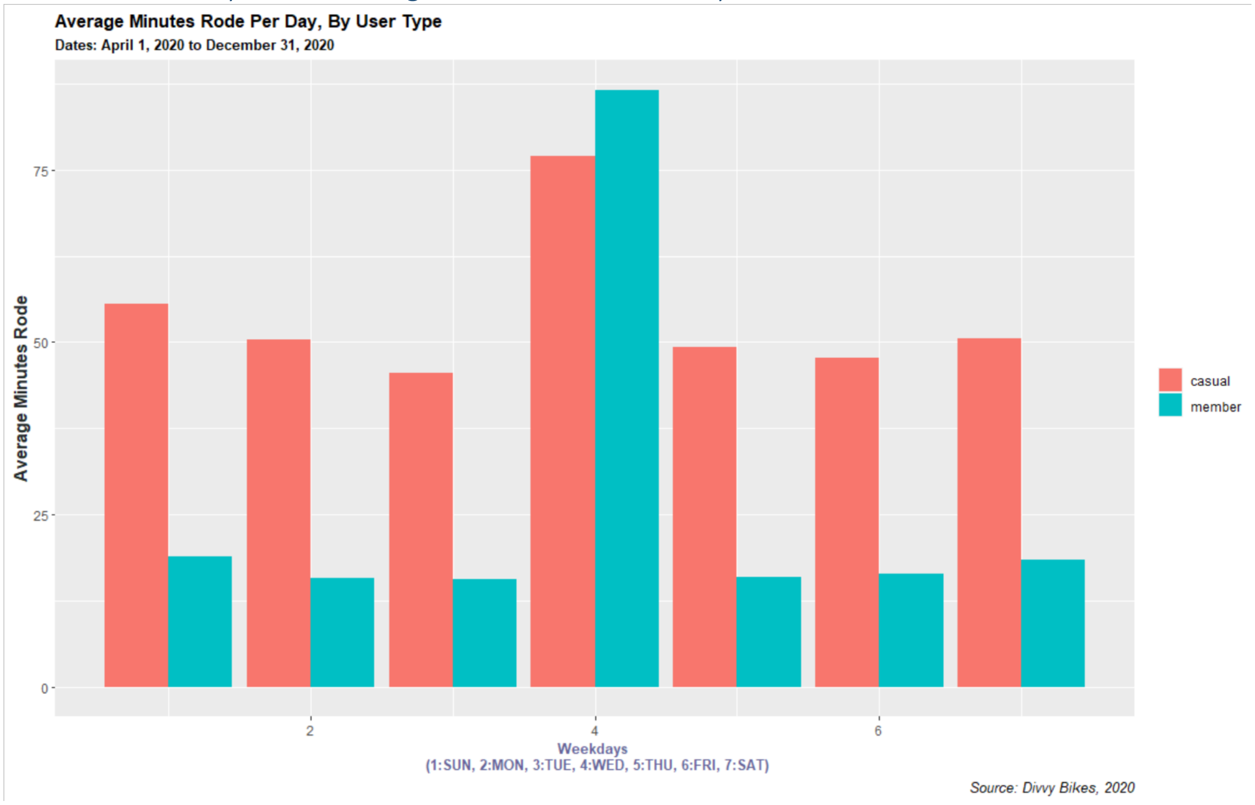
Bar Plot: Weekday Data - Total Minutes Rode Per Day



(Figure 2.1)

Similar patterns are illustrated here with annual members following a slight normal distribution. The peak total minutes rode being on Wednesday, while the remaining minutes are distributed rather evenly across the other days. Casual riders seem to again be riding the most on the weekends, potentially pointing towards more of a leisure activity with the bikes, rather than commuting.

Bar Plot: Weekday Data - Average Minutes Rode Per Day

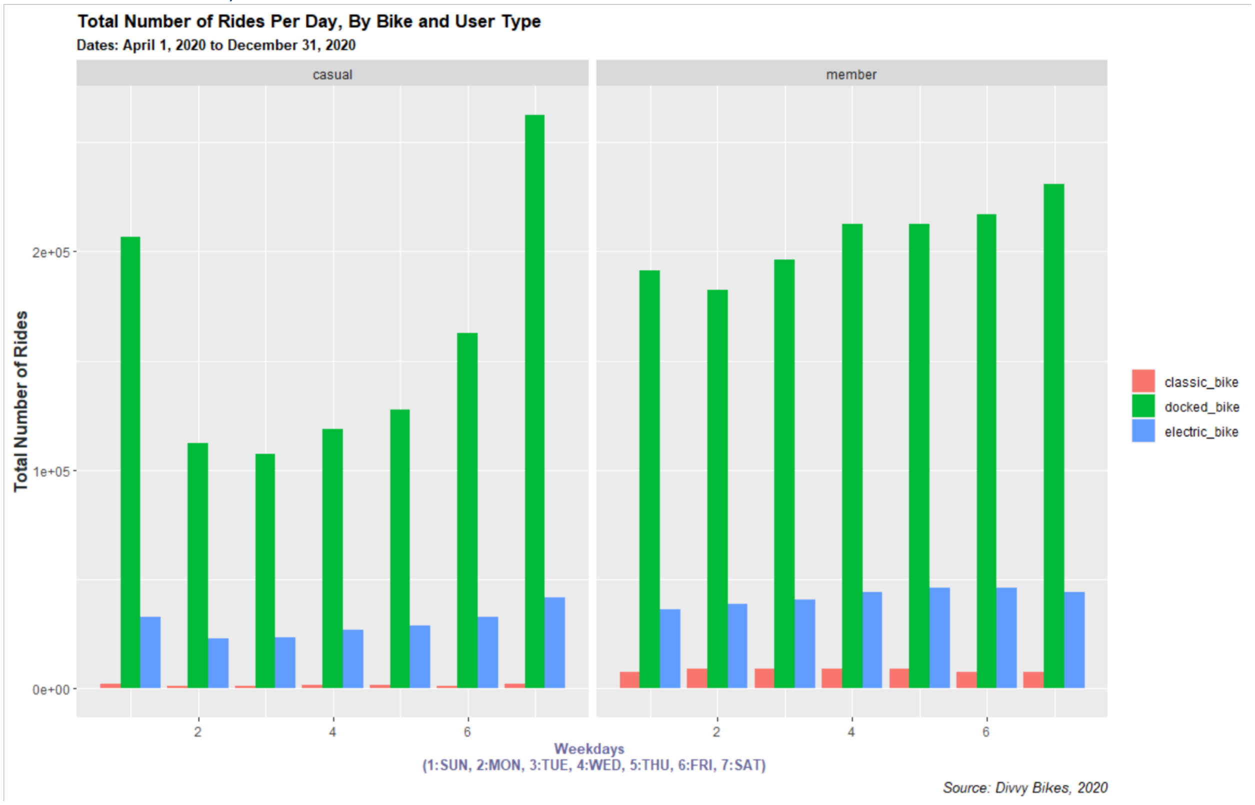


(Figure 2.2)

The average minutes rode per day by both user types has their closest difference being on Wednesday. This could be due to a calculation outcome, with total minutes being higher compared to total rides for casual riders, thus boosting their average minutes in comparison. However, both user types seem to show a relatively uniform distribution of average minutes rode per day with the only key spike on Wednesday. Casual riders again seem to be riding more on average closer to the weekends than during the week. Illustrating leisure activity preferences. Investigating the casual riders’ behavioral patterns would be insightful.

Weekday Summary Data - By Bike Type

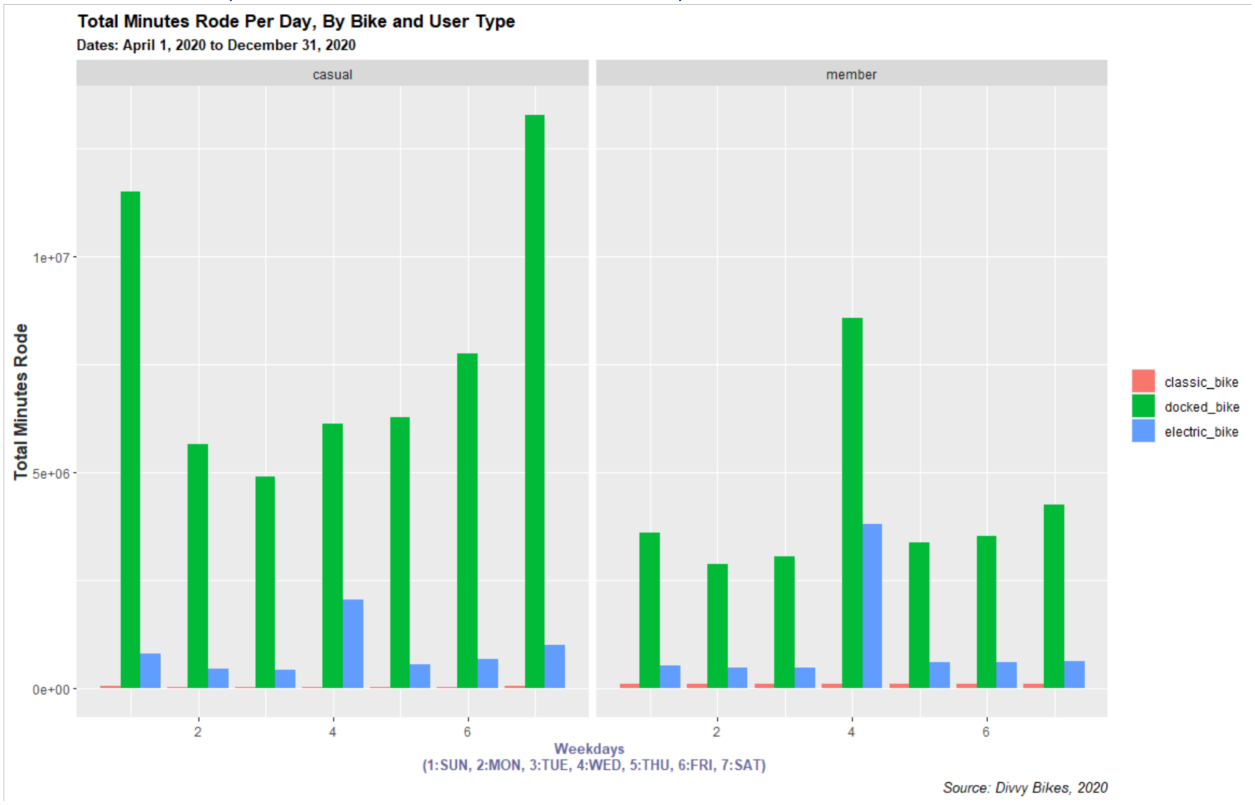
Bar Plot: Weekday Data - Total Number of Rides



(Figure 3.0)

This graph illustrates the total number of rides per day for each bike type by member type. The docked bikes between both groups show a clear dominant usage in comparison to the other two bike types. In second for both groups the preference goes to electric bikes. With very minimal usage again between both groups is the classic bike, with casual riders barely reporting in anything in comparison to the other types. In terms of usage by day, for casual riders the trend pattern shows to primarily focused on the weekend with Saturday the highest. This goes similar for casual riders for the electric bikes as well. For annual members, the difference is illustrated in its distribution of usage, with it being more evenly distributed. Across all bike types, there is a relatively steady number of rides across the weekday. Potential behavioral pattern is that these are being used for everyday or commuting purposes.

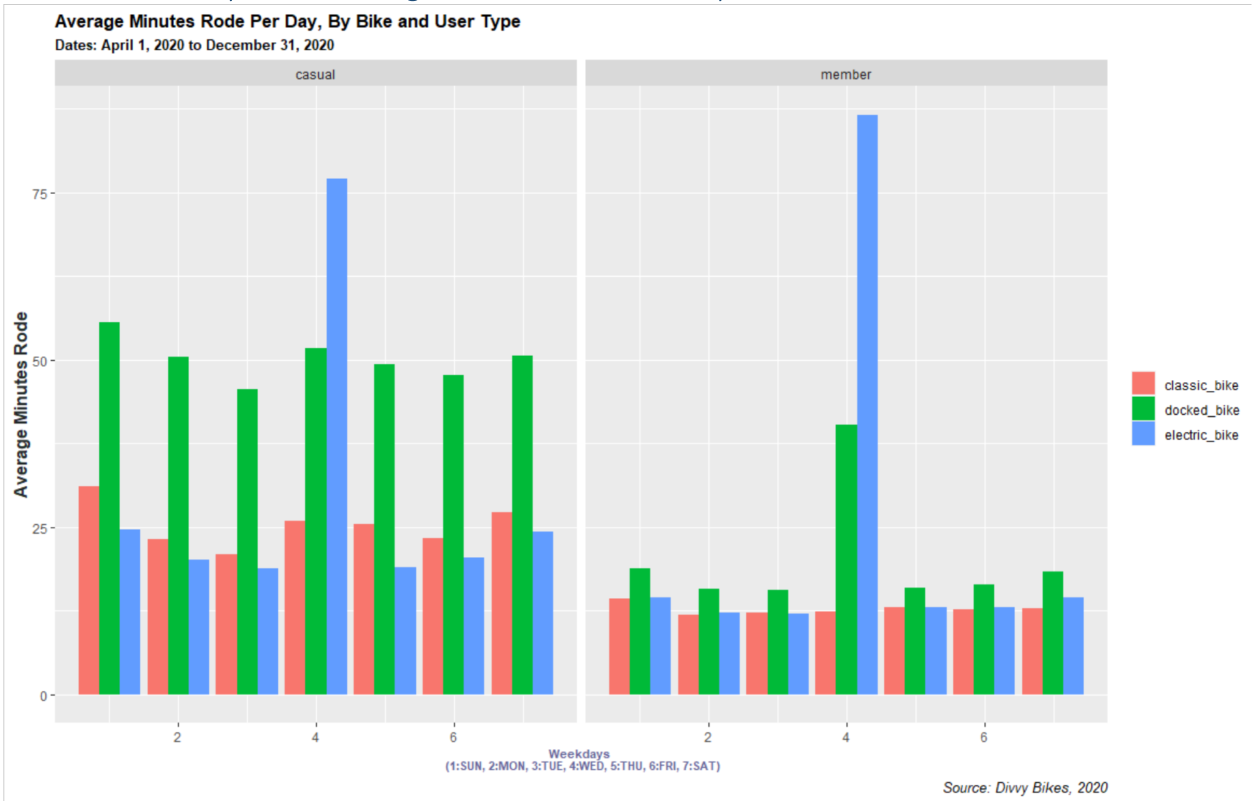
Bar Plot: Weekday Data - Total Minutes Rode Per Day



(Figure 3.1)

Viewing the total minutes rode per day by bike type for each member type, we can see again that docked bikes are the most rode in terms of absolute minute for both users. Casual riders, again, riding the most on weekends, with a gradual increase throughout the weekday. For electric bikes, casual riders instead seem to illustrate a potential normal distribution for the total minutes rode. With Wednesday being a spike in electric bike usage compared to the other days, which fall into a similar sum of minutes. For annual members, the pattern seems to follow a more normal(ish) distribution for the docked and electric bike, with both bikes spiking in total minutes rode on Wednesday. It seems that on Wednesday for an unbeknownst reason the annual members increase their total minutes rode that day relative to the rest of the week. A further investigation on behavioral patterns on why annual members do so could prove to be insightful. For both user types, the classic bike is barely again used at all in terms of usage by minute. However, the stark difference in usage by minutes between the groups again shows that potentially casual riders are using it for leisure and annual members are using for non-leisure/commuting reasons.

Bar Plot: Weekday Data - Average Minutes Rode Per Day

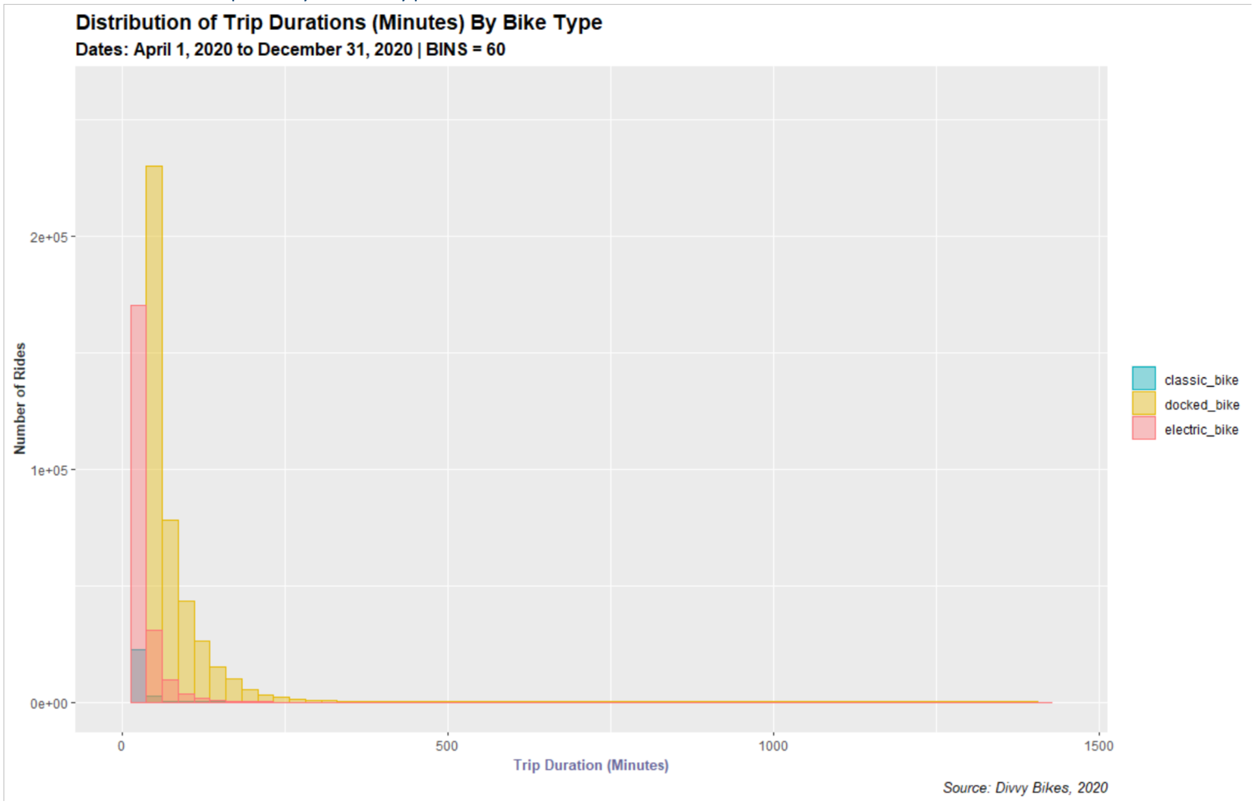


(Figure 3.2)

After viewing the visualizations for total minutes and number of rides, the final assessment is through the average minutes rode per day for each bike type by user type. Here the patterns viewed are not necessarily following the same trend as the other two plots. For casual riders, the docked bike is more uniformly distributed in average minutes per day, while for the electric bike it follows more a normal(ish) distribution. Although, in relative sense the classic bikes were seldom used in comparison to the other two bike types, casual riders stilled averaged around 25 minutes or so per day, which is on most days greater than the electric bikes. However, this could be that although they are not using them in higher frequency, when they do use them, they on average ride them longer. For annual members, the docked and electric bikes follow a relative uniform distribution of average minute usage, until again for Wednesday. This is somewhat expected, as the total number of minutes rode per day matches similar trends. For classic bike again it seems to be almost at par with the other two bike types, however, again this is for potentially the same reason as the casual rider's usage on average. Across both groups though on average in terms of each day, the docked bike is most preferred compared to the other bike types.

Distribution of Usage Less Than 24 Hours (1,440 minutes)

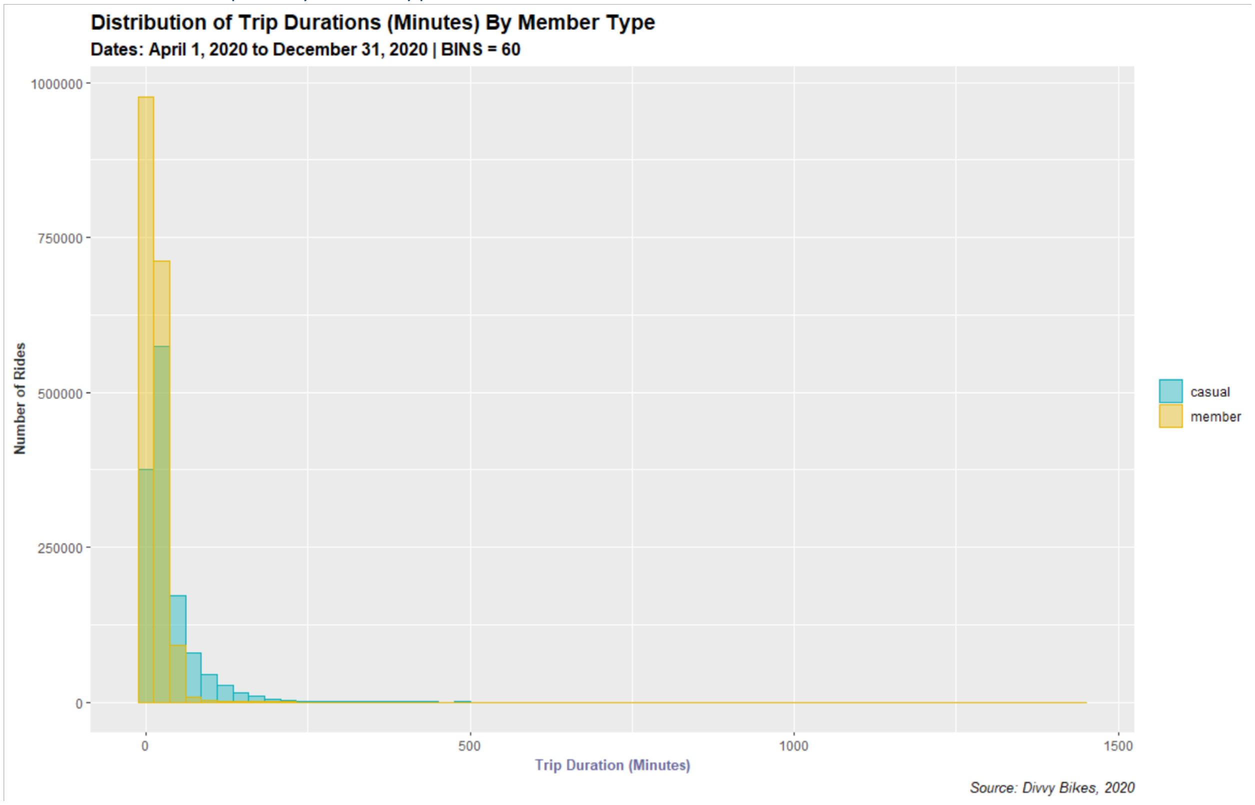
Distribution of Trips - By Bike Types



(Figure 4.0)

Analysis of the distribution of trip duration (bin = 60 minutes) by bike type, reveals that it is right skewed across all types. Docked bikes having their peak usage with two hours and electric for a single hour. It seems, classic bike, compared to the other bikes is used significantly less, but also within a single hour the most.

Distribution of Trips - By User Types



(Figure 4.1)

Distribution of trip duration (bin = 60 minutes) reveals a right skewed distribution for both user types. In terms of behavioral differences, it seems that the majority of annual members are using the bikes at most for one hour, while casual riders are using them for at least two hours. Further investigation on usage reasons could prove to be insightful for behavioral differences between the user types.

To access and inspect the code for the visualizations, please click below:

- Sharing: File Directory 07
- Sharing: File Directory 08

Phase: Act

Recommendations

Target Ads

Stakeholder expectations are set to increase the conversion rates of casual riders into annual members, to increase the revenue stream of the company. As such, a recommendation for the marketing analytics team and the director of marketing would be to place ads when the casual riders are using the service the most. Looking at figures 1.0 and 1.3, the recommended time to place the highest frequency of advertisements would be in the summer months (June-September), with August being the most targeted month. Given it's the peak ridership month for casual riders. To narrow down further within which days to target, looking at figure 2.0, casual riders are using the service the most on the weekends, with Saturday being the peak. Therefore, it is recommended that the highest frequency of ads should be set on that day.

Surveys

For the content of the advertisements, it is recommended to understand what drives people to sign up for such services, by understanding their usage preferences for Cyclistic. Figure 2.2 and figure 4.1, respectively show that annual members are using the bikes uniformly for docked bikes and (slightly) normally distributed for electric bikes. Figure 4.1 shows that they are using the bikes mostly within a one-hour period. A survey with a sufficient randomized sample size could be sent out to annual members to understand why they signed up and what they use the service for primarily. This way it will potentially provide insights as to the reasons of use, and then see if casual riders would align with such preferences as well. Targeting the ads closer to what could help convert riders into an annual membership.

Cost-Efficiency: Restructuring Bike Type Availability

For a potential cost-saving structure in the availability of bikes, figure 3.0 and figure 4.0, both show a clear trend of docked and electric bikes being used the most. If casual riders are enticed to increase the total number of their rides with a given increase in either bike types of availability, a potential possibility could be to shift resources from classic bikes to docked and electric bikes. A further analysis on whether this is a cost-effective possibility could be insightful for later projects.

Appendix

Links

[Capstone Project Code: GitHub]
[Capstone Project Graphs: GitHub]
[Capstone Project Data Sets: Google Drive]
[Capstone Project Original Raw Data: AWS]