

## Case Study 1: How Does a Bike-Share Navigate Speedy Success?

### Introduction

I will perform many real-world tasks of a junior data analyst in this case study. In this study, I will work for a fictional company, Cyclistic, and meet different characters and team members. I will answer the key business questions using the data analysis process: ask, prepare, process, analyze, share, and act. By the end of this lesson, I will have a portfolio-ready case study.

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

### Scenario

I am a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

### Characters & teams

- Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can't use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.
- Lily Moreno: The director of marketing and your manager. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels. •
- Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy. You joined this team six months ago and have been busy learning about Cyclistic's mission and business goals — as well as how you, as a junior data analyst, can help Cyclistic achieve them.
- Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

### About the company

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

## Ask

### Guiding Questions

1. How do annual members and casual riders use Cyclistic bikes differently?
  - a. Annual members – I think annual members are people who ride Cyclistic bike's on a repeated basis or as their main form of transportation.
  - b. Casual riders – I think casual riders are riders that may ride the Cyclistic bike's every once in a while for leisure or minor transportation. Casual riders may not be using Cyclistic bike's for their main form of transportation. Or they could be, and can't afford the annual memberships due to living on budget on budget.
2. Why would casual riders buy Cyclistic annual memberships?
  - a. Casual riders might buy Cyclistic annual memberships if the casual riders are repeatedly using these bike's as their main form of transportation. Casual riders may also buy the annual memberships if their repeated expense of the single-ride/full-day passes exceeds the annual memberships cost.
3. How can Cyclistic use digital media to influence casual riders to become members?
  - a. Can market towards the casual riders. Cyclistic can send promotions or deals to get casual riders to be annual members.

Moreno has assigned me the first question to answer: How do annual members and casual riders use Cyclistic bikes differently?

- I am initially thinking here, I will be analyzing the frequency and duration (in distance and time) of bike rides. From here, I expect that annual members will have a certain trend that is different from the whole of the casual riders. If this is the case, I will look for a subset (if there is any) of the casual riders that compare in similarity to the annual members. 30% of the cyclist use the bikes to commute to work everyday. Will we find a correlation of cyclists commuting to work

everyday with annual memberships? In other words, can we expect to see 30% of the cyclists to be in the annual membership plan? If so, how can we convert casual riders to annual membership (Moreno's plan) or introduce a new audience to the Cyclistic bikes (probably not likely due to Moreno's goals).

#### Deliverables:

1. Clear statement of the business task
  - a. I will be looking for differences in annual members and casual riders of the Cyclistic bikes in order to provide data for insight for the following two questions.
    - i. Why would casual riders buy Cyclistic annual memberships?
    - ii. How can Cyclistic use digital media to influence casual riders to become members?
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of your analysis
5. Supporting visualizations and key findings
6. Your top three recommendations based on your analysis

#### Prepare

- I will be using 2019 Q1-Q4 datasets.
- We have full license to use this data under Motivate International Inc.'s Data License Agreement.

#### Potential bias or credibility in this data

- I will be checking for bias or credibility in this data using the ROCCC method
  - Reliable
    - This is publicly available data provided by Motivate International Inc. Data includes certain City of Chicago's Divvy system data.
  - Original
    - The dataset is public data.
  - Comprehensive
    - The dataset is understandable
  - Current
    - The dataset we are using is from 2020-2021. Relatively current. In the midst of covid so the data might not reflect previous years. That said, consumer behavior might have changed years following COVID due to covid being such an impactful phase in human society.
  - Cited
    - Cited from Motivate International Inc.

#### Process

1. Downloaded August 2020 – 2021 July divvy trip data onto personal desktop.
2. Unzipped files, renamed files, organized files to .xls, zipped, and extracted files.

3. Created two new columns in all the spreadsheets
  - a. ride\_length
    - i. calculated by subtracting the "started\_at" from the column "ended\_at"
    - ii. reformatted column to HH:MM:SS format
  - b. day\_of\_week
    - i. calculated by using "WEEKDAY" command. 1 = Sunday and 7 = Saturday
4. Cleaning data/Sort/Filter
  - a. Checking for duplicates using the "Remove Duplicates"
  - b. Sorted data through the started\_at date.

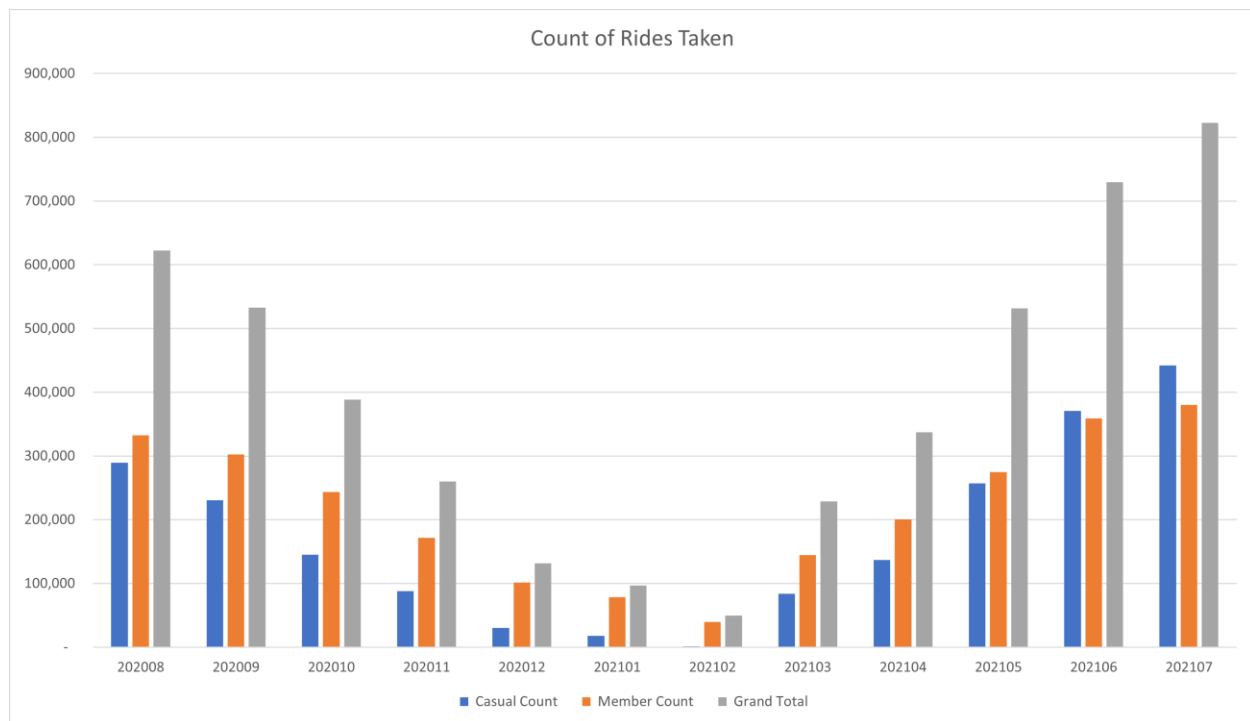
## Analyze

- For this step, I would use Excel and SQL for my analysis.

## Excel

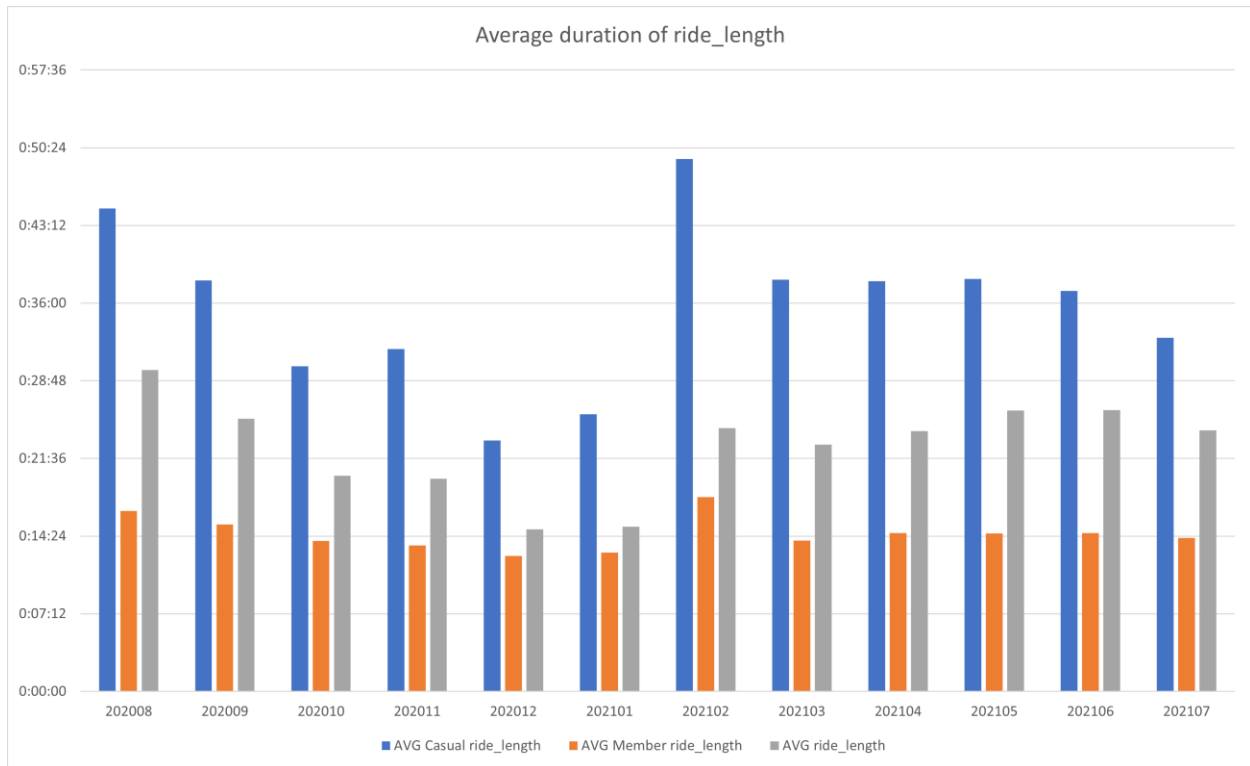
1. Where relevant, make columns consistent and combine them into a single worksheet.
  - a. Will combine some datasets for easier analysis because managing 12 is difficult
  - b. Will be splitting the spreadsheets into 4 sections. 2020 and 2021
2. Clean and transform your data to prepare for analysis.
3. Conduct descriptive analysis.
4. Run a few calculations in one file to get a better sense of the data layout. Options:
  - a. Calculate the mean of ride\_length
  - b. Calculate the max ride\_length
  - c. Calculate the mode of day\_of\_week
  - d. Calculate the number of rides
5. Create a pivot table to quickly calculate and visualize the data. Options:
  - a. Calculate the average ride\_length for members and casual riders. Try rows = member\_casual; Values = Average of ride\_length.
  - b. Calculate the average ride\_length for users by day\_of\_week. Try columns = day\_of\_week; Rows = member\_casual; Values = Average of ride\_length.
  - c. Calculate the number of rides for users by day\_of\_week by adding Count of trip\_id to Values.
6. Open another file and perform the same descriptive analysis steps. Explore different seasons to make some initial observations.
7. merged into a full-year view
8. Created visualizations for the data collected
9. Export a summary file for further analysis.

## Summary File



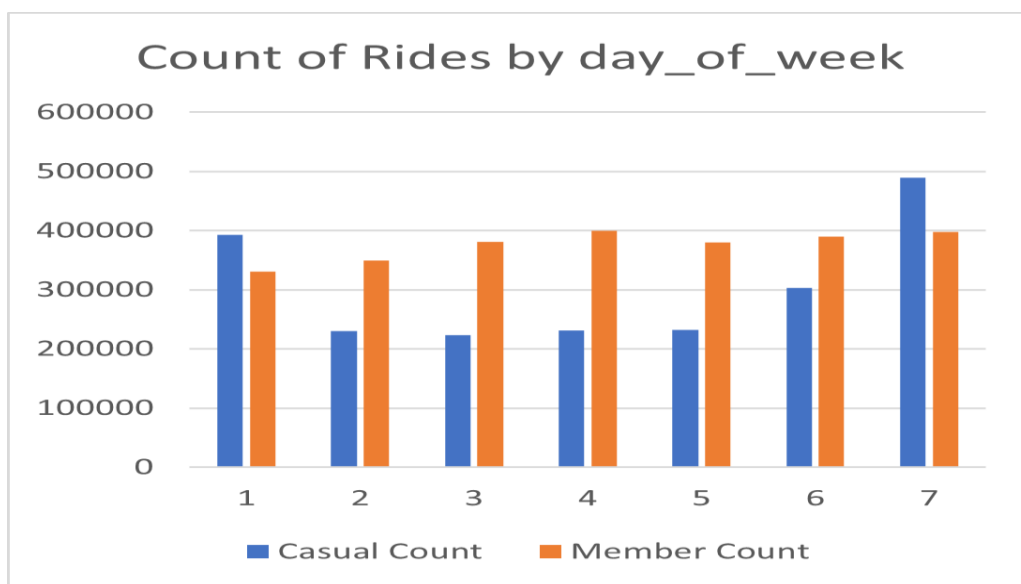
We have a visualization of the count of bike rides taken by casual and member riders spread out by the past 12 months. We see big dips in the count around 202011 – 202103. These months include some of the biggest holidays: thanksgiving, Christmas, new years, and valentines day which can have some affect on the count due to people spending time with family. Looking at the timeline of the city of Chicago health-orders, the city was in phase 4 of covid during the start of this time-period. Things started slowly reopening. However, around mid-November 2020 and late January 2021 there was a stay-at home-advisory sent out due to the rapid rise of COVID-19 cases which can potentially explain the dip we see in 202011-202103. During mid-May, a couple of factors such as, restrictions on bars and restaurants get lifted, more people getting vaccinated, and COVID cases dropping, can potentially attribute to the increase of rides. From 202008 – 202105, we see more member counts to casual counts. This starts changing to more casual counts around 202105 – 202107. I speculate this may be due to the warming climate and more casual riders willing to ride in Chicago. The cold months of the city of Chicago is December 2 to March 9, with an average daily high temperature below 43 degrees. This cold may deter riders on both sides to find other more protected forms of transportation.

- a. I will be looking for differences in annual members and casual riders of the Cyclistic bikes in order to provide data for insight for the following two questions.



We have here the average duration of ride\_length throughout the whole months. Even while the whole covid situation was going on, we don't see too wide of duration drop off especially for the member group.

The differences between casual riders and member riders is that casual riders seem to ride at least twice as long as member riders. Seeing that member riders are more consistent durations as well, it seems like members ride to commute back and forth to places such as work, gym, etc. Casual riders seem to ride for leisure.



1 = Sunday, 7 = Saturday

We see here that the member count is relatively consistent throughout the weeks which can mean the same people are riding the bikes in a consistent day to day basis. For the casual members, we have relatively low usage Monday – Thursday and a ramp in usage on Friday – Sunday with the most rides on Saturday. This lines up with our theory that casual riders ride for leisure when they have time on Friday – Sunday. If we want to convert casual riders to members, I would suggest focusing on the casual riders on Monday – Thursday, because we see a stable amount of counts meaning this group of people may potentially be riding for commuting purposes. Or we can make a package called the weekend deal that includes fri – sun to have a mid level between casual and the members.

## SQL

yyymm	casual_count	member_count	total_count	casual_avg_ride_length	member_avg_ride_length	avg_ride_length
202008	289661	332700	622361	0:44:46	0:16:45	0:29:47
202009	230692	302266	532958	0:38:07	0:15:28	0:25:16
202010	145012	243641	388653	0:30:09	0:13:57	0:19:59
202011	88099	171617	259716	0:31:45	0:13:32	0:19:42
202012	30080	101493	131573	0:49:22	0:18:01	0:24:25
202101	18117	78717	96834	0:25:41	0:12:52	0:15:16
202102	10131	39491	49622	0:49:22	0:18:01	0:24:25
202103	84033	144463	228496	0:38:10	0:13:58	0:22:52
202104	136601	200629	337230	0:38:01	0:14:41	0:24:08
202105	256916	274717	531633	0:38:14	0:14:38	0:26:02
202106	370681	358914	729595	0:37:07	0:14:41	0:26:05
202107	370681	358914	729595	0:32:47	0:14:14	0:24:13

We have here a table that I exported from SQL into EXCEL. Same data points from excel. I imported all the datasets into bigquery and created this table.

## Share

Link to Presentation

[https://docs.google.com/presentation/d/1KAZpr-Is3pju-IIQsKTA5UbbVoXt0nwhDVI4HFV\\_c/edit?usp=sharing](https://docs.google.com/presentation/d/1KAZpr-Is3pju-IIQsKTA5UbbVoXt0nwhDVI4HFV_c/edit?usp=sharing)

## Act

## Deliverables summed up

Deliverables:

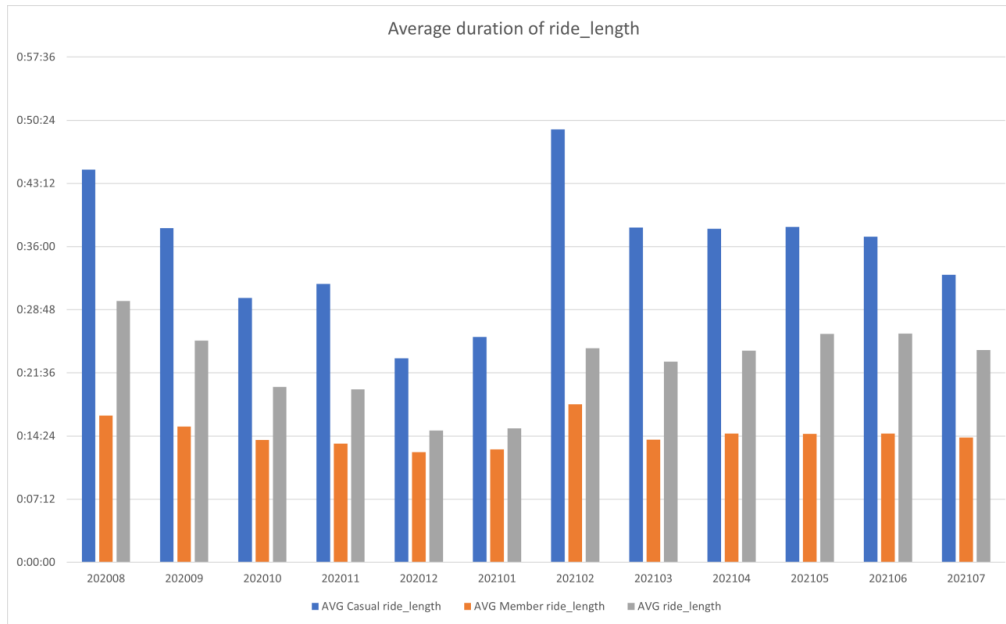
1. Clear statement of the business task
  - a. I will be looking for differences in annual members and casual riders of the Cyclistic bikes in order to provide data for insight for the following two questions.
    - i. Why would casual riders buy Cyclistic annual memberships?
    - ii. How can Cyclistic use digital media to influence casual riders to become members?
2. A description of all data sources used
  - a. Excel data files got obtained from a publicly available data storage provided by Motivate International Inc. Data

- b. Each file contains bicycling records for one month
  3. Documentation of any cleaning or manipulation of data
    - a. Downloaded August 2020 – 2021 July divvy trip data onto personal desktop.
    - b. Unzipped files, renamed files, organized files to .xls, zipped, and extracted files.
    - c. Created two new columns in all the spreadsheets
      - i. ride\_length
        1. calculated by subtracting the “started\_at” from the column “ended\_at”
        2. reformatted column to HH:MM:SS format
      - ii. day\_of\_week
        1. calculated by using “WEEKDAY” command. 1 = Sunday and 7 = Saturday
    - d. Cleaning data/Sort/Filter
      - i. Checking for duplicates using the “Remove Duplicates”
      - ii. Sorted data through the started\_at date.
  4. A summary of your analysis & 5. Supporting visualizations and key findings



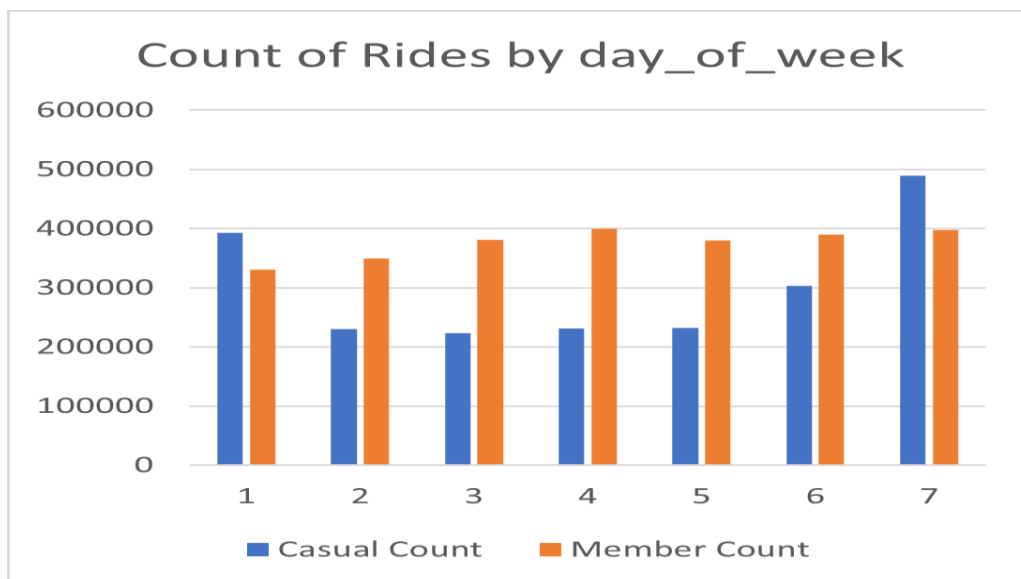
We have a visualization of the count of bike rides taken by casual and member riders spread out by the past 12 months. We see big dips in the count around 202011 – 202103. These months include some of the biggest holidays: thanksgiving, Christmas, new years, and valentines day which can have some affect on the count due to people spending time with family. Looking at the timeline of the city of Chicago health-orders, the city was in phase 4 of covid during the start of this time-period. Things started slowly reopening. However, around mid-November 2020 and late January 2021 there was a stay-at home-advisory sent out due to the rapid rise of COVID-19 cases which can potentially explain the dip we see in 202011-202103. During mid-May, a couple of factors such as, restrictions on bars and restaurants get lifted, more people getting vaccinated, and COVID cases dropping, can potentially attribute to the increase of rides. From 202008 – 202105, we see more member counts to casual counts. This starts changing to more casual counts around 202105 – 202107. I speculate this may be due to the warming climate and more casual riders willing to ride in Chicago. The cold months of the city of Chicago is December 2 to March 9, with an average daily high temperature below 43 degrees. This cold may deter riders on both sides to find other more protected forms of transportation.





We have here the average duration of ride\_length throughout the whole months. Even while the whole covid situation was going on, we don't see too wide of duration drop off especially for the member group.

The differences between casual riders and member riders is that casual riders seem to ride at least twice as long as member riders. Seeing that member riders are more consistent durations as well, it seems like members ride to commute back and forth to places such as work, gym, etc. Casual riders seem to ride for leisure which can explain the longer duration.



1 = Sunday, 7 = Saturday

We see here that the member count is relatively consistent throughout the weeks which can mean the same people are riding the bikes in a consistent day to day basis. For the causal members, we have relatively low usage Monday – Thursday and a ramp in usage on Friday – Sunday with the most rides on

Saturday. This lines up with our theory that casual riders ride for leisure when they have time on Friday – Sunday. If we want to convert casual riders to members, I would suggest focusing on the casual riders on Monday – Thursday, because we see a stable amount of counts meaning this group of people may potentially be riding for commuting purposes. Or we can make a package called the weekend deal that includes fri – sun to have a mid level between casual and the members.

6. Your top three recommendations based on your analysis

- 1) Create different membership options
  - Weekend memberships, weekday memberships
    - Can tailor to the needs of the wide amount of casual riders that are using single-use and day passes
    - There may be Monday-Friday casual riders that don't think the annual membership is worth it because they don't ride on the weekends. Creating a sperate tier membership or them maybe a good choice
- 2) Give an option to opt out on winter months
  - The count of riders in the winter are small for casual and members. This may be holding riders back from getting memberships. So we are missing out on membership passes due to these 3 months. We are operating poorly in these months anyways so better to take the member profits from the other months than potentially losing member profits completely.
- 3) Offer discounted prices in the winter months to boost sales/deep check the durability and condition of bikes
  - It is better to get some cashflow rather than none in these months
  - Since bikes are not used as much here, it will be good to check the condition of the bikes to prevent malfunctions and accidents with bikers.

## Log

### Excel

1. Downloaded August 2020 – 2021 July divvy trip data onto personal desktop.
2. Unzipped files, renamed files, organized files to .xls, zipped, and extracted files.
3. Created two new columns in all the spreadsheets
  - a. ride\_length
    - i. calculated by subtracting the “started\_at” from the column “ended\_at”
    - ii. reformatted column to HH:MM:SS format
  - b. day\_of\_week
    - i. calculated by using “WEEKDAY” command. 1 = Sunday and 7 = Saturday
4. Cleaning data/Sort/Filter
  - a. Checking for duplicates using the “Remove Duplicates”
  - b. Sorted data through the started\_at date.
5. Create a pivot table to quickly calculate and visualize the data. Options:

- a. Calculate the average ride\_length for members and casual riders. Try rows = member\_casual; Values = Average of ride\_length.
  - b. Calculate the average ride\_length for users by day\_of\_week. Try columns = day\_of\_week; Rows = member\_casual; Values = Average of ride\_length.
  - c. Calculate the number of rides for users by day\_of\_week by adding Count of trip\_id to Values.
6. Combined Pivot tables into one Excel Page for easier analysis

## SQL

1. Uploaded all edited excel files into SQL Big Query
2. Files over 1,000,000 kb was uploaded to the cloud and then transferred over to Big Query
3. These are the queries I ran in SQL

```
CREATE TABLE `case-study-1-323322.divvy_trips_202008_202107.divvyTrip_12_month` (
  yyyyymm INT,
  casual_count INT,
  member_count INT,
  total_count INT,
  casual_avg_ride_length TIME,
  member_avg_ride_length TIME,
  avg_ride_length TIME,
)
```

## INSERT INTO

```
`case-study-1-323322.divvy_trips_202008_202107.divvyTrip_12_month`
  (yyyyymm, casual_count, member_count, total_count,
   casual_avg_ride_length, member_avg_ride_length, avg_ride_length)
VALUES
  (202008,
   (SELECT COUNT(member_casual) FROM divvy_trips_202008_202107.divvyTrip_202008 WHERE member_
casual = 'casual'),
   (SELECT COUNT(member_casual) FROM divvy_trips_202008_202107.divvyTrip_202008 WHERE member_
casual = 'member'),
   (SELECT COUNT(member_casual) FROM divvy_trips_202008_202107.divvyTrip_202008),
   (SELECT AVG(ride_length) FROM divvy_trips_202008_202107.divvyTrip_202008 WHERE member_casu
al = 'casual'),
   (SELECT AVG(ride_length) FROM divvy_trips_202008_202107.divvyTrip_202008 WHERE member_casu
al = 'member'),
   (SELECT AVG(ride_length) FROM divvy_trips_202008_202107.divvyTrip_202008))
```