# Case Study of Cyclistic Bike Sharing Company

Stephen Hartley



## Introduction

This case study is the Google Data Analytics Capstone Project that can be found on [coursera.org](coursera.org) and is part of the Google Data Analytics Professional Certificate course. This case study is for the fictitious Bike Sharing company called Cyclistic based on the Chicago Divvy Bike sharing company. The data analysis will follow the 6-step process of Ask, Prepare, Process, Analyze, Share, and Act. Along the way, I will be referring to the Case Study Roadmap guiding questions and key tasks.

## Background

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of over 6,000 bicycles that are geo-tracked and locked into a network of over 1500 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

The Director of Marketing, Lily Moreno, has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, 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. Three questions will guide the future marketing program:

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

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

# 1. Ask

Following the case study roadmap's first section is "Ask", which refers to who is asking the questions and who are the major stakeholders.

Who is the Key Stakeholders: Lily Moreno

This case study asks the question: How do annual members and casual riders use the Cyclistic bikes differently?

# 2. Prepare

Following the case study roadmap's second section is "Prepare", which refers to the data.  Where and how has the data been made available, how was it collected, and if not internal to the company, then under what license.

The data for this case study has been provided by the Motivate International Inc. under license from the Divvy Bikes and Scooters, LLC. Here are the links to the data and license. I am using the data for the last 12 months - July 2022 to June 2023.  For privacy all customer data was never included in the date files.

Downloaded the 12 monthly zip files and unzipped the CSV files into a separate folder, these files when unzipped ranged in size from 36MB to 141MB or 181 thousand rows to over 800 thousand rows. Each file was named with year-month-tripdata.csv.  Each row in the file described a rental trip by either a member or casual customer with a starting date timestamp and ending date timestamp, along with station names and/or geo tracking Latitude and Longitude.  At first look these files they were too large to combine into one spreadsheet with some missing data, like station names and geo data.

# 3. Process

Following the case study roadmap's third section is "Process", which refers to cleaning the data, including what steps taken and using what tools.

To start, I used Excel to do a preliminary pass through the data.

Created a column "ride length" by subtracting the ending at date timestamp from the starting at date timestamp. I noticed that some of these ride lengths were either zero or a negative number, so I eliminated those rows.

I also noticed that some of the rows were missing both the ending station names and the geo data, so I assumed that these bikes were taken out of service or stolen, I removed those rows.

Next, I wanted to fill in the missing station names by using the Geo data to locate the name of the station. To do this, created a separate sheet with the list of only station names and their Geo data (longitude and latitude). What I found was that each station did not have a single geo data location. It appears that the unit collecting the geo data was not exact even though some of the data went out 9 decimal points, so I reduced the data width to only 5 decimal places which helped align some of the data. I also collected multiple geo data for each station to allow for better matching of geo data to station names. I used a VLOOKUP function to fill in the missing station names from my master station name sheet to each of the month's spreadsheets.

The spreadsheet was getting very large so I removed unnecessary columns, all the longitude and latitude fields, the station ID fields, and ride ID columns. I added a column describing the day of the week each ride took place on using the =WEEKDAY(C2,1) function for number day of week "nday_of_week". Saved the spreadsheet and repeated for the other 11 months of files.

Now to combine all the months together I used the software R Studio which was able to read in all twelve spreadsheets and combine them into one data frame which allowed me to run calculations on the full year's worth of data. Creating reports of how many rides took place by the two customer categories (member and casual) breaking out the by days of the week and hour of the day. Placing this data into other data frames and then saving them into CSV files to be analyzed in the next section.

For complete list of these steps used and those in the next analyze phase, please refer to separate document "Cyclistic Bike Sharing - Technical Steps .docx"

# 4. Analyze

When looking at the ridership you can see that both types of customers (Casual and Members) take less rides during the cold winter months.

## Number of Rides per Month
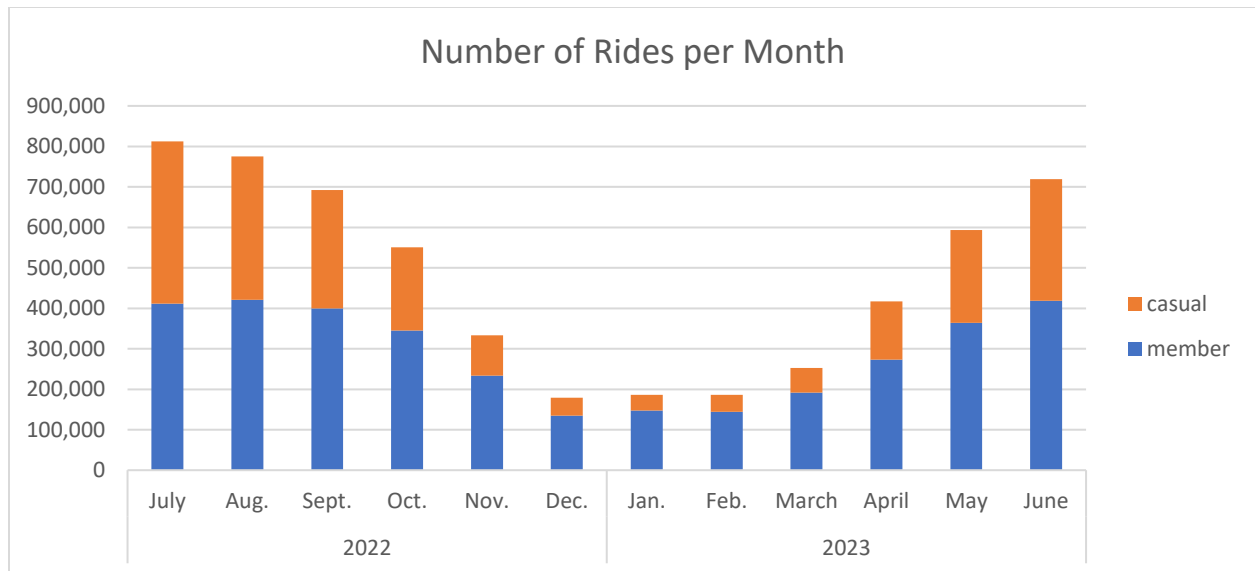
| Rides per Month | | | |
|---|---:|---:|---:|
| | Casual | Member | Total |
| **2022** | **1395804** | **1945634** | **3341438** |
| July | 400381 | 411877 | 812258 |
| Aug. | 353684 | 421225 | 774909 |
| Sept. | 292561 | 399359 | 691920 |
| Oct. | 205876 | 344738 | 550614 |
| Nov. | 99188 | 233761 | 332949 |
| Dec. | 44114 | 134674 | 178788 |
| **2023** | **816356** | **1538265** | **2354621** |
| Jan. | 39206 | 147157 | 186363 |
| Feb. | 42177 | 144370 | 186547 |
| March | 60861 | 191871 | 252732 |
| April | 144017 | 272902 | 416919 |
| May | 229677 | 363717 | 593394 |
| June | 300418 | 418248 | 718666 |
| **Grand Total** | **2212160** | **3483899** | **5696059** |

The Customers use the bikes differently during the week, with Member customers taking the most rides mid-week and the least over the weekend. The Casual customers are the opposite with the weekend (Friday, Saturday, and Sunday) having the most rides.
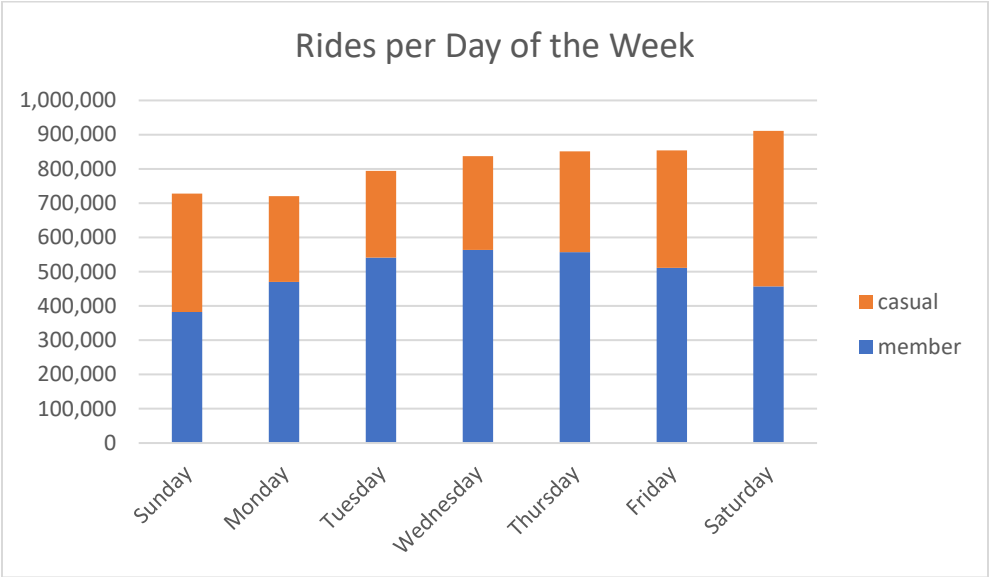
## Rides per Day of the Week

| | Casual | Member | Totals |
|---|---|---|---|
| Sunday | 346057 | 382167 | 728224 |
| Monday | 249541 | 470690 | 720231 |
| Tuesday | 253302 | 540953 | 794255 |
| Wednesday | 273294 | 563678 | 836972 |
| Thursday | 294188 | 557418 | 851606 |
| Friday | 342291 | 511378 | 853669 |
| Saturday | 453487 | 457615 | 911102 |
| **Grand Total** | **2212160** | **3483899** | **5696059** |

Heatmap Legend

| | |
|---|---|
| Most | |
| In Between | |
| Least | |

Looking at the Grand Totals of 5,696,059 trips of which Casual customers took 2,212,160 rides (39%) and Member customers took 3,483,899 rides (61%).

My data has found that the busiest time for Casual customers is Saturday at 3pm and Saturday is also the busiest day.

## Casual Rides per Hour

| | Sun. | Mon. | Tues. | Wed. | Thurs. | Fri. | Sat. |
|---|---|---|---|---|---|---|---|
| 0am | 13527 | 3454 | 2597 | 2789 | 3522 | 5228 | 11349 |
| 1am | 8945 | 2021 | 1509 | 1554 | 1942 | 3221 | 8541 |
| 2am | 5949 | 1355 | 891 | 925 | 1186 | 1861 | 4769 |
| 3am | 3124 | 1013 | 646 | 665 | 802 | 1180 | 2407 |
| 4am | 1813 | 915 | 720 | 729 | 674 | 841 | 1163 |
| 5am | 1663 | 1693 | 1862 | 1958 | 1798 | 1864 | 1322 |
| 6am | 2499 | 4127 | 5352 | 5254 | 5363 | 6334 | 2476 |
| 7am | 3960 | 6863 | 9675 | 9873 | 9808 | 8698 | 4568 |
| 8am | 6647 | 9045 | 11912 | 12475 | 12700 | 10073 | 9121 |
| 9am | 12263 | 8340 | 8386 | 8813 | 9122 | 9798 | 15369 |
| 10am | 18622 | 9804 | 8716 | 8629 | 9563 | 11833 | 22872 |
| 11am | 24069 | 12475 | 10719 | 11361 | 11898 | 15763 | 29385 |
| 12pm | 27075 | 15485 | 13179 | 13205 | 14659 | 19863 | 33952 |
| 1pm | 28436 | 16270 | 13515 | 13680 | 15239 | 21058 | 35911 |
| 2pm | 29820 | 17104 | 14356 | 15004 | 16720 | 23131 | 36872 |
| 3pm | 31256 | 18916 | 17420 | 17949 | 20213 | 27362 | 37926 |
| 4pm | 29213 | 21908 | 23185 | 24618 | 26039 | 30578 | 37221 |
| 5pm | 25160 | 25337 | 29033 | 31821 | 32963 | 33216 | 34581 |
| 6pm | 21173 | 22247 | 24698 | 27135 | 29564 | 30055 | 31020 |
| 7pm | 16001 | 17012 | 18050 | 20676 | 21775 | 22568 | 23948 |
| 8pm | 11484 | 12395 | 12744 | 14970 | 15428 | 16346 | 18344 |
| 9pm | 9671 | 9654 | 11144 | 13247 | 13321 | 13803 | 16955 |
| 10pm | 8208 | 7322 | 8198 | 9947 | 12243 | 14345 | 17235 |
| 11pm | 5479 | 4786 | 4795 | 6017 | 7646 | 13272 | 16180 |
| Total | 346057 | 249541 | 253302 | 273294 | 294188 | 342291 | 453487 |

| Heatmap Legend | Most | In Between | Least |
|---|---|---|---|

The busiest time for Member customers is Wednesday at 5pm and is also the busiest day of the week.

It is also noticeable that a lot of the Members are using the bike to commute to and from work with rush hours of 7am and 8am along with 4, 5, 6 pm.

## Member Rides per Hour

| | Sun. | Mon. | Tues. | Wed. | Thurs. | Fri. | Sat. |
|---|---|---|---|---|---|---|---|
| 0am | 9822 | 2979 | 2580 | 3019 | 3619 | 4786 | 9219 |
| 1am | 6634 | 1647 | 1374 | 1456 | 1790 | 2702 | 6431 |
| 2am | 4143 | 1066 | 753 | 785 | 1062 | 1504 | 3443 |
| 3am | 2311 | 784 | 613 | 577 | 702 | 1012 | 1994 |
| 4am | 1485 | 1204 | 1287 | 1154 | 1205 | 1299 | 1156 |
| 5am | 1892 | 4933 | 6319 | 6277 | 5733 | 5138 | 1922 |
| 6am | 4156 | 15351 | 20416 | 19365 | 18359 | 15942 | 5003 |
| 7am | 7160 | 27323 | 38449 | 37866 | 35807 | 26176 | 9084 |
| 8am | 10710 | 33011 | 45627 | 44424 | 43756 | 30049 | 16156 |
| 9am | 16883 | 19679 | 23921 | 24584 | 25056 | 20814 | 22893 |
| 10am | 23234 | 16378 | 17698 | 18477 | 18691 | 18716 | 28355 |
| 11am | 28006 | 19835 | 20444 | 21750 | 21887 | 22891 | 32317 |
| 12pm | 29665 | 23380 | 24475 | 25809 | 26347 | 27786 | 33772 |
| 1pm | 30190 | 23488 | 23475 | 25566 | 25593 | 29016 | 33508 |
| 2pm | 30042 | 23917 | 23239 | 25295 | 25481 | 30433 | 33930 |
| 3pm | 31203 | 29790 | 32673 | 33464 | 34174 | 38258 | 33644 |
| 4pm | 30373 | 43465 | 52118 | 53670 | 50723 | 46581 | 33667 |
| 5pm | 28459 | 55918 | 65708 | 68147 | 64693 | 51103 | 33695 |
| 6pm | 25877 | 43783 | 48576 | 51002 | 50455 | 42828 | 31098 |
| 7pm | 19812 | 30862 | 33609 | 36085 | 35273 | 30585 | 25090 |
| 8pm | 14759 | 21261 | 22803 | 25047 | 23815 | 21062 | 18484 |
| 9pm | 11522 | 15552 | 17599 | 19374 | 19303 | 16272 | 15813 |
| 10pm | 8596 | 9690 | 11280 | 13288 | 15057 | 14686 | 14380 |
| 11pm | 5233 | 5394 | 5917 | 7197 | 8837 | 11739 | 12561 |
| Total | 382167 | 470690 | 540953 | 563678 | 557418 | 511378 | 457615 |

Heatmap Legend   Most   In Between   Least

The busiest time for all customers is Wednesday at 5pm, but Saturdays are the busiest day.

## Total Rides per Hour

| | Sun. | Mon. | Tues. | Wed. | Thurs. | Fri. | Sat. |
|---|---|---|---|---|---|---|---|
| 0am | 23349 | 6433 | 5177 | 5808 | 7141 | 10014 | 20568 |
| 1am | 15579 | 3668 | 2883 | 3010 | 3732 | 5923 | 14972 |
| 2am | 10092 | 2421 | 1644 | 1710 | 2248 | 3365 | 8212 |
| 3am | 5435 | 1797 | 1259 | 1242 | 1504 | 2192 | 4401 |
| 4am | 3298 | 2119 | 2007 | 1883 | 1879 | 2140 | 2319 |
| 5am | 3555 | 6626 | 8181 | 8235 | 7531 | 7002 | 3244 |
| 6am | 6655 | 19478 | 25768 | 24619 | 23722 | 22276 | 7479 |
| 7am | 11120 | 34186 | 48124 | 47739 | 45615 | 34874 | 13652 |
| 8am | 17357 | 42056 | 57539 | 56899 | 56456 | 40122 | 25277 |
| 9am | 29146 | 28019 | 32307 | 33397 | 34178 | 30612 | 38262 |
| 10am | 41856 | 26182 | 26414 | 27106 | 28254 | 30549 | 51227 |
| 11am | 52075 | 32310 | 31163 | 33111 | 33785 | 38654 | 61702 |
| 12pm | 56740 | 38865 | 37654 | 39014 | 41006 | 47649 | 67724 |
| 1pm | 58626 | 39758 | 36990 | 39246 | 40832 | 50074 | 69419 |
| 2pm | 59862 | 41021 | 37595 | 40299 | 42201 | 53564 | 70802 |
| 3pm | 62459 | 48706 | 50093 | 51413 | 54387 | 65620 | 71570 |
| 4pm | 59586 | 65373 | 75303 | 78288 | 76762 | 77159 | 70888 |
| 5pm | 53619 | 81255 | 94741 | 99968 | 97656 | 84319 | 68276 |
| 6pm | 47050 | 66030 | 73274 | 78137 | 80019 | 72883 | 62118 |
| 7pm | 35813 | 47874 | 51659 | 56761 | 57048 | 53153 | 49038 |
| 8pm | 26243 | 33656 | 35547 | 40017 | 39243 | 37408 | 36828 |
| 9pm | 21193 | 25206 | 28743 | 32621 | 32624 | 30075 | 32768 |
| 10pm | 16804 | 17012 | 19478 | 23235 | 27300 | 29031 | 31615 |
| 11pm | 10712 | 10180 | 10712 | 13214 | 16483 | 25011 | 28741 |
| Total | 728224 | 720231 | 794255 | 836972 | 851606 | 853669 | 911102 |

Heatmap Legend ■ Most ■ In Between ■ Least

The next question is which stations are the busiest?  (For both starting and ending trip locations)

| Top Ten Locations | | | | | | | | Grand Total |
|---|---|---|---|---|---|---|---|---|
| | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | |
| Grand Total | 1330675 | 1326507 | 1464228 | 1538439 | 1560678 | 1555840 | 1658209 | 10434576 |
| Streeter Dr & Grand Ave* | 24714 | 16428 | 14418 | 14697 | 15969 | 20365 | 33686 | 140277 |
| DuSable Lake Shore Dr & Monroe St* | 15258 | 8869 | 7382 | 7572 | 8427 | 11238 | 20699 | 79445 |
| DuSable Lake Shore Dr & North Blvd* | 13294 | 9533 | 8833 | 9915 | 9982 | 9616 | 18124 | 79297 |
| Michigan Ave & Oak St* | 13593 | 8707 | 8560 | 8935 | 9671 | 11354 | 17472 | 78292 |
| Wells St & Concord Ln* | 10263 | 8804 | 9100 | 10063 | 10543 | 11264 | 14926 | 74963 |
| Clark St & Elm St | 9423 | 9126 | 10075 | 10420 | 10286 | 10288 | 10711 | 70329 |
| Kingsbury St & Kinzie St | 6504 | 10472 | 11455 | 11206 | 10716 | 10021 | 7773 | 68147 |
| Millennium Park* | 11164 | 8521 | 7130 | 7037 | 9040 | 10246 | 14221 | 67359 |
| Theater on the Lake* | 11944 | 7381 | 6763 | 6827 | 7208 | 8187 | 15383 | 63693 |
| Halsted St & Fulton St | 7031 | 7001 | 8521 | 9631 | 10343 | 9859 | 9255 | 61641 |

It is interesting that 7 of the top 10 locations are all along the edge of Lake Michigan*.

(Not near the lake: Clark St & Elm St, Kingsbury St & Kinzie St. Halsted St & Fulton St)

Where is everyone going?

| Top Ten Casual Ending Locations | |
|---|---|
| Streeter Dr & Grand Ave* | 54925 |
| DuSable Lake Shore Dr & Monroe St* | 28132 |
| Michigan Ave & Oak St* | 25122 |
| Millennium Park* | 24959 |
| DuSable Lake Shore Dr & North Blvd* | 23994 |
| Theater on the Lake* | 18319 |
| Shedd Aquarium* | 17262 |
| Wells St & Concord Ln* | 14592 |
| Dusable Harbor* | 13116 |
| Clark St & Armitage Ave* | 12683 |

For the casual riders the top ten ending station are all near Lake Michigan.  Compared to the member riders few of their locations are near the lake.

| Top Ten Member Ending Locations | |
| --- | --- |
| Kingsbury St & Kinzie St | 25300 |
| Clark St & Elm St | 23270 |
| Wells St & Concord Ln | 23247 |
| Clinton St & Washington Blvd | 23154 |
| University Ave & 57th St | 21186 |
| Loomis St & Lexington St | 21080 |
| Clinton St & Madison St | 21054 |
| Ellis Ave & 60th St | 20364 |
| Wells St & Elm St | 19393 |
| Broadway & Barry Ave | 19162 |

In evaluating the data, I found a number of rides (274,715) that started and ended at the same stations, I call these rides "joy" rides.  When compared to the total number of rides (5,696,059) this amounts to about 5 % of the total, the casual riders spent more time riding the bikes.

| Breakdown of Joy Rides | | | |
| --- | --- | --- | --- |
| | Number of Joy Ride | Percentage of Total Rides | Average Ride Length in Minutes |
| casual | 163,856 | 3% | 34.7 |
| member | 110,859 | 2% | 15.3 |
| Totals | 5,696,059 | 274,715 | 5% | 26.9 |

## Top Ten Locations for Casual Joy Rides

|  | Number of Ride | Average Ride Length in Minutes |
|---|---|---|
| Streeter Dr & Grand Ave* | 9324 | 41.2 |
| DuSable Lake Shore Dr & Monroe St* | 6360 | 34.6 |
| Michigan Ave & Oak St* | 4148 | 46.1 |
| Millennium Park* | 3725 | 39.2 |
| Montrose Harbor* | 2640 | 49.6 |
| Dusable Harbor* | 2369 | 35.3 |
| Shedd Aquarium* | 2284 | 22.2 |
| DuSable Lake Shore Dr & North Blvd* | 2146 | 38.8 |
| Theater on the Lake* | 1949 | 42.7 |
| Adler Planetarium* | 1803 | 36.0 |

## Top Ten Locations for Member Joy Rides

|  | Number of Ride | Average Ride Length in Minutes |
|---|---|---|
| Halsted St & Fulton St | 1548 | 4.2 |
| Streeter Dr & Grand Ave* | 1285 | 23.1 |
| Loomis St & Lexington St | 1139 | 12.1 |
| Michigan Ave & Oak St* | 886 | 24.4 |
| N Clark St & W Elm St | 840 | 11.7 |
| State St & 33rd St | 766 | 12.5 |
| DuSable Lake Shore Dr & North Blvd* | 758 | 17.6 |
| Clark St & Elm St | 706 | 16.1 |
| N Sheffield Ave & W Wellington Ave | 697 | 4.4 |
| W Armitage Ave & N Sheffield Ave | 697 | 4.0 |

It appears that both riders like to go near the lake parks and go for a relaxing spin through the parks along the waterfront.

# 5. Share

This analysis discovered three differences between the casual customers and the member customers of the Cyclistic bike sharing community.

- Days used:
    - Casual riders used the bikes more on Friday, Saturday, and Sunday with Saturday being the busiest.
    - Member riders used the bikes more mid-week on Tuesday, Wednesday, and Thursday with Wednesday being the busiest.
- Locations used:
    - Casual riders used the bikes more going to and along the waterfront parks of Lake Michigan.
    - Member riders used the bikes in a mix of commuting around the city and at the waterfront parks.
- Time spent on the bikes:
    - Casual riders used the bikes for longer trips on average 21 minutes.
    - Member riders used the bikes for shorter trips on average 12 minutes.

# 6. Act

While this analysis uncovered some key insights, further analysis would be beneficial.

- Such as track riders to discover usage patterns of individual customers. The question is of the casual customers which ones are "tourists" enjoying the city for a short period of time and then never use the service again. By eliminating the tourist this would uncover the casual customers that reside in the city and would be the prime targets of any ad campaign. The difficulty of this added level of information would need to be done in a way to adhere with the privacy license.
- With most of the Casual customers using the bike sharing on the weekends, the marketing department could consider a weekend only membership. Along the same line of thought, offer a membership for 9 months instead of the full year allowing customers to not feel like they are wasting their membership during the winter months in Chicago.
- Continue tracking of the busiest stations to see if customers usage changes over time. For example, how busy are the stations near the Universities to see if new classes of students change the bike usage.