

## Table of Contents

Introduction .....	2
Problem Statement .....	2
About Company .....	2
Some High-level Questions .....	2
Case Study Roadmap: ASK .....	3
Identifying Key Business Task .....	3
Key Stakeholders .....	3
Case Study Roadmap: PREPARE .....	3
Download data and store it appropriately .....	3
Identify how it's organized. ....	3
Determine the credibility the data. ....	4
Case Study Roadmap: PROCESS .....	4
Check the data for errors. ....	4
Tools used. ....	4
Transform the data. ....	4
Document Cleaning Process .....	5
Case Study Roadmap: ANALYSE .....	5
Aggregation & data organization .....	5
Identify trends and relationships .....	5
1. Number of Bike rides by different user-type of Cyclistic .....	5
2. Distribution of Bike-type across user-type .....	6
3. Usage Trends – Ride count over days of week .....	8
4. Usage Trends – Ride count based on hour of day .....	9
5. Usage Trends - Quarterly count of bike ride .....	10
6. Usage Trends – Average Trip Duration for ride in minute .....	12
7. Active Start and End Stations .....	12
Case Study Roadmap: Share .....	16
Case Study Roadmap: ACT .....	17

# Cyclistic Case Study

## Introduction

In this case study, I will present various insights drawn based on data analysis done for Cyclistic bike share company. These analysis and insights are meant to assist data-driven managerial decision for future growth of company by answering key business question.

This case study is a capstone project done towards completion of [Google Data Analytics Professional Certification](#) course.

In further sections, I will introduce the problem statement, analysis steps and bring forwards various insights.

## Problem Statement

### About Company

Cyclistic is a bike-share company launched in 2016 and based in Chicago. The company has established 692 docking stations and 5,824 geo-tracked and station-locked bicycles. The company sets itself apart by offering various types of bicycles and assistive options for people with disabilities. Approximately 8% of riders utilize the assistive option, and 30% use Cyclistic bikes regularly for everyday commutes to work. The demographic of Cyclistic users compose of two user groups: casual riders and annual subscription members. The director of marketing believes that the company's future success relies heavily on maximizing the number of annual subscribers.

### Some High-level 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 annual members?

**Project Purpose:** Support the decision-making process behind supporting or opposing a marketing strategy that attempts to convert casual riders into member riders.

# Case Study Roadmap: ASK

## Identifying Key Business Task

The director of marketing believes the company's future success depends on maximizing the number of annual memberships. As member of data analytics team, I am expected to:

- understand how casual riders and annual members use Cyclistic bikes differently.
- present compelling data insights and professional data visualizations which can help Cyclistic design a new marketing strategy to convert casual riders into annual members.

## Key Stakeholders

- Lily Moreno – director of marketing, responsible for the development of campaigns and initiatives to promote the bike-share program.
- Cyclistic marketing analytics team – responsible for collecting, analysing, and reporting data that help guide Cyclistic marketing strategy.
- Cyclistic executive team – will decide whether to approve the recommended marketing program.

# Case Study Roadmap: PREPARE

## Download data and store it appropriately.

For this case study, historical data is shared [here](#). (Disclaimer: The datasets have a different name because Cyclistic is a fictional company. The dataset is shared in the course. This is public dataset shared by [Divvy, bike sharing company based in Chicago](#).)

## Identify how it's organized.

The website has historical data, segregated month-wise. I have used data for past 11 months, Jan-2022 to Nov-2022 for analysis. Data for each month is available in Comma Separated values(.csv) file. These data are download and stored in local PC for further processing and analysis. 11 .csv files are saved in local directory on PC.

Each .csv file consist of 13 columns, as explained in Table-1. Each row is an entry for one ride.

*Table-1 Details of Columns*

Column Name	Description
<b>ride_id</b>	<ul style="list-style-type: none"><li>• Unique identifier for each ride.</li><li>• Type: 8-Byte alpha-numeric string.</li></ul>
<b>rideable_type</b>	<ul style="list-style-type: none"><li>• Type of bike used for ride.</li><li>• Type: String</li><li>• Values: <i>Classic_bike/Electric_bike/Docked_bike</i></li></ul>
<b>started_at</b>	<ul style="list-style-type: none"><li>• Date and time when ride started.</li><li>• Type: <i>dd-mm-yyyy hh:mm:ss</i></li></ul>

<b>ended_at</b>	<ul style="list-style-type: none"> <li>• Date and time when ride ended.</li> <li>• Type: <i>dd-mm-yyyy hh:mm:ss</i></li> </ul>
<b>start_station_name</b>	<ul style="list-style-type: none"> <li>• Name of place where ride started.</li> <li>• Type: String</li> </ul>
<b>start_station_id</b>	<ul style="list-style-type: none"> <li>• Numeric number to uniquely identify station</li> <li>• Type: <i>Alpha-numeric</i></li> </ul>
<b>end_station_name</b>	<ul style="list-style-type: none"> <li>• Name of place where ride ended.</li> <li>• Type: String</li> </ul>
<b>end_station_id</b>	<ul style="list-style-type: none"> <li>• Numeric number to uniquely identify station</li> <li>• Type: <i>Alpha-numeric</i></li> </ul>
<b>start_lat</b>	<ul style="list-style-type: none"> <li>• Latitude of place where ride started.</li> <li>• Type: Float</li> </ul>
<b>start_lng</b>	<ul style="list-style-type: none"> <li>• Longitude of place where ride started.</li> <li>• Type: Float</li> </ul>
<b>end_lat</b>	<ul style="list-style-type: none"> <li>• Latitude of place where ride ended.</li> <li>• Type: Float</li> </ul>
<b>end_lng</b>	<ul style="list-style-type: none"> <li>• Longitude of place where ride ended.</li> <li>• Type: Float</li> </ul>
<b>member_casual</b>	<ul style="list-style-type: none"> <li>• Type of membership.</li> <li>• Type: String</li> <li>• Values: <i>casual/member</i> (who has taken subscription).</li> </ul>

### Determine the credibility the data.

Data is made public from [Divvy](#) for public use. These data are from actual usage, hence we can assume that the data is credible.

## Case Study Roadmap: PROCESS

### Check the data for errors.

Data was properly formatted, and I did not find need to pre-process data. For Analysis of station, there were trip which had station name missing. It was handled during analysis of active start and end stations.

### Tools used.

I have used following tools:

- Data Processing: MS Excel (Pivot table), Power Query
- Visualization: MS Excel

### Transform the data.

It is required to merge data in single file to perform the analysis. As per [Microsoft specification](#), Excel has limitation of 1,048,576 rows. In current dataset, there are 5,485,911 rows, which exceeds excel limitation. Therefore, in order to process data in excel I have used

power query tool to import and perform data transformation such as extracting data, time information, trip duration etc. These are explained in Analysis section in details.

Months	Row_count
Jan-22	103770
Feb-22	115609
Mar-22	284042
Apr-22	371249
May-22	634858
Jun-22	769204
Jul-22	823488
Aug-22	785932
Sep-22	701339
Oct-22	558685
Nov-22	337735
<b>Total</b>	<b>5485911</b>

## Document Cleaning Process

Columns which are not required for analysis were removed, viz. ride\_id, latitude and longitude columns.

# Case Study Roadmap: ANALYSE

## Aggregation & data organization

After importing data in power query and removing unnecessary columns, I created few more columns need to create more column,

New Column added	Description
Started_date	Date when trip started.
Started_hour	Hour-of-day when trip started (in 24hr Format)
Trip_duration in minute	Trip duration in minutes
Started_months	Month in which ride was taken.
Started_Quarter	Year was divided into 4-Quarters (3 months each Quarter starting Jan)

Analysis is done using Pivot table in MS Excel.

## Identify trends and relationships.

### 1. Number of Bike rides by different user-type of Cyclistic

- There are two types of users on Cyclistic, viz. **Casual user** and **Member user**.

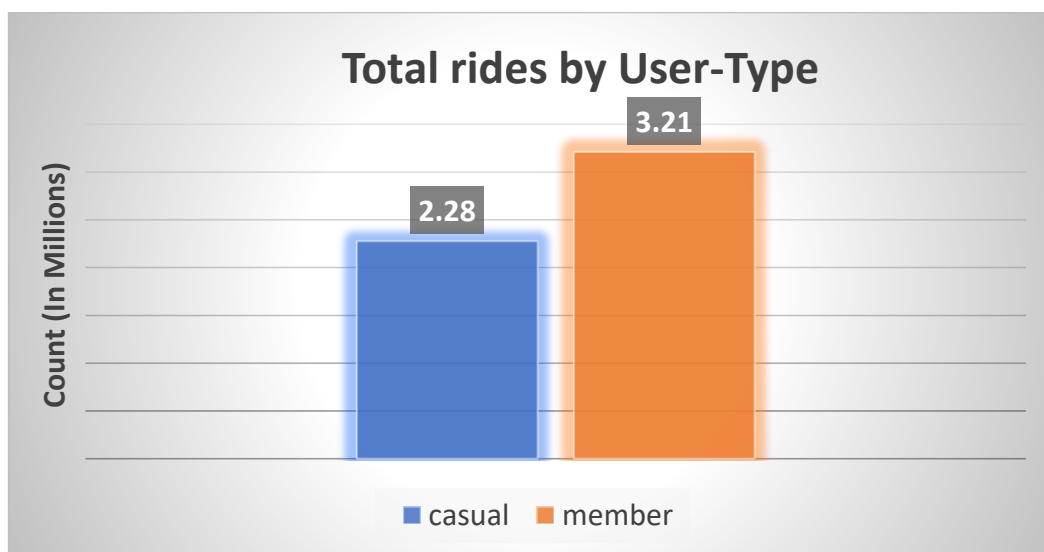
- To start with analysis, we will check how many rides are taken by casual and member user over 11 months.

Summary of analysis:

- Number of rides taken by **casual users** are approximately **2.27 million**.
- Number of rides taken by **member users** are approximately **3.20 million**.
- There are 41.5% Casual users and 58.5% Member users using Cyclistic bikes.

Table 2 Count of rides taken by casual and member user

	Column Labels		Grand Total
	casual	member	
Count of member_casual	2277138	3208773	5485911



## 2. Distribution of Bike-type across user-type

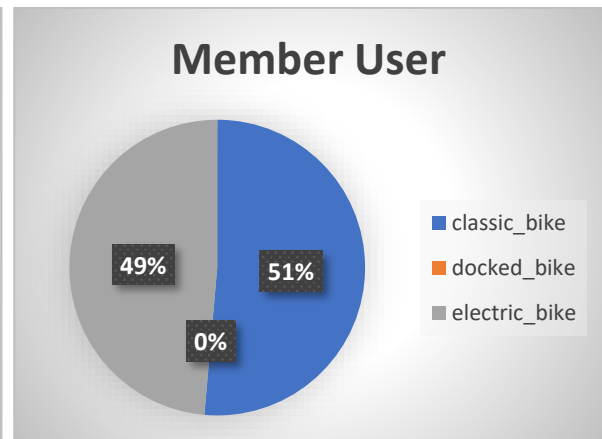
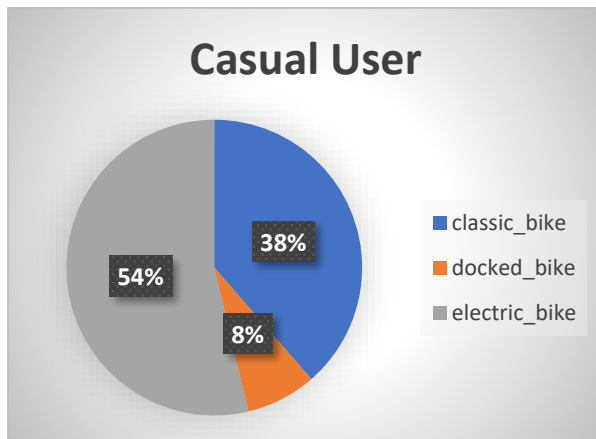
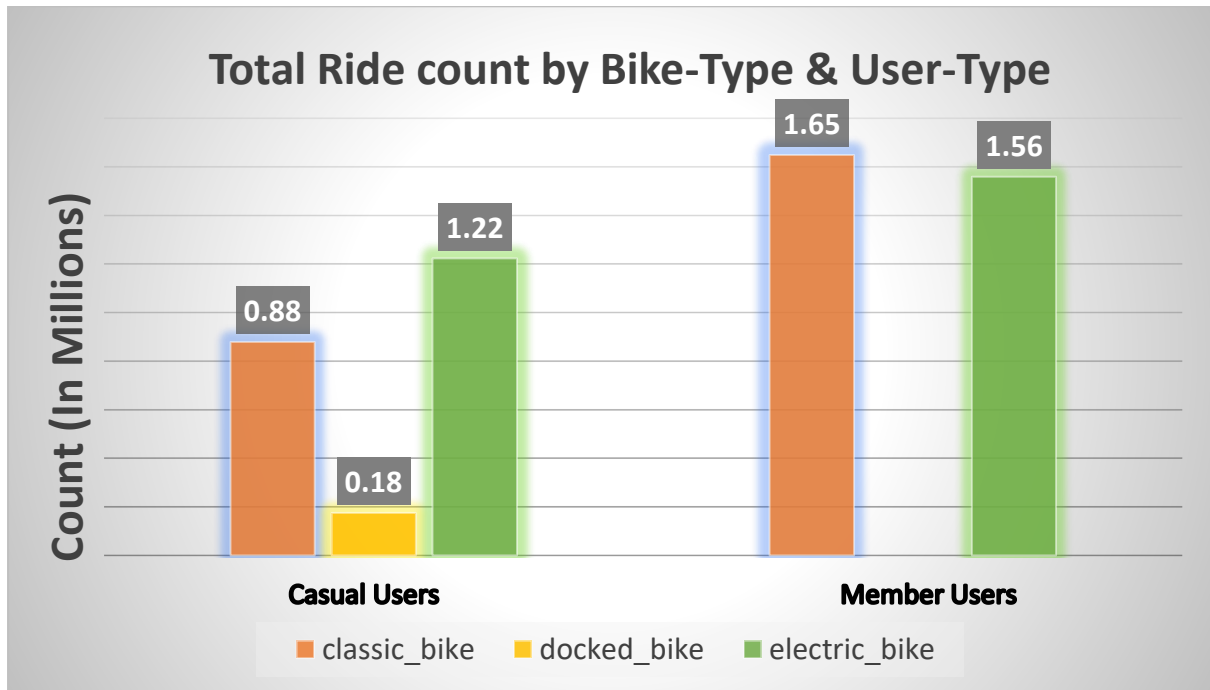
- Since Cyclistic provides three type of Bikes, viz. **Classic, Docked and Electric**. We will check which type of bikes are used by casual and member users.

Summary of analysis:

- Docked bikes are used only by casual users.
- Usage of Electric Bike is highest at 50.72%. Classic bike and docked bike usage is 46.07% and 3.21% respectively.
- Casual users prefer to use electric bike more. This could be attributed to fact that casual users use bike occasionally for fun ride.
- Member users usage of classic bike and electric bike is 51% and 49% respectively.

Count of member_casual	Column Labels			Grand Total
	classic_bike	docked_bike	electric_bike	
Row Labels				

casual	878807	175549	1222782	2277138
member	1649057		1559716	3208773
<b>Grand Total</b>	<b>2527864</b>	<b>175549</b>	<b>2782498</b>	<b>5485911</b>



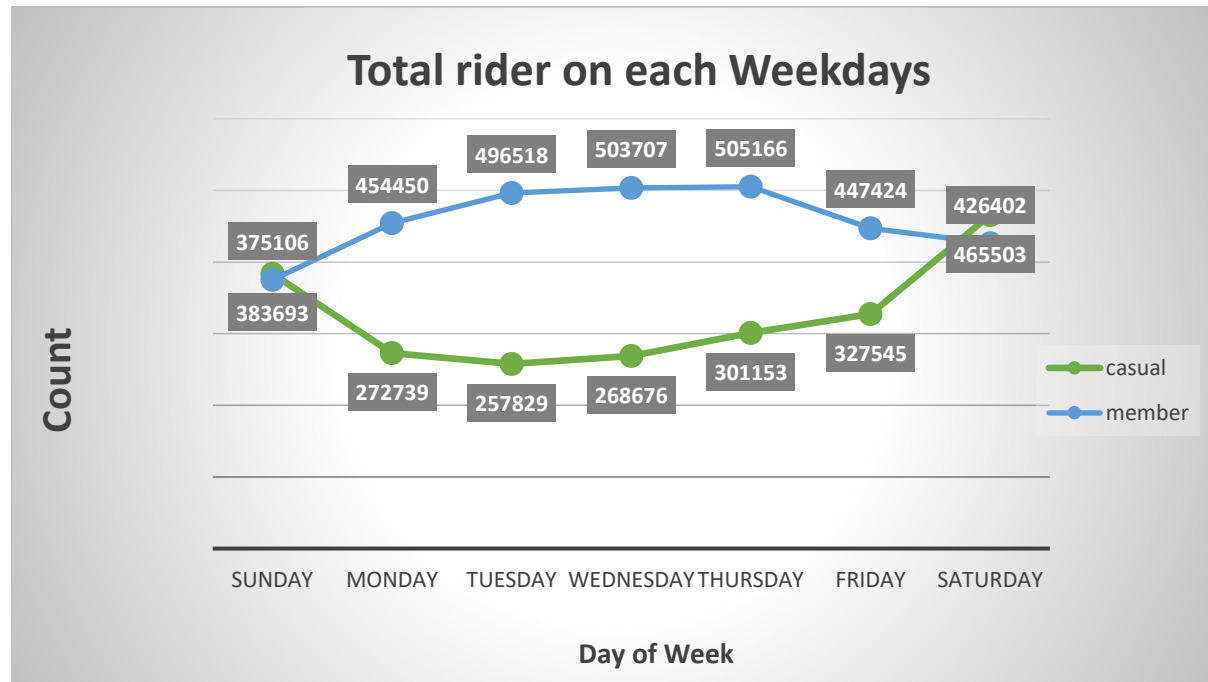
### 3. Usage Trends – Ride count over days of week

- We will check the usage of bike on 7 days of week.

Summary of analysis:

- Casual users usage bike more on weekends i.e. Saturday and Sunday.
- Member users are active on five working days. This could happen because members would be using bike to commute to office/work place.

Count of member_casual	Column Labels		
Row Labels	casual	member	Grand Total
Sunday	383693	375106	758799
Monday	272739	454450	727189
Tuesday	257829	496518	754347
Wednesday	268676	503707	772383
Thursday	301153	505166	806319
Friday	327545	447424	774969
Saturday	465503	426402	891905
Grand Total	2277138	3208773	5485911





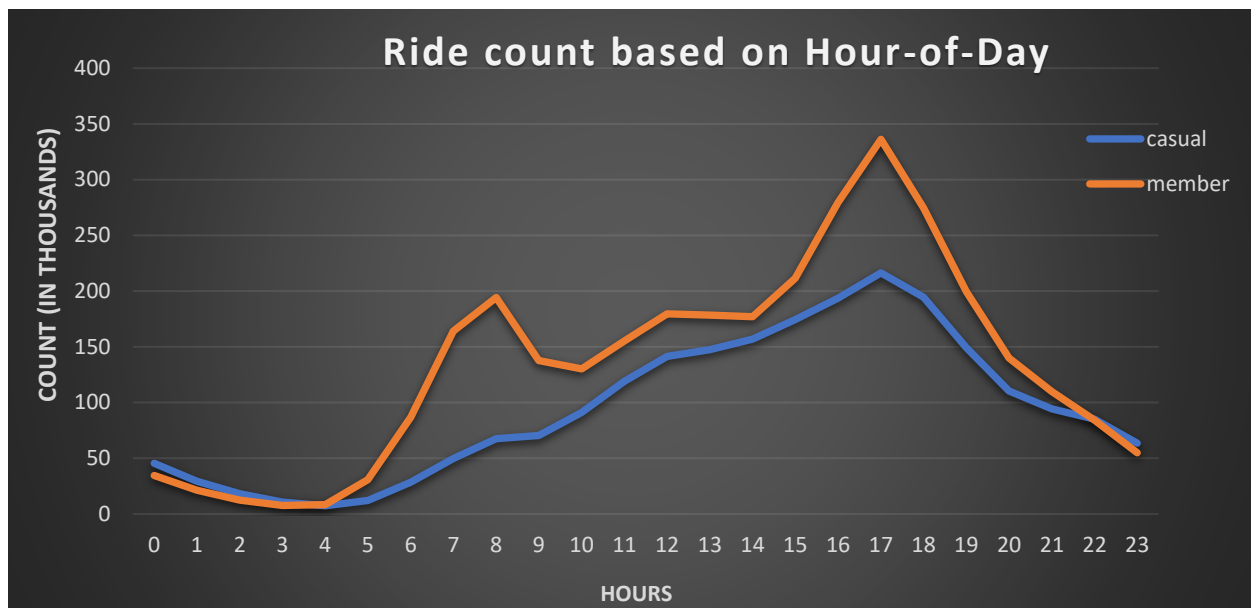
#### 4. Usage Trends – Ride count based on hour of day

- We will investigate how many rides are taking place across hour-of-day.

Row Labels	casual	member	Grand Total
0	45567	34573	80140
1	29498	21197	50695
2	18274	12307	30581
3	10826	7681	18507
4	7378	8486	15864
5	12040	30851	42891
6	28484	86864	115348
7	49889	164208	214097
8	67576	194374	261950
9	70491	137575	208066
10	91189	130113	221302
11	118906	155303	274209
12	141384	179504	320888
13	147518	178355	325873
14	156940	177049	333989
15	174389	211577	385966
16	193787	279608	473395
17	216365	336146	552511
18	194514	274704	469218
19	149017	199338	348355
20	110280	140044	250324
21	94324	109879	204203
22	85074	84114	169188
23	63428	54923	118351
<b>Grand Total</b>	<b>2277138</b>	<b>3208773</b>	<b>5485911</b>

#### Summary of Analysis:

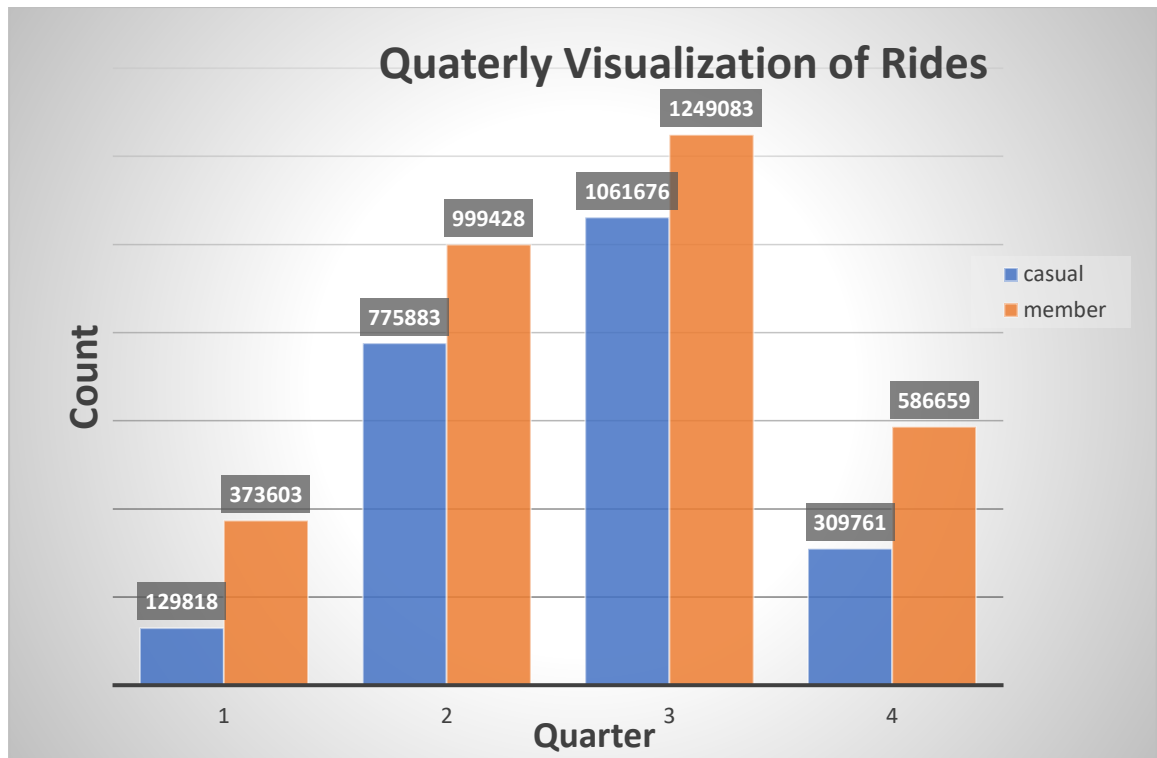
- Users are active across each hour of day including late night and early morning.
- Member users rides more bike during morning (8:00AM) and evening (4:00-6:00PM) time.
- Member users take more rides, however, Casual users ride more bike during late hours.



## 5. Usage Trends - Quarterly count of bike ride

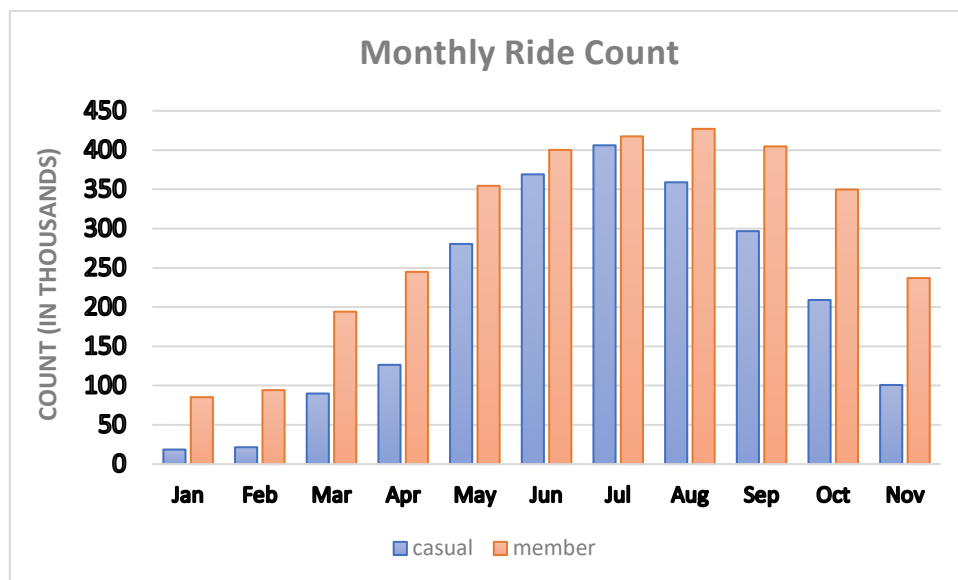
- From previous two analysis, we concluded that Member users uses bike more. Now we will analysis usage trend as how users are using bike. For this, I have analysed the rides taken in each quarter. Dividing the time frame in quarter will help get an insight of usage trend in months.
- Q1 is first quarter showing rides taken in month of Jan, Feb and March.
- Q2 is second quarter showing rides taken in month of Apr, May and June.
- Q3 is third quarter showing rides taken in month of July, Aug. and Sept.
- Q4 is fourth quarter showing rides taken in month of Oct, and Nov. Since at time of analysis, December data was not available.

Count of member_casual		Column Labels	
Row Labels	casual	member	Grand Total
Q1	129818	373603	503421
Q2	775883	999428	1775311
Q3	1061676	1249083	2310759
Q4	309761	586659	896420
Grand Total	2277138	3208773	5485911



#### Summary of Analysis:

- First quarter is least active duration.
- Third Quarter see a greater number of rides.
- Based on monthly analysis, we can conclude that rides are decreasing in 4<sup>th</sup> quarter.



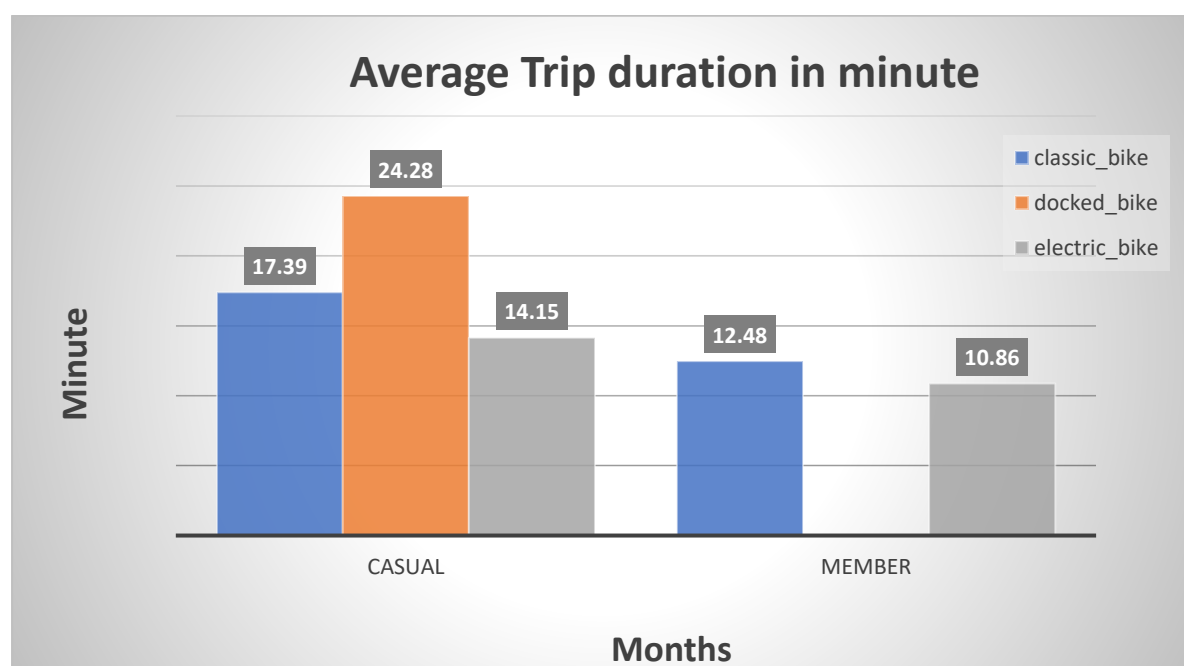
## 6. Usage Trends – Average Trip Duration for ride in minute

- To understand how casual and member users are using bike differently, we need to check the duration of rides. We will take average trip duration in minute for both users for casual and member.

Summary of Analysis:

- Average trip duration for casual users are higher than Member users.

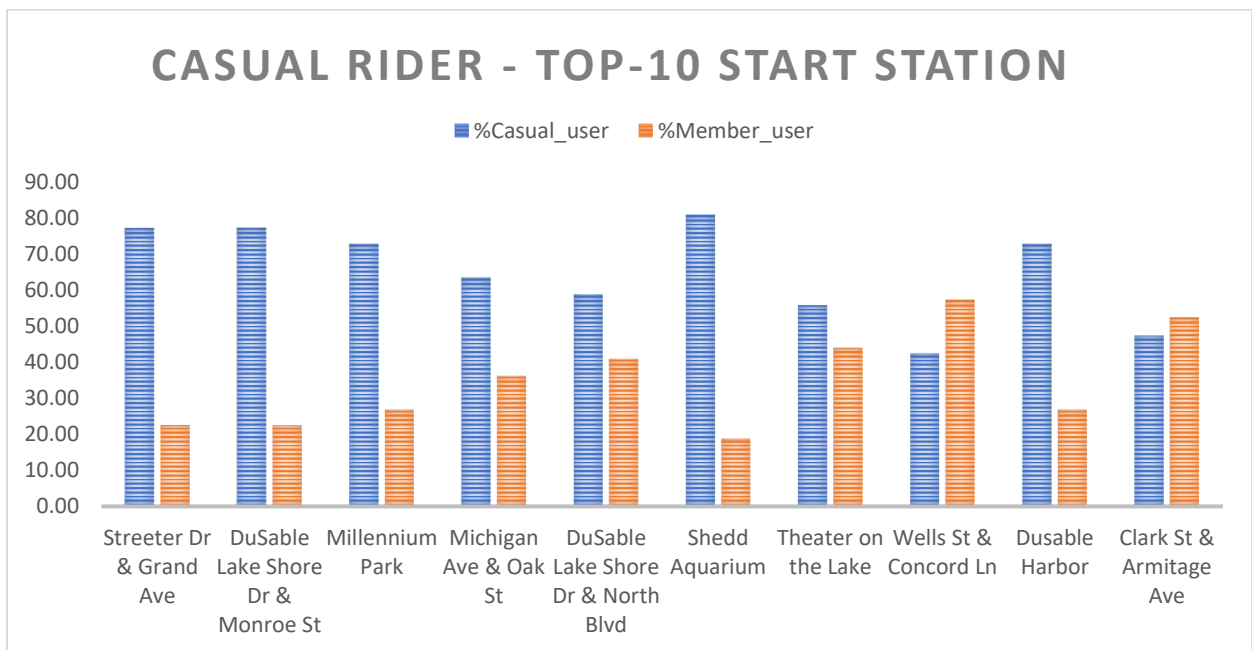
Average of Trip_duration_minute	Column Labels			
Row Labels	classic_bike	docked_bike	electric_bike	Grand Total
casual	17.39	24.28	14.15	16.18
member	12.48		10.86	11.69
Grand Total	14.19	24.28	12.31	13.56



## 7. Active Start and End Stations

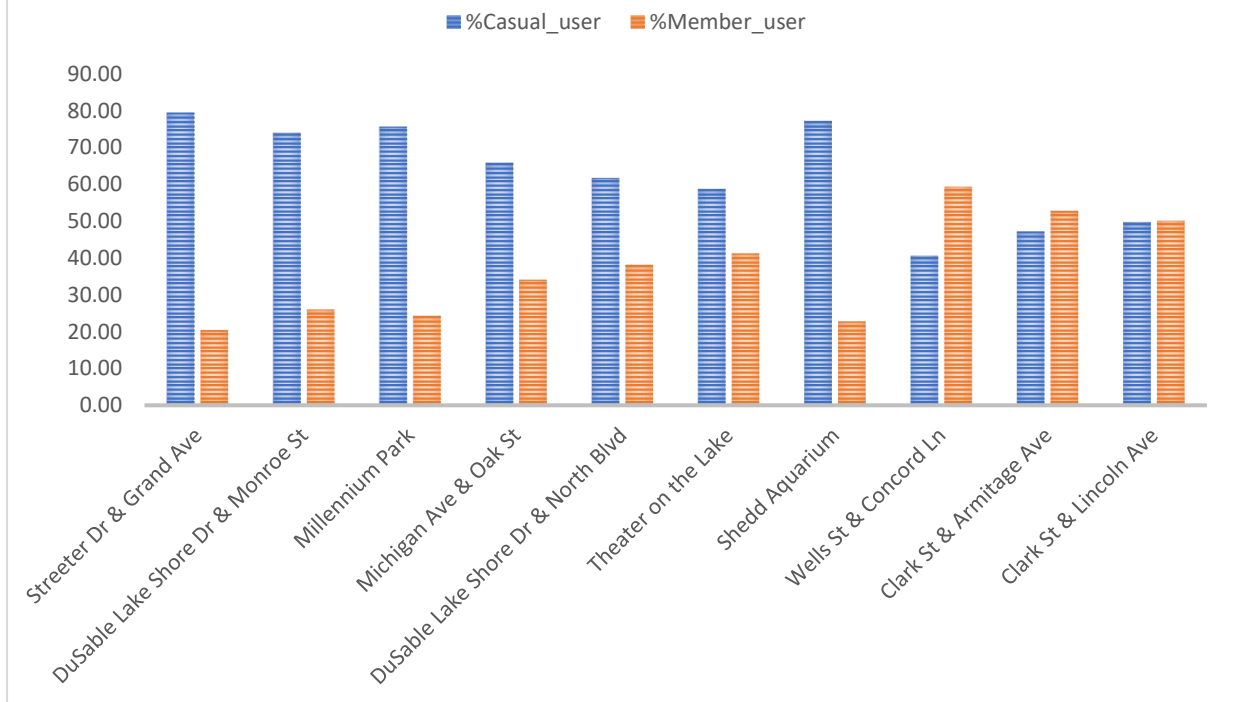
- There are 1644 stations from where ride had started and 1661 stations where ride had ended.
- In order to convert casual users to member, let us investigate stations where there are casual users are more than member users and vice-versa.
- List of Top-10 start stations where casual riders are more than member riders are:

Casual Top 10 Start_station ride	Casual	Member	Total	%Casual_user	%Member_user
Streeter Dr & Grand Ave	48072	14083	62155	77.34	22.66
DuSable Lake Shore Dr & Monroe St	26425	7697	34122	77.44	22.56
Millennium Park	21276	7869	29145	73.00	27.00
Michigan Ave & Oak St	20638	11794	32432	63.63	36.37
DuSable Lake Shore Dr & North Blvd	18797	13111	31908	58.91	41.09
Shedd Aquarium	16960	3957	20917	81.08	18.92
Theater on the Lake	15004	11821	26825	55.93	44.07
Wells St & Concord Ln	13134	17752	30886	42.52	57.48
Dusable Harbor	11507	4261	15768	72.98	27.02
Clark St & Armitage Ave	11205	12405	23610	47.46	52.54



Casual Top 10 End_station rides	Casual	Member	Total	%Casual_user	%Member_user
Streeter Dr & Grand Ave	49516	12732	62248	79.55	20.45
DuSable Lake Shore Dr & Monroe St	24477	8595	33072	74.01	25.99
Millennium Park	22156	7124	29280	75.67	24.33
Michigan Ave & Oak St	21669	11204	32873	65.92	34.08
DuSable Lake Shore Dr & North Blvd	20558	12711	33269	61.79	38.21
Theater on the Lake	15756	11045	26801	58.79	41.21
Shedd Aquarium	15567	4593	20160	77.22	22.78
Wells St & Concord Ln	12511	18228	30739	40.70	59.30
Clark St & Armitage Ave	11157	12478	23635	47.21	52.79
Clark St & Lincoln Ave	11021	11083	22104	49.86	50.14

## CASUAL RIDER - TOP-10 END STATION

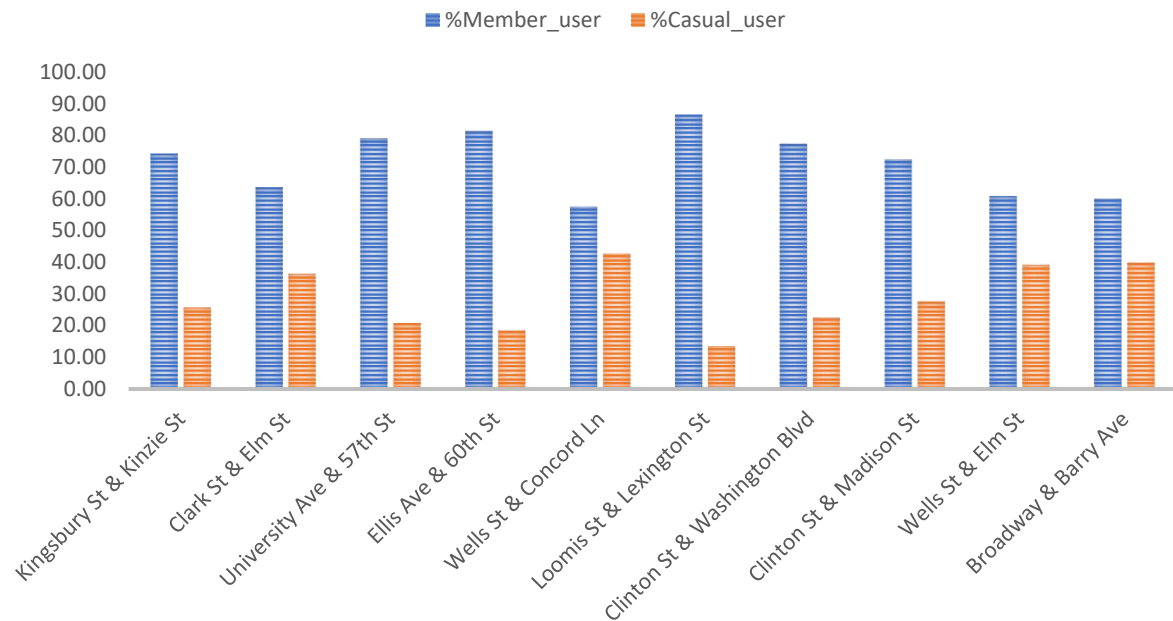


We notice that stations have higher number of casual users and less number of member users. These stations can be targeted for ads and offers which can motivate casual users to take membership. Notice that except Dusable Harbor (In start station) and Clark St & Lincoln Ave(in end station), stations are same in list. Thus we can target these 11 stations for such ad.

We can target another set of ads where members users prefer to take the ride more. High volume of member user indicates, these stations have potential to attract more member users. Here we can target ads/offers where members users can use referral code to add users similar to their behaviour/job nature.

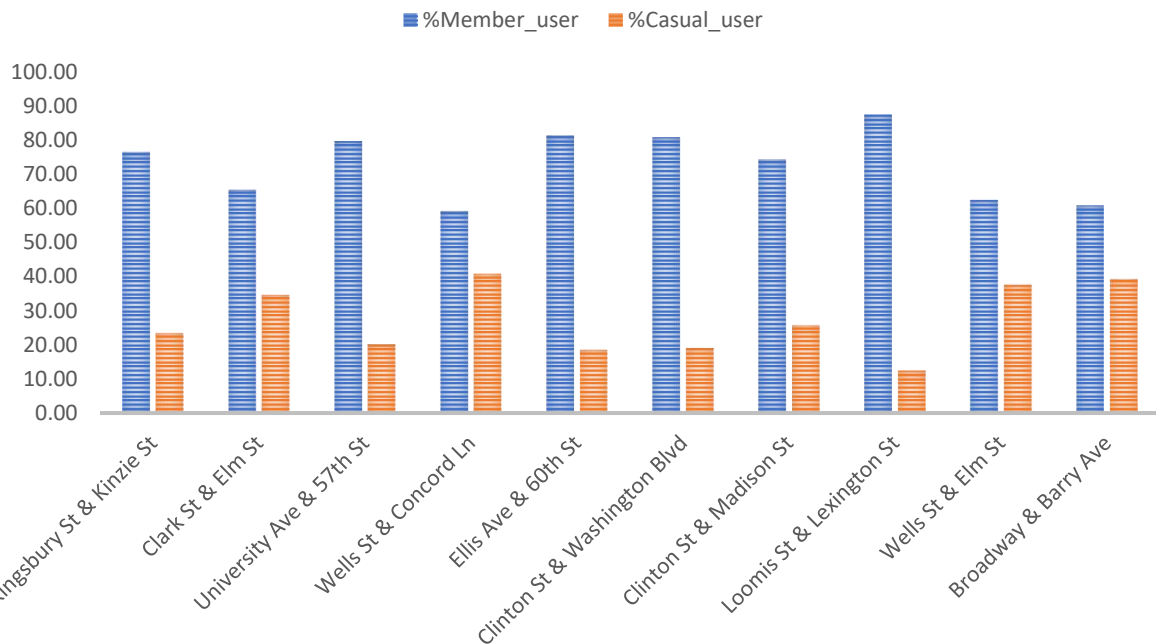
Member Top 10 Start_station ride	Member	Casual	Total	%Member_user	%Casual_user
Kingsbury St & Kinzie St	20818	7199	28017	74.30	25.70
Clark St & Elm St	18445	10507	28952	63.71	36.29
University Ave & 57th St	17873	4705	22578	79.16	20.84
Ellis Ave & 60th St	17854	4074	21928	81.42	18.58
Wells St & Concord Ln	17752	13134	30886	57.48	42.52
Loomis St & Lexington St	16708	2593	19301	86.57	13.43
Clinton St & Washington Blvd	16407	4778	21185	77.45	22.55
Clinton St & Madison St	15941	6092	22033	72.35	27.65
Wells St & Elm St	15584	10011	25595	60.89	39.11
Broadway & Barry Ave	14803	9820	24623	60.12	39.88

## MEMBER RIDER - TOP-10 START STATION



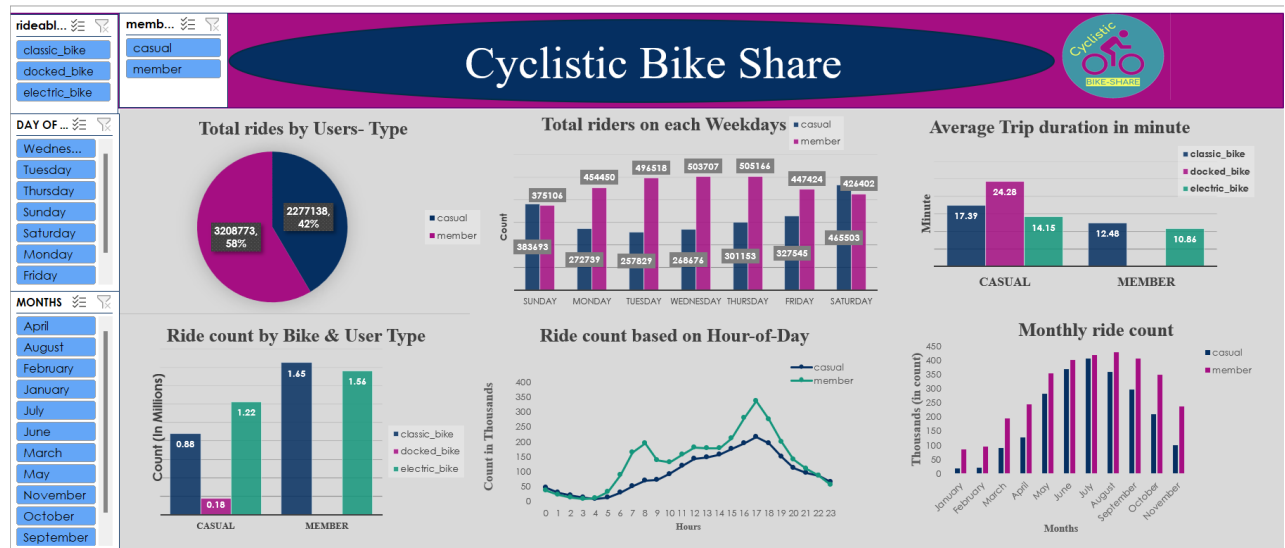
MemberTop 10 End_station ride	Member	Casual	Total	%Member_user	%Casual_user
Kingsbury St & Kinzie St	20538	6298	26836	76.53	23.47
Clark St & Elm St	18686	9848	28534	65.49	34.51
University Ave & 57th St	18354	4645	22999	79.80	20.20
Wells St & Concord Ln	18228	12511	30739	59.30	40.70
Ellis Ave & 60th St	17579	3993	21572	81.49	18.51
Clinton St & Washington Blvd	17047	4036	21083	80.86	19.14
Clinton St & Madison St	16582	5720	22302	74.35	25.65
Loomis St & Lexington St	16573	2359	18932	87.54	12.46
Wells St & Elm St	15437	9251	24688	62.53	37.47
Broadway & Barry Ave	15141	9735	24876	60.87	39.13

## MEMBER RIDER - TOP-10 END STATION



## Case Study Roadmap: Share

Following is the Dashboard which presents the summary of all analysis result we have seen above.



Interactive version of dashboard is available on [Github](#).



# Case Study Roadmap: ACT

After analysing, we reached to the following conclusion:

- Casual riders take less number of rides but for longer durations.
- Casual Riders are most active on weekends, and the second and third quarter.
- Casual riders mostly use bikes for recreational purposes.

Here are my top recommendations based on above key findings:

1. Design riding packages by keeping recreational activities, weekend contests, and summer events in mind and offer special discounts and coupons on such events to encourage casual riders get annual membership.
2. Design seasonal packages, It allows flexibility and encourages casual riders to get membership for specific periods if they are not willing to pay for annual subscription.