Intro

This is a capstone analysis case study of a fictional company, Cyclistic, for the Google Data Analytics Professional Certification.

It's March of 2023 and the director of marketing believes the company's future success depends on maximizing the number of annual memberships. To do this, the goal for the marketing analyst team is to design new marketing strategies to convert casual riders into members.

Ask

The marketing director has assigned me with the following analysis question to answer to help my team fulfill the objective mentioned above:
> **How do annual members and casual riders use Cyclistic bikes differently?**

Prepare

The dataset was obtained from a link of the company's historical public data. The data has been made available by Motivate International Inc. under this [license](#). Data-privacy issues have prohibited riders' personally identifiable information from being included.

The dataset contained an index of several different zip files that ranged from 2013 to 2023. Some files were labeled by quarters of the year and the other half were titled with the month and year. To identify the trends of my business task I will analyze the last 12 months of Cyclistic bike trip data. Thus, I downloaded 12 files from March of 2022 to February of 2023 and stored them in a folder on a desktop.

Process

The zip files were too large to be used with spreadsheets so I used BigQuery SQL to process the data. The smallest file contained 190,301 rows and the largest file contained 823,488 rows.

The following are the steps I took to see if the data was clean and extract some data to analyze further in spreadsheets.

1. I checked to see if there were any misspellings in the member_casual and rideable_type columns. I also made sure there were not any duplicate trips recorded.

```
SELECT DISTINCT member_casual
FROM `bike-case-study41223.bike_data.bikes_322`

SELECT DISTINCT ride_id
FROM `bike-case-study41223.bike_data.bikes_322`
```

2. I made sure there were no null values in the member_casual, rideable_type, ride_id, started_at, and ended_at columns.

```
SELECT member_casual
FROM `bike-case-study41223.bike_data.bikes_322`
WHERE member_casual is null
```

After doing some cleaning queries, I thought of some questions I wanted to answer to help me determine what kind of information I needed to analyze to complete my business task.

3. How many trips did members & casual riders make?

```
SELECT member_casual,
 COUNT(member_casual) AS num_of_trips,
FROM `bike-case-study41223.bike_data.bikes_322`
GROUP BY member_casual
```
*Extracted results into March 2022 spreadsheet

4. Which bike type is most popular with each rider?

```
SELECT member_casual, rideable_type,
 COUNT(rideable_type) AS num_of_bikes_used,
FROM `bike-case-study41223.bike_data.bikes_322`
GROUP BY member_casual, rideable_type
ORDER BY member_casual
```
*Extracted results into a new tab of the March 2022 spreadsheet

5. Where do riders get their bikes from the most? Top 10 most visited locations?
There were visible null values in the start_station_name column but there was still sufficient information recorded in the other rows to be used for analysis.

```
SELECT member_casual, start_station_name,
COUNT(start_station_name) AS trips_from_station
FROM `bike-case-study41223.bike_data.bikes_722`
WHERE start_station_name is not null AND member_casual = "casual"
GROUP BY start_station_name, member_casual
ORDER BY COUNT(start_station_name) DESC
LIMIT 10
```
*Extracted results into a new tab of the March 2022 spreadsheet

6. How long do the riders keep the bike?

```
SELECT member_casual,
SUM(travel_mins) AS total_travel_mins,
AVG(travel_mins) AS avg_travel_mins
FROM `bike-case-study41223.bike_data.With_travel_min_322`
GROUP BY member_casual
```
        *Extracted results into a new tab of the March 2022 spreadsheet

7. How many rides were taken on each day of the week?

```
SELECT rideable_type, started_at, ended_at, member_casual,
EXTRACT(DAYOFWEEK from started_at) AS day_of_week_num
FROM `bike-case-study41223.bike_data.bikes_322`
```
        *Made a new table of the results to work with for the next step

```
SELECT *,
CASE WHEN day_of_week_num=1 THEN 'Sunday'
WHEN day_of_week_num=2 THEN 'Monday'
WHEN day_of_week_num=3 THEN 'Tuesday'
WHEN day_of_week_num=4 THEN 'Wednesday'
WHEN day_of_week_num=5 THEN 'Thursday'
WHEN day_of_week_num=6 THEN 'Friday'
WHEN day_of_week_num=7 THEN 'Saturday'
END AS day_of_the_week
FROM `bike-case-study41223.bike_data.day_of_week_322`
```
        *Made a new table of the results to work with for the next step

```
SELECT member_casual, day_of_the_week,
COUNT(day_of_the_week) AS bike_trips
FROM `bike-case-study41223.bike_data.Day_of_week_322`
GROUP BY member_casual, day_of_the_week
ORDER BY member_casual
```
        *Extracted results into a new tab of March 2022 spreadsheet

Once I extracted all of those data subsets from SQL, I repeated the same steps for the other 11 months. Then, I went through and formatted each spreadsheet; I added a month column, bold and capitalized titles, capitalized values, changed title names, formatted the numbers properly, and spell checked each sheet.

Analyze

I aggregated all of the information from the 12 spreadsheets into a year-review spreadsheet, formatted in the same way as the individual months spreadsheets, having a separate tab for each of the questions above.

- **Trips tab:**
  - sorted sheet by rider
  - manually arranged the rows from march to february
  - calculated the number of trips taken by members and casual riders for the year

=SUMIF(B:B,"Casual",C:C)
=SUMIF(B:B,"Member",C:C)

E2	▼	fx =SUMIF(B:B,"Casual",C:C)				
	A	B	C	D	E	F
1	Month	Rider	Trips		Casual Yearly	Mbr Yearly
2	March	Casual	89,882		2,365,120	3,046,531
3	April	Casual	126,417			
4	May	Casual	280,415			
5	June	Casual	369,051			
6	July	Casual	406,055			
7	August	Casual	358,924			
8	September	Casual	296,697			
9	October	Casual	208,989			
10	November	Casual	100,772			
11	December	Casual	44,894			
12	January	Casual	40,008			
13	February	Casual	43,016			
14	March	Member	194,160			
15	April	Member	244,832			
16	May	Member	354,443			

- **Minutes tab:**
  - sorted sheet by rider
  - sort range of average minutes traveled in descending order of each rider,
  - calculate total hours traveled
  - calculated the sum of the total hours traveled for each rider
  - calculated the average trip time of the year for each rider

=D2/60
=SUMIF(B2:B25,"Casual",E2:E25)
=SUMIF(B:B,"Member",E:E)
 =AVERAGEIF(B:B,"Casual",C:C)
=AVERAGEIF(B:B,"Member",C:C)

	A	B	C	D	E	F
1	**Month**	**Rider**	**Avg Minutes Traveled**	**Total Minutes Traveled**	**Hours Traveled**	**Casual Av**
2	March	Casual	32	2,888,301	48,138	
3	June	Casual	32	11,664,784	194,413	
4	May	Casual	30	8,519,112	141,985	
5	April	Casual	29	3,671,686	61,195	
6	August	Casual	29	10,344,269	172,404	
7	July	Casual	29	11,689,724	194,829	
8	September	Casual	27	8,157,802	135,963	
9	October	Casual	26	5,412,452	90,208	
10	February	Casual	23	976,677	16,278	

Average Travel Time
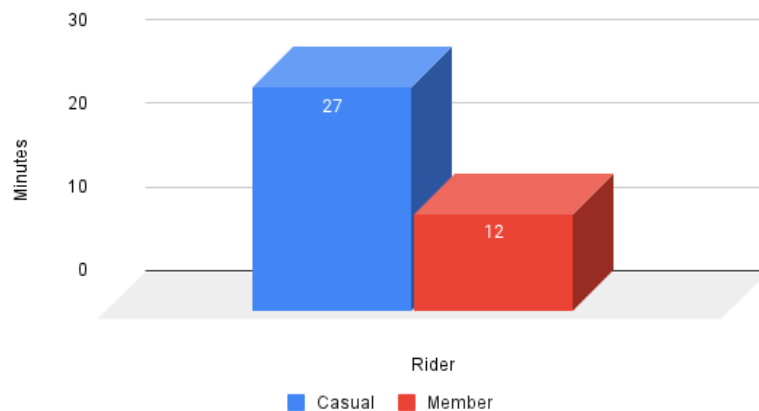
- **Bike type tab:**
  - sorted sheet by rider
  - calculated the sum of docked bikes used
  - calculated how many bikes of each type were used by each rider

```
=SUMIF(C:C,"Docked",D:D)
 =SUMIF(C2:C37,"Classic",D2:D37)
=SUMIF(C2:C37,"Electric",D2:D37)
=SUMIF(C38:C61,"Classic",D38:D61)
=SUMIF(C38:C61,"Electric",D38:D61)
```

G2 ▾ | *fx* =SUMIF(C2:C37,"Electric",D2:D37)

	A	B	C	D	E	F	G
1	Month	Rider	Bike	Trips	Cas Docked	Cas Classic	Cas Electric
2	July	Casual	Electric	218,905	179,085	905,821	1,280,214
3	August	Casual	Electric	203,966			
4	September	Casual	Electric	171,496			
5	June	Casual	Classic	169,996			
6	June	Casual	Electric	168,415			
7	July	Casual	Classic	156,095			
8	October	Casual	Electric	134,807			

- **Casual days of the week tab & Member days of the week tab:**
  - calculated the sum of trips taken on each day of the week for each rider

=SUMIF(C:C,"Sunday",D:D)

F2 ▾ | *fx* =SUMIF(C:C,"Monday",D:D)

	A	B	C	D	E	F	
1	Month	Rider	Day of the Week	Trips	Casual Sun	Casual Mon	
2	July	Casual	Saturday	95,230	398,647	283,327	
3	July	Casual	Sunday	78,251			
4	August	Casual	Saturday	66,190			
5	June	Casual	Sunday	65,851			
6	June	Casual	Saturday	65,137			
7	September	Casual	Saturday	64,077			
8	June	Casual	Thursday	57,978			

- **Casual locations tab & Member locations tab:**
  - sorted sheet by start station
  - count function of each station name
  - applied conditional formatting to the count function cells for greater than 7, then deleted all of the stations whose count was not highlighted
  - calculated sum of trips taken of each station left

=COUNTIFS(C:C,C2)
=SUM(D2:D11)

F2 ▾ | *fx* =SUM(D2:D11)

	A	B	C	D	E	F
1	Month	Rider	Start Station	Trips	Count	Sum of Station
2	July	Casual	Wells St & Concord Ln	3,002	10	16,012
3	June	Casual	Wells St & Concord Ln	2,851		
4	August	Casual	Wells St & Concord Ln	2,555		
5	May	Casual	Wells St & Concord Ln	2,037		
6	September	Casual	Wells St & Concord Ln	1,997		
7	October	Casual	Wells St & Concord Ln	1,486		
8	April	Casual	Wells St & Concord Ln	752		
9	November	Casual	Wells St & Concord Ln	566		
10	March	Casual	Wells St & Concord Ln	534		
11	January	Casual	Wells St & Concord Ln	232		
12	July	Casual	Theater on the Lake	4,240	8	18,035
13	June	Casual	Theater on the Lake	3,396		
14	August	Casual	Theater on the Lake	3,238		
15	May	Casual	Theater on the Lake	2,634		
16	September	Casual	Theater on the Lake	2,003		
17	October	Casual	Theater on the Lake	1,140		
18	April	Casual	Theater on the Lake	791		
19	March	Casual	Theater on the Lake	593		
20	July	Casual	Streeter Dr & Grand Ave	11,794	12	58,610
21	August	Casual	Streeter Dr & Grand Ave	9,651		
22	June	Casual	Streeter Dr & Grand Ave	9,608		
23	May	Casual	Streeter Dr & Grand Ave	7,743		
24	September	Casual	Streeter Dr & Grand Ave	7,016		

__Analysis Findings__
Casual Riders:
1) Used the bikes the most in the summer months (July, June, and August)
2) Bikes are used primarily on the weekends (Saturday, Sunday, and Friday)
3) Most often used the electric bikes more than the classic bikes & only 7.6% of their total trips were taken with the docked bikes
4) Casual riders kept the bike for a longer period of time than members
5) Visited these stations the most to get the bikes
   a) Streeter Dr & Grand Ave
   b) DuSable Lake Shore Dr & Monroe St
   c) Millennium Park
   d) Michigan Ave & Oak St
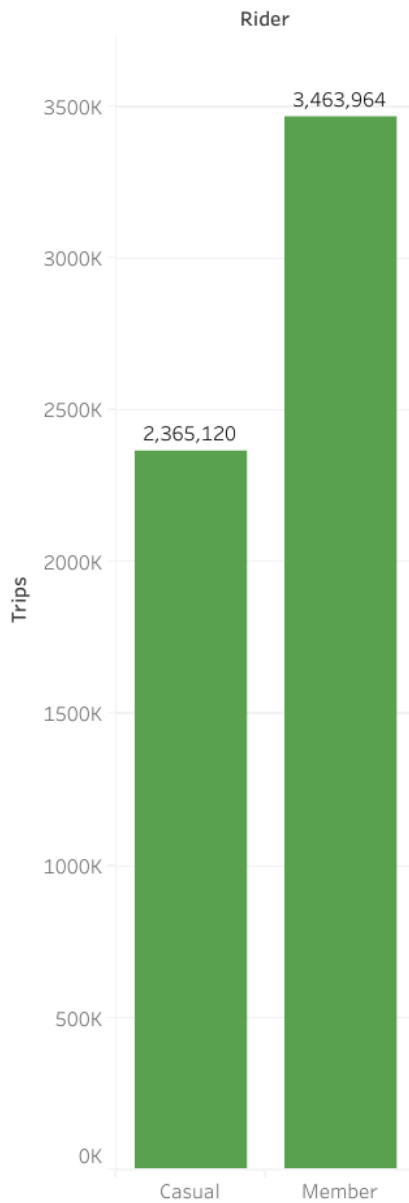   e) DuSable Lake Shore Dr & North Blvd

Members:
1) Took more trips than casual riders every month
2) Primarily at the end of summer leading into fall (August, July, and September)
3) Bikes are used primarily during the week (Thursday, Tuesday, and Wednesday)
4) Members most often used the classic bikes and they have not used the docked bikes in the last 12 months
5) Visit these stations the most to get the bikes
   a) Kingsbury St & Kinzie St

b) Clark St & Elm St
c) Wells St & Concord Ln
d) Ellis Ave & 60th St
e) University Ave & 57th St

Share
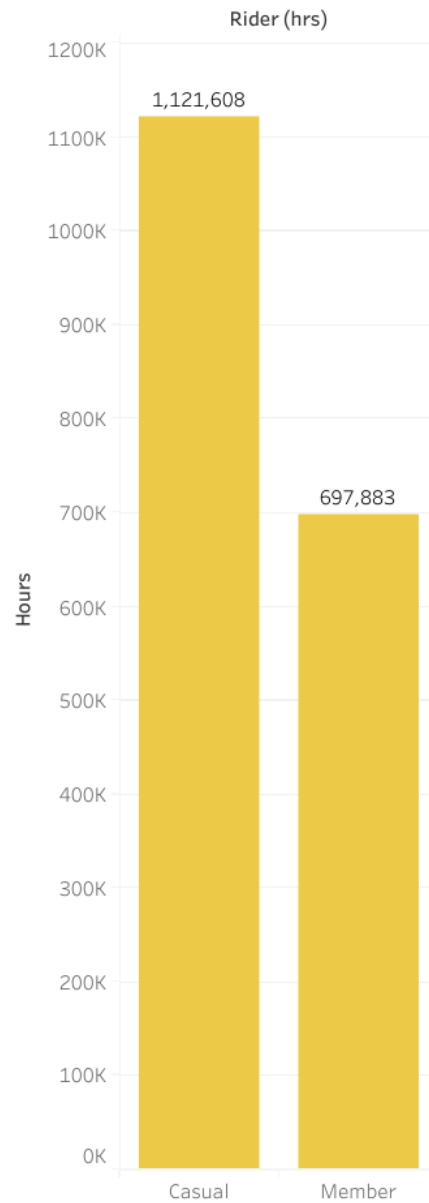Analysis visualizations were made using Tableau

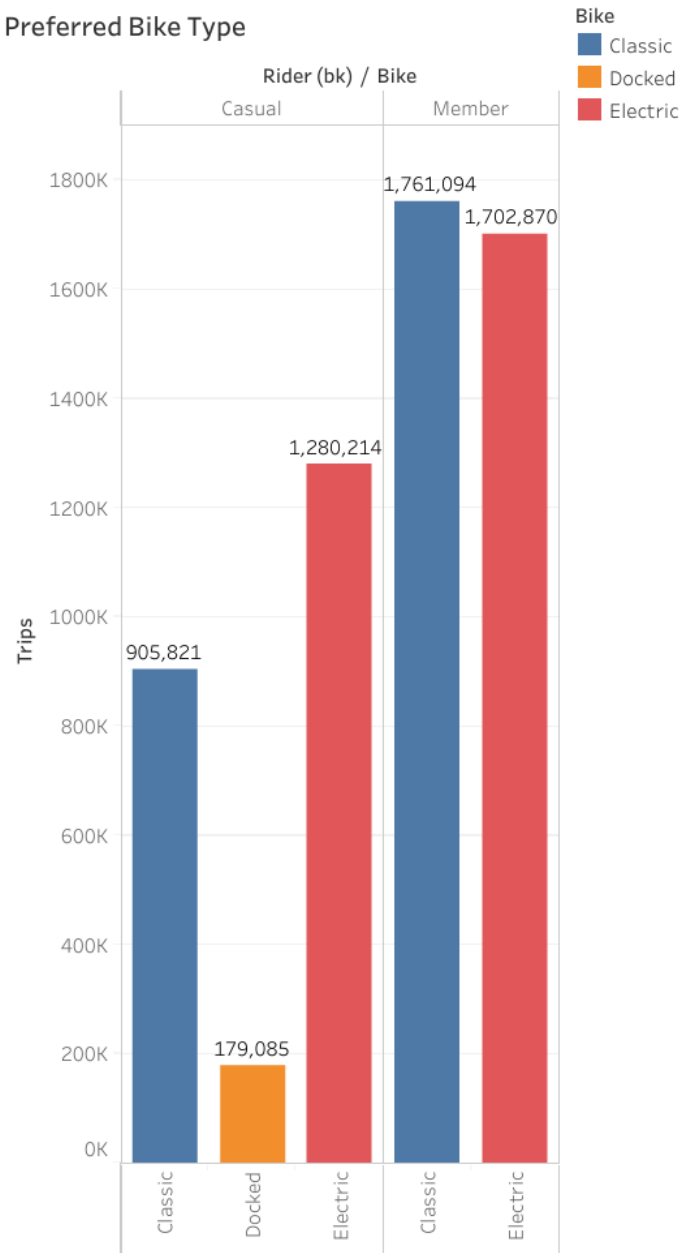### Trips Taken Over the Past Year
Rider

Total Hours Traveled
Rider (hrs)

Preferred Bike Type
Rider (bk) / Bike

Bike
- Classic
- Docked
- Electric

Trips Taken by Month

| | Month | | | | | | | | | | | |
Rider	January	February	March	April	May	June	July	August	Septemb..	October	November	December
Casual	40,008	43,016	89,882	126,417	280,415	369,051	406,055	358,924	296,697	208,989	100,772	44,894
Member	150,293	147,429	194,160	244,832	354,443	400,153	417,433	427,008	404,642	349,696	236,963	136,912

Casual Riders Trips by Weekday

Day of the Week

Day	Trips
Sunday	398,647
Monday	283,327
Tuesday	272,322
Wednesday	279,897
Thursday	313,642
Friday	338,806
Saturday	478,479

Members Trips by Weekday

Day of the Week (mbr)

Day	Trips
Sunday	402,869
Monday	488,630
Tuesday	546,620
Wednesday	542,433
Thursday	547,400
Friday	481,194
Saturday	454,818
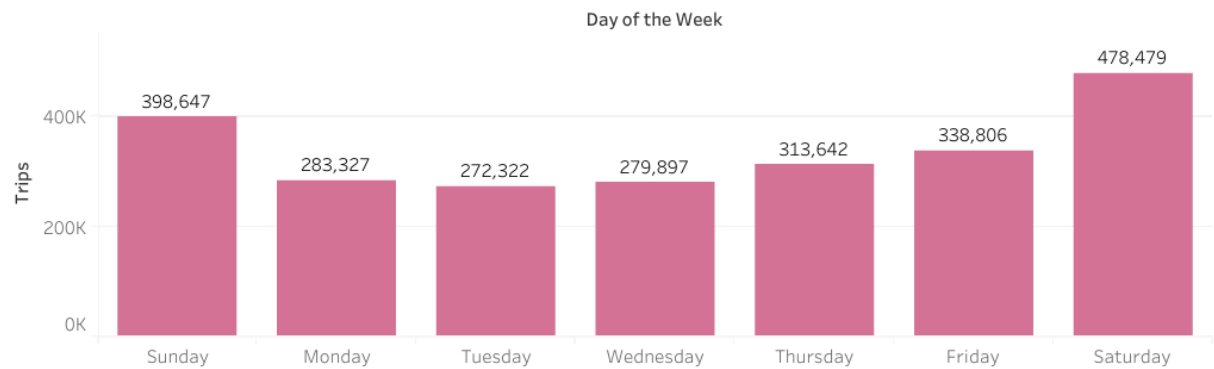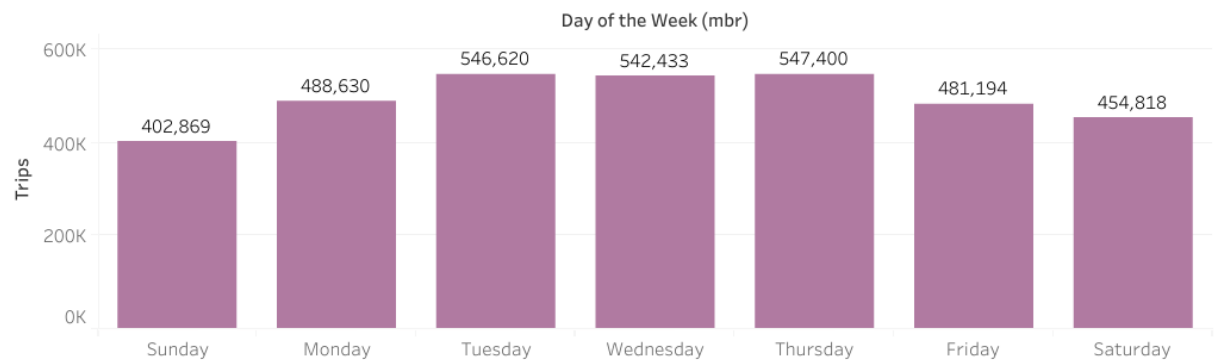
Top 8 Stations Where Casual Riders Obtained the Bikes

Start Station

Station	Trips
Streeter Dr & Grand Ave	58,610
DuSable Lake Shore Dr & Monroe St	31,969
Millennium Park	25,847
Michigan Ave & Oak St	25,056
DuSable Lake Shore Dr & North Blvd	23,159
Shedd Aquarium	20,681
Theater on the Lake	18,035
Wells St & Concord Ln	16,012

Trips (axis: 0K–65K)
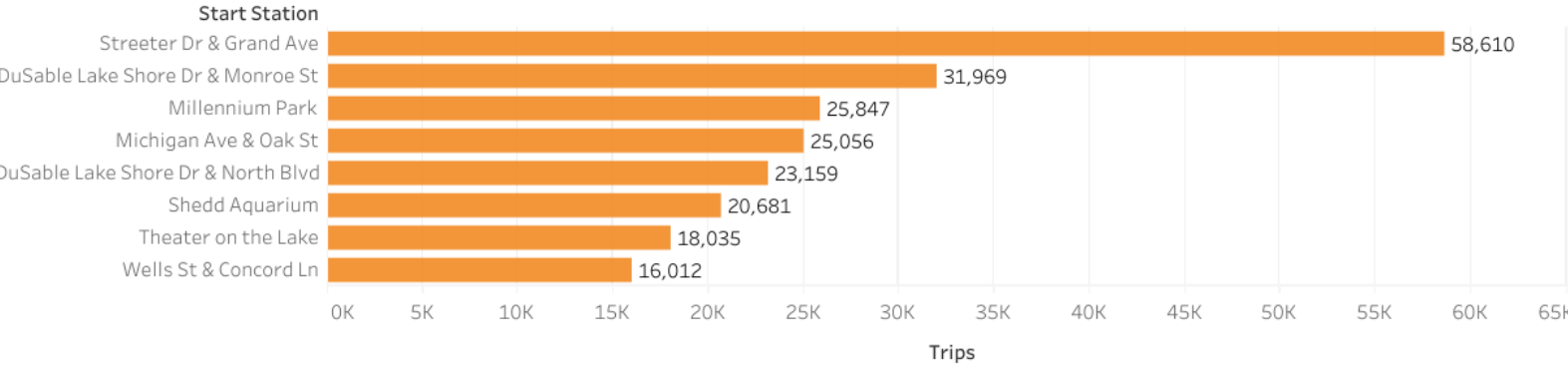
Top 8 Stations Where Members Obtained the Bikes

Start Station (mbr)

Station	Trips
Kingsbury St & Kinzie St	25,365
Clark St & Elm St	22,877
Wells St & Concord Ln	18,136
Ellis Ave & 60th St	15,865
University Ave & 57th St	15,628
Clinton St & Washington Blvd	12,821
Wells St & Elm St	12,735
Clinton St & Madison St	11,619

Trips (axis: 0K–28K)
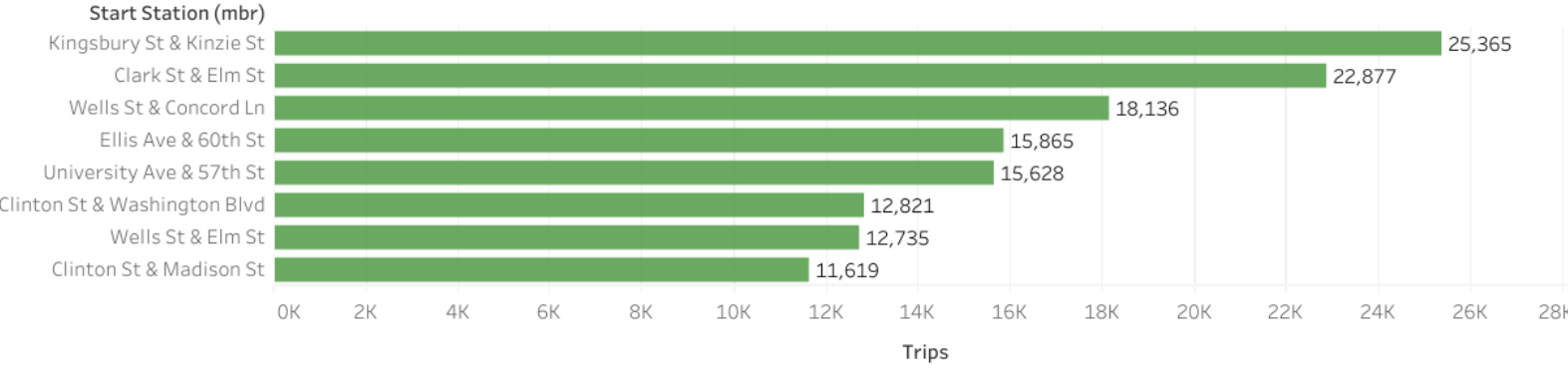
Act

My top three recommendations based on my analysis would be:

1. Create a summer promotion of discounted bikes for new members
2. Discount the docked and electric bikes for members to get more casual riders to become members (and for more members to start using these bikes more)
3. Advertise all new promotions aimed at converting casual riders into members heavily in the areas where casual riders visit certain bike stations the most