Cyclistic: This Year's Membership Overview (Sept 2023 - Sept 2024)

By Alyssa Ayala
Google Data Analytics Capstone Project

Agenda

Based on Google's **Data Analytics steps**

1. **Ask:** The statement of the business task
2. **Prepare:** The data sources I used
3. **Process:** My data cleaning and manipulation process
4. **Analyze:** A summary of my analysis
5. **Share:** My supporting visualizations and key findings
6. **Act:** My top three recommendations based on my analysis

Executive summary

- Cyclistic customers with **annual memberships** mainly do **short commutes** to school or work, while **casual riders** do occasional **long trips** for **recreational** purposes.
- From September 2023 to September 2024:
  - Most members did **short trips**, totaling their trip count to 4,105,894. The 1 to 2 hour trips were done by casual customers, totaling 71,433 and 14,382 trips, respectively.
  - Most members preferred bikes for their short trips, but **electric scooters** were more ideal for casual customers with 1-2 hour trips, totaling their trip count to 85,215.
  - The **weekends** were especially popular among **casual riders** totaling their trip count to 498,570 for **Saturday**, and 411,038 for **Sunday**.
  - The most popular docking stations like Streeter Dr and Grand Ave are dense and favored by **casual riders** but the least popular are spread out and may only have one trip count each.
- To convert casual riders into annual members, **Lily Moreno** should focus on Cyclistic's advertising efforts on electric scooters, offer membership benefits for weekend rides, and offer membership benefits for different docking stations.

Ask: The business task statement

Cyclistic stakeholders

- **Cyclistic** is a bike sharing service based in Chicago, Illinois, that was founded in 2016. They offer bikes and electric scooters, including those for the disabled that can't use traditional bikes or scooters.
- Current staff:
  - **Alyssa Ayala:** The presenter, junior data analyst for Cyclistic's marketing analytics team. I help them come up with the best marketing strategy according to their company's needs.
  - **Lily Moreno:** The director of marketing and my manager. She develops the campaigns that the marketing team uses to promote Cyclistic's services by email, social media, etc.

The problem we're trying to solve

- Cyclistic has three different types of passes: **single-ride, full-day,** and **annual memberships.**
  - Single-ride or full-day: **Casual riders**
  - Annual memberships: **Members**
- The financial analysts at Cyclistic have concluded that **annual memberships** are the most profitable passes for the company.
- **Lily Moreno** wants to figure out how the marketing team can convert more **casual riders** into **members**.
- The question this presentation is aiming to answer: **How do annual members and casual riders use Cyclistic bikes differently?**

Prepare: Data sources used

Sources of the datasets

- I got the previous 12 months of data from this website that has Cyclistic trip data for every month.
- They were all .csv files enclosed in ZIP files. I downloaded and extracted them on my computer, keeping the datasets in a folder separate from my other files.

All files downloaded:

- **Sept 2023:** 202309-divvy-tripdata.zip
- **Oct 2023:** 202310-divvy-tripdata.zip
- **Nov 2023:** 202311-divvy-tripdata.zip
- **Dec 2023:** 202312-divvy-tripdata.zip
- **Jan 2024:** 202401-divvy-tripdata.zip
- **Feb 2024:** 202402-divvy-tripdata.zip
- **Mar 2024:** 202403-divvy-tripdata.zip
- **Apr 2024:** 202404-divvy-tripdata.zip
- **May 2024:** 202405-divvy-tripdata.zip
- **Jun 2024:** 202406-divvy-tripdata.zip
- **Jul 2024:** 202407-divvy-tripdata.zip
- **Aug 2024:** 202408-divvy-tripdata.zip
- **Sept 2024:** 202409-divvy-tripdata.zip

The datasets' specs

- Each ride is identified by its **ride_id.**
- The dates and times of the trips are in **start_time** and **end_time,** where the dates are listed by month, day, and year. The times are listed by hour, minutes, and seconds.
- The docking stations are listed by their start and end station ids and names in the variables **start_station_name/id** and **end_station_name/id**, respectively.
- The docking stations also have physical locations by their latitude and longitude in the variables **start/end_lat** and **start/end_lng**, respectively.

Process: Data cleaning and manipulation

Using Microsoft Excel to make formulas

- I made a column called **ride_length** to get the length of each trip by subtracting **start_time** in column C from **end_time** in column D like this: `=D(num)-C(num)`
    - Since the responses were in decimals, I had to format column C into how much time passed (37:30:55).
- I made a column called **day_of_week** by using the WEEKDAY() function on column C to get the day of the week of the starting day and time like this: `WEEKDAY(C(num),1)`, where 1 = Sunday, 2 = Monday, etc.
- I did this process for each of the 12 datasets. At this point, nothing had been combined yet.

Importing the data into RStudio

- I used **RStudio** for the rest of the steps until the visualization process.
- First, I installed and loaded the **tidyverse** library using **packages.install()** and **library()**, respectively.
- I used tidyverse's **read.csv()** function to import all of the 12 datasets in this format: *month_year*. The months were abbreviated, and just the last two digits of the year were shown.

```
# Install and load tidyverse library
packages.install("tidyverse")
library(tidyverse)

# Import datasets
sept_23 <- read.csv("202309-divvy-tripdata.csv")
oct_23 <- read.csv("202310-divvy-tripdata.csv")
nov_23 <- read.csv("202311-divvy-tripdata.csv")
dec_23 <- read.csv("202312-divvy-tripdata.csv")
jan_24 <- read.csv("202401-divvy-tripdata.csv")
feb_24 <- read.csv("202402-divvy-tripdata.csv")
mar_24 <- read.csv("202403-divvy-tripdata.csv")
apr_24 <- read.csv("202404-divvy-tripdata.csv")
may_24 <- read.csv("202405-divvy-tripdata.csv")
jun_24 <- read.csv("202406-divvy-tripdata.csv")
jul_24 <- read.csv("202407-divvy-tripdata.csv")
aug_24 <- read.csv("202408-divvy-tripdata.csv")
sept_24 <- read.csv("202409-divvy-tripdata.csv")
```

Combining multiple datasets with RStudio

- I used tidyverse's **bind_rows()** function to combine each month into a new dataset called **trips**.
- Since all of the datasets had the same column names and format, everything was combined with no conflicts.

```
# Combine datasets into one
trips <- sept_23 %>%
  bind_rows(oct_23)

trips <- trips %>%
  bind_rows(nov_23)

trips <- trips %>%
  bind_rows(dec_23)

trips <- trips %>%
  bind_rows(jan_24)

trips <- trips %>%
  bind_rows(feb_24)

trips <- trips %>%
  bind_rows(mar_24)

trips <- trips %>%
  bind_rows(apr_24)

trips <- trips %>%
  bind_rows(may_24)

trips <- trips %>%
  bind_rows(jun_24)

trips <- trips %>%
  bind_rows(jul_24)

trips <- trips %>%
  bind_rows(aug_24)

trips <- trips %>%
  bind_rows(sept_24)
```

Data cleaning with RStudio

- Just to make sure, I used tidyverse's **drop_na()** function to delete any entries with null values to only keep verified trips.
- Since the days of the week were integers (1 = Sunday, etc.), I changed the values to their corresponding weekday depending on the number.
- I used tidyverse's **write_csv()** function to save the dataset into a file called **trips.csv.**

```
# Drop entries with null values, incomplete trip data
drop_na(trips)

# Now assign days of week based on numerical values
trips$day_of_week[trips$day_of_week == 1] <- 'Sunday'
trips$day_of_week[trips$day_of_week == 2] <- 'Monday'
trips$day_of_week[trips$day_of_week == 3] <- 'Tuesday'
trips$day_of_week[trips$day_of_week == 4] <- 'Wednesday'
trips$day_of_week[trips$day_of_week == 5] <- 'Thursday'
trips$day_of_week[trips$day_of_week == 6] <- 'Friday'
trips$day_of_week[trips$day_of_week == 7] <- 'Saturday'

# Finally, save the dataset
write_csv(trips, "trips.csv")
```

Analyze: Summary of my analysis

	ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng	end_lat	end_lng	member_casual	ride_length	day_of_week
1	011C1903BF4E2E28	classic_bike	9/23/2023 0:27	9/23/2023 0:33	Halsted St & Wrightwood Ave	TA1309000061	Sheffield Ave & Wellington Ave	TA1307000052	41.92914	-87.64908	41.93625	-87.65266	member	0:05:37	Saturday
2	87DB80E048A1BF9F	classic_bike	9/2/2023 9:26	9/2/2023 9:38	Clark St & Drummond Pl	TA1307000142	Racine Ave & Fullerton Ave	TA1306000026	41.93125	-87.64434	41.92557	-87.65842	member	0:11:36	Saturday
3	7C2EB7AF669066E3	electric_bike	9/25/2023 18:30	9/25/2023 18:41	Financial Pl & Ida B Wells Dr	SL-010	Racine Ave & 15th St	13304	41.87506	-87.63314	41.86127	-87.65663	member	0:11:28	Monday
4	57D197B010269CE3	classic_bike	9/13/2023 15:30	9/13/2023 15:39	Clark St & Drummond Pl	TA1307000142	Racine Ave & Belmont Ave	TA1308000019	41.93125	-87.64434	41.93974	-87.65887	member	0:08:29	Wednesday
5	8A2CEA7C8C8074D8	classic_bike	9/18/2023 15:58	9/18/2023 16:05	Halsted St & Wrightwood Ave	TA1309000061	Racine Ave & Fullerton Ave	TA1306000026	41.92914	-87.64908	41.92557	-87.65842	member	0:06:06	Monday
6	03F7044D1304CD58	electric_bike	9/15/2023 20:19	9/15/2023 20:30	Southport Ave & Wrightwood Ave	TA1307000113			41.92884	-87.66387	41.90000	-87.64000	member	0:11:02	Friday
7	672503E0FC0835EC	electric_bike	9/27/2023 16:52	9/27/2023 17:03	Kedzie Ave & Milwaukee Ave	13085			41.92956	-87.70796	41.93000	-87.66000	member	0:11:04	Wednesday
8	1D806492F95973AC	electric_bike	9/17/2023 11:07	9/17/2023 11:13	Jeffery Blvd & 71st St	KA1503000018			41.76659	-87.57645	41.77000	-87.57000	member	0:06:34	Sunday
9	40D9EF382CC6C53D	classic_bike	9/17/2023 11:58	9/17/2023 12:08	Kedzie Ave & Milwaukee Ave	13085	California Ave & Milwaukee Ave	13084	41.92957	-87.70786	41.92269	-87.69715	member	0:09:46	Sunday
10	C60CE661AF7ECC93	electric_bike	9/7/2023 20:52	9/7/2023 21:06	Southport Ave & Wrightwood Ave	TA1307000113			41.92882	-87.66391	41.90000	-87.63000	member	0:14:08	Thursday
11	3812B98E9406040E	classic_bike	9/12/2023 16:01	9/12/2023 16:17	Financial Pl & Ida B Wells Dr	SL-010	Adler Planetarium	13431	41.87502	-87.63309	41.86610	-87.60727	member	0:16:19	Tuesday
12	EBA56298CB3C803F	classic_bike	9/24/2023 13:17	9/24/2023 13:50	Clark St & Schreiber Ave	KA1504000156	Oakley Ave & Touhy Ave	RP-004	41.99990	-87.67007	42.01234	-87.68824	member	0:33:20	Sunday
13	C6BD5AF648F11D11	electric_bike	9/28/2023 18:09	9/28/2023 18:15	Halsted St & Wrightwood Ave	TA1309000061	Halsted St & Roscoe St	TA1309000025	41.92919	-87.64914	41.94367	-87.64895	member	0:05:24	Thursday
14	585C82FA2E006DE9	classic_bike	9/22/2023 12:30	9/22/2023 12:42	Halsted St & Wrightwood Ave	TA1309000061	Halsted St & Roscoe St	TA1309000025	41.92914	-87.64908	41.94367	-87.64895	member	0:11:40	Friday
15	95E72C49D692F822	classic_bike	9/7/2023 16:28	9/7/2023 16:31	Clark St & Drummond Pl	TA1307000142	Clark St & Wellington Ave	TA1307000136	41.93125	-87.64434	41.93650	-87.64754	member	0:03:08	Thursday
16	26094B920577B434	classic_bike	9/4/2023 16:53	9/4/2023 16:56	Clark St & Drummond Pl	TA1307000142	Clark St & Wellington Ave	TA1307000136	41.93125	-87.64434	41.93650	-87.64754	member	0:03:05	Monday
17	A13DECC07C318A6F	classic_bike	9/14/2023 10:37	9/14/2023 10:40	Clark St & Drummond Pl	TA1307000142	Clark St & Wellington Ave	TA1307000136	41.93125	-87.64434	41.93650	-87.64754	member	0:02:22	Thursday
18	BE665598CC823457	electric_bike	9/9/2023 22:42	9/9/2023 22:49	Halsted St & Wrightwood Ave	TA1309000061	Lincoln Ave & Addison St	TA1309000050	41.92917	-87.64915	41.94618	-87.67331	member	0:07:05	Saturday
19	34AB2700F7062615	classic_bike	9/30/2023 14:04	9/30/2023 14:04	Wabash Ave & Grand Ave	TA1307000117	Wabash Ave & Grand Ave	TA1307000117	41.89147	-87.62676	41.89147	-87.62676	member	0:00:29	Saturday

The final dataset, which will be used for analysis and visualizations

Initial observations

- There are **15** columns and **65209** rows, meaning there are 15 columns * 65209 rows = **978135** cells in total.
- The columns we could do calculations with to answer the business question are **rideable_type, member_casual, ride_length,** and **day_of_week.**
- Some cells in **end_station_name** and **end_station_id** are empty, notably at **Southport Ave & Wrightwood Ave.** Was this docking station somehow corrupted?
  - We should keep this station in mind in case it comes up in the visualizations.
- Some entries even have no information on the start or end docking stations. Did the locations get corrupted? Maybe someone was using a VPN?
  - These trips automatically won't be counted when it comes to analyzing docking stations since there's no information on them.

length() and which(): number of occurrences in a column

```
> length(which(trips$rideable_type=='classic_bike'))
[1] 3132850
> length(which(trips$rideable_type=='electric_bike'))
[1] 3243728
> length(which(trips$rideable_type=='electric_scooter'))
[1] 144337
```

Classic bike trip count: **3132850**
Electric bike trip count: **3243728**
Electric scooter trip count: **144337**

The **electric scooter** is clearly **less popular** than the other two.

```
> length(which(trips$member_casual=='casual'))
[1] 2392528
> length(which(trips$member_casual=='member'))
[1] 4128387
```

Casual rider count: **2392528**
Member count: **4128387**

So the annual membership is already popular since most trips have been from paid memberships.

According to [Statology]

tail(), names(), sort(), and table() for finding the mode

```
> tail(names(sort(table(trips$day_of_week))),7)
[1] "Monday"    "Sunday"    "Tuesday"   "Thursday"  "Friday"    "Wednesday" "Saturday"
```

The least popular to most popular days of the week are sorted from left to right.
The least popular day for riding is **Monday,** while the most popular is **Saturday.**

```
> tail(names(sort(table(trips$start_station_name))),1)
[1] ""
> tail(names(sort(table(trips$start_station_name))),2)
[1] "Streeter Dr & Grand Ave" ""
```

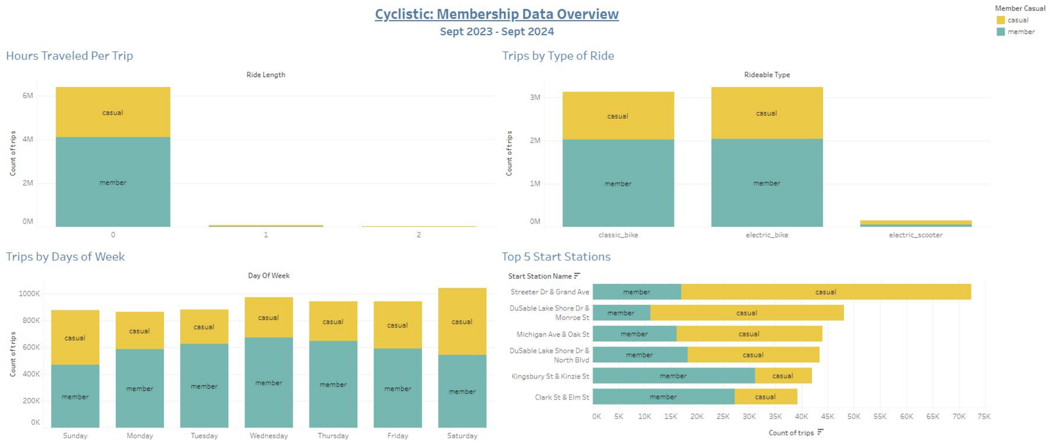
The most popular starting docking station is of course empty since some entries don't have them logged. So the top docking station is **Streeter Dr & Grand Ave.**

According to [Stack Overflow]

Share: Visualizing the data

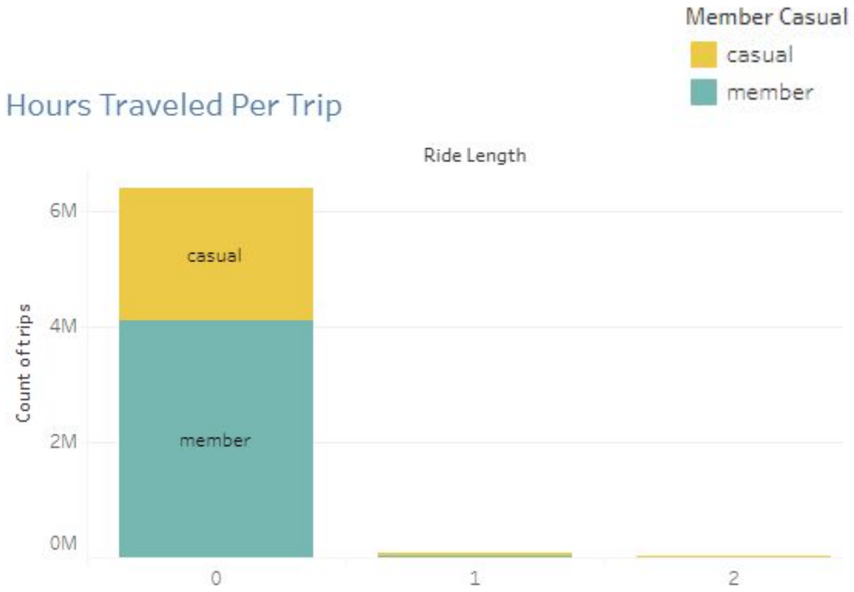
Visualizing the data

- I used **Tableau** to make the visualizations that were used to make my insights.
- I uploaded **trips.csv** on Tableau and made several visualizations like bar graphs and a heat map of the most popular docking stations.
- I combined these visualizations into an interactive dashboard and published it on **Tableau Public**, which can be viewed here.



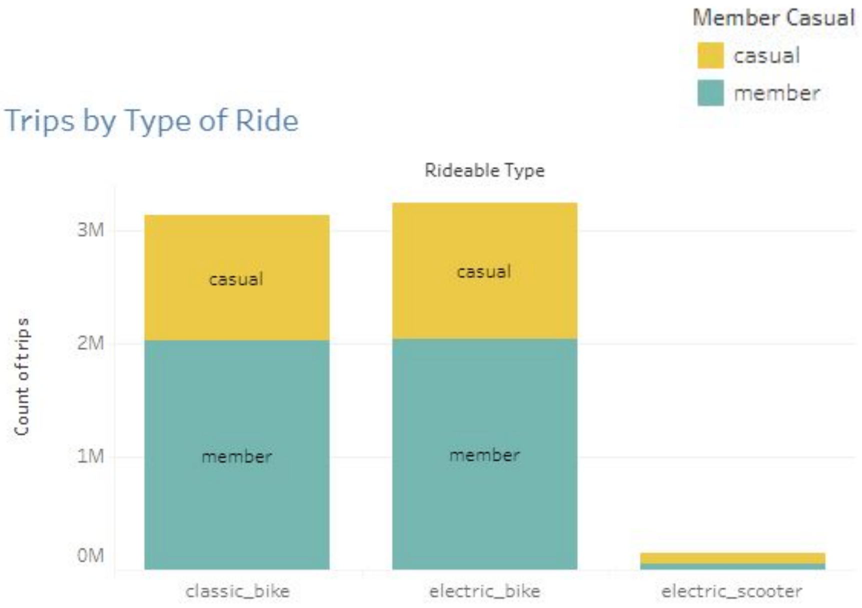
Most annual members do not travel for long

- Most of the trips are **less than one hour** with an **annual member** trip count of 4,105,894.
- If you look deeper into the graph on Tableau Public, the 1-2 hour trips are mainly done by **casual riders** with trip counts of 71,433 and 14,382, respectively.
- **Casual riders** with **longer trips** may find the annual membership plan **inefficient** for their travel needs. Let’s look further into why.



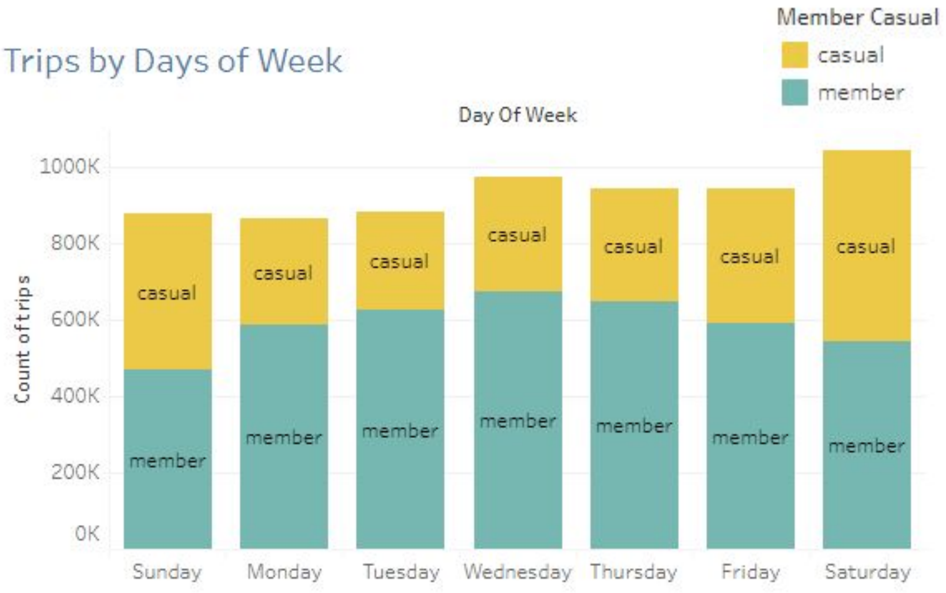
Electric scooters are perfect for long-range travelers

- Most of the **electric scooter** trips are from **casual riders**, with a trip count of 85,215.
- This makes sense since some electric scooters do not require **strenuous exercise** compared to bikes, making them ideal for longer trips.
- But if bikes are more popular than scooters, then **casual riders** may be **discouraged** from investing in a membership plan since their desired vehicle may be **scarce**.



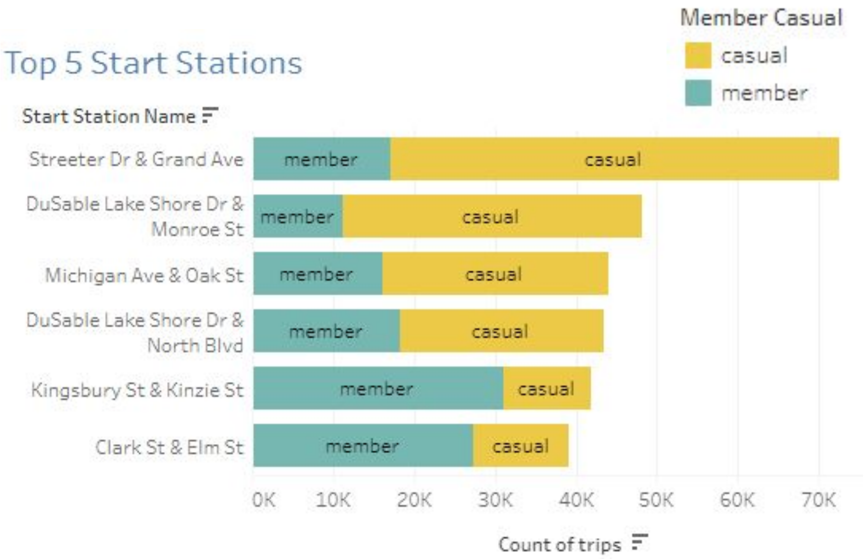
The weekends are popular amongst casual riders

- One reason could be their desire to do recreational activities outside of work.
- Casual riders have a total trip count of 498,570 for **Saturday**, and 411,038 for **Sunday**.
- Meanwhile, **Members** may be using their bikes to **commute** everyday, making them take up more of the weekday (Mon-Fri) trips.



Casual riders use the top docking stations the most

- It is the **start station** that travelers think about first, and casual riders take up 55,400 of the trips from **Streeter Dr & Grand Ave**.
- But members take up the majority of the least popular start stations, such as **Clark St & Elm St.**



Tourist sites make popular docking stations

- Casual-dominated docking stations like **Streeter St & Grand Ave** hold attractions like Chicago Children's Museum, making them popular for those **traveling a long time** to tourist sites.
- However, member-dominated docking stations like **Clark St & Elm St** do not have any notable attractions, so **members** may travel to them to run errands or for work.

Streeter St & Grand Ave

Clark St & Elm St

Act: Top three recommendations

Key takeaways

- **Casual riders** may use Cyclistic's services for occasional long trips, while **members** may use their annual membership plan for short commutes to school or work.
- Casual riders may choose to not subscribe to the **annual membership** plan because it favors **short-range travelers**.
- This answer is the result of my analysis into the **ride length, type of ride, day of the week,** and **docking stations** of Cyclistic's trip data**.**

Top three recommendations

1. **Focus your advertising efforts on electric scooters.**
    a. Since electric scooters are not nearly as popular as the bikes, making this type of ride known may even encourage members to start using them. Long-range travelers may also get to know about Cyclistic and get an annual membership plan if they do such travels regularly.
2. **Offer membership benefits for riding on the weekends.**
    a. Casual riders may not use Cyclistic services for commuting to school or work, making weekend trips more popular for them. If they're encouraged to ride on the weekends, they may decide that the annual membership plan is for them.
3. **Offer membership benefits for using different docking stations.**
    a. Casual riders may prefer popular tourist destinations over those used just for daily commutes. If they got benefits for making these trips, they may even plan more fun trips!

Call to action

What should **Lily Moreno** do after this presentation ends?

- Start developing plans to **advertise electric scooters** with the marketing team.
  - Think about how you'll do it, the costs, etc.
- Discuss how the marketing team could provide **membership benefits**, and decide where they could apply.
  - For the weekends and certain docking stations, for example.
  - Would you give credits or discounts? Free rides for future trips?

Thanks for listening

Sources

- [How to Count Number of Occurrences in Columns in R]
- [How to retrieve the most repeated value in a column present in a dataframe]
- [Cyclistic: Membership Data Overview (Sept 2023 - Sept 2024)]
- [Streeter St & Grand Ave]
- [Clark St & Elm St]