

CYCLISTIC RIDE-SHARING ANALYSIS

Differentiating casual riders from annual members to inform the upcoming marketing campaign.

Karl Madl, updated for June 2022

OBJECTIVE

Cyclistic is a fictitious alias for a bicycle ridesharing company in Chicago. The goal of this mock is to help inform the marketing team, whose campaign is aimed towards converting “casual riders” into annual members. Towards this aim, I identify key differences between the groups and general patterns in casual riding to provide opportunities for effective marketing.

DATA

The data used was provided in csv files by month. We only consider the last 12 months in our analysis. Because the analysis is to focus on Cyclistic riders only we have the population data and need not use statistical methods to make inferences about the data.

DATA

- The 12 separate csv files were loaded into dataframes and subsequently concatenated together for a master dataframe with ~5.75 million rows and 13 columns:
 - Ride ID
 - Rideable type: “classic”, “electric”, or “docked”
 - Datetimes of when ride started and ended (2 separate columns)
 - Station names and IDs for starting station and ending station (4 cols)
 - Latitude and longitude coordinates for each starting and ending station (4 cols)
 - Membership status: “member” or “casual”

DATA CLEANING

- 140 rides had the start time later than the end time, resulting in negative time spent on the bike.
- **Solution:** I swapped the start and end times for these rides. Due to the small relative amount and impossibility of accuracy, I could've dropped these observations instead but chose to potentially save as much data as possible.
- 790,000 observations were missing starting station names while 843,000 observations were missing ending station names, with a good deal of overlap. 4,766 observations were missing ending latitude and longitude coordinates, all these observations overlapped with those missing ending station names.
- **Solution:** There's no way to know where the 4,766 rides ended so those observations were dropped. For the rest of the missing data, a dictionary was created matching station names with known latitude and longitude coordinates and then using this dictionary to determine the missing station names based on their coordinates.
 - Some bias may be introduced as all the 4,766 dropped rides used electric bikes

ANALYSIS

There are **four** key differences and/or patterns that were identified:

1. **What**
2. **When**
3. **How Long**
4. **Where**

WHAT

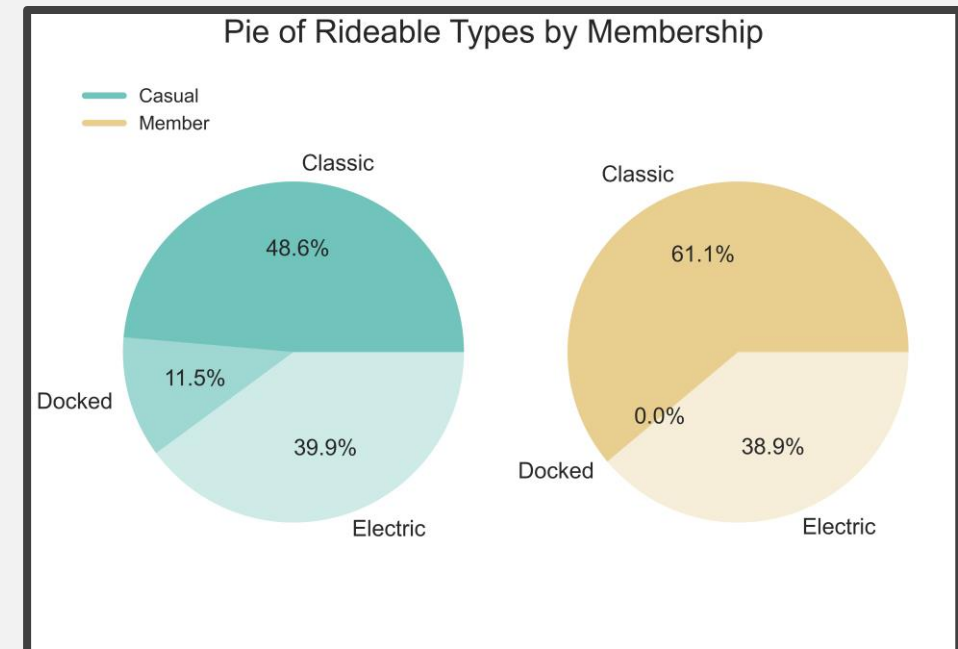
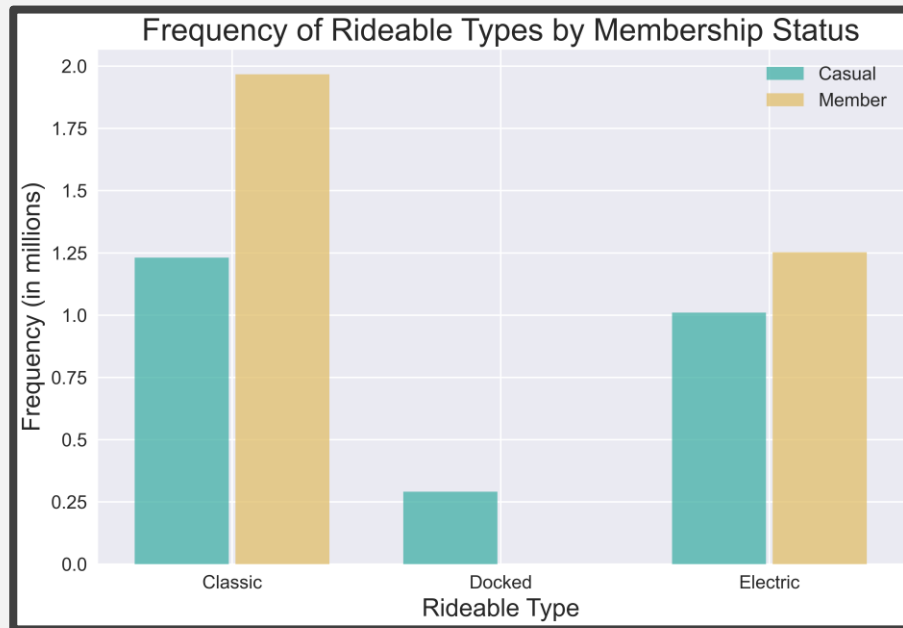
Cyclistic offers 3 rideables:

- Classic bike
- Electric bike
- Docked bike

How are member and casual rides distributed over these?

WHAT

- While members and casuals ride electric at the same rate, only one member ride was on a docked bike this past year. The difference is made up in classic bike rides.



ANALYSIS

There are **four** key differences and/or patterns that were identified:

1. What
2. **When**
3. **How Long**
4. **Where**

WHEN

When do casual riders and members ride bikes?

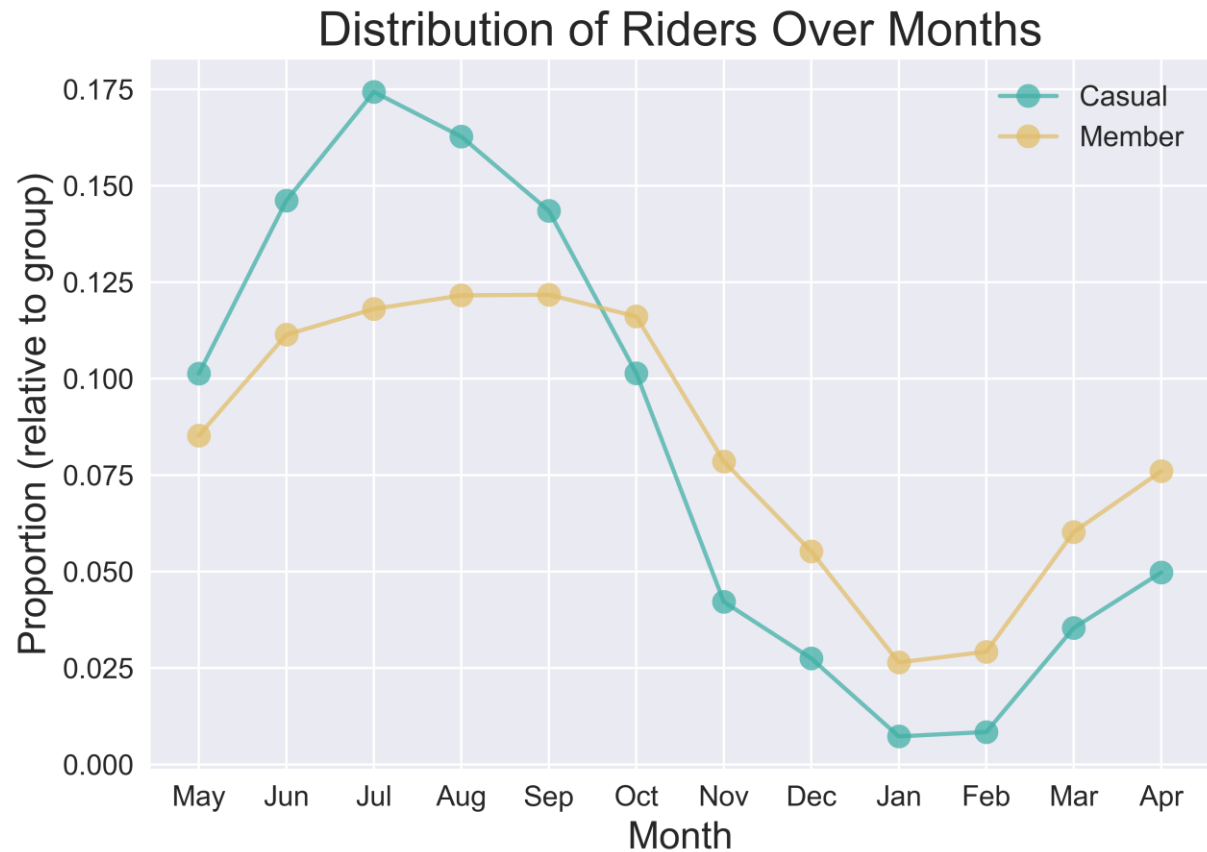
There are 3 levels of “when”:

- Yearly scale over months
- Weekly scale over days
- Daily scale over hours/minutes

WHEN

There are 3 levels of “when”:

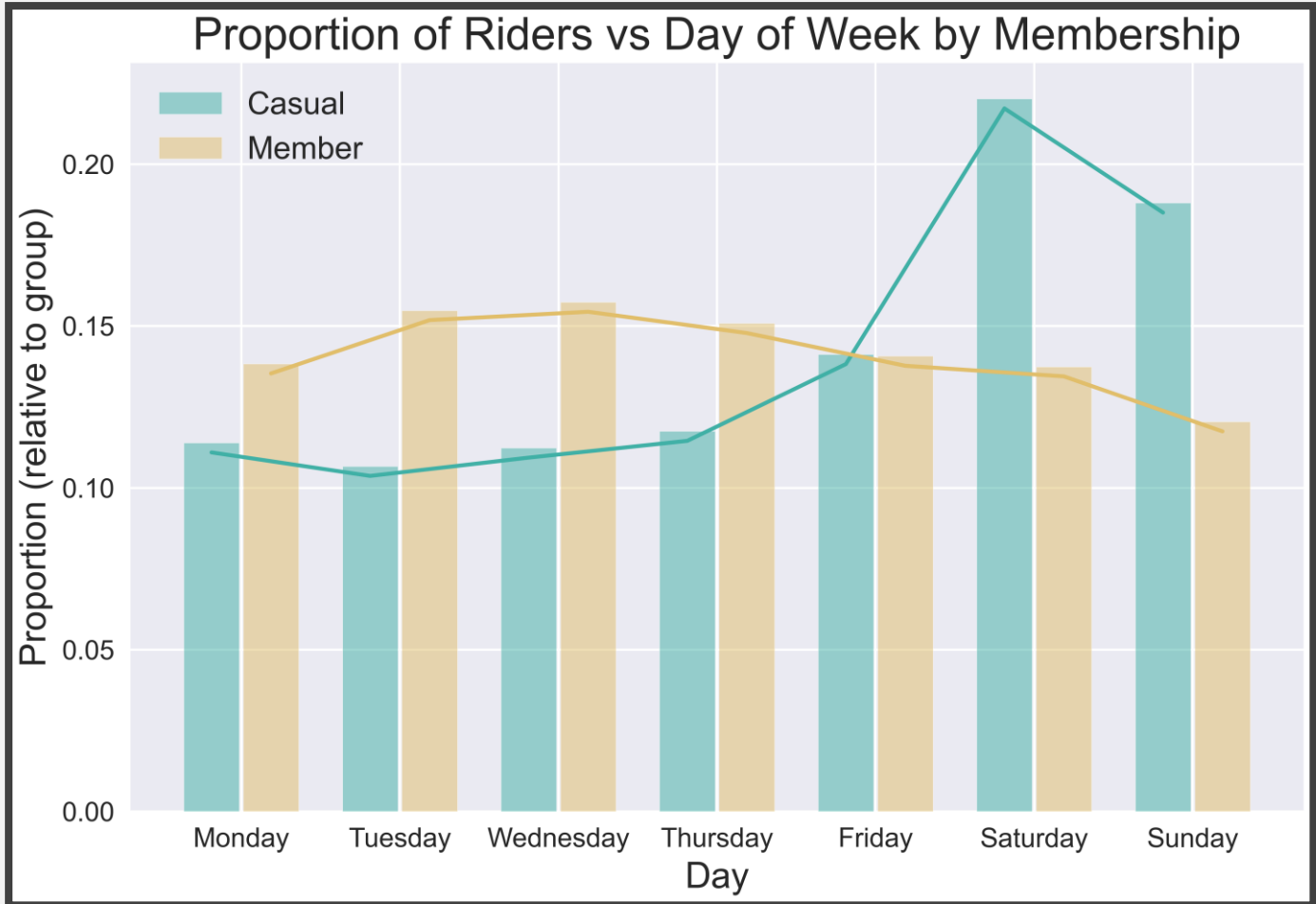
- **Yearly scale over months**
- Weekly scale over days
- Daily scale over hours/minutes



WHEN

There are 3 levels of “when”:

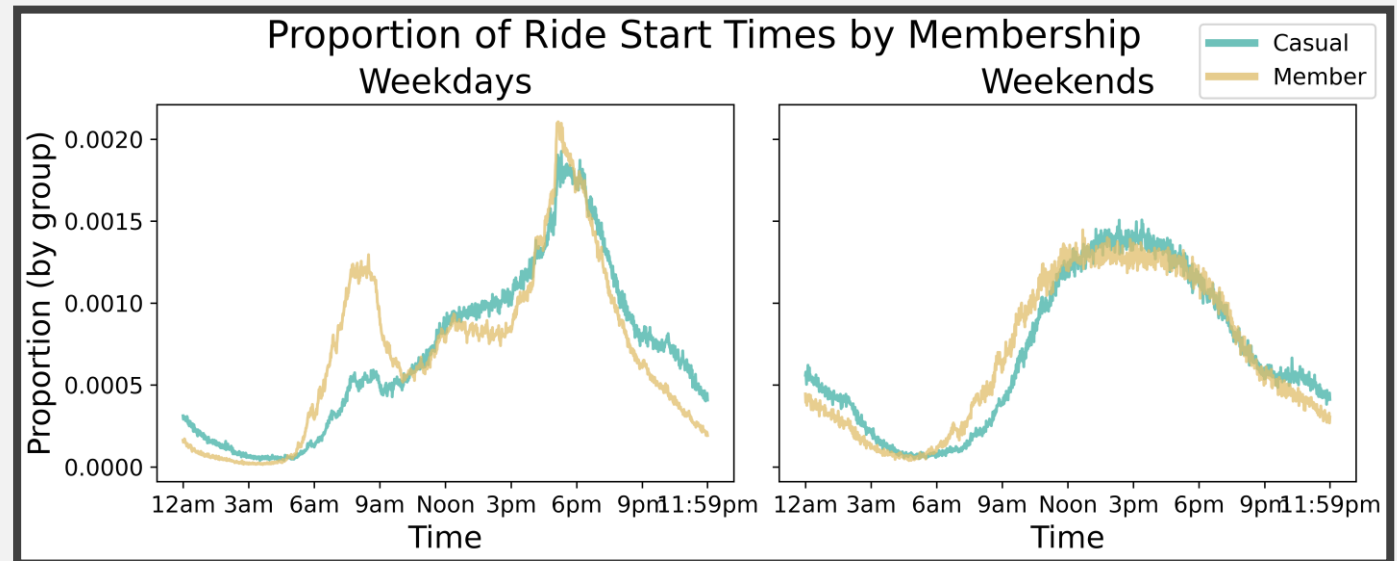
- Yearly scale over months
- **Weekly scale over days**
- Daily scale over hours/minutes



WHEN

There are 3 levels of “when”:

- Yearly scale over months
- Weekly scale over days
- **Daily scale over hours/minutes**



ANALYSIS

There are **four** key differences and/or patterns that were identified:

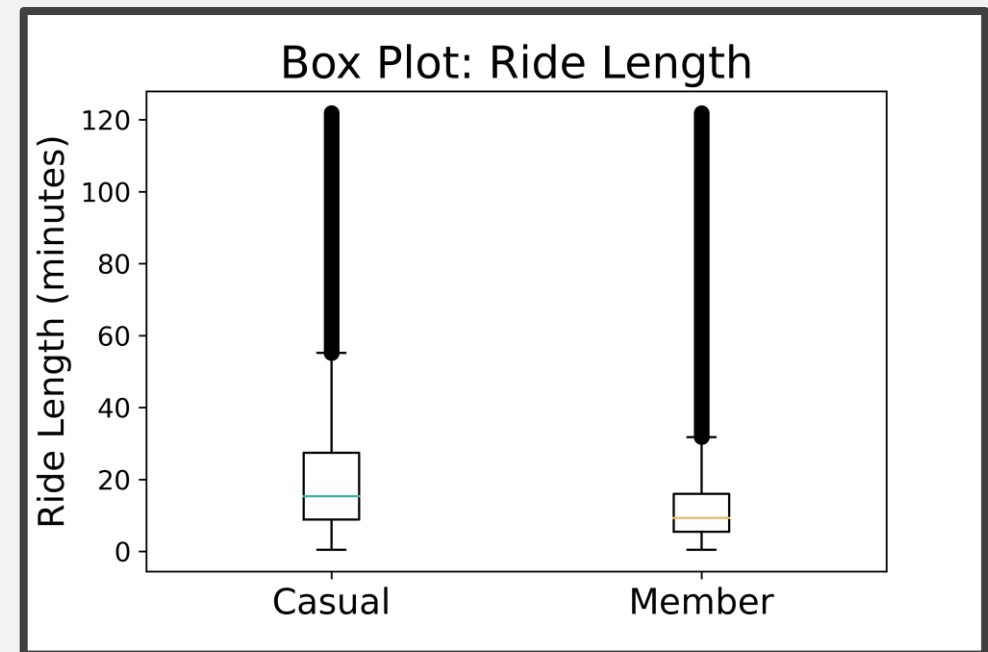
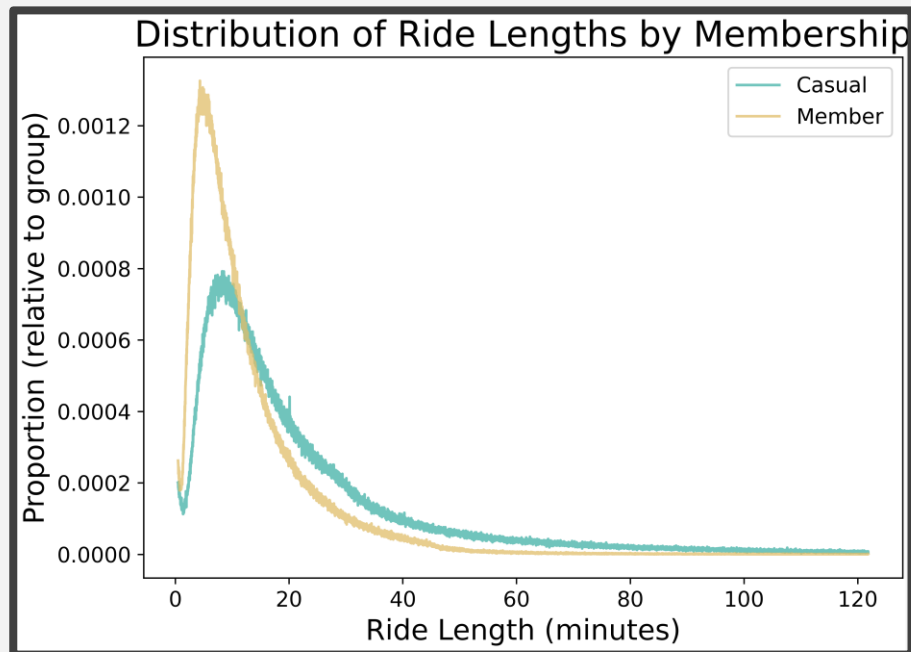
1. What
2. When
3. **How Long**
4. **Where**

HOW LONG

- The *ride_length* column was created by taking the difference between start and end times, kept in minutes.
- There are some large outliers in the data here (e.g., times well over 30 days) so 1% trimmed statistics were used
- The *ride_length* data was analyzed against multiple categorical variables

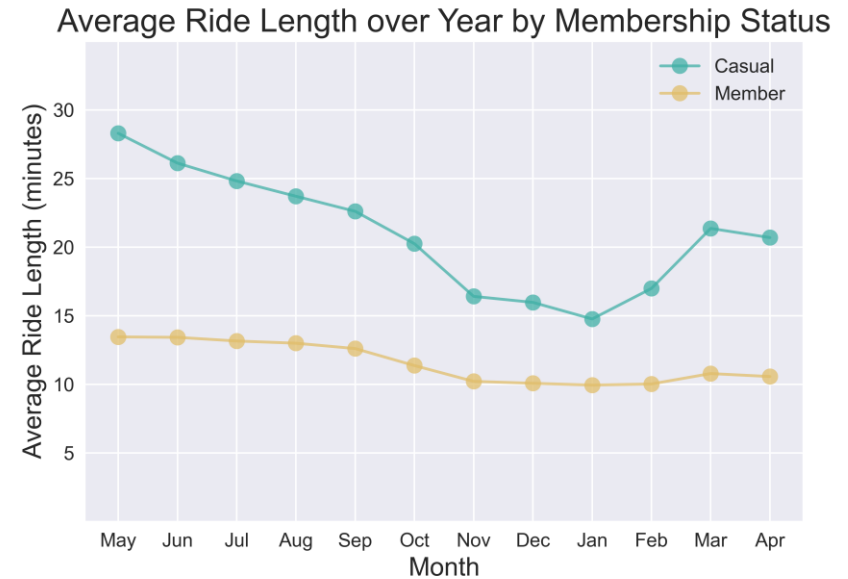
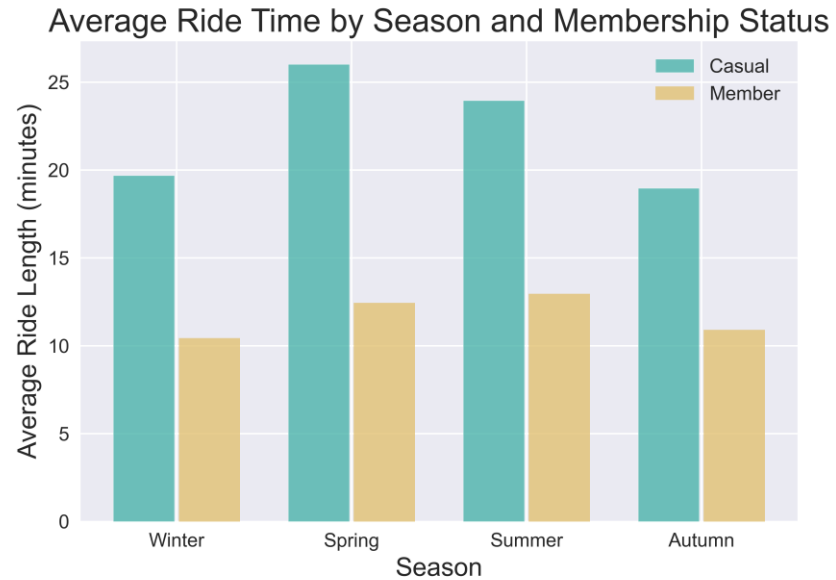
HOW LONG

- Ride time lengths distribution is more positively skewed for casual riders than members
- Average ride time length for
 - Casual riders is 23 minutes
 - Members is 12 minutes



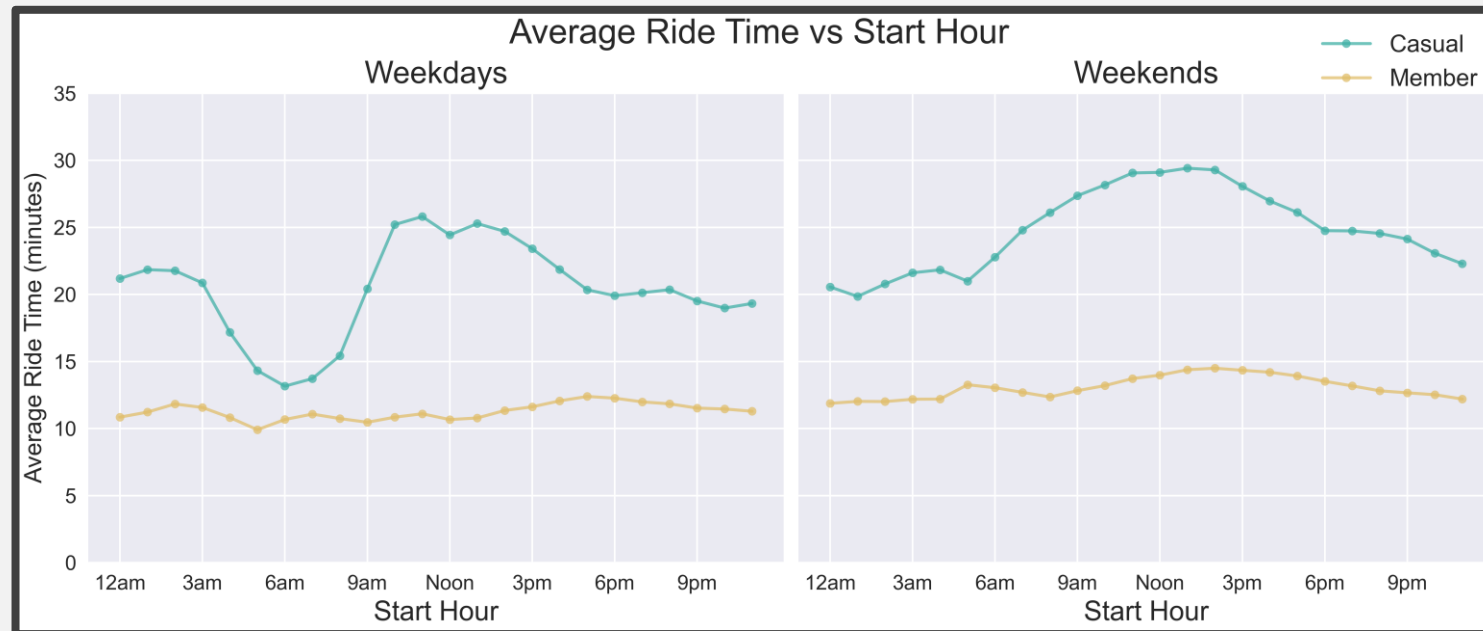
HOW LONG

- The greatest differences in ride times occur in the spring summer months.
- Members are consistent throughout the year while casual riders show much more variance.



HOW LONG

- Casual riders have the shortest trips weekday mornings and the longest around noon on weekends.
- Members ride slightly longer on weekends and are consistent across all times of day



ANALYSIS

There are **four** key differences and/or patterns that were identified:

1. What
2. When
3. How Long
4. **Where**

WHERE

What areas were riders in most often?

We analyzed which stations casual riders and members interacted with most frequently to inform possible physical marketing locations.

We used machine learning to separate stations into 3 areas:

- University of Chicago
- Navy Pier
- North Side

WHERE

To analyze the spread of station interactions, we create a distribution of the Euclidean Norms from the average location of all interactions to each station interaction. We choose standard deviation as a measure of spread.

The distribution for members has a standard deviation almost **4x** as large as the standard deviation for casual riders. Using different norms leads to various differences between the standard deviations.

In every case, the standard deviation of the members distribution is greater than that of the casuals → Casual Riders are more condensed around the center of the distribution (near Navy Pier)

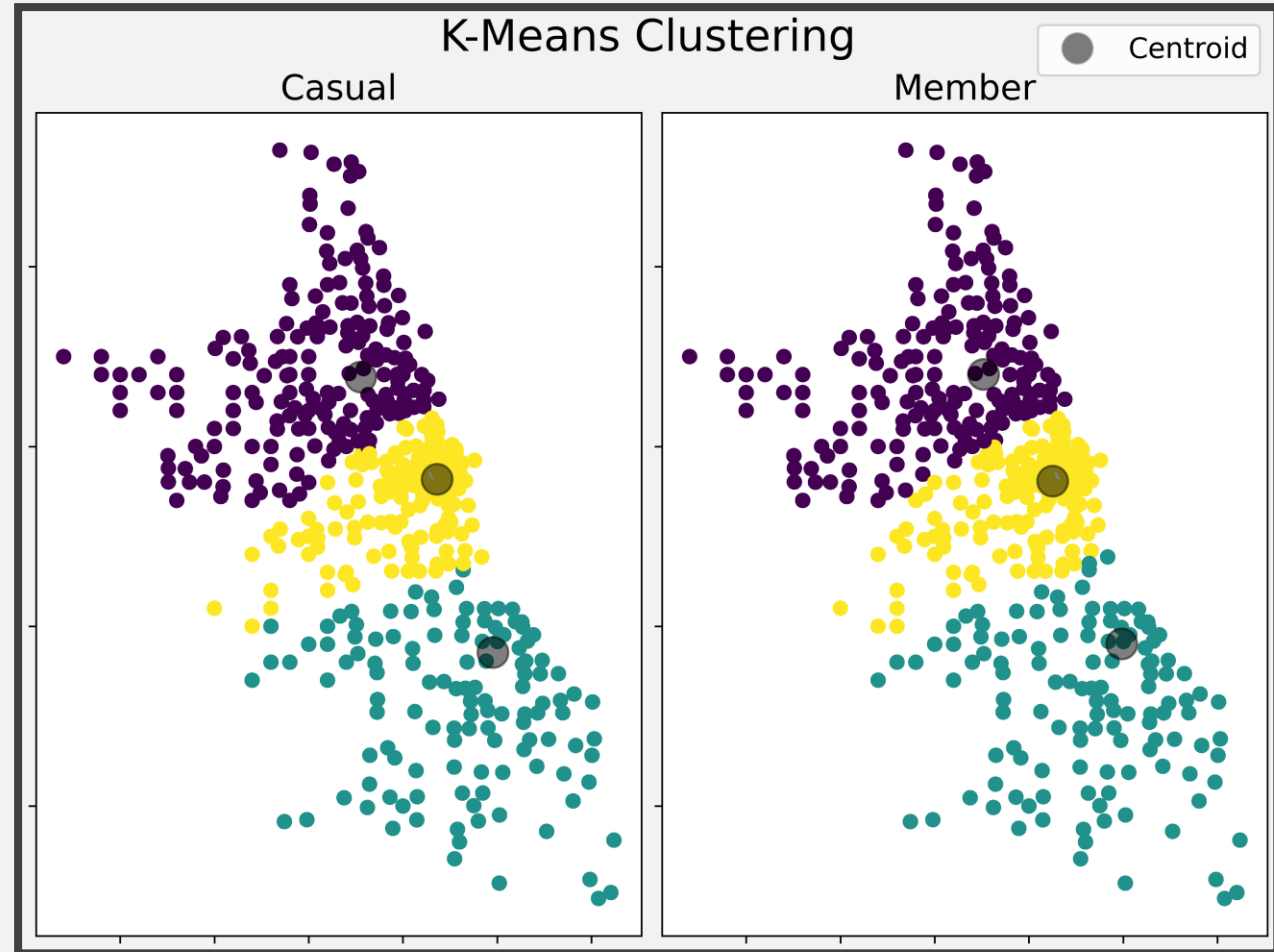
WHERE

Elbow method with the sum of squares was used to determine the optimal number of clusters

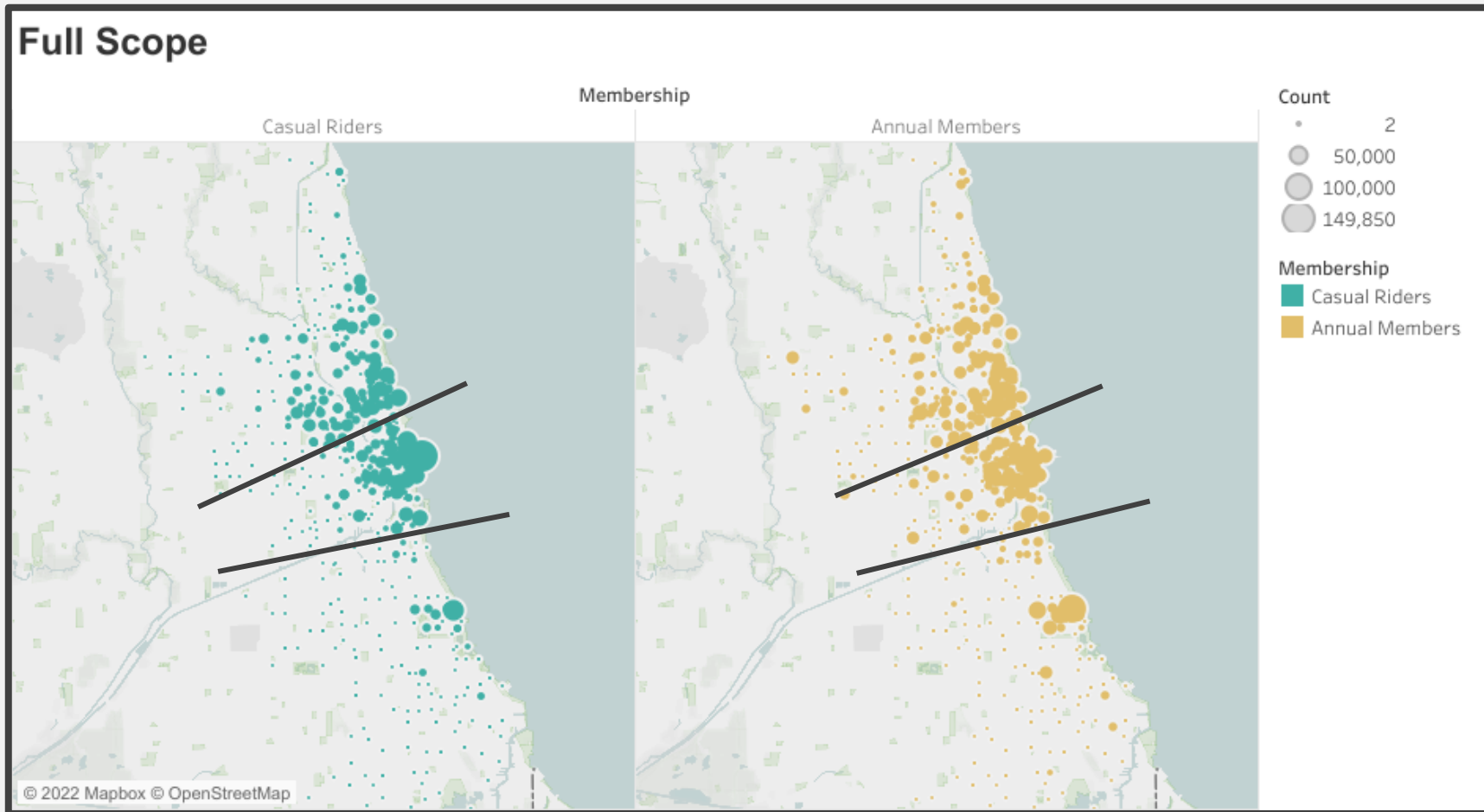
- 3 for casual riders and members

K-Means clustering algorithm grouped the stations similarly, but not identically.

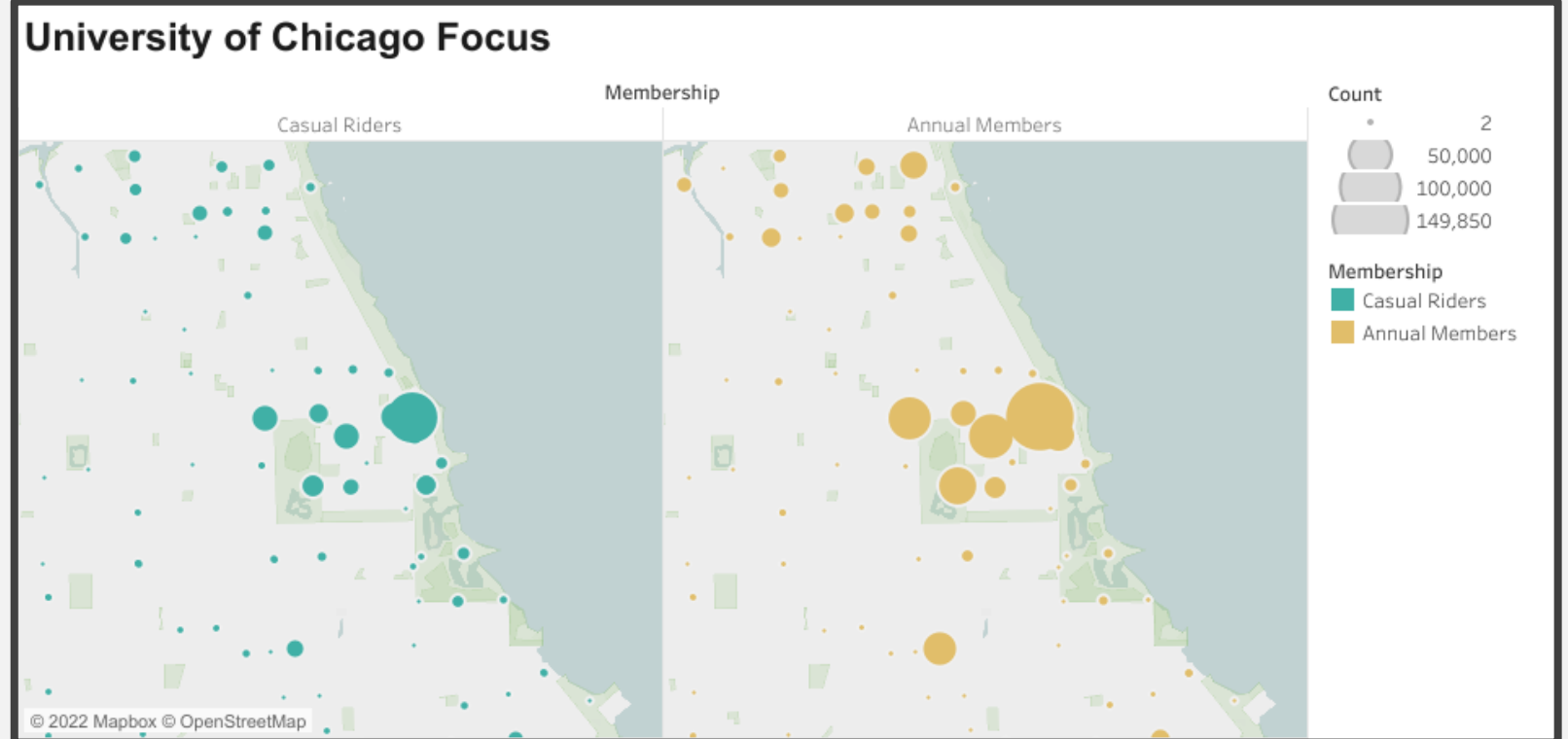
Member cluster sizes are more equivalent in size, matching the expected behavior from being more spread out



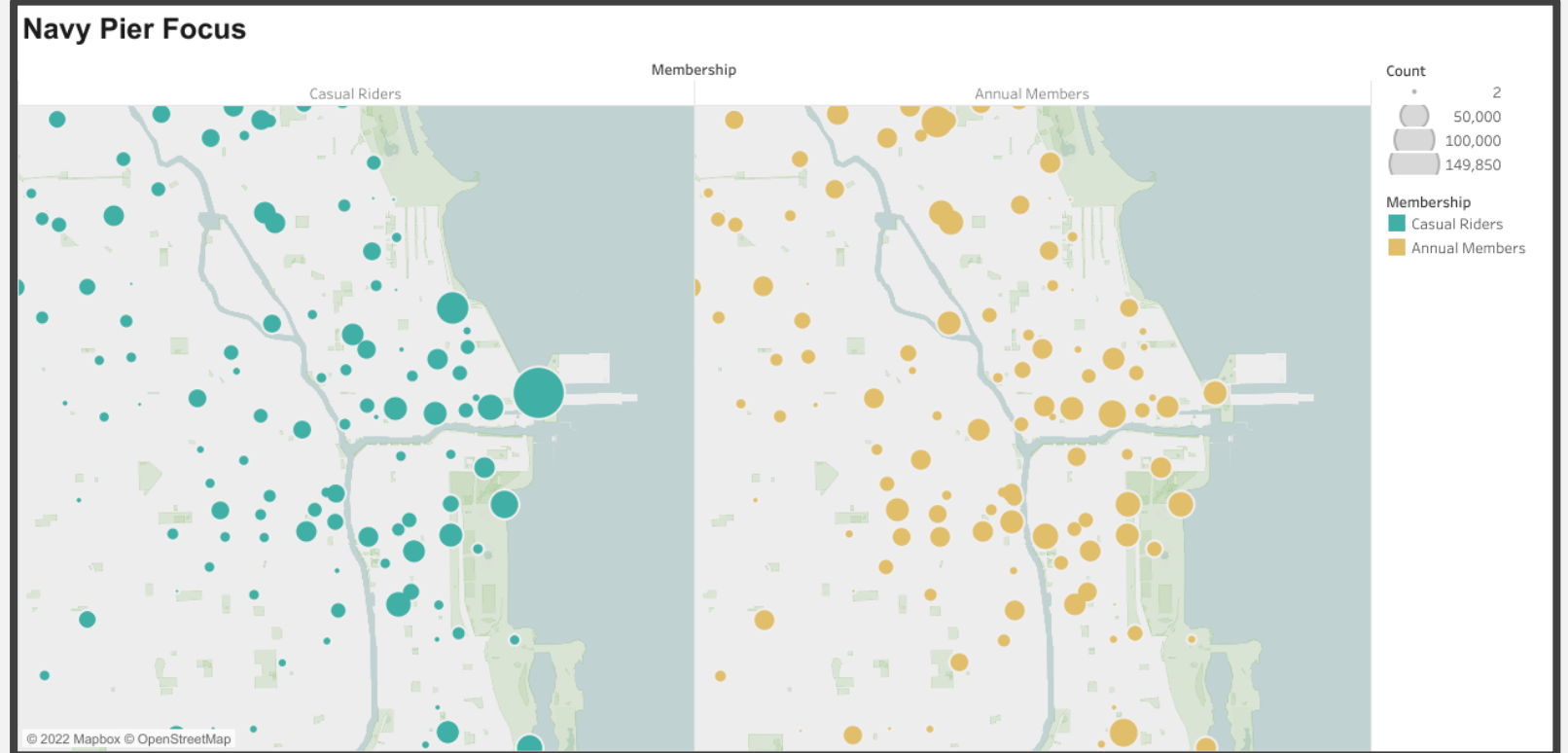
WHERE



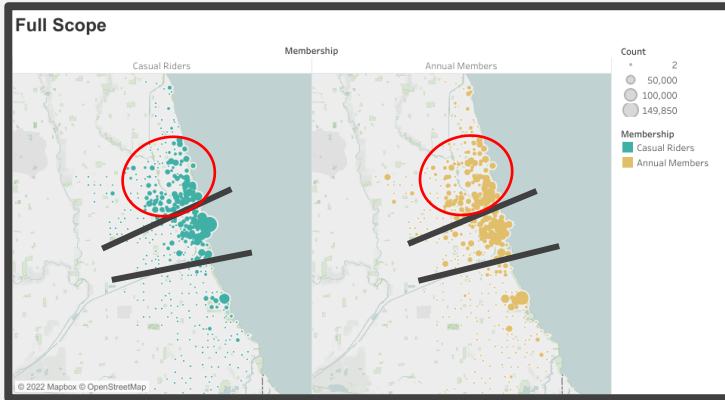
WHERE



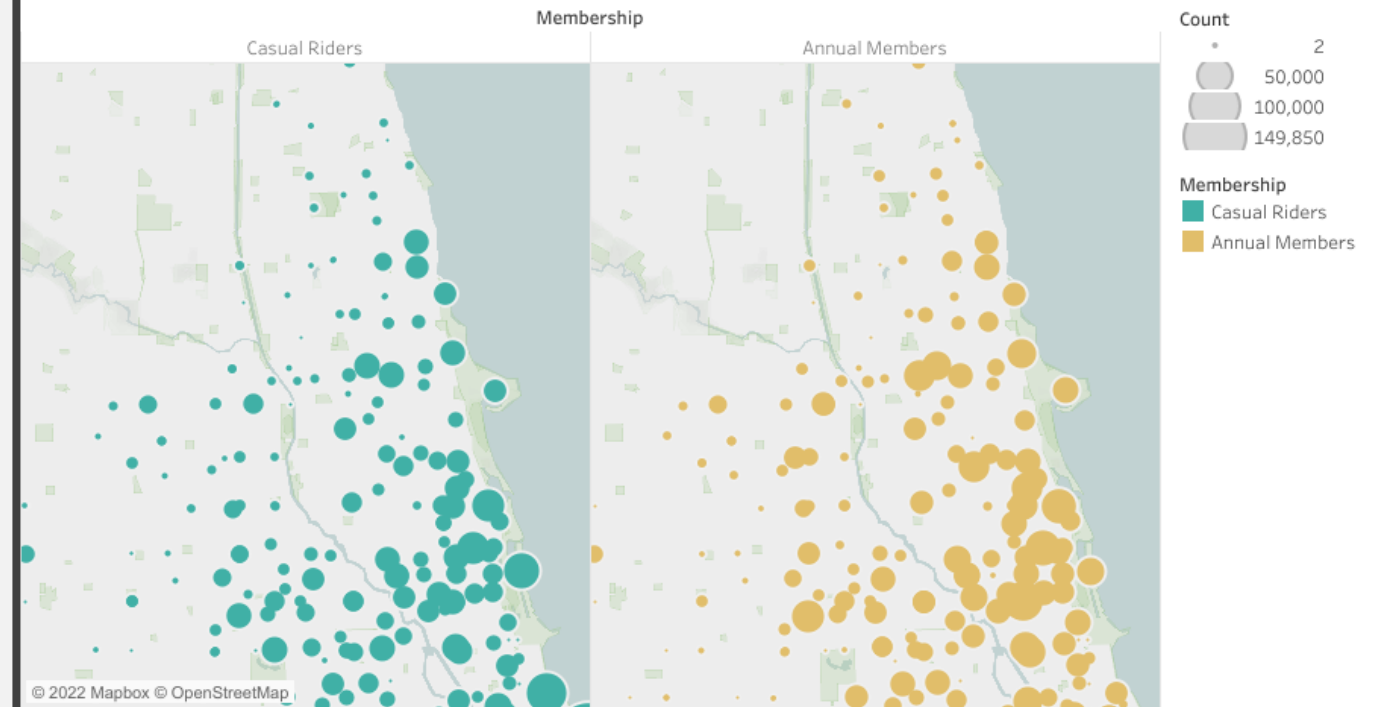
WHERE



WHERE



North Side Focus



WHERE

“Routes” were created as origin-destination pairs (we do not have access to during-ride location tracking, so these aren’t really routes)

10.6% of all casual riders began and ended their ride at the same station;
5.9% of all members did the same.

The top 10 most popular “routes” are given on the right.

- Most of the casual rides top 10 are in the Navy Pier and lower North Side regions
- Most of the member rides top 10 are in the University of Chicago region

Streeter Dr & Grand Ave --> Streeter Dr & Grand Ave	13290
Millennium Park --> Millennium Park	5546
Michigan Ave & Oak St --> Michigan Ave & Oak St	5511
DuSable Lake Shore Dr & Monroe St --> DuSable Lake Shore Dr & Monroe St	3927
Streeter Dr & Grand Ave --> Millennium Park	3596
DuSable Lake Shore Dr & Monroe St --> Streeter Dr & Grand Ave	3551
Millennium Park --> Streeter Dr & Grand Ave	3329
Montrose Harbor --> Montrose Harbor	3256
Theater on the Lake --> Theater on the Lake	3236
Shedd Aquarium --> Streeter Dr & Grand Ave	3150

CASUALS

University Ave & 57th St --> University Ave & 57th St	20127
Ellis Ave & 60th St --> University Ave & 57th St	11000
University Ave & 57th St --> Ellis Ave & 60th St	10254
Blackstone Ave & Hyde Park Blvd --> Blackstone Ave & Hyde Park Blvd	5973
University Ave & 57th St --> Blackstone Ave & Hyde Park Blvd	5406
University Ave & 57th St --> Ellis Ave & 53rd St	5338
Harper Ave & 59th St --> University Ave & 57th St	5289
University Ave & 57th St --> Harper Ave & 59th St	5218
Ellis Ave & 53rd St --> University Ave & 57th St	5117
Ellis Ave & 60th St --> Ellis Ave & 55th St	5089

MEMBERS

FINAL CONCLUSIONS

Summary of findings and general recommendations for marketing strategy based on the data.

Targeting docked bikes would yield a high rate of reaching the targeted audience of the campaign (casual riders)

Late spring through the summer are when casual riders are most active. Marketing during these months, and in the early spring to incentivize ridership, is recommended to reach as many riders as possible.

Casual riders spend twice as much time per ride as members; providing information to the potential savings a member may see is an option.

Focusing physical advertising space in areas of high station interactions for casual members is advised; namely, the most popular stations are Streeter Dr & Grand Ave, Millennium Park, and Michigan Ave & Oak St. A full list of stations and their interaction counts is available in the full report.

The general patterns of members indicate that they use Cyclistic bikes to make their daily commute and embark on some leisure activities.

Casual riders seem to be predominantly made up of leisure riders – tourists or locals choosing a different way to get around on a given day, particularly in the warmer months and off days.

For these casual riders, an annual membership may not make sense. I propose the possibility of a temporary membership/pass to draw in more casual riders to the membership program. These passes can be seasonal, monthly, weekly, or for just a weekend.

Lastly, although Cyclistic does well at gaining members at/near the University of Chicago, there is room for growth both there and at Loyola University. Continuing marketing to students, faculty, and staff of local colleges is another potential avenue at attracting more members.