Cyclistic, bike-share service

Case Study

Overview

5,824 bikes across 692 stations

Pricing plans:

- Single-ride pass
- Full-day pass
- Annual membership

Customers are divided into “casual riders” and “annual Cyclistic members”.

Task

Question: How do annual members and casual riders use Cyclistic bikes differently?

The end goal is to create a marketing strategy to **convert casual riders into annual members**, which have shown to be more profitable.

Data

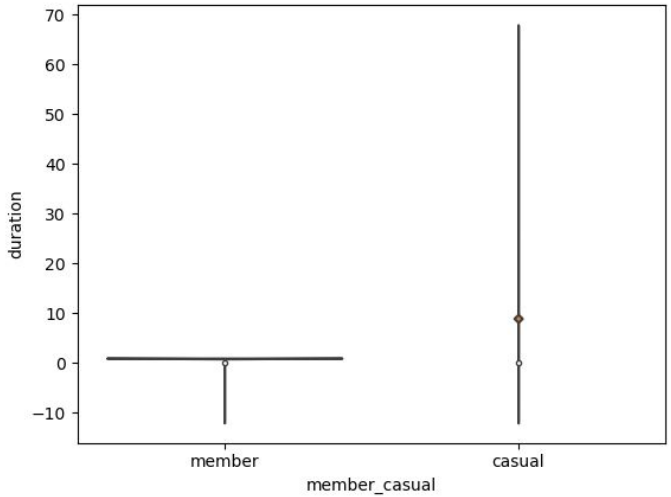
5,734,381 observations dated between July 2023 and June 2024:

ride_id	rideable_type	started_at	ended_at	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng
CC60058DD1DDD488	electric_bike	2023-10-20 21:17:54	2023-10-20 21:18:13	NaN	NaN	NaN	NaN	41.940000	-87.650000
77D007DA4633F239	classic_bike	2024-04-19 12:35:25	2024-04-19 12:59:38	Greenview Ave & Fullerton Ave	TA1307000001	Rush St & Cedar St	KA1504000133	41.925330	-87.665800
D95194B1F01BD550	classic_bike	2023-10-21 15:53:17	2023-10-21 16:00:44	Wells St & Huron St	TA1306000012	Wells St & Evergreen Ave	TA1308000049	41.894722	-87.634362

There are 13 columns in total.

Preprocessing

Removed outliers (some rentals lasted -12 and 68 days, latter accounting for bikes never returned):



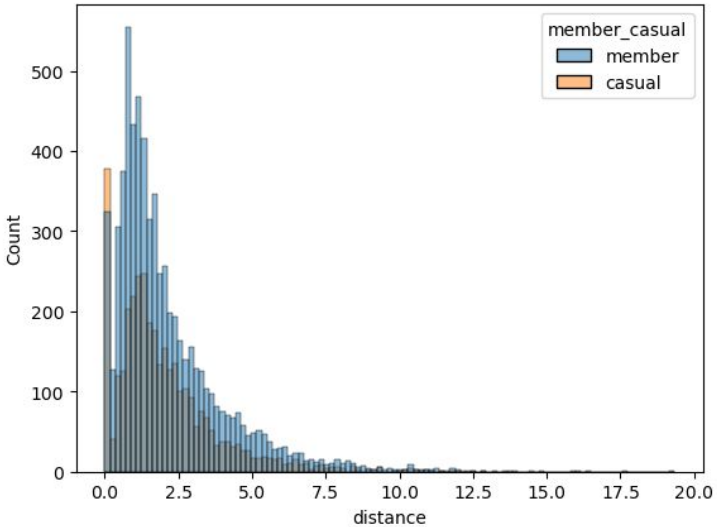
Removed duplicates (same rentals with difference only in milliseconds):

	ride_id	rideable_type	started_at	ended_at	start_station_name
4069781	FEA150E7A56F187E	classic_bike	2024-05-31 23:09:06.280	2024-06-01 08:47:44.728	Sangamon St & Lake St
5309580	FEA150E7A56F187E	classic_bike	2024-05-31 23:09:06	2024-06-01 08:47:44	Sangamon St & Lake St

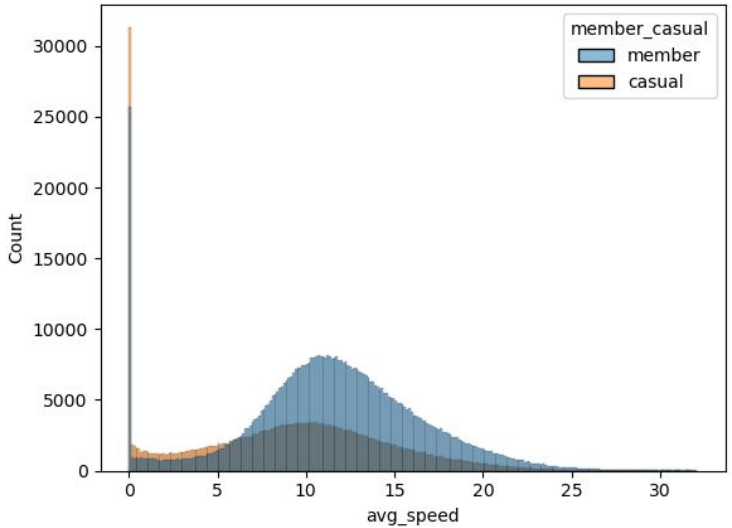
	ride_id	rideable_type	started_at	ended_at	start_station_name
4525586	F2E698ECB05C43D7	electric_bike	2024-05-31 23:50:48.241	2024-06-01 00:04:35.916	Bissell St & Armitage Ave*
5050049	F2E698ECB05C43D7	electric_bike	2024-05-31 23:50:48	2024-06-01 00:04:35	Bissell St & Armitage Ave*

Feature engineering

Used Haversine formula to calculate `distance` between two coordinates:



Created `avg_speed` column out of ride distance and duration:



Seasonal trends

Just from the file sizes, we can observe that there was a **decline in bike shares during winter** season (since more data, or file size, would mean more rentals).

```
data/202307-divvy-tripdata.csv 153861112
data/202308-divvy-tripdata.csv 155106942
data/202309-divvy-tripdata.csv 134160122
data/202310-divvy-tripdata.csv 107890751
data/202311-divvy-tripdata.csv 73030585
data/202312-divvy-tripdata.csv 45031306
data/202401-divvy-tripdata.csv 29458763
data/202402-divvy-tripdata.csv 45854560
data/202403-divvy-tripdata.csv 60761710
data/202404-divvy-tripdata.csv 82116053
data/202405-divvy-tripdata.csv 120615989
data/202406-divvy-tripdata.csv 144699212
```

First thoughts

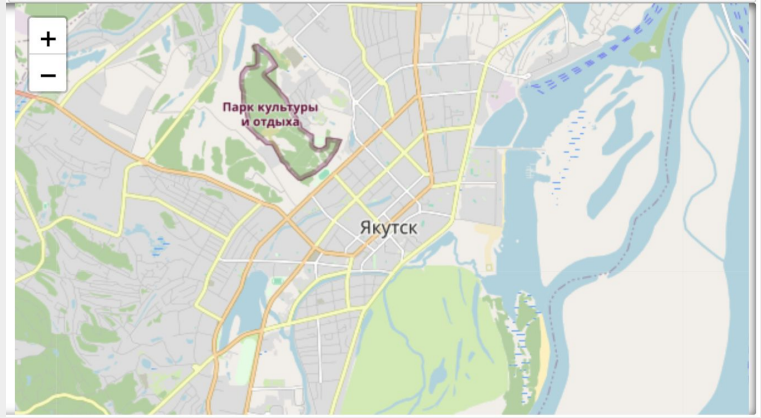
- Very short rentals (< 1 min)
  - Could have problems with setting up => might want to make UI simpler for new users
  - Observe cases when starting and ending point was the same. Hypothesize reasons
- Distances that rides were taken for
  - Do annual members tend to rent for longer (timewise/distance-wise) rides or more often rides?
- Location
  - Check if annual members commuted back and forth (on the same day, e.g. to work)
    - Cannot be checked, since rides are anonymous
  - Does location of start (i.e. a station) influence purchasing annual membership?
    - No data as well

Since all points involve analyzing **geodata**, let's use Folium library.

Geodata with Folium

Folium is a Python library made to visualize geographical data with Leaflet.js framework. It is easily customizable and convenient for use.

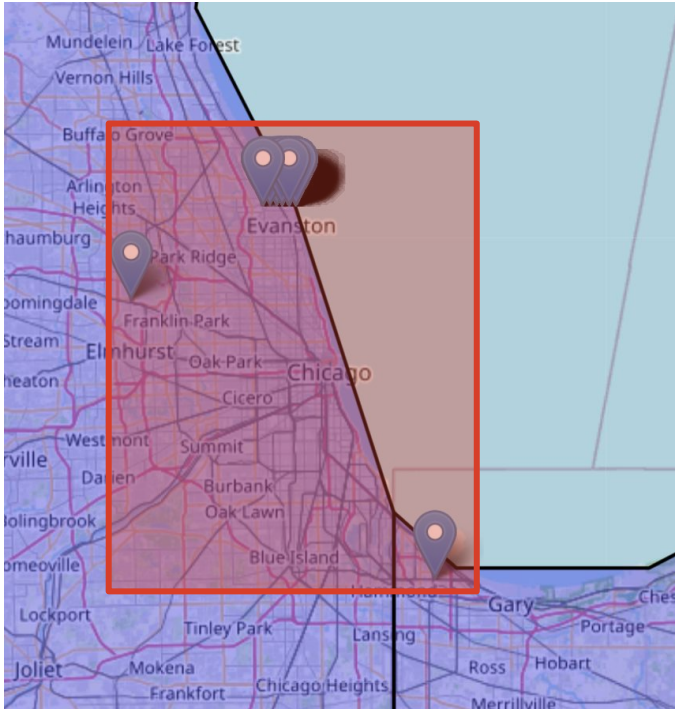
```
import folium

map = folium.Map(
    location=[62.035454, 129.675476], # Yakutsk (Якутск)
    zoom_start=11
)
map
```



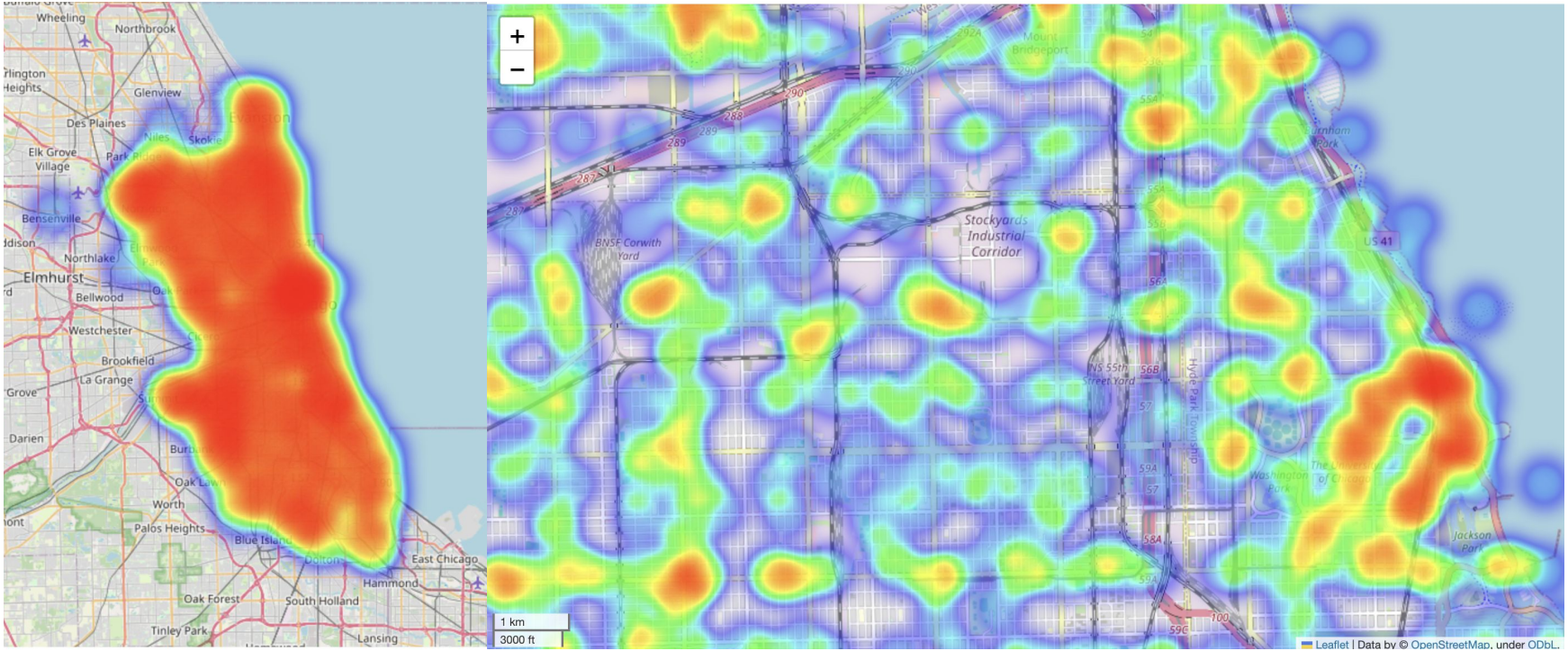
Area borders with Choropleth

The whole area can be classified as **Chicagoland**, the largest metropolitan area in the USA.

It lays on the borders of *two* states: Illinois and Indiana.



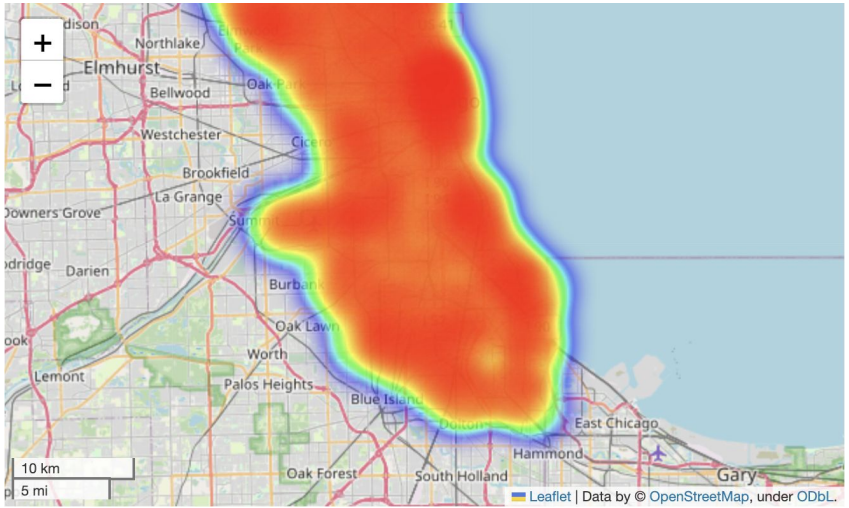
Total stations' usage with Heatmap



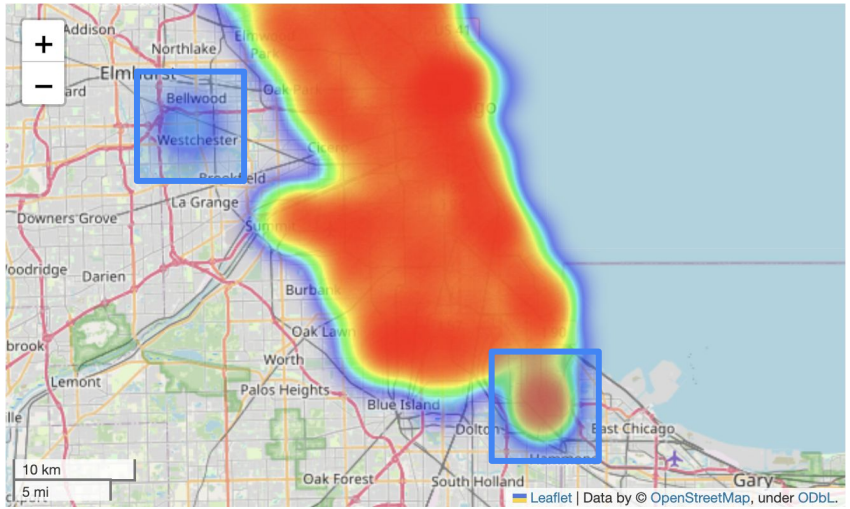
Stations' usage by membership types

Annual members clearly used farther stations more often.

Casual riders



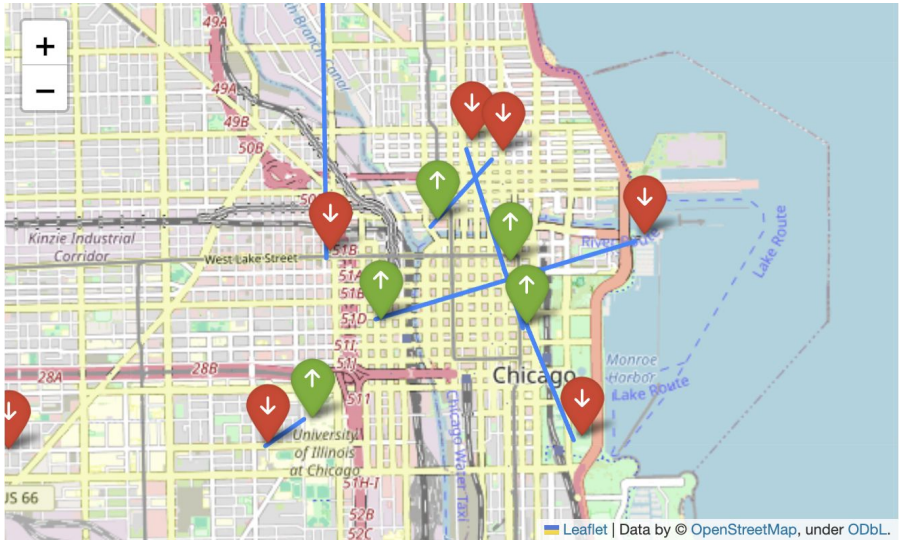
Annual members



Routes with Marker and PolyLine

Markers for stations are jittered so that they would not overlay.

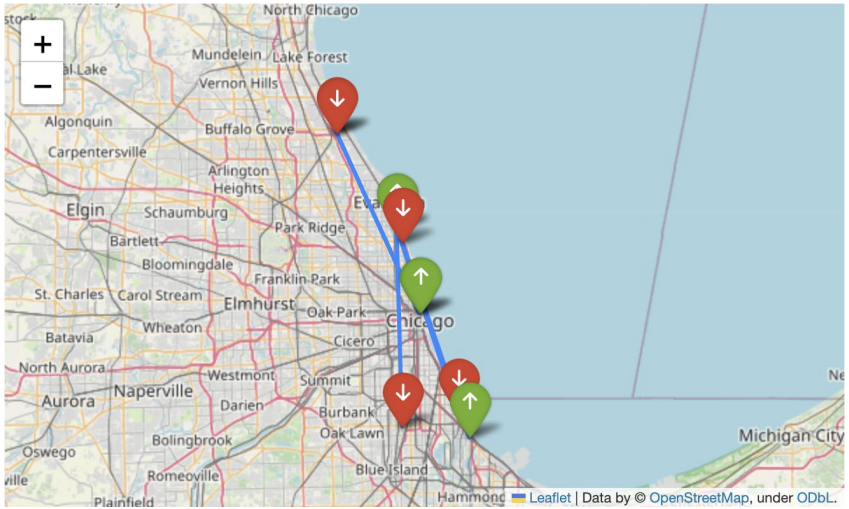
A route indicates where a rider **started** (green arrow up) and **ended** (red arrow down).



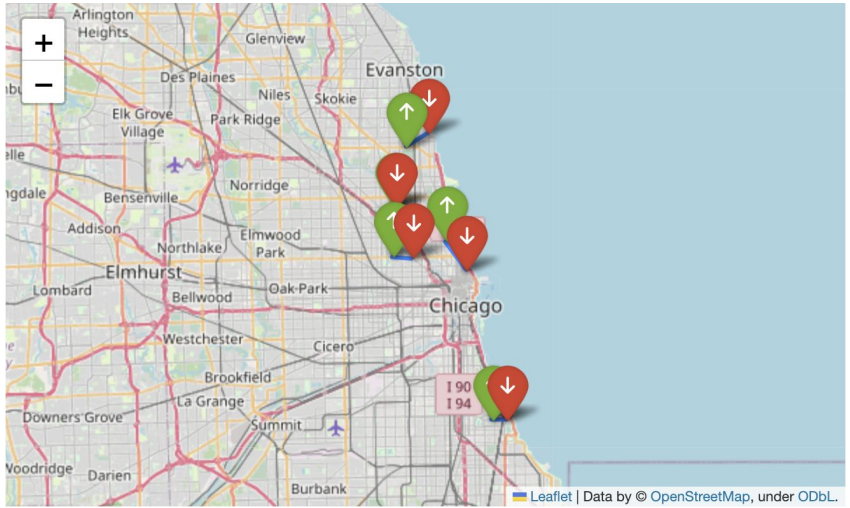
Top routes by distance and duration

It is notable that rentals lasting long are not necessarily longest ones!

Top 5 longest **distance** routes:

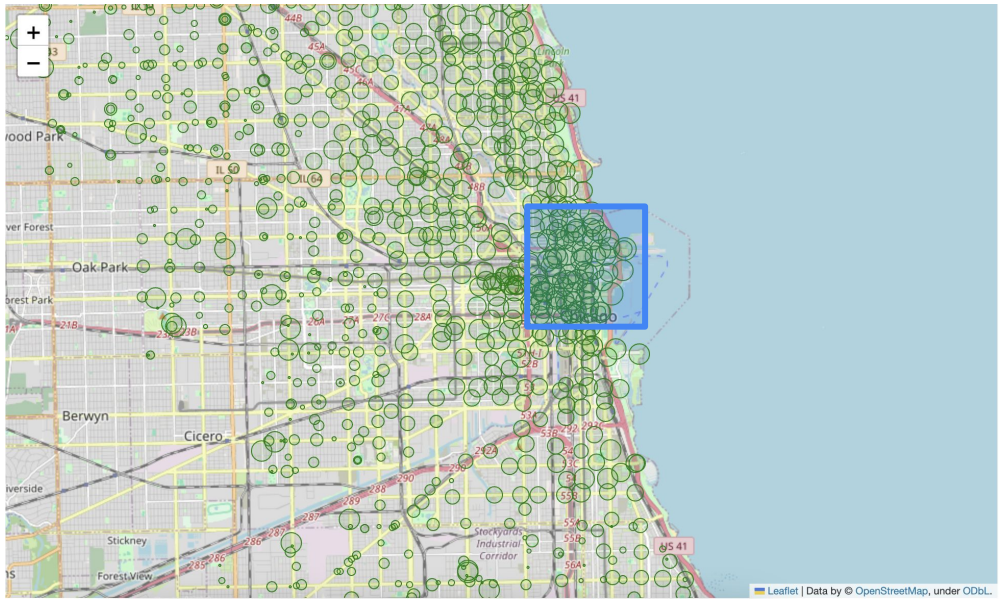


Top 5 longest **duration** routes:



Total stations' usage with CircleMarker

High density closer to the coastline.

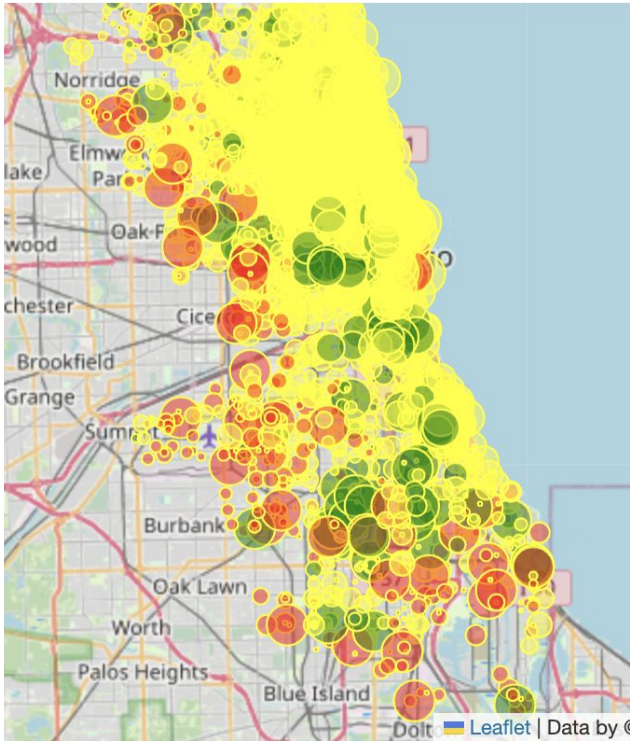


Stations' usage by membership types

Green circles are stations used prevalently by annual members (over 75%), **yellow** neutral, while **red** are those used more by casual riders.

As we can see, casual riders tend to ride around the downtown area, while annual members outnumber in farther locations.

**The circle size is proportional to the number of rentals started at that station.*



Conclusion

Marketing strategy should focus on **promoting accessibility to any point of the Chicago metropolitan area** and **speediness** in order to make users buy an annual membership. These features seem to be characteristics of Cyclistic annual members, and could be the main reason for the purchase of membership.

Casual riders	Annual members
Tend to ride in downtown area	Ride everywhere
Average speed 9 km/hour (slower)	Average speed 11 km/h (faster)

What else could be done?

We could consider some factors that could influence rentals:

- Weather
  - Import weather conditions on specified date and area (*define area*)
- Crime rates, traffic congestion, etc.

For testing out hypotheses, some estimators and predictors could be built as well.

Thank you! Mahtal!

Contacts:

eginovasardaana@berkeley.edu

Sardaana Eginova