



Opening a Gym a Chicago

COURSERA CAPSTONE

SAHIL DESAI

BUSINESS PROBLEM

- Location of the gym will be one of the deciding factors in determining opening success
- Objective: To analyze and select the best suburbs or neighborhoods in Chicago
- Find an area where gyms are not available, and avoid areas that have too many gyms

Business Question

Where would be the best area to build a new gym in Chicago as an aspiring gym owner?

DATA REQUIREMENTS

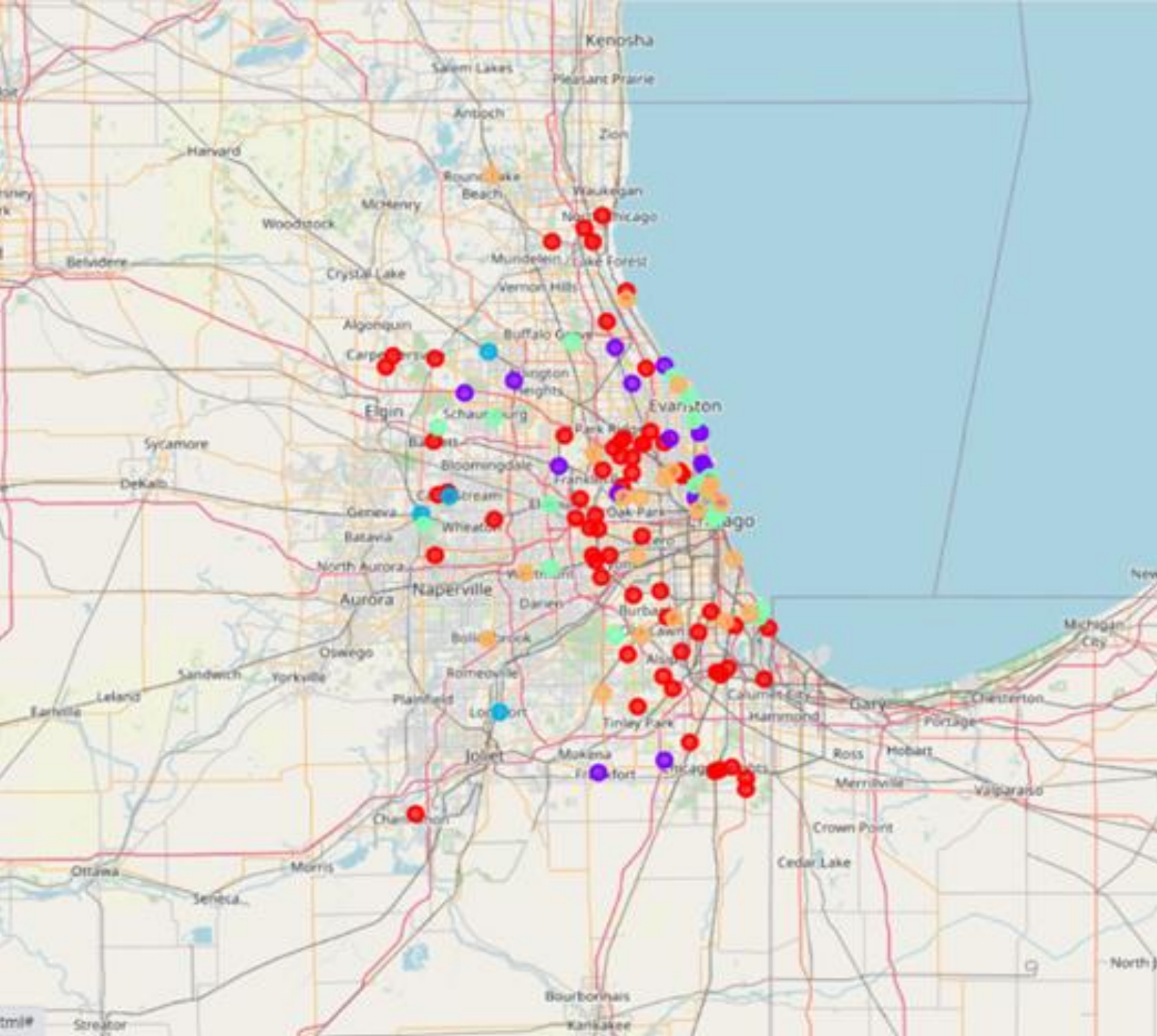
- Neighborhoods/Suburbs in Chicago
- Latitude and Longitude coordinates of each neighborhood that is webscrapped
- Venue data with emphasis on gyms

Data Source

- Wikipedia page [https://simple.wikipedia.org/wiki/Category:Suburbs of Chicago, Illinois](https://simple.wikipedia.org/wiki/Category:Suburbs_of_Chicago,_Illinois)
- Foursquare to make API calls and retrieve needed venue data
- Geocoder package to get coordinates

METHODOLOGY

- Webscrape off Wikipedia page to get list of neighborhoods
- Get the latitude and longitude coordinates for each neighborhood using Geocoder
- Use Foursquare to make API calls for venue data
- Grouping data by neighborhood and taking average frequency of each venue occurrence in neighborhoods
- Filter venue category for gyms
- Perform unsupervised learning algorithm k-means-clustering for analysis
- Visualize clusters using Folium



RESULTS

Categorized neighborhoods into 5 clusters

- Cluster 0: Neighborhood with no number of gyms (red)
- Cluster 1: Neighborhood with moderate concentration of gyms (blue)
- Cluster 2: Neighborhood with high concentration of gyms (light blue)
- Cluster 3 Neighborhood with moderate concentration of gyms (light green)
- Cluster 4 Neighborhood with low concentration of gyms (green)

DISCUSSION

- Cluster 0 takes a heavy portion of the neighborhoods and is widely scattered across near inner city and outer suburbs. However, I interpreted this to be inaccurate as many of these areas are close to other clusters. Foursquare didn't have any gym data over locations in this cluster unfortunately.
- Clusters 1 and 3 which have moderate amount of gyms and would remotely difficult to open a gym there
- . Cluster 2 is the smallest and has most gyms out of any cluster
- . Cluster 4 has low amount of gyms

CONCLUSION

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing unsupervised machine learning by clustering the data into 5 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders (gym contractors, gym owners, property developers etc.) regarding the best locations to open a new gym. I would have liked my API call to capture more information about Chicago venues that were categorized as gyms, because I had a large cluster unrealistically portray a big part of the city not having gyms. But based off my findings, we still found clusters that had high and low competition. As to answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 4 are the most preferred locations to open a new gym. The findings of this project will help those to make an informed decision on choosing high potential locations while avoiding over competitive areas in opening a new gym.

THE END