Applied Data Science Capstone

Choosing an opening location for a Chinese restaurant

BY: XIAN DONG

FEBRUARY 2020

Business Problem

- Chinese Restaurants are opened throughout Chicago and are very profitable.
- Choosing a restaurant's location plays a key role in reducing the chance of it failing within the next few years.
- A good location should not be suffocated with competitions from other Chinese Restaurants and the area should be relatively safe.
- "In which Chicago neighborhood should someone look to open a Chinese Restaurant?"

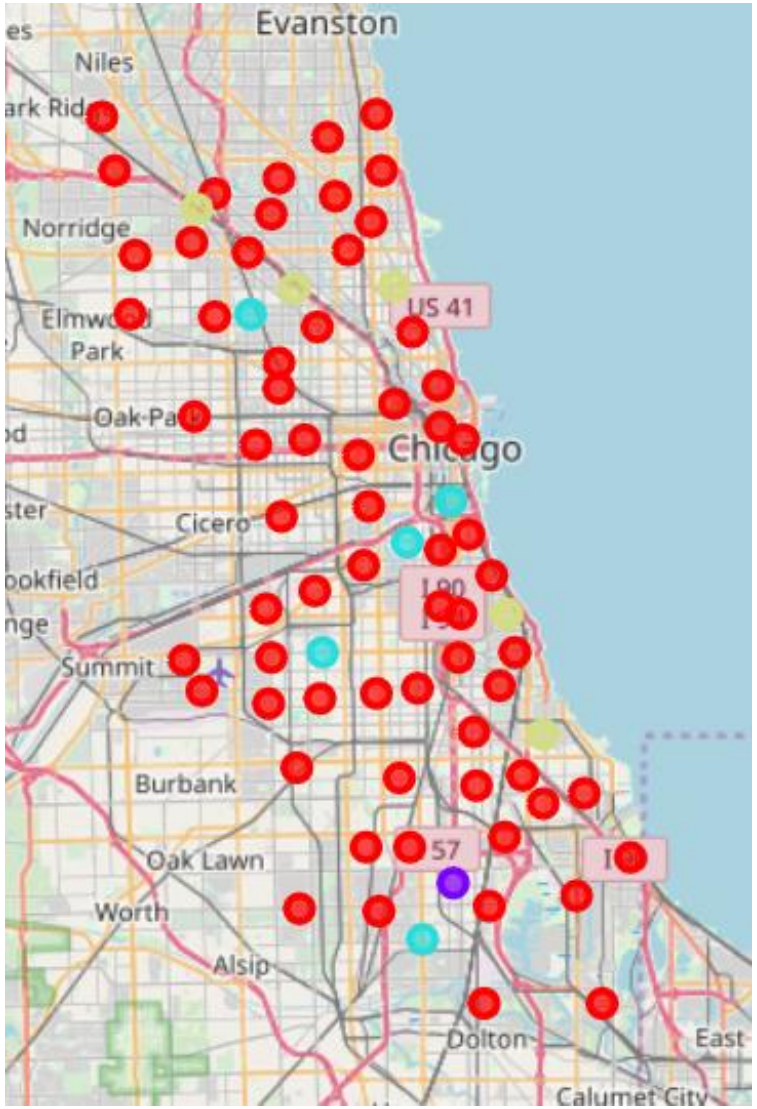
Data

Required Data:

- Names of the 77 neighborhoods in Chicago, Illinois (USA)
- The community number that correspond with each of the 77 neighborhoods
- The coordinate for each neighborhood
- Venue Data for Chinese Restaurants

Methodology

- Web scraping Wikipedia using the *Requests* and *BeautifulSoup4* Python package
- Geolocator to obtain the coordinates for each of the neighborhoods
- Using the coordinates and FourSquare API, obtain the venue data
- Group data by neighborhood and take the mean frequency for each venue category
- Keep only the category of Chinese Restaurant
- Find the optimal K and then perform K-Mean clustering
- Visualize the Clustering
- Compare the clusters and then evaluate the best neighborhoods with the help of the crime rate heat map



Results

Cluster 1 (Red) – Close to no Chinese Restaurants in the area

Cluster 2 (Purple) – Highest number of Chinese Restaurants in the area

Cluster 3 (Blue) – Moderate number of Chinese Restaurants in the area.

Cluster 4 (Yellow) – Low to moderate number of Chinese Restaurants in the area.

Discussions

Chinese restaurants seem to be randomly distributed throughout Chicago with no patterns shown on the map

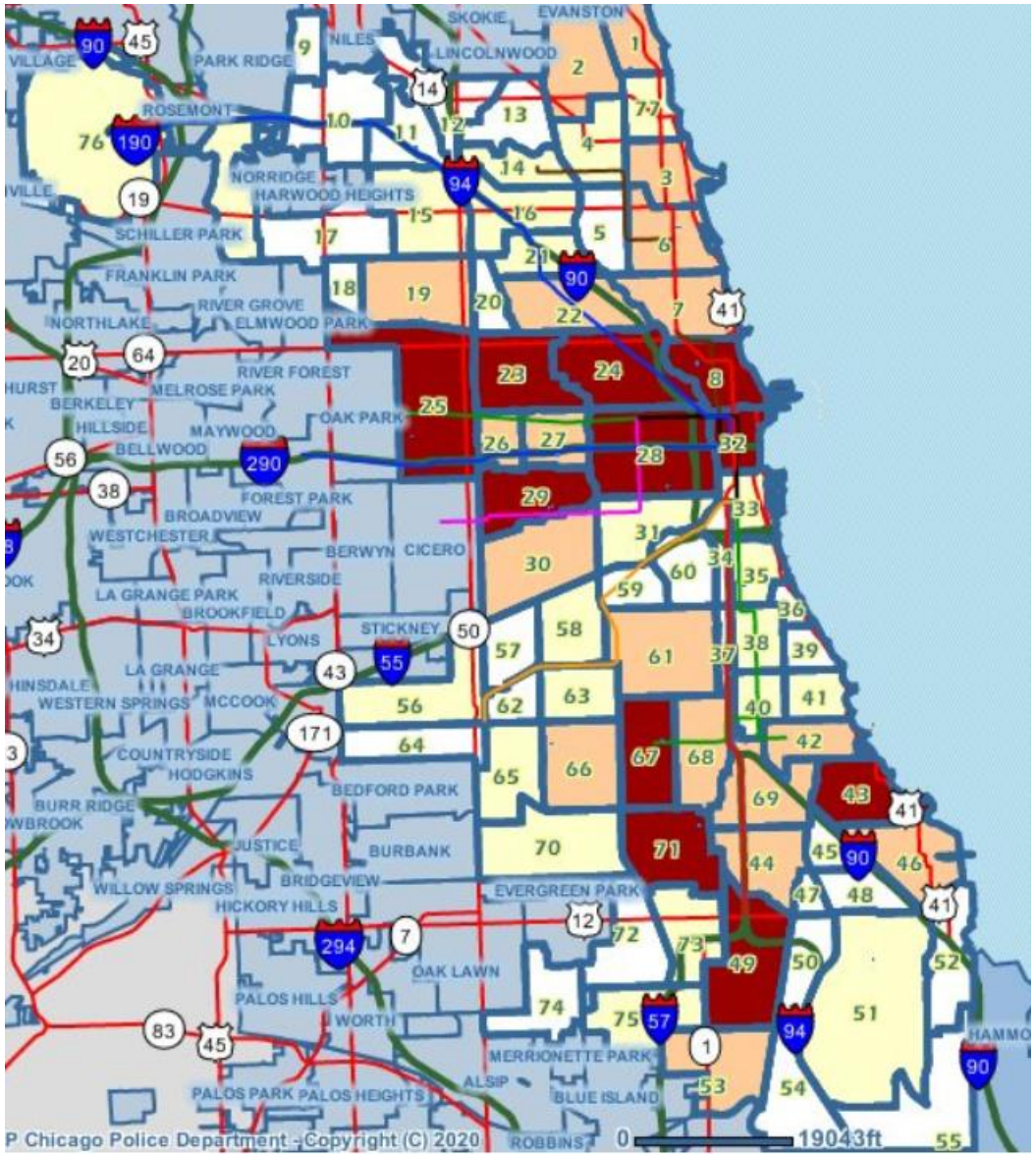
The optimal cluster to choose the neighborhood to begin the location search would be cluster 3 and 4

Cluster 1 have close to no Chinese restaurants in the area but given the long history of Chicago that would mean those area are either not profitable or too dangerous and should be avoided

Cluster 2 have too many Chinese restaurants so opening one there would leave the new restaurant vulnerable to too many competitions

Number	Neighborhood	Chinese Restaurant
21	Avondale, Chicago	0.071429
11	Irving Park, Chicago	0.045455
39	Jefferson Park, Chicago	0.034483
6	Kenwood, Chicago	0.047619
43	South Shore, Chicago	0.052632

Number	Neighborhood	Chinese Restaurant
60	Bridgeport, Chicago	0.105263
63	Fuller Park, Chicago	0.100000
20	Hegewisch, Chicago	0.111111
33	Near South Side, Chicago	0.162500
53	West Pullman, Chicago	0.100000



Conclusion

Best Neighborhoods in terms of competition and safety

- Cluster 3: Bridgeport, Fuller Park, Hegewisch, and Near South Side
- Cluster 4: Avondale, Irving Park, and Jefferson Park.

Future Improvements

- Combine the heat map and clustering map if that is possible
- Create a map of median earning in the neighborhoods for better decision making