



FEBRUARY 1, 2020

# IBM CAPSTONE PROJECT

PICKING A RESTAURANT LOCATION IN CHICAGO, ILLINOIS

BY: XIAN DONG



## Chicago, Illinois (USA)



# **Introduction**

## **Background**

Chicago, also known as the Windy City, is one of the largest cities in the United States of America and is ranked 2<sup>nd</sup> as the most visited city in the country after New York. Chicago is famous for its skyscraper, food and tourist locations. Chicago, unfortunately, is also known for its high crime rate. That is not to say all of Chicago is dangerous, just certain neighborhoods are. Taking that into consideration, starting a restaurant business can be a rewarding investment. Americanized Chinese food is one of the top cuisines consumed in Chicago and there are so many of them opened there. This means two things, one, that it is profitable and two, there is a lot of competitions.

## **Business Problem**

Opening a restaurant comes with great financial risks. Studies have shown that in the first year of a newly opened restaurant around 26% of them fails follow by 19% in the second and 14% in the third year. To reduce the risk of your restaurant failing, it is important to have a good menu, good staffs, and perhaps the most important of all, a good location that serves as your solid foundation. Having a good location is a key point that you can't overlook. While choosing a location on a little-traveled side street or back road might save you on rent, but consequently your business would suffer from the lack of visibility. Setting up the restaurant near an area with a lot of vehicle traffic and foot traffic puts your name out to a lot more people and offers you the opportunity for a lot cheaper marketing. The demographic of the location is also important when designing your menu. Having parking space available could also improve business. At the same time, it is important that the location you choose does not have too many other competitions and that the location is safe. The question we will be answering will be “In which Chicago neighborhood should someone look to open a Chinese Restaurant?”

## **Audience**

Anyone who is looking to start a Chinese restaurant in Chicago but is not sure which neighborhood they should start looking in.

# DATA

## **To solve this problem, we are going to need the following data:**

- Neighborhoods or communities in Chicago
- The community's assigned number
- The coordinates (longitude and latitude) for plotting out the map
- Venue data, in this case just the Chinese restaurants

## **Sources**

- [https://en.wikipedia.org/wiki/Category:Community\\_areas\\_of\\_Chicago](https://en.wikipedia.org/wiki/Category:Community_areas_of_Chicago) for web scraping and obtaining the necessary data frame to get this project started
- [http://gis.chicagopolice.org/CLEARMap\\_crime\\_sums/startPage.htm](http://gis.chicagopolice.org/CLEARMap_crime_sums/startPage.htm) for the crime rate heat map

## **Data Content:**

- Neighborhood – The name of the 77 neighborhoods in Chicago
- Latitude – Latitude of the neighborhood
- Longitude – Longitude of the neighborhood
- Number – Number that was assigned to the neighborhood in Chicago

## Methodology

To start, we need to get the list of Chicago's neighborhoods and their numbers. This can be easily done by web scraping the data from Wikipedia using the Python packages *requests* and *beautifulsoup4*. After turning the data scraped into a data frame, there were some cleaning to do as not all the data scraped were names of the neighborhoods. I blame Wikipedia for this but nothing too much to fix using Python. There is, however, a problem when trying to obtain the numbers that correlate with each individual neighborhood due the lack of public data sets online and my lack of data scraping skills. So, for time efficiency and common sense, the numbers would be put in manually.

Next, we need the coordinates for each of the individual communities in order to find the venues in those area later using the FourSquare API. Using the Python package *Geocoder*, the neighborhoods were converted into coordinates given their latitude and longitude. The coordinates were then stored in a data frame. Plotting them out ensures the coordinates were correct and they are all located in Chicago.

For the next step, the FourSquare API is going to come into play. Using the coordinates inside the data frame, FourSquare API can find venues around that area given a radius. Using Python, a loop was created to get the top 100 venues for all the neighborhoods in Chicago from the FourSquare API. The retrieved data, venue name, venue category, venue latitude and longitude, are turned into a data frame. Then, the neighborhoods are grouped, taking the mean of the frequency for each venue category for clustering. Since we are only looking for Chinese restaurants, we can filter and save the data frame with only the Chinese Restaurant as the venue.

Finally, we apply the machine learning technique, K-Mean Clustering to cluster the neighborhood together base on how frequently there are Chinese Restaurants. The optimal K was chosen using the "elbow" method for the accuracy. The resulting 4 clusters show which of the neighborhoods were suitable as a location candidate base on competition in the area. Next, we need a heat map showing which of the neighborhoods in Chicago are safe. Creating the heat map for the crime rate in Chicago is possible but inefficient since there is already one for it online and so the Python package *IPython.display* will be used to show the heat map that I found.

## Results

The K-mean clustering results in 4 different clusters based on the frequency of occurrence for “Chinese Restaurant”

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

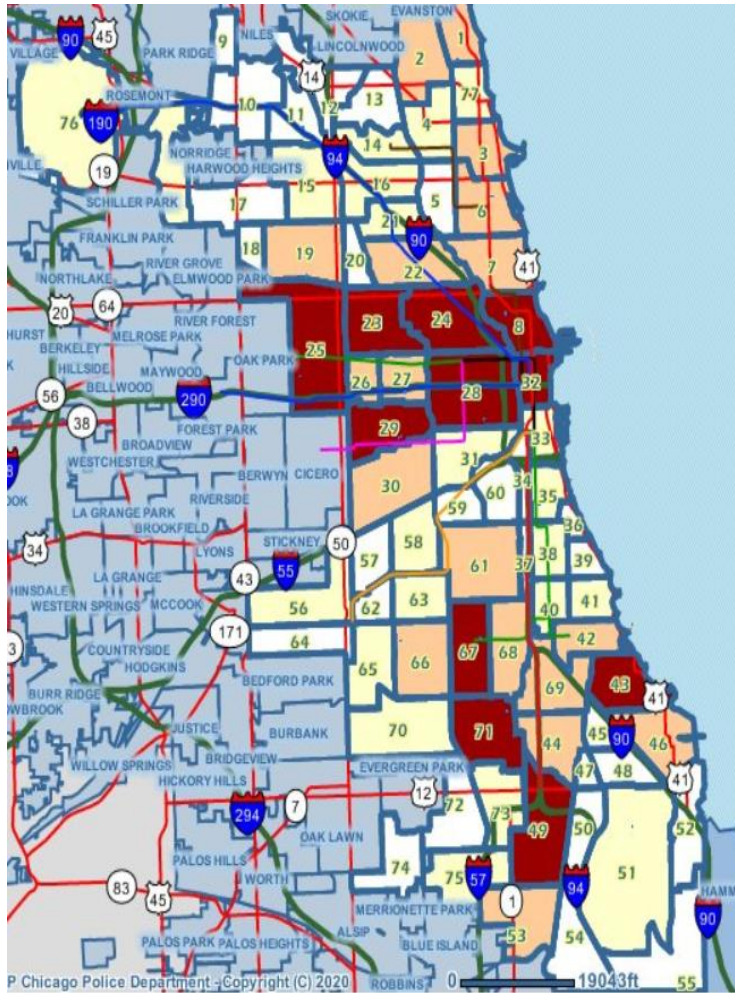
Cluster 2 (Purple) – Highest number of Chinese Restaurants in the area.

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

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







### Cluster 3

| 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           |

### Cluster 4

| 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           |

## Discussion

From the K-mean Clustering map above we can see that the Chinese Restaurants seems to be distributed randomly throughout Chicago. Cluster 2, the purple cluster, is where the most Chinese Restaurants are so opening one there would result in a lot of competitions. Cluster 1 have the least number of Chinese Restaurants so logically those neighborhoods would be the best options to open one. Unfortunately, this is Chicago and there is a reason why there is no Chinese Restaurants opened in those areas. Either those areas are too dangerous or is not profitable. If those areas really are good, then by the long history of Chicago there would be a few Chinese Restaurants there. Yet, there isn't, which should raise a red flag and choose those areas. Cluster 3, blue, and cluster 4, yellow, would be the best neighborhoods to look for a spot. Cluster 3 seem to have a moderate amount already but perhaps there would be a good spot in the neighborhood. Cluster 4 seems to be the better option since there is less competition in those neighborhoods. When deciding the neighborhood to look for the ideal restaurant spot, it is important to look at both the crime rate heat map and the listed neighborhoods in cluster 3 and 4.

## Conclusion

From looking at the Clustering result and the heat map, for **Cluster 3**, I believe the neighborhood with the ideal spot should be in Bridgeport, Fuller Park, Hegewisch, and Near South Side. For **Cluster 4** the best neighborhoods would be Avondale, Irving Park, and Jefferson Park. The results from this project has its limitations and only advise which neighborhood should someone do their search base on established competition in the area and safety. There are other factors involved in choosing a good location, as mentioned before, such as if it is visible, what is the demographic, and if there are perks such as parking. Nevertheless, it shortens the list and made the searching range a lot smaller, saving time.