

# Analysis of Location Choice for Opening New Restaurants in Chicago

## Introduction

Chicago is one of the most populous cities in United States, with an 2018 estimated population of ~2.7 million.[1] It is a central city for finance, commerce, industry, education, transportation and attracts various people to work, and live there. Opening a small business-like restaurant in Chicago will an attractive opportunity for a lot people.

Here, I examine a hypothetical situation to choose location for opening a new business in Chicago. To narrow down the problem, I specifically focus on location choice for starting a new Chinese restaurant. As a common sense, the location is an essential factor of the success or failure of a business and need a comprehensive analysis. This study will be potentially useful for small business owner who want to start a new Chinese restaurant.

## Data Description

The Foursquare API is used to retrieve most common venues in each community area of Chicago.[2] The data will be used to cluster different community areas to identify attractive locations.

The other data I used are mainly from the data portal of city of Chicago (<https://data.cityofchicago.org>)

### (1) Boundaries - Community Areas (current)[3]

This data defines the boundaries of 77 community areas in Chicago. I will use this data to make visualization of different data and clustering of different community areas

### (2) Census Data - Languages spoken in Chicago, 2008 – 2012 [4]

This census data focus of population spoken language and will be used to identity populous areas where Chinese are widely used.

### (3) Crimes - One year prior to present [5]

This data records crimes reported are used to identify safer location to start new business. Furthermore, I use Foursquare API to retrieve most common venues in each community area of Chicago. Based on this, a clustering is performed to shed light on potential good new Chinese restaurant location.

## Methodology

The whole analysis is done in Jupyter Notebook and has been uploaded in the GitHub repository.

In the analysis, I use community area as the basic unit to do the analysis. Chicago has 77 community areas. The community area data is obtained from the census data mentioned

above. The latitude and Longitude data were obtained by using python Geocoder library. The combined community area data frame is made as shown in Figure 1.

	Neighbourhood	Latitude	Longitude
0	Rogers Park	42.00882	-87.66618
1	West Ridge	41.99948	-87.69266
2	Uptown	41.98123	-87.66000
3	Lincoln Square	41.97570	-87.68914
4	North Center	41.95411	-87.68142

Figure 1 Community Area DataFrame

In this study, I use python folium library to make the geographic visualization. In the choropleth map, the Chicago community area geojson file is used to find the boundary of different community areas.

When analyzing the Chinese-spoken population, a choreopleth map is made from the Census Data - Languages spoken in Chicago dataset shown in Figure 2. As we can see, most Chinese spoken population are concentrated in certain community areas in central Chicago.

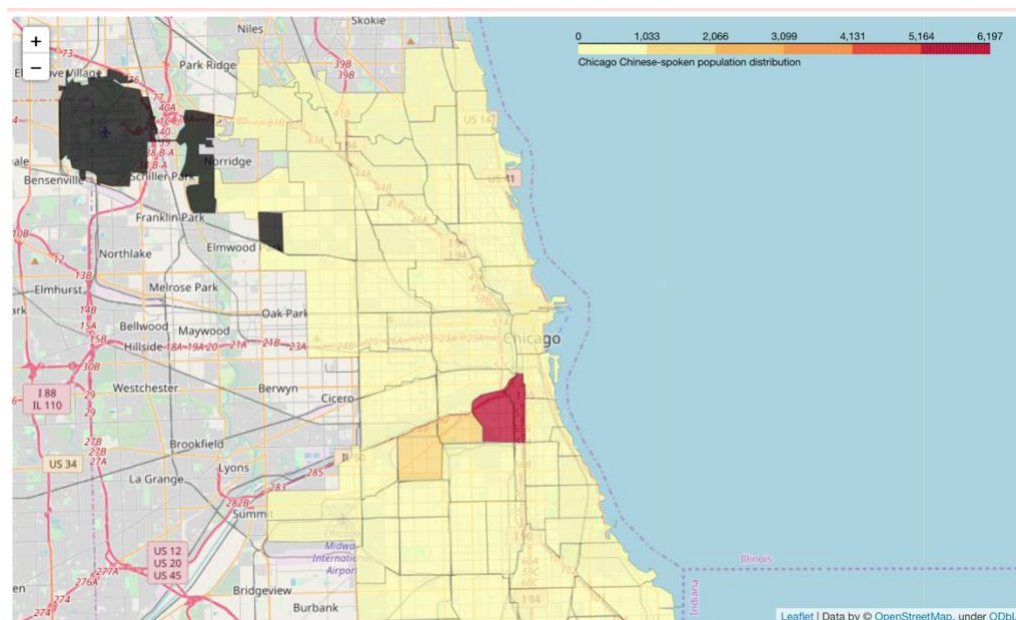


Figure 2 Chinese-spoken population distribution by community area

To study the crime incidents distribution, the crime dataset is first grouped by community area and then made into a choreopleth map. Using this data, we can avoid certain community area when choosing the new business location.

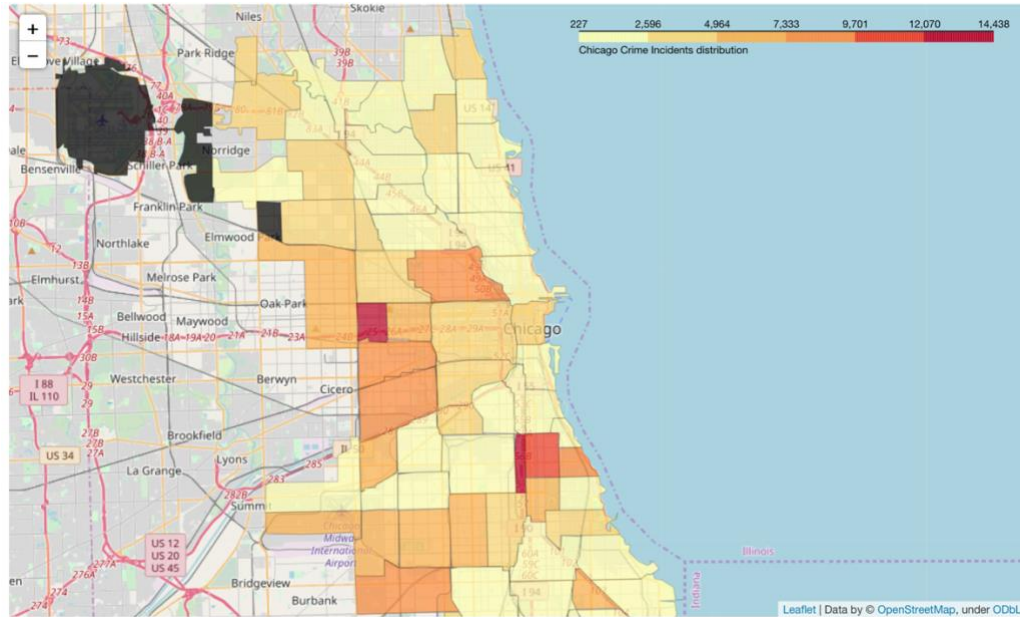


Figure 3 Crime incident distribution by community area

A further clustering was performed to explore the location choice. I used the Foursquare API to explore the community area and segment them. The venue limit is set to 100 and the radius is set to 500 m from the given latitude and longitude of each community area. Then one-hot key encoding is used to transform the data.

In the final clustering, the K-means algorithm is used as it is the most common one and what I have learned from this course. To simplify the analysis, I only use the Chinese Restaurant as the feature for the clustering. The cluster number is set to 5 for simplicity.

## Results

The main result of the clustering is the map shown. Each dot represents a community area and the same color dot are within the same cluster. The red color dot represents the cluster with no Chinese restaurant presence according the Foursquare API. Similar to the Chinese-spoken population, the Chinese restaurant are concentrated in some community areas.

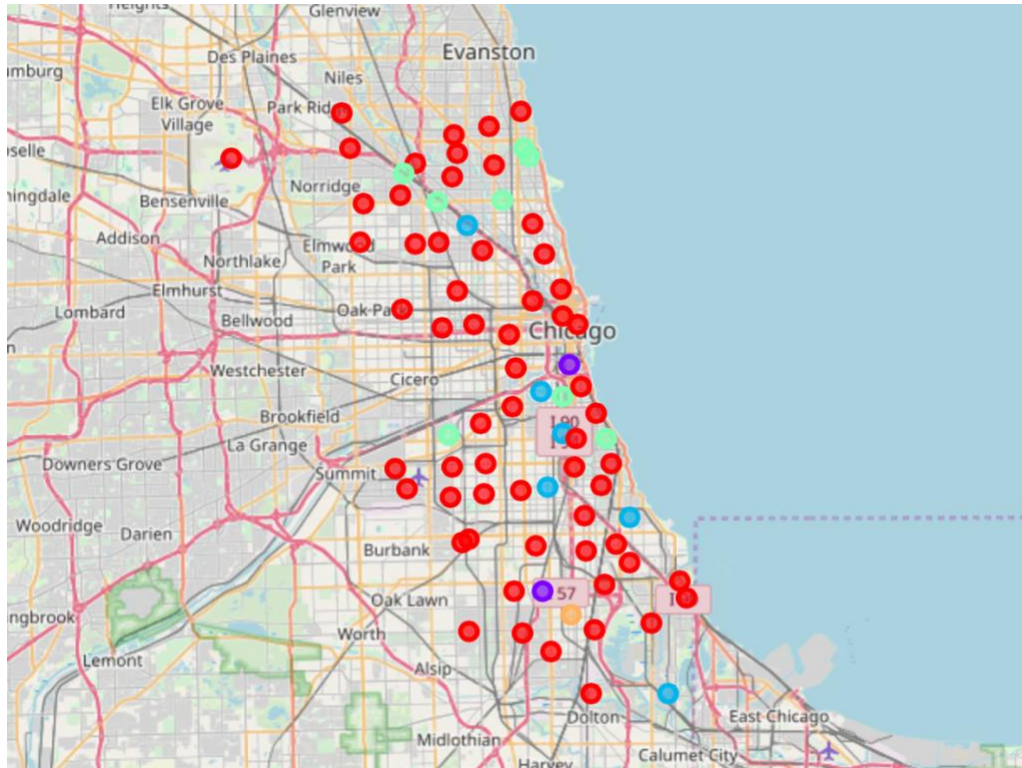


Figure 4 Clustering of Chicago Community Area

Further, an intersection between cluster red community areas (with no Chinese restaurant presence) and the top 10 Chinese-spoken population community area was obtained: McKinley Park, West Ridge, and Brighton Park.

A further check on the crime incidents was performed on the three candidates are performed and all of them are ranking in the middle the crime incidents lists. Therefore, all three make the final recommendation list.

## Discussion

Though Chicago is a diverse city, the Chinese-spoken population are rather small and concentrated in community areas like Bridgeport, Armour Square (where the Chicago Chinatown locates), McKinley Park, and Brighton Park. In this analysis, I assumed that Chinese-spoken people are more likely go to the Chinese restaurant. However, the leave out those who love Chinese food but do not speak Chinese. It is to be decided if this will skew the location analysis.

Crime issue is a concern for people who live in Chicago. In the rough analysis using the total number of crime incidents, I tried to avoid locations that rank high in the crime incident number. In the future study, a crime incident in category (e.g. crime incidents per restaurant by community area) are likely reveal more insights.

In this simply clustering of this study, I only use the Chinese restaurant as the feature. Since competition between different restaurant, a further study can be done with all the restaurants or Asian restaurants.

### **Conclusion**

Following the simple analysis and clustering, I find Mckinley Park, West Ridge, and Brighton Park could be possible community area to open a new Chinese restaurant in Chicago based on Chinese-spoken population, Chinese restaurant clustering, and crime incidents. In the real situation, more factors should be considered, and a more complex analysis should be conducted before the investor can decide.

### **References**

- [1] <https://en.wikipedia.org/wiki/Chicago>
- [2] <https://developer.foursquare.com>
- [3] <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6>
- [4] <https://data.cityofchicago.org/Health-Human-Services/Census-Data-Languages-spoken-in-Chicago-2008-2012/a2fk-ec6q>
- [5] <https://data.cityofchicago.org/Public-Safety/Crimes-One-year-prior-to-present/x2n5-8w5q>