Title  
Analysis of Chicago City for Opening a Restaurant using Machine Learning

# Introduction

An entrepreneur is looking up for opening a restaurant in one of the seventy-seven community areas in the city of Chicago. The entrepreneur has not yet decided on the type of cuisine his restaurant will be serving but need analysis and recommendations based on overview of competition level, socio-economic conditions and suitability of community areas for opening a restaurant in the city of Chicago. We will consider the following factors in order to carry out the analysis:

1. **Population Density/Square Mile:** We have calculated this ratio by dividing population of a community area by the area in square miles of that community area. This ratio will be used to examine how densely populated a community area is. The assumption is that the highly populous a community area is the greater the demand will be for restaurants.
2. **Per Capita Income Ratio:** Per Capita Income is a measure of standard of living of an area.We have calculated the Per Capita Income Ratio by dividing per capita income of a community area by median value of per capita income of all the communities. This ratio will give us indication of how a community area compares in terms of standard of living to other communities in the city of Chicago. The assumption is that the wealth a community is, the greater the demand will be for eating from restaurants.
3. **Restaurant per Thousand of Population:** We have computed this ratio by dividing the total number of restaurants in a given community area by total population of that area multiplied by thousand i.e. number of restaurants available per thousand of people in a community area. This ratio gives us indication of competition among restaurants in a given community area. The assumption is that the greater this ratio is the greater will be the level of competition.
4. **Car Parking per Thousand of Population:** We have calculated this ratio by dividing the number of car parkings in a community area by total population of that area multiplied by thousand. Another way to say this is the number of car parkings available per thousand of population in a given community area. Higher ratio will imply greater number of car parkings available in a given community area. The assumption here is that people tend to prefer to go to restaurants with car parking.
5. **Crimes per Hundred of Population:** We have calculated this ratio by dividing the number crimes in a community area by population of that community area multiplied by hundred i.e. the number of crimes committed per hundred of population in a given community area. Higher ratio will indicate higher crimes rate in a community area. The assumption is that people refrain to visit areas with higher crimes rates, hence, there will be lower demand for restaurants in areas with higher crime rates.

It is tempting to use average rating of restaurants in each community area of the city that would indicate the quality of service offered by restaurants in the respective community and would be indicative of the level of competition an investor would face in the restaurant business in a community area. We tried to obtain rating of restaurants in the city of Chicago using Foursquare API but the API returned large number of restaurants without rating. Hence, we dropped the rating factor from our analysis. Had we included the rating result from Foursquare API, this would have caused noise in the analysis and clustering based on machine learning system that would not be useful for our purpose. Further, obtaining rating of venues from Foursquare is limited to five hundred venues per day only for standard account. There’s cost associated with obtaining rating for more than five hundred venues.

# Data Sources

We obtained data about the number of communities, their population and population density per square mile from Wikipedia**1**. We also obtained names of neighborhood in each community area from Wikipedia**2**. We used this data to determine population density per square and calculate the other ratios as defined in the introduction section above.

We used Chicago government data portal to obtain data about crimes3 and per capita income4 of each community area. We used these data in conjunction with population data to compute Crimes per Hundred of Population and Income per Capita ratios discussed above.

To determine the number of restaurants and car parkings in each community, we used Foursquare’s5 API to explore each community area for restaurants and car parkings. The data from the API was used in conjunction with population data to calculate ratios of Restaurant per Thousand of Population and Car Parking per Thousand of Population as defined in the introduction section.

Appendices

1. <https://en.wikipedia.org/wiki/Community_areas_in_Chicago>
2. <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago>
3. <https://data.cityofchicago.org/api/views/w98m-zvie/rows.csv?accessType=DOWNLOAD>
4. <https://data.cityofchicago.org/api/views/kn9c-c2s2/rows.csv?accessType=DOWNLOAD>
5. <https://foursquare.com/>
6. <https://kite.com/python/answers/how-to-find-the-distance-between-two-lat-long-coordinates-in-python>