**Predicting the best neighborhood in Chicago, United States to start an Indian Restaurant**

**Leeja Bijoy**

**July 2019**

1. **Introduction**

Chicago is the most populous city in Illinois, as well as the third most populous city in the United States. Chicago is an international hub for finance, culture, commerce, industry, technology, telecommunications, and transportation.

Chicago has a large community of Indians working mainly in Information Technology industry.

Chicago has a wide range of fine dine restaurants – Asian, Middle Eastern, Latin , African, American etc. Asian cuisine includes several major regional cuisines: Central Asian, East Asian, North Asian, South Asian, Southeast Asian, and West Asian. South Asian cuisine includes the cuisines from the Indian subcontinent and when included in the definition, also that of Afghanistan. Foods in this area of the world are flavored with various types of chili, black pepper, cloves, and other strong herbs and spices along with the flavored butter and ghee. Turmeric and cumin are often used to make curries.

* 1. **Discussion of the Background**

A client, a successful restaurant chain in India is looking to expand operation into United States. They want to begin with Chicago as it has a large Indian community. They want to create a high-end restaurant that serves all Indian cuisines like Punjabi, Bengali, Rajashtani, Andhra, Tamil, Kerala, Maharashtrian, Goan etc. Their target is not only Indians but all Asians who would love Indian cuisine.

The client requires a deeper insight into the demographics of Chicago so that they can decide on where to establish their first 'Royal Indian' restaurant that would provide a wide variety of Indian cuisine and the ambience of the restaurant would represent variety of Indian cultural attributes.

* 1. **Target Audience**

Chicago being a city where lots of Asians reside, it is very important to the client to decide on a location to set up an Indian that is likely to guarantee a return on investment in Chicago.

Also he wants to know the major factors that contribute to the setup of restaurants in a particular location within Chicago. Understanding the neighborhood thoroughly would help him in taking a decision on the neighborhood where he would set up his restaurant.

1. **Data** 
   1. **Data acquisition and cleaning**

This project will rely on public data from Wikipedia and Foursquare. In this project, data is taken from the wikipedia link <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago> . There are 246 community areas ( neighborhoods) in Chicago as per the above link.

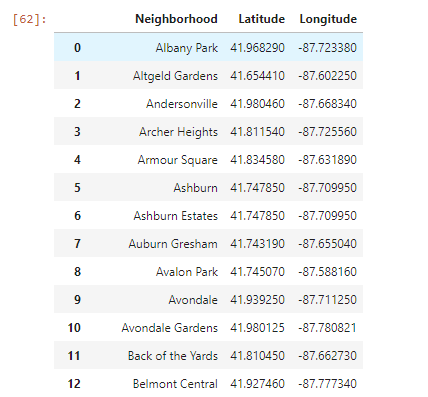
A sample of the data web scrapped of the Wikipedia page for the Chicago Area is provided below:



We are interested only the Neighborhood. Hence the ‘Community Area’ column is dropped.

In obtaining the location data of the locations, the Geocoder package is used with the arcgis\_geocoder to obtain the latitude and longitude of the needed locations.

These will help to create a new dataframe that will be used subsequently for the Chicago area. This data frame has the latitudes and longitudes for neighborhood from the Chicago neighbor hood . Both the dataframes are merged together to get a new dataframe as shown below.



The Foursquare API will be used to obtain the Chicago Area venues for the geographical location data. These will be used to explore the neighborhoods of Chicago accordingly.

The venues within the neighborhoods Chicago like the area’s restaurants and proximity to amenities would be correlated. Also, accessibility and ease of supplies would be considered as it relates to venues. The data from the both the datasets will be explored by considering the venues within different neighborhood of Chicago. These areas' restaurants would be checked in terms of the types of restaurants within a certain mile radius. Due to Foursquare restrictions, the number of venues will be limited to 100 venues. The proximity to transport connection and other amenities would be correlated. Also, accessibility and ease of supplies of ingredients would be considered.

1. **Methodology**
   1. **Data Exploration**

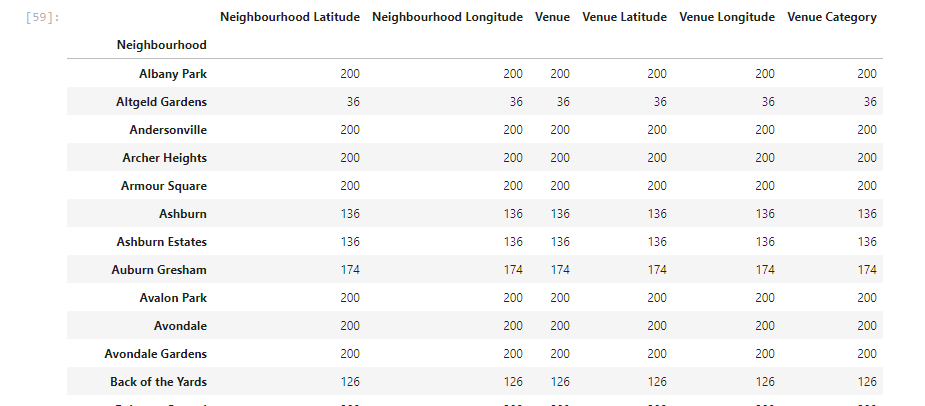
An initial exploration of a single neighborhood within the Chicago area is done to examine the Foursquare workability. 'Albany Park' neighborhood is chosen here. The result is then cleaned up from json to a structured pandas dataframe as shown below:



This data is then grouped to find the unique categories. There are 59 categories in total.



Then multiple neighborhoods are explored. The number of venues returned for each neighborhoods is then explored as follows:

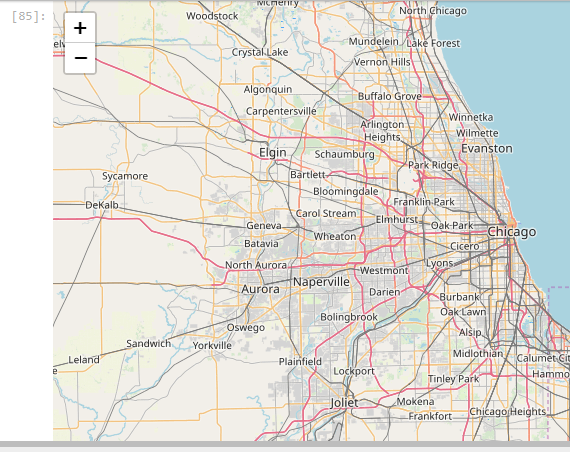


The next step is to check how many unique categories can be returned for the venues. See as follows:

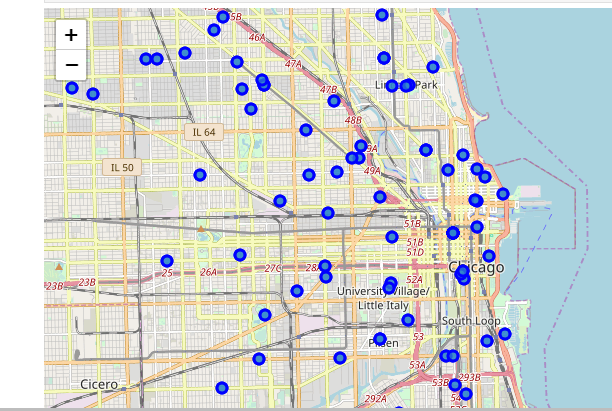


Next step is to cluster the above data. For this section, the neighbourhoods in Chicago will be clustered based on the processed data obtained above.

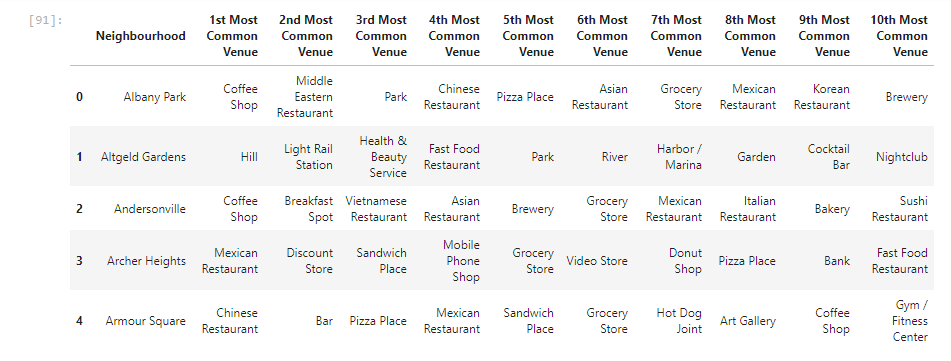
Using the geopy library, the latitude and longitude values of Chicago is obtained. The folium library is then used to obtain the coordinates.   
The Chciago neighbourhoods are then superimposed on top as shown below, still using the folium library.



Add markers to the map



One-hot encoding is done and then Grouping of each Neighbourhoods with 10 common venues is done.

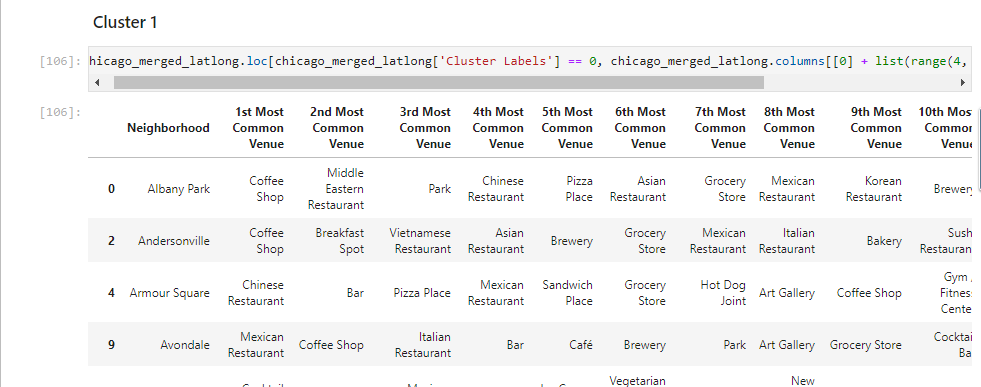


Created clusters of the neighborhood using the k-means to cluster the neighborhood into 5 clusters



Each clusters are then examined

Eg.



**4. Result**

On analyzing 5 clusters it can be observed that:

1. Fast food restaurants, Pizza places and Coffee Shops are popular in the Chciago neighborhoods

2. If we consider restaurants, Mexican resataurants are very popular in Cluster 5

3. Chrysler Village, West Ridge, West Rogers Park, South East Ravenswood and Eden Green already have popular Indian restaurants

4. Cluster 1 has more Asian restaurants

5. Cluster 5 has lots of popular Mexican restaurants

**5. Discussion and Conclusion**

Clusters 3 and 4 are the most feasible options for starting 'Royal Indian' restaurants. They don't have restaurants that serve any kind of Asian cuisines. So including other Asian cuisines like Thai, Japanese, Arab along with Indian menu is also a good option. Neighborhoods like Auburn Gresham, Beverly View, East Beverly, and Oakland in Cluster 3 doesn't have restaurants as 1st most common venue and also has convenience stores, grocery stores, discount stores and farmer's markets in the vicinity. So they can be considered as good options to start Indian restaurants. Cluster 4 has more of hotels and gyms as the most common venues. So it looks more like a business location rather than a location where people would love to hang out and dine out. Considering this, Cluster 3 seems to be the best cluster for the client to explore options of starting an Indian restaurant.