Battle of neighborhoods

Shylaja Koodappan

April 28, 2021

Introduction

Background

Chicago is the third most populous city in United States. Chicago is an international hub for businesses.
The city is also very diverse and provides an umlimited opportunity for expansion of any business. However, the location has to be decided after complete analysis of the data for the business to be successful

Problem

The objective of this project is to find a best location for starting a new Indian Restaurant. Indian restaurants could be a great opportunity given the diversity of the population. This project mainly focusses on the analysis of the geospatial data of Chicago neighborhoods to define a suitable location for new opportunities.
The location is determined using data science technologies and machine learning algorithms to answer the business question - What is the best location in Chicago neighborhoods to start an Indian Restaurant?

Interest

The target audience are the clients that want to open a restaurant business in Chicago area

Data

Data sources

To solve this problem we require the below data

- List of neighborhoods in Chicago
  We use the wikipedia
  page https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago to get the list of neighborhoods in Chicago. We use beautiful soup to scrap the data
- Lattitue and Longitude coordinates to plot the map and determine the nearby venues
  Using python geocoder package, we the get lattitude and longitude coordinates for the list of neighborhoods Folium will be used for map visualization

- Venue data
  FourSquarae API will be used to fetch the nearby venue data

Data processing

Beautiful soup data was used to scrap the neighborhood data from the website and it was observed there were 246 neighborhoods in Chicago.The corresponding coordinates were fetched using geocoder and merged together.

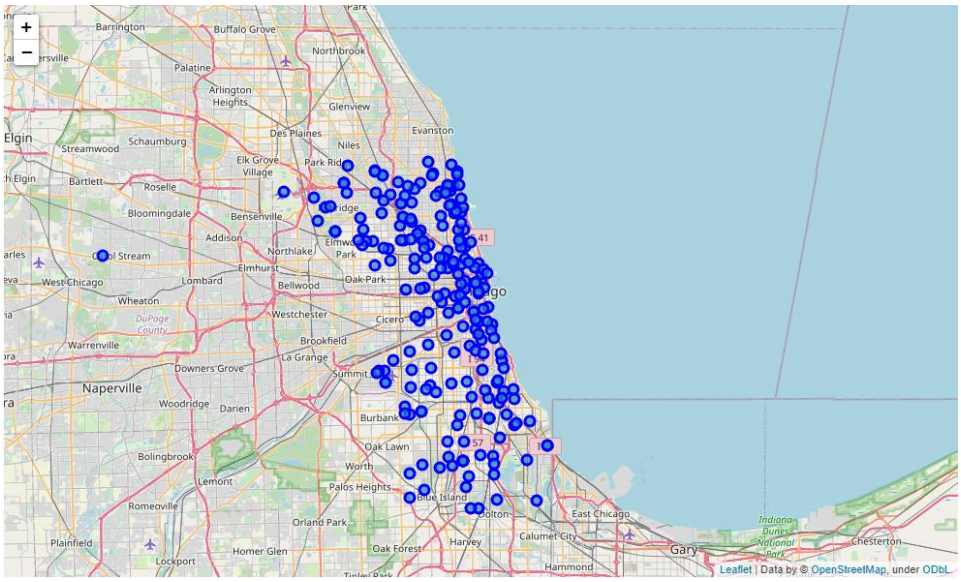
	Neighborhood	Community	Latitude	Longitude
0	Albany Park	Albany Park	41.968290	-87.72338
1	Altgeld Gardens	Riverdale	41.654480	-87.60225
2	Andersonville	Edgewater	41.980460	-87.66834
3	Archer Heights	Archer Heights	41.811540	-87.72556
4	Armour Square	Armour Square	41.834580	-87.63189

The geogrpahical coordinates of Chicago will be passed as input to FourSquare API and the list of venues along with the venue category are retrieved.

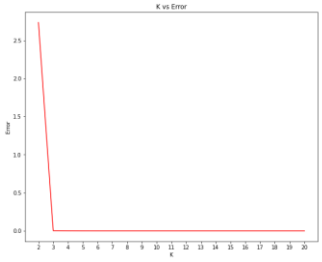
	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Albany Park	41.96829	-87.72338	Lawrence Fish Market	41.968280	-87.726250	Seafood Restaurant
1	Albany Park	41.96829	-87.72338	Starbucks	41.968911	-87.728817	Coffee Shop
2	Albany Park	41.96829	-87.72338	Chicago Kalbi Korean BBQ	41.968314	-87.722771	Korean Restaurant
3	Albany Park	41.96829	-87.72338	Nighthawk	41.967974	-87.713415	Cocktail Bar
4	Albany Park	41.96829	-87.72338	Ssyal Korean Restaurant and Ginseng House	41.968172	-87.733207	Korean Restaurant

Methodology

Below is the map of Chicago that represents the 246 neighborhoods that is visualized using folium. We segment this data and form clusters using K means clustering algorithm, which is an unsupervised algorithm and group these neighborhoods into clusters.



As a first step, elbow plot was created to determine the optimum number of clusters. As observed in the below figure, the optimum number of clusters were 3

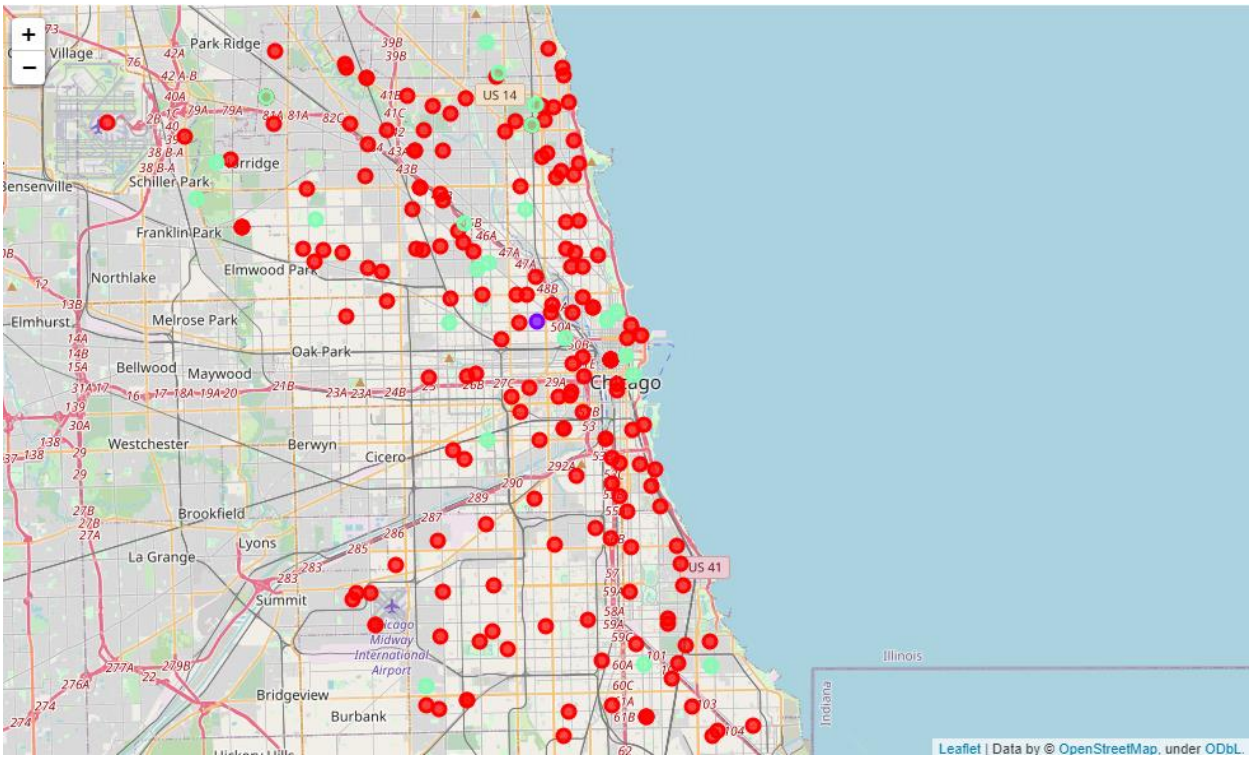


The clusters were now created

	Neighborhood	Indian Restaurant	Cluster Labels
0	Albany Park	0.00	0
1	Altgeld Gardens	0.00	0
2	Andersonville	0.00	0
3	Archer Heights	0.00	0
4	Armour Square	0.00	0
5	Ashburn	0.00	0
6	Ashburn Estates	0.00	0
7	Auburn Gresham	0.00	0
8	Avalon Park	0.00	0
9	Avondale	0.01	2

Results

Here is the visualtion of the map after creating the clusters

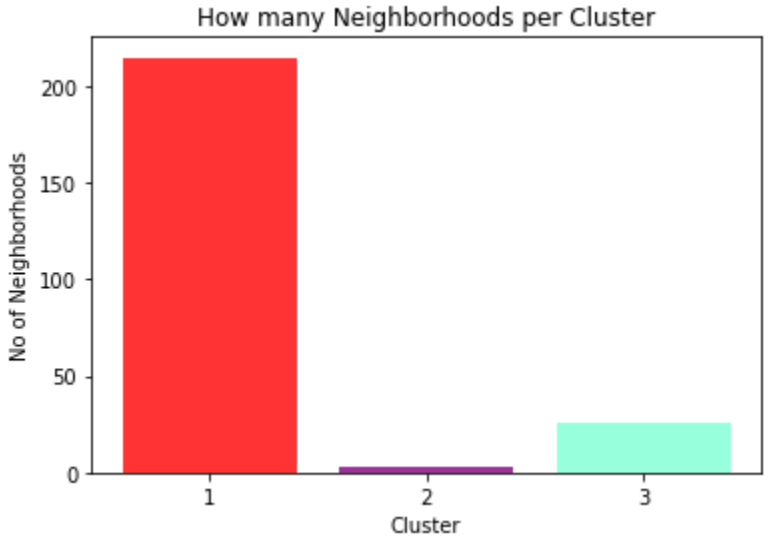


The results from the k means clustering shows that the neighborhoods are categorized into 3 clusters based on the number of Indian Restaurants in each neighborhood.

Cluster 1 represents the neighborhoods with very less or no Indian Restaurants

Cluster 2 represents neighborhoods with high concentration of Indian Restaurants

Cluster 3 represents neighborhoods with medium concentration of Indian Restaurants



Discussion

As noted, cluster 1 has no Indian restaurants and there are around 214 neighborhoods in this cluster. Cluster 2 has more Indian restaurants and Cluster 3 has moderate level

There is a high level of competition in Cluster 2 and the client has to avoid that area.

There are high potentials in Cluster 3 and Cluster 1 where there are no or less Indian Restaurants

Cluster 1

	Neighborhood	Community	Latitude_x	Longitude_x	Indian Restaurant	Cluster Labels	Latitude_y	Longitude_y
0	Albany Park	Albany Park	41.968290	-87.72338	0.0	0	41.968290	-87.723380
1	Altgeld Gardens	Riverdale	41.654480	-87.60225	0.0	0	41.654480	-87.602250
2	Andersonville	Edgewater	41.980460	-87.66834	0.0	0	41.980460	-87.668340
3	Archer Heights	Archer Heights	41.811540	-87.72556	0.0	0	41.811540	-87.725560
4	Armour Square	Armour Square	41.834580	-87.63189	0.0	0	41.834580	-87.631890
...	...	...	...	...	...	...	...	...
210	Wildwood	Forest Glen	42.001350	-87.77537	0.0	0	41.762390	-87.595761
211	Woodlawn	Woodlawn	41.780460	-87.60135	0.0	0	41.910360	-87.682960
212	Wrightwood	Ashburn	41.928979	-87.65619	0.0	0	42.001350	-87.775370
213	Wrightwood Neighbors	Lincoln Park	41.928979	-87.65619	0.0	0	41.780460	-87.601350
214	Wrigleyville	Lake View	41.947250	-87.65320	0.0	0	41.928979	-87.656190

Cluster 2

	Neighborhood	Community	Latitude_x	Longitude_x	Indian Restaurant	Cluster Labels	Latitude_y	Longitude_y
0	Eden Green	Riverdale	42.011757	-87.699985	0.12	1	41.899640	-87.672430
1	West Ridge	West Ridge	41.999480	-87.692660	0.15	1	41.667419	-87.646462
2	West Rogers Park	West Ridge	41.997680	-87.694140	0.15	1	41.679510	-87.641890

Cluster 3

	Neighborhood	Community	Latitude_x	Longitude_x	Indian Restaurant	Cluster Labels	Latitude_y	Longitude_y
0	Avondale	Avondale	41.939250	-87.711250	0.010000	2	41.939250	-87.711250
1	Edgebrook	Forest Glen	41.997320	-87.764230	0.010000	2	42.011757	-87.699985
2	Forest Glen	Forest Glen	41.976400	-87.753610	0.010000	2	41.753151	-87.732240
3	The Gap	Douglas	41.931593	-87.712179	0.010000	2	41.761580	-87.578020
4	Goose Island	Near North Side	41.903285	-87.653001	0.010000	2	41.903950	-87.628860
5	Gresham	Auburn Gresham	41.935956	-87.714608	0.010000	2	41.762060	-87.614580
6	Hollywood Park	North Park	41.989300	-87.710730	0.050000	2	41.653680	-87.546220
7	Illinois Medical District	Near West Side	41.863506	-87.680931	0.010000	2	41.899070	-87.719470
8	Kosciuszko Park	Logan Square	41.930040	-87.724530	0.010000	2	41.681090	-87.607000
9	Legends South (Robert Taylor Homes)	Grand Boulevard	31.226166	-85.418333	0.016949	2	41.885760	-87.624310
10	Loyola	Rogers Park	42.001549	-87.658629	0.020000	2	41.923280	-87.698100
11	Margate Park	Uptown	41.972493	-87.652450	0.010000	2	41.878340	-87.619970
12	Merchant Park	Irving Park	41.927816	-87.706418	0.010000	2	41.852026	-87.699066
13	Noble Square	West Town	41.903400	-87.664510	0.020000	2	41.900340	-87.634330
14	Old Edgebrook	Forest Glen	41.997320	-87.764230	0.010000	2	41.989551	-87.818670
15	Peterson Park	West Ridge	41.985820	-87.728480	0.010000	2	41.920870	-87.704780
16	River West	West Town	41.948815	-87.855969	0.014706	2	46.314930	-88.419690
17	Rogers Park	Rogers Park	42.008970	-87.666190	0.020000	2	41.978400	-87.675073
18	Rosehill	West Ridge	41.986691	-87.672655	0.010000	2	41.948815	-87.855969
19	Sauganash	Forest Glen	41.989970	-87.742270	0.010101	2	41.944590	-87.678490
20	Schorsch Forest View	O'Hare	41.963399	-87.845134	0.014493	2	41.986691	-87.672655
21	Sheridan Station Corridor	Lakeview	41.998250	-87.658110	0.010000	2	41.963399	-87.845134
22	Sleepy Hollow	Garfield Ridge	41.085330	-73.858900	0.013158	2	41.940460	-87.791650
23	Tri-Taylor	Near West Side	41.869320	-87.686080	0.010000	2	39.808120	-75.512510
24	West Town	West Town	41.893290	-87.657430	0.010000	2	41.999480	-87.692660
25	Wicker Park	West Town	41.910360	-87.682960	0.010000	2	41.893290	-87.657430

Conclusion

In this study, I analyzed the the neighborhood data to form clusters based on the number of Indian Restaurants in the nieghborhood. I identified the list of neighborhoods, gathered the georgraphical coordinates and fetched the nearby venues using Foursquare API. I built K means clustering models to categorize the neighborhoods. These models can be very useful in helping the client decide the location for a restaurant in a number of ways. For example, the client the choose a location from the cluster which has less number of restaurants.