December 30th, 2020

OPENING A FRENCH RESTAURANT IN CHICAGO, IL, USA

Aline Campos Camargo

Table of Contents

NOTE: in this report the words VARIABLE and FEATURE are used interchangeably.

Introduction

The owner of a successful French restaurant in New York is interested in opening a similar restaurant unit in Chicago, IL, USA. The first step he would like to take in this direction is to select a location in Chicago, but he does not know where to start. He is requesting some overall guidance on what areas to explore first.

His French restaurant is of medium-to-high price tier ($$$) and is considered more a formal dining restaurant than the average casual, fast food or bar experience. Consequently, this French restaurant is better designed to address fewer guests per day in longer dining experiences than high volumes of clients usually seen in very casual restaurants, fast food venues and bars. Hence, we should look for areas where there is some indication that similar sit-down dining experiences are of interest for local population.

Chicago is a large city with large diversity of people and interests. The city is divided into 77 different communities for statistical studies such as the Census. Some communities directly correspond to neighborhoods, others comprise of two or more neighborhoods.

The objective in this Data Science project is to identify which communities this French Restaurant owner should explore first when looking for a location to open his new restaurant in Chicago, IL, USA.

Data

Data requirements

For this analysis, I wanted to evaluate the differences between the various communities in Chicago by analyzing data related to population size and restaurants present in each of them. More specifically, the following are the variables (or features) I would like to leverage for this analysis:

	Numerical	**Categorical**
Population	Population size Population density	N/A
Ratio population/restaurants	People / Restaurant	N/A
Restaurants	Number of restaurants Price tier (on a 1$ to 4$ scale)	Restaurant categories

Data collection

After some online research, the below Wikipedia page proved to be a good source for three types of information:

Wikipedia page: https://en.wikipedia.org/wiki/Community_areas_in_Chicago

Relevant information:

- Names of Chicago's 77 communities
- Population in each community
- Population density per km2 in each community

The main table in this Wikipedia page, where the above information can be found, also provides duplicate or additional information that is not relevant for this analysis and will be removed after scraping the webpage, which I will explain in the Methodology section.

Continuing the data collection process, we need to get information on existing restaurants in each community and, for that, we were asked to leverage the Foursquare Places API. The following relevant variables are available through the Foursquare API

- Restaurant Names
- Restaurant Categories
- Restaurant Price Tier (on a 1 to 4 scale)

One limitation encountered in this data collection process is that Price Tier is considered a Premium endpoint in Foursquare which would require a Premium subscription, but our Budget for this project does not allow us to proceed with that subscription. So, we will drop this feature from the analysis and use only Restaurant Names and Category.

Finally, after collecting data from both Wikipedia and Foursquare, we can calculate two additional relevant variables:

- Count of restaurants per community
- Ratio average number of people per restaurant

With that, we conclude the process of collecting data and we can then proceed to cleaning, preparing and modelling the data, which will be discussed in the next section.

Methodology

For this data science project, I decided to use the K-means Clustering machine learning modelling technique to group communities with similar features, to then analyze each of them and decide which one the French Restaurant owner should prioritize. To run a Clustering model, I need to prepare a dataset with normalized data, where categorical variables are converted to numerical variables, and numerical variables are normalized to account for differences in scale between all of them.

Our Clustering dataset should contain the following 5 variables for each community:

- Population
- Population Density
- Number of restaurants
- Ratio people/restaurant
- Frequency of each restaurant category

Preparation

I started preparing our data by scraping the table in the Wikipedia page. I then removed both columns with Area information (Area sq mi and Area km2), because we only needed Density (Population divided by Area) for our analysis. Finally, as the table provided Density in two different measurement systems, I decided to only keep the one in Km2, removing the other column.

Next, I wanted to search restaurants in all 77 communities leveraging the Foursquare Places API. To do so, I first needed to get latitude and longitude coordinates for all communities, which I successfully achieved using Geopy.geocoders.

I then proceeded with calling the Foursquare Places API to get IDs, names and categories of all venues returned after calling a search query "Restaurant" across all 77 communities.

The first thing I noticed is that the search did not return many French Restaurants – only 1 in Lincoln Park - which didn't seem realistic. Indeed, a quick search of "French Restaurants" in Chicago, IL in the Foursquare App user interface retrieved 30 results, so it is surprising that the API could not find more French restaurants through a search by community. Trying to resolve this situation would take too much time and considering that the count of French Restaurants per community was not specifically a variable that we wanted to analyze in our Clustering model, I decided to proceed with this resulting dataset but will discuss this limitation in the Discussion section for this report for further exploration.

I also observed that many of the Categories retrieved with the API call where not related to Restaurants or eating venues – such as "Park", "Conference Room", and "Residential Building". I consolidated a list of all non-restaurant categories and removed from the API resulting dataset all venues that fell into this list of categories. Note however that certain categories – such as "Bars", "Bagel Shop", and "Hot Dog Joint" – even though not containing the word "Restaurant", were still kept as part of the analysis because those are places that people go to eat and are relevant to illustrate the type of eating experiences that exists in each community.

The resulting dataset contained 69 different Categories and 64 Communities, which means 13 communities were not retrieved in this process: 10 of them were not found when calling the Foursquare API, and another 3 were eliminated because only listed venues under non-restaurant categories.

Finally, with the Foursquare data cleaned, I could calculate the number of restaurants in each Community, as well as the ratio of people/restaurants in each community, dividing Population from the Wikipedia dataset by the Count of restaurants calculated using the dataset from the Foursquare API call. With that, I concluded the process of gathering all variables in one dataset.

I then needed to normalize our data to make it ready for Clustering.

The variable Restaurant Category is Categorical data and therefore has to be converted to numerical and consolidated by community. I used get_dummies function for transforming categorical into numerical data, and then calculated the frequency of each restaurant category in each community. Consequently, all transformed values, float type, fall in the range 0 to 1.

I then normalized all Numerical variables - Population, Density, Restaurant Count and Ratio People/Restaurant - by dividing each of them by the respective maximum value. Those values also consequently fit the range 0 to 1, with float values.

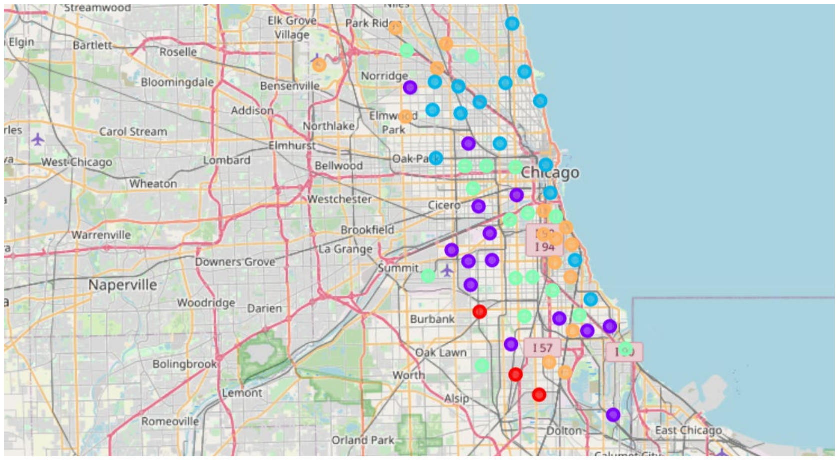
The combination of all normalized data across 5 key features – Population, Density, Restaurant Count, ratio People/Restaurant, and Frequency of Restaurant Categories – created the cluster dataset ready to be fitted in a K-means Clustering model.

Machine Learning Modelling

For this data science project, I used K-Means Clustering machine learning model to group communities into similar clusters.

The key parameter when initiating a K-means Cluster model is the choice of K number of clusters. I ran the model with K values ranging from 3 to 8 and, for each value, did the steps I describe next.

Following each model fitting process with K number of clusters, I looked at the geographical distribution of clusters in a Chicago map, like the one below (with 5 clusters). The advantage of visualizing the size of the clusters in terms of number of communities in each, and also starting to understand where each cluster is located.



Then, for each cluster, I created a dataframe showing the 5 key features (below) for each community, so that I could understand the common characteristics of communities clustered together:

- Population
- Density
- Restaurant Count
- Ratio People/Restaurant
- Frequency of restaurant categories

For this last feature – Frequency of categories – instead of looking at the frequency of all 69 categories, I decided to look at Top 5 categories in each community. That proved to be enough to understand the results calculated by the machine learning model. When reading this information in each cluster, I assumed that if a restaurant category is seen in top 5 categories, it means that people living, working or visiting those communities are interested in that cuisine and therefore sustaining the presence and popularity of that category of restaurants in the area. Otherwise, restaurants under those categories would be shutting down or present as minority in the region.

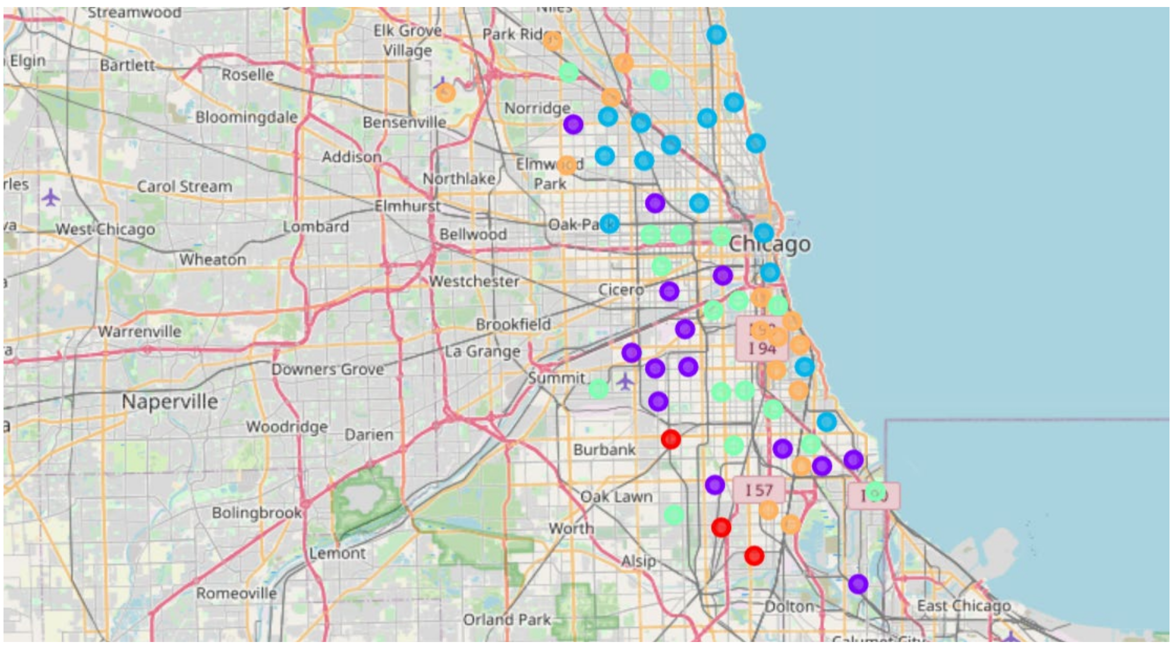
I finally proceeded to analyze and compare the characteristics of each cluster to understand intra-community similarities and inter-cluster differences. For this process of interpretation, I also compared the values in 4 first Numerical features – Population, Density, Restaurant Count and ratio People/Restaurant – to the descriptive statistics for the entire dataset, shown below:

	Restaurant Count	Population	Density (/km2)	People/Restaurant
count	64.000000	64.000000	64.000000	64.000000
mean	9.796875	32471.546875	4676.187500	5781.675779
std	10.486847	20489.634950	2191.099279	5379.199691
min	1.000000	2254.000000	358.230000	492.083333
25%	4.000000	18974.750000	3136.237500	2249.946429
50%	6.000000	27284.500000	4336.920000	4455.000000
75%	11.500000	42747.000000	6309.310000	6941.694444
max	48.000000	95260.000000	11554.110000	27742.000000

I repeated the process described above for Clustering results using different K number of clusters – from 3 to 8 – and found 5 to be the optimal number. The choice of 3 clusters resulted in undefined cluster profiles, mixing communities with very different values across the 5 key features, and a choice of 6, 7 and 8 clusters resulted in more than one cluster with 1 or 2 communities only each, which was not optimal and probably an overfitting problem. Therefore, our optimal K number of clusters was 5 and the results will be detailed in the next session.

Results

After running the K-means Clustering model with K=5, I started by mapping the 5 clusters in the Chicago area, as shown below:



Red dots = Cluster 0, **Purple dots** = Cluster 1, **Blue dots** = Cluster 2, **Green dots** = Cluster 3, **Orange dots** = Cluster 4

The first cluster (Cluster 0) has 3 communities with a high ratio of People/Restaurant compared to the entire set of 64 clustered communities (above 3rd quartile = 6,941 people/restaurant). Those communities are also located far from Chicago downtown and, based on the list of top 5 categories in those communities, there is no indication that a French restaurant would be particularly successful there.

	Community	Cluster Labels	Restaurant Count	Population	Density (/km2)	People/Restaurant	1st Category	2nd Category	3rd Category	4th Category	5th Category
1	Ashburn	0	6	43792	3479.05	7298.666667	American Restaurant	Sports Bar	Hot Dog Joint	Italian Restaurant	Yemeni Restaurant
38	Morgan Park	0	1	22394	2620.11	22394.000000	American Restaurant	Yemeni Restaurant	Food	Deli / Bodega	Dim Sum Restaurant
61	West Pullman	0	1	27742	3008.78	27742.000000	Seafood Restaurant	Yemeni Restaurant	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega

Cluster 1, as shown below, reunites 14 communities where there seems to be high interest for Mexican food, as those are the most frequent categories of restaurants open in the region. It also seems that Caribbean, American and Fast Food are popular cuisines in those locations. I also noticed that most communities (10 out of 14) have Population values above the average of 27,284 people per community, seen across all 64 analyzed communities.

	Community	Cluster Labels	Restaurant Count	Population	Density (/km2)	People/Restaurant	1st Category	2nd Category	3rd Category	4th Category	5th Category
7	Beverly	1	1	20822	2528.12	20822.000000	Mexican Restaurant	Yemeni Restaurant	Fast Food Restaurant	Cuban Restaurant	Deli / Bodega
9	Brighton Park	1	7	44813	6361.18	6401.857143	Mexican Restaurant	Seafood Restaurant	Chinese Restaurant	Breakfast Spot	Deli / Bodega
11	Calumet Heights	1	3	13188	2909.67	4396.000000	Diner	Burger Joint	Mexican Restaurant	Yemeni Restaurant	Fast Food Restaurant
12	Chatham	1	3	31120	4073.05	10373.333333	Mexican Restaurant	Caribbean Restaurant	Yemeni Restaurant	Fast Food Restaurant	Deli / Bodega
15	Dunning	1	8	43689	4534.52	5461.125000	Mexican Restaurant	Eastern European Restaurant	Bar	American Restaurant	Seafood Restaurant
22	Gage Park	1	10	40873	7173.25	4087.300000	Mexican Restaurant	Asian Restaurant	Coffee Shop	Gay Bar	Gastropub
23	Garfield Ridge	1	6	36396	3322.12	6066.000000	Mexican Restaurant	Restaurant	Food	Ethiopian Restaurant	Cuban Restaurant
26	Hegewisch	1	1	9418	693.95	9418.000000	Mexican Restaurant	Yemeni Restaurant	Fast Food Restaurant	Cuban Restaurant	Deli / Bodega
28	Humboldt Park	1	6	56427	6051.83	9404.500000	Mexican Restaurant	Caribbean Restaurant	Taco Place	Breakfast Spot	Fast Food Restaurant
35	Lower West Side	1	21	32888	4333.83	1566.095238	Mexican Restaurant	Food	Restaurant	Taco Place	Food Court
52	South Chicago	1	7	28263	3267.19	4037.571429	Mexican Restaurant	Italian Restaurant	Caribbean Restaurant	Food	African Restaurant
53	South Lawndale	1	16	74851	6296.33	4678.187500	Mexican Restaurant	Chinese Restaurant	Diner	Restaurant	Yemeni Restaurant
57	West Elsdon	1	9	19237	6348.25	2137.444444	Mexican Restaurant	Food	American Restaurant	Chinese Restaurant	Deli / Bodega
60	West Lawn	1	14	33108	4333.24	2364.857143	Mexican Restaurant	Pizza Place	Seafood Restaurant	Food	Chinese Restaurant

The next cluster, number 2, gathers 15 communities with high Population and Population Density, compared to the total averages seen across all 64 communities. It also comprises a set of communities with very diverse cuisines, which indicates that restaurant guests in the area probably appreciate having variety of food and dining experiences. This is also the cluster where European cuisine is more frequently seen in top 5 categories: Greek, Eastern European, Scandinavian, Italian, and Ukrainian cuisines are featured in the top 5 categories. This could be an indication that French cuisine – also European – could be successful in the location.

	Community	Cluster Labels	Restaurant Count	Population	Density (/km2)	People/Restaurant	1st Category	2nd Category	3rd Category	4th Category	5th Category
3	Austin	2	6	95260	5144.07	15876.666667	Southern / Soul Food Restaurant	Greek Restaurant	Breakfast Spot	Food	Taco Place
5	Avondale	2	31	37368	7286.80	1205.419355	Mexican Restaurant	Chinese Restaurant	Food	Vietnamese Restaurant	Diner
6	Belmont Cragin	2	20	79910	7890.90	3995.500000	Mexican Restaurant	Restaurant	American Restaurant	Food	Eastern European Restaurant
27	Hermosa	2	17	24144	7967.57	1420.235294	Mexican Restaurant	Restaurant	Latin American Restaurant	American Restaurant	Pizza Place
29	Hyde Park	2	14	26827	6433.52	1916.214286	Food	Japanese Restaurant	Mexican Restaurant	American Restaurant	Gastropub
30	Irving Park	2	19	54606	6568.06	2874.000000	Food	Pizza Place	Mexican Restaurant	Chinese Restaurant	Mediterranean Restaurant
33	Lincoln Park	2	15	67710	8273.10	4514.000000	Chinese Restaurant	Bagel Shop	Noodle House	Scandinavian Restaurant	Seafood Restaurant
34	Loop	2	46	35880	8395.97	780.000000	Food	Chinese Restaurant	American Restaurant	Italian Restaurant	Restaurant
40	Near South Side	2	48	23620	5123.44	492.083333	Chinese Restaurant	Food	Asian Restaurant	American Restaurant	Cantonese Restaurant
42	North Center	2	19	35789	6740.59	1883.631579	Chinese Restaurant	Thai Restaurant	Latin American Restaurant	American Restaurant	Mexican Restaurant
48	Portage Park	2	10	64307	6285.84	6430.700000	Chinese Restaurant	Restaurant	Italian Restaurant	Mexican Restaurant	Latin American Restaurant
50	Rogers Park	2	20	55062	11554.11	2753.100000	Food	Mexican Restaurant	Chinese Restaurant	American Restaurant	Wings Joint
54	South Shore	2	11	50418	6643.86	4583.454545	Food	African Restaurant	Breakfast Spot	Gay Bar	Mexican Restaurant
55	Uptown	2	47	57973	9648.06	1233.468085	Vietnamese Restaurant	Food	Chinese Restaurant	African Restaurant	American Restaurant
62	West Town	2	25	84502	7123.67	3380.080000	Latin American Restaurant	Mexican Restaurant	Caribbean Restaurant	Ukrainian Restaurant	Food

Cluster number 3 below reunites 17 communities. Data values in Population, Density and People/Restaurant ratio vary across communities, hence I we cannot see a pattern regarding those numerical features. Cuisines in this area seem concentrated around General and American Food, Fast Food and Delis, as well as Cuban and Asian Food. Those cuisines are generally very different from French cuisine, so this is probably not a strong location for a French restaurant.

	Community	Cluster Labels	Restaurant Count	Population	Density (/km2)	People/Restaurant	1st Category	2nd Category	3rd Category	4th Category	5th Category
2	Auburn Gresham	3	8	46278	4739.53	5784.750000	Food	Southern / Soul Food Restaurant	Chinese Restaurant	Mexican Restaurant	American Restaurant
4	Avalon Park	3	5	9985	3084.18	1997.000000	Food	Fast Food Restaurant	Diner	Comfort Food Restaurant	Gay Bar
8	Bridgeport	3	10	33637	6214.03	3363.700000	Food	American Restaurant	Restaurant	Asian Restaurant	Japanese Restaurant
13	Clearing	3	4	25891	3920.22	6472.750000	Food	Pizza Place	Asian Restaurant	Latin American Restaurant	Fast Food Restaurant
14	Douglas	3	3	20781	4862.78	6927.000000	African Restaurant	Café	Food	German Restaurant	Gay Bar
16	East Garfield Park	3	4	19996	4000.26	4999.000000	Food	Diner	Fast Food Restaurant	Cuban Restaurant	Deli / Bodega
17	East Side	3	3	23737	3075.47	7912.333333	Food	Bar	Italian Restaurant	Deli / Bodega	Dim Sum Restaurant
19	Englewood	3	3	25075	3153.59	8358.333333	Food	Restaurant	Ethiopian Restaurant	Comfort Food Restaurant	Cuban Restaurant
25	Greater Grand Crossing	3	4	31766	3454.91	7941.500000	Food	Southern / Soul Food Restaurant	Yemeni Restaurant	Ethiopian Restaurant	Cuban Restaurant
36	McKinley Park	3	7	15767	4317.50	2252.428571	Food	Diner	American Restaurant	Chinese Restaurant	Fast Food Restaurant
39	Mount Greenwood	3	4	19277	2746.45	4819.250000	Food	Breakfast Spot	Italian Restaurant	Mexican Restaurant	Deli / Bodega
41	Near West Side	3	9	62872	4266.26	6985.777778	Food	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Italian Restaurant
43	North Lawndale	3	2	35947	4323.74	17973.500000	Food	Fast Food Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant
44	North Park	3	13	18842	2886.88	1449.384615	Food	Korean Restaurant	Yemeni Restaurant	American Restaurant	Chinese Restaurant
45	Norwood Park	3	7	37089	3276.92	5298.428571	Pizza Place	Diner	Latin American Restaurant	Eastern European Restaurant	Food
58	West Englewood	3	4	29929	3668.46	7482.250000	Food	American Restaurant	Fast Food Restaurant	Yemeni Restaurant	Deli / Bodega
59	West Garfield Park	3	5	17163	5177.09	3432.600000	Food	Taco Place	Coffee Shop	Ethiopian Restaurant	Cuban Restaurant

The last Cluster, number 4, comprises 15 communities where there seems to be a concentration of American, Southern and Fast Food, as well as Chinese, Middle Eastern and African Food. The Population in most of those communities are below average of the entire community dataset. Similarly, the average number of people per restaurant seems generally below average of the entire dataset. This may be an indication that the area is already filled with many restaurants and opening a new dining venue there could be more challenging.

	Community	Cluster Labels	Restaurant Count	Population	Density (/km2)	People/Restaurant	1st Category	2nd Category	3rd Category	4th Category	5th Category
0	Armour Square	4	6	13455	5195.00	2242.500000	Chinese Restaurant	Food	Fast Food Restaurant	American Restaurant	Deli / Bodega
10	Burnside	4	1	2254	1426.68	2254.000000	Comfort Food Restaurant	Fast Food Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant
18	Edison Park	4	1	11605	1635.30	11605.000000	Steakhouse	Yemeni Restaurant	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega
20	Forest Glen	4	4	19019	2294.78	4754.750000	Restaurant	Indian Restaurant	Chinese Restaurant	Yemeni Restaurant	Ethiopian Restaurant
21	Fuller Park	4	4	2439	1326.34	609.750000	Fast Food Restaurant	Italian Restaurant	Diner	Southern / Soul Food Restaurant	Ethiopian Restaurant
24	Grand Boulevard	4	8	22313	4951.20	2789.125000	Southern / Soul Food Restaurant	Chinese Restaurant	BBQ Joint	Restaurant	Caribbean Restaurant
31	Jefferson Park	4	7	26808	4442.33	3829.714286	Chinese Restaurant	Restaurant	Middle Eastern Restaurant	Food	Eastern European Restaurant
32	Kenwood	4	3	17189	6381.45	5729.666667	Chinese Restaurant	Caribbean Restaurant	BBQ Joint	Yemeni Restaurant	Fast Food Restaurant
37	Montclare	4	5	13830	5393.73	2766.000000	Restaurant	Mexican Restaurant	Pizza Place	Eastern European Restaurant	Ethiopian Restaurant
46	O'Hare	4	6	12377	358.23	2062.833333	Bar	Sports Bar	Greek Restaurant	Italian Restaurant	Pub
47	Oakland	4	4	6645	4423.53	1661.250000	Restaurant	Caribbean Restaurant	BBQ Joint	Southern / Soul Food Restaurant	Ethiopian Restaurant
49	Pullman	4	4	6613	1204.38	1653.250000	Mexican Restaurant	Fast Food Restaurant	Diner	Restaurant	Cuban Restaurant
51	Roseland	4	3	42433	3399.06	14144.333333	Chinese Restaurant	Southern / Soul Food Restaurant	Fast Food Restaurant	Gay Bar	Gastropub
56	Washington Park	4	3	11502	2921.68	3834.000000	American Restaurant	Chinese Restaurant	Restaurant	Yemeni Restaurant	Ethiopian Restaurant
63	Woodlawn	4	5	23268	4340.01	4653.600000	Fast Food Restaurant	American Restaurant	Caribbean Restaurant	Diner	Restaurant

From the analysis above, I would recommend the French Restaurant owner to prioritize communities found in Cluster number 2: Austin, Avondale, Belmont Cragin, Hermosa, Hyde Park, Irving Park, Lincoln Park, Loop, Near South Side, North Center, Portage Park, Rogers Park, South Shore, Uptown, West Town.

In this Cluster, the diversity of cuisines and a distinct higher presence of European cuisines not found in the other four clusters, seem to indicate that people living, working and/or visiting the area are interested in various food tastes and therefore more curious to try a new restaurant in the area with European origins. This is also a cluster of communities with large Population and high population Density, which is a promising situation for opening a new restaurant and having better chances of a higher daily occupancy, which is a crucial metric for any restaurant trying to sustain its business.

In all other clusters there was either an indication that a more formal French dining experience would not be successful, with casual and fast dining experiences being at the top 5 restaurant categories, or small Population and Population/Restaurant ratio that could present an occupancy problem for a new restaurant in their locations.

Discussion

This data analysis provides the French Restaurant owner with high-level guidance on what communities to prioritize when searching for a location. We found however a few limitations while collecting and analyzing the data, that should be considered as described below.

First, using Foursquare Places API by community did not return restaurant information for 13 of the 77 communities: Albany Park, Archer Heights, Chicago Lawn, Edgewater, Lake View, Lincoln Square, Logan Square, Near North Side, New City, Riverdale, South Deering, Washington Heights, West Ridge. Maybe working closely with Foursquare to understand this limitation would help address this problem, but we did not have time for that in this analysis. Therefore, collecting and analyzing data about those 13 communities would have to be addressed in the future.

Moreover, the calls to Foursquare API by community surprisingly did not return French Restaurants. After a quick manual search on Foursquare App for the entire city of Chicago, I could find 30 French restaurants listed, but unfortunately when calling the API to retrieve restaurant information by community, only 1 French Restaurant was found. I tried calling the API using “French Restaurant” as query and tested different radius values (500, 1000, 1500, 2000) as well, but still could not get more than 1 French restaurant. While this was not a crucial variable for this high-level clustering analysis, it’s is crucial for any business to understand its direct competition in any location. Therefore, the French Restaurant owner should spend some time understanding where other French Restaurants are located and where his restaurant would have a competitive advantage.

Finally, given that I didn’t have a Budget set up to potentially calling Premium endpoints of the Foursquare API, I had to drop Price Tier as a feature for the Clustering analysis. This is not an essential variable, but it would have been interesting to see the average restaurant Price Tier per cluster as an indication of what restaurants guests in each community are willing to pay.

Conclusion and Next Steps

This analysis successfully identified a cluster of 15 communities in Chicago, IL, USA that the French Restaurant owner should prioritize for next steps in the search for a location for his first restaurant in that city.

As a next step, I would recommend the restaurant owner to, if possible, engage a Research company to conduct quantitative customer research with restaurant guests in those 15 communities, or even with all Chicago communities, to collect and analyze information about restaurant guests' income, average value spent in dining experiences, cuisine preferences, and opinion about existing French restaurants in Chicago. That information would both supplement the analysis presented in this report and, if done across all 77 communities, compensate for missing information about 13 communities and French Restaurants not retrieved when calling the Foursquare Places API. Data collected from this new research could be once again analyzed under a Clustering model or any other modelling technique suggested by the Research company.

The analysis presented in this report, combined with this recommended future customer research, would equip the French Restaurant owner with quality information to make educated decisions about what communities he could more actively start exploring to open a restaurant, looking for a specific location, evaluating space rental or purchase cost, and all other specific restaurant opening steps.