Finding the best spot for opening a new restaurant in Calgary

IBM Data Science Capstone Project, May, 2020

1 Introduction

Opening a new restaurant is a challenge that those who love food and serving it to others take on. There are so many factors leading to a successful restaurant and, location is one the most important ones. A restaurant can have a very delicious food and warmest atmosphere but, in order to maintain a healthy financial status, it should be placed in the best location to attract the largest number of customers possible. In this project I am going to use data science techniques to find the most suitable neighborhoods for opening a new restaurant in Calgary.

1.1 Background

Calgary is the largest city in Alberta with 1.24 million population. It has more than 180 neighborhoods and more than 3000 restaurants. If we compare Calgary with Toronto which has around 16000 restaurants for its 2.93 million population, we can see that there is a lot of room for growth in the restaurant business in Calgary. There are two main categories for neighborhoods in Calgary which are residential and industrial. there could be a case for opening a new restaurant in an industrial neighborhood, but my main focus would be on residential neighborhoods for this project.

2 Data

For this project I have used the following sources :

- Calgary neighborhood data from Wikipedia
- Calgary neighborhood venue data from Foursquare

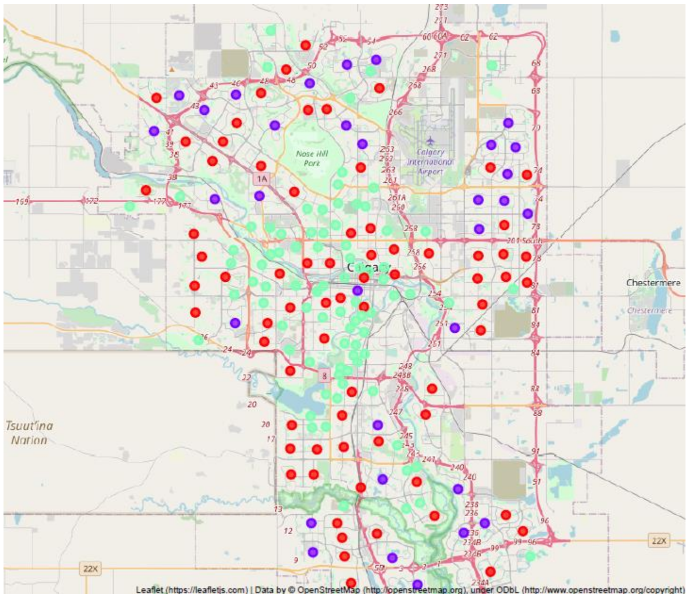
These sources are used to extract data for neighborhoods and restaurants in Calgary. after analysing and modeling the data we will have an insight on the proper locations to open a new restaurant as well as what type of restaurant might have a better chance in each neighborhood.

2.1 Calgary neighborhoods data

Calgary neighborhoods extracted from Wikipedia is presented in the table below:

	Name	Quadrant	Labels	Population_2012	Dwelings	Area	Population_Density	Long	Lat
0	Abbeydale	NE/SE	0	5917	2023	1.7	3480.6	51.058836	-113.929413
1	Acadia	SE	1	10705	5053	3.9	2744.9	50.968655	-114.055587
2	Albert Park	SE	0	6234	2709	2.5	2493.6	51.024400	-113.592200
3	Altadore	SW	0	9116	4486	2.9	3143.4	51.015104	-114.100756
5	Applewood Park	SE/NE	0	6498	2215	1.6	4061.3	51.044658	-113.928930
...	...	...	...	...	...	...	...	...	...
232	Willow Park	SE	0	5229	2282	3.4	1537.9	50.960293	-114.054645
233	Windsor Park	SW	2	4126	2421	1.3	3173.8	51.006165	-114.076187
234	Winston Heights	NE	2	3891	1883	3.0	1297.0	51.072303	-114.047588
235	Woodbine	SW	0	9131	3371	3.2	2853.4	50.942554	-114.128853
236	Woodlands	SW	0	6201	2397	2.8	2214.6	50.942435	-114.109359

After applying K-means clustering on this data the neighborhoods in 5 clusters can be seen on the following map:



2.2 Calgary restaurants data

Calgary restaurants data extracted from Foursquare is presented in the table below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.058836	-113.929413	roadside pub	51.059277	-113.934529	Wings Joint
1	Abbeydale	51.058836	-113.929413	Ginger Garden	51.058900	-113.934600	Chinese Restaurant
2	Altadore	51.015104	-114.100756	Pegasus Restaurant	51.010767	-114.100119	Greek Restaurant
3	Aspen Woods	51.043119	-114.210185	Edo Japan	51.041333	-114.210438	Japanese Restaurant
4	Aspen Woods	51.043119	-114.210185	Diner Deluxe Aspen	51.039636	-114.209193	Restaurant
...	...	...	...	...	...	...	...
395	West Springs	51.058822	-114.204254	Vin Room	51.060566	-114.210780	Tapas Restaurant
396	West Springs	51.058822	-114.204254	Brekkie Cafe	51.060558	-114.209951	Breakfast Spot
397	Whitehorn	51.088808	-113.970020	Little Caesars Pizza	51.092557	-113.966641	Pizza Place
398	Windsor Park	51.006165	-114.076187	Browns Socialhouse Britannia	51.008041	-114.081513	Restaurant
399	Woodlands	50.942435	-114.109359	3 Crowns	50.940765	-114.109430	Pub

3 Methodology

The data gathered from Wikipedia by Beautiful Soup library was used to create a data frame of all Calgary neighborhoods. This data included the name of the neighborhoods, the section of the city for each of them and information about the population and area of each neighborhood.

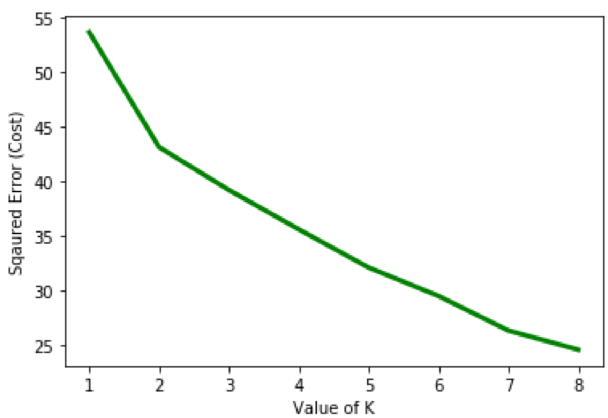
In order to create a map of the city, the longitude and latitude of each neighborhood was extracted using Geocoder. I tried to get all the location information at once but I would get error. To resolve this error, I split the neighborhoods in 20 or 30 batches and I successfully retrieved the data. In the next step I concatenated the two tables containing neighborhood information and location information and created a map using Folium.

Science I am not trying to train the model using target labels I had to use unsupervised learning techniques and for this project I decided to use K-means clustering. The goal is to divide the neighborhoods into separate clusters based on the venue types and also based on the venue count versus population density and finally try to get meaningful insight from the results to be able to help decide the best location for a new restaurant as well as the type of restaurant to open.

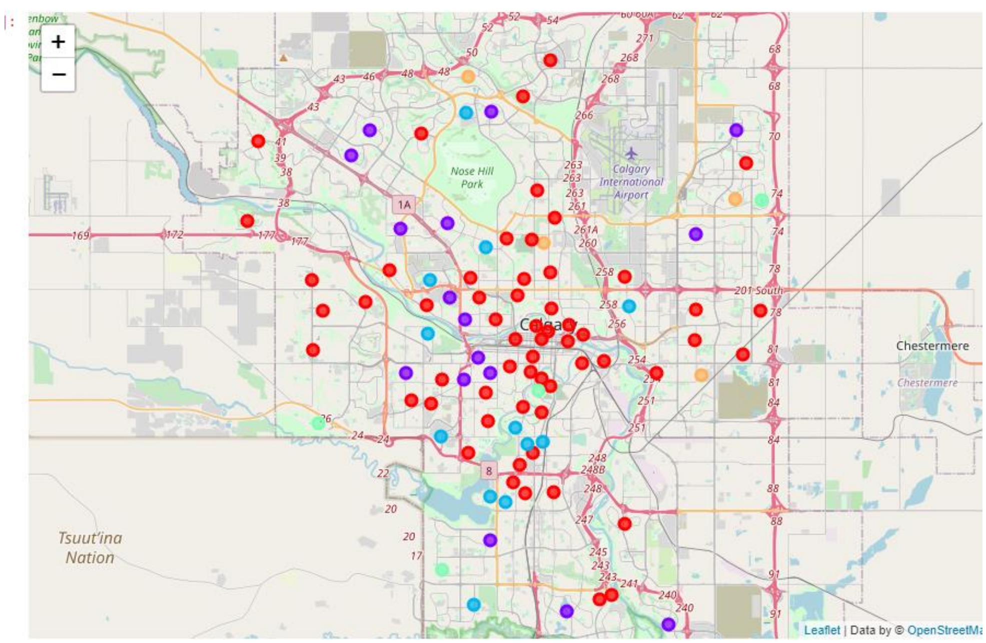
4 Results and Discussion

4.1 Clustering based on restaurant type

Using the cleaned and prepared data and Elbow method the optimum number of clusters is 5.



After applying the K-Means clustering method with 5 clusters the results are as followed:



Cluster0 Cluster1 Cluster2 Cluster3 Cluster4

Cluster0: Cluster zero is the biggest cluster and as we can see in the following table we cant see a common property for this cluster. What we can say about this cluster is that we can see eastern style food venues in this cluster more than the rest of the clusters.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Abbeydale	Wings Joint	Chinese Restaurant	Falafel Restaurant	Hotpot Restaurant
1	Altadore	Greek Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
2	Aspen Woods	Restaurant	Japanese Restaurant	Juice Bar	Mexican Restaurant
3	Banff Trail	Pizza Place	Vietnamese Restaurant	BBQ Joint	Breakfast Spot
5	Beltline	Pub	Bar	Restaurant	Vietnamese Restaurant
6	Bonavista Downs	Chinese Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
8	Bridgeland	Pizza Place	Italian Restaurant	Sushi Restaurant	Breakfast Spot
10	CFB Lincoln Park PMQ	Italian Restaurant	Japanese Restaurant	Seafood Restaurant	Greek Restaurant
12	Cedarbrae	Vietnamese Restaurant	Bar	Wings Joint	Falafel Restaurant
14	Chinatown	American Restaurant	Vietnamese Restaurant	Brazilian Restaurant	Breakfast Spot

Cluster1: the second biggest cluster is the cluster one. We can see a great similarity between the most common restaurants in this cluster which are pizza place, wing joint and Italian restaurant for the first three most common venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
4	Bankview	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant
7	Brentwood	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant
13	Chaparral	Sushi Restaurant	Pizza Place	Wings Joint	Falafel Restaurant
25	Douglasdale	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant
39	Glendale	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant
41	Hawkwood	Chinese Restaurant	Pizza Place	Wings Joint	Fast Food Restaurant
51	Lake Bonavista	Restaurant	Pizza Place	Wings Joint	Falafel Restaurant
60	Millrise	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant
69	Pump Hill	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant
72	Ranchlands	Pizza Place	Wings Joint	Italian Restaurant	Hotpot Restaurant

Cluster2: in this cluster we can see that the most common venue are pubs. The other common venues are wing joints, falafel restaurant and Chinese and Vietnamese restaurants.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
9	Britannia	Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
15	Chinook Park	Pub	Italian Restaurant	Hotpot Restaurant	Hotel Bar
17	Collingwood	Pub	Italian Restaurant	Hotpot Restaurant	Hotel Bar
29	Eagle Ridge	Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
52	Lincoln Park	Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
54	MacEwan Glen	Pub	Sports Bar	Eastern European Restaurant	Hotel Bar
55	Manchester	Pub	Bar	Restaurant	Eastern European Restaurant
57	Mayland Heights	Pub	Vietnamese Restaurant	Eastern European Restaurant	Hotel Bar
67	Parkland	Pub	Fast Food Restaurant	Italian Restaurant	Hotpot Restaurant

Cluster3: the most common venue in this cluster is Indian restaurants. The second and third common venues are falafel and hotpot restaurants.

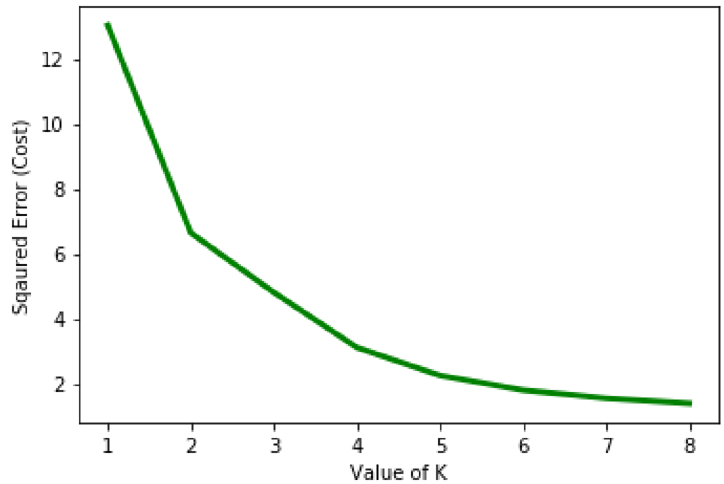
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
11	Cambrian Heights	Indian Restaurant	Falafel Restaurant	Hotpot Restaurant	Hotel Bar
18	Coral Springs	Indian Restaurant	Falafel Restaurant	Hotpot Restaurant	Hotel Bar
24	Discovery Ridge	Indian Restaurant	Falafel Restaurant	Hotpot Restaurant	Hotel Bar
77	Roxboro	Indian Restaurant	Bar	Falafel Restaurant	Hotpot Restaurant

Cluster4: this cluster is the most homogenous cluster and the most common venues respectively are fast food restaurant, falafel restaurant and hot pot restaurants.

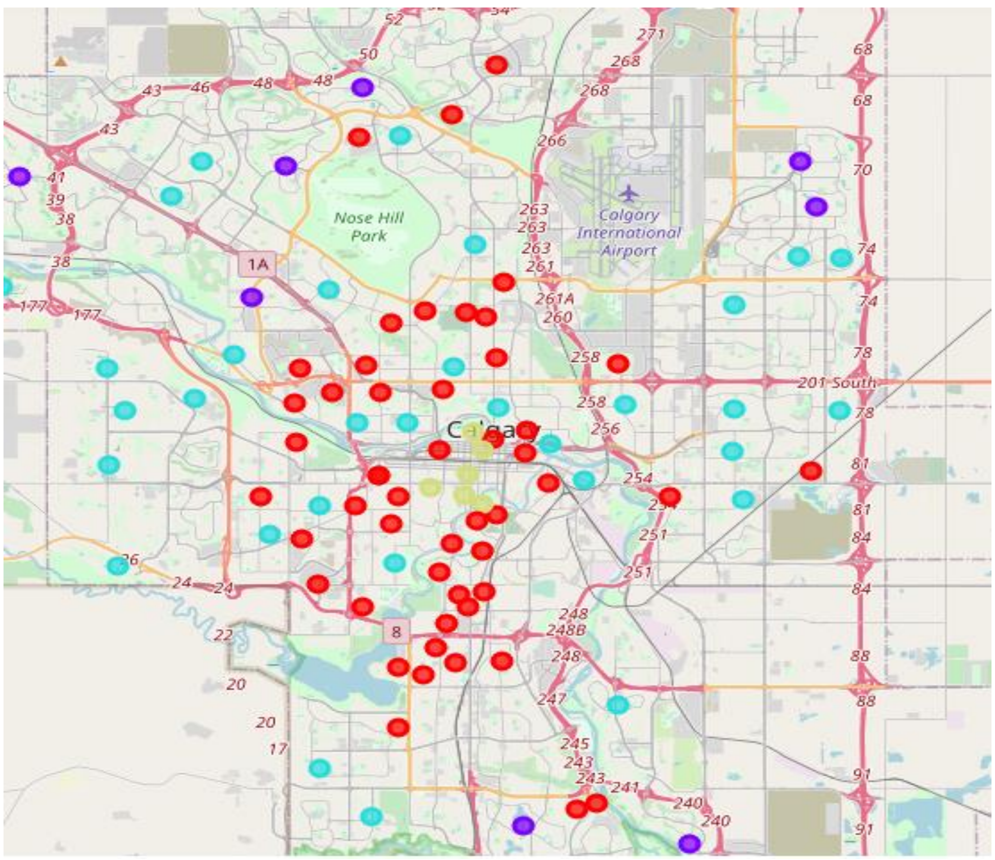
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
35	Falconridge	Fast Food Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
37	Forest Lawn	Fast Food Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
42	Hidden Valley	Fast Food Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant
43	Highland Park	Fast Food Restaurant	Wings Joint	Falafel Restaurant	Hotpot Restaurant

4.2 Clustering based on venue count and population density

Using the cleaned and prepared data and Elbow method the optimum number of clusters is 4.

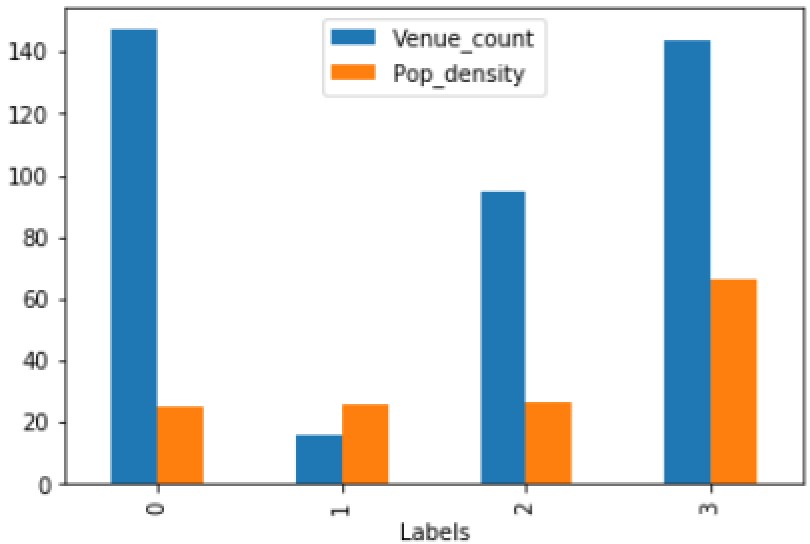


After applying the K-Means clustering method with 4 clusters the results are as followed:



Count_Cluster0 Count_Cluster1 Count_Cluster2 Count_Cluster3

From the modeled data we can see the venue count and population density for each cluster in the following chart:



From the chart we can see that clusters 0 has the biggest venue per population density and clusters 2 and 3 follows cluster 1 in that trend but, cluster 1 has very low number of venues compared to its population density.

5 Conclusion

From the results of *Clustering based on restaurant type* we can conclude that cluster 1 and 2 are saturated with pizza places and pubs respectively and unless the investor is ready to compete with the existing venues they are not suitable to open a pizza place in cluster 1 and a pub in cluster2.Cluster 3 can be defined as Indian food neighborhoods, so if you have great Indian food recipes you might want to try those neighborhoods for your Indian restaurant.

From the results of *Clustering based on venue count and population density* we can conclude that neighborhoods from Count_Cluster0 which are mostly in the inner-city area has lots of venues compared to their population density and opening a new restaurant would be very competitive. Count_Cluster2 is like Count_Cluster0 but, with less venues so the neighborhoods in this cluster might be less competitive.

Count_Cluster3 neighborhoods are all in the downtown area and although they have less venues per population density compared to clusters 0 and 2 but there are lots of day visitors and night life happening in this area which is not included in population density data. Opening a restaurant in this cluster could be a good idea if you can handle the high rents of downtown area and have experience in food industry.

Count_Cluster1 is the outlier between the cluster since it has very few venues per population density. These neighborhoods are all suburban area. From the map we can see that they are dispersed and surrounded by cluster0 and 2 neighborhoods. Any new restaurant in those neighborhoods should appeal to that exact neighborhood and finding that is beyond the scope of this project.