



BATTLE OF NEIGHBORHOODS

A Data Science Project

Olumide Jones

July 19, 2020

IBM Data Science Certification on Coursera

1. Introduction

1.1 BACKGROUND

My client, who is new to the country has asked me to help in deciding where to invest his capital in the province of Alberta in Canada. He is interested in a specialized restaurant business which is similar to what he owes back home in Indonesia. As Alberta is a big province with low population density, it is critical to analyze and decide where a restaurant business will be a success, hence, this project.

As they say in food business, location is KEY. So, project will be looking at best location to open a restaurant in the province.

1.2 BUSINESS PROBLEM

The challenge is to find a suitable neighborhood in any borough in Alberta where a restaurant will thrive. Location of interest will be a fairly densely populated area with few restaurants to minimize the impact of competition. I will analyze and recommend a neighborhood using recently acquired data science skills by exploring Alberta neighborhood datasets and extracting necessary information.

2. Data Acquisition and Cleaning

2.1 DATA SOURCES

Two different data sources were used for the project. Both sources can be found [here](#) and [here](#):

Alberta data from Wikipedia has Alberta's Borough with associated Neighborhood and geo coordinates while the Canadian census data from Canadian government website has each neighborhood by postal code with associated populations.

Foursquare API will be used to explore neighborhoods

2.2 DATA CLEANING

Data downloaded or scraped from the internet were combined into one table. There were lot of missing values that needed to be cleaned up. For example, dataset of Alberta from Wikipedia page has lots of “not assigned” Boroughs.

[4]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T1A	Medicine Hat	Central Medicine Hat	50.036460	-110.679250
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
3	T4A	Airdrie	East Airdrie	51.272450	-113.986980
4	T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413
5	T6A	Edmonton	North Capilano	53.5483	-113.408
6	T7A	Drayton Valley	<u>Not assigned</u>	53.2165	-114.9893
7	T8A	Sherwood Park	West Sherwood Park	53.519	-113.3216
8	T9A	Wetaskiwin	<u>Not assigned</u>	52.9741	-113.3646
9	T1B	Medicine Hat	South Medicine Hat	50.0172	-110.651

Fig. 1: Alberta table read into a data frame showing “Not assigned” in some Borough.

After cleaning up the data sets, I proceeded to merge both data frames (Population and Alberta Neighborhood) and group by Borough and summing up their population to get needed table for analysis. Resulting data frame is shown in figure 2.

[16]:

Borough	Population
Calgary	1249824.0
Edmonton	932650.0
Red Deer	102877.0
Sherwood Park	96569.0
Lethbridge	96277.0
Grande Prairie	72646.0
Fort McMurray	69667.0
Medicine Hat	67569.0
Airdrie	64602.0
Spruce Grove	44071.0
Leduc	31710.0

Fig. 2: Merged data frame showing borough with their respective population

The result above showed that Calgary and Edmonton are the two most populated Borough in Alberta with population of 1.2million and 932K respectively. Since, my stakeholder is interested in a city with relatively large population to support intended business, I have chosen to concentrate on the top 2 borough for my analysis.

Further exploring each borough neighborhood to see their population per neighborhood as shown in figure 3 and 4 confirms they are indeed suitable for locating a new restaurant and satisfy stake holder criteria of fairly populated neighborhood. Top 30 neighborhood in both Calgary and Edmonton has over 10,000 population.

[20]:

	Borough	Neighborhood	Latitude	Longitude	Population
0	Calgary	Sandstone, MacEwan Glen, Beddington, Harvest H...	51.127	-114.0787	80792.0
1	Calgary	Martindale, Taradale, Falconridge, Saddle Ridge	51.0999	-113.9422	77605.0
2	Calgary	Discovery Ridge, Signal Hill, West Springs, Ch...	51.0566	-114.1815	71251.0
3	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.9023	-113.9873	68438.0
4	Calgary	Millrise, Somerset, Bridlewood, Evergreen	50.9093	-114.0721	61344.0
5	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320	59641.0
6	Calgary	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	51.1147	-114.1796	59025.0
7	Calgary	Rundle, Whitehorn, Monterey Park	51.0759	-114.0015	57237.0
8	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158	53224.0
9	Calgary	Queensland, Lake Bonavista, Willow Park, Acadia	50.9693	-114.0514	46394.0
10	Calgary	Midnapore, Sundance	50.9204	-114.0674	43339.0
11	Calgary	Braeside, Cedarbrae, Woodbine	50.9604	-114.1001	42377.0
12	Calgary	Lakeview, Glendale, Killarney, Glamorgan	51.0227	-114.1342	41415.0

Fig. 3: Snapshot of Top 30 Neighborhood population in Calgary (All > 10,000) - Full list in the notebook

[21]:

	Borough	Neighborhood	Latitude	Longitude	Population
0	Edmonton	West Jasper Place, West Edmonton Mall	53.5157	-113.6339	69914.0
1	Edmonton	East Mill Woods	53.4681	-113.4339	51562.0
2	Edmonton	Heritage Valley	53.4129	-113.4957	50904.0
3	Edmonton	Kaskitayo, Aspen Gardens	53.4822	-113.5269	47109.0
4	Edmonton	Horse Hill, East Lake District	53.6026	-113.3837	45947.0
5	Edmonton	Southgate, North Riverbend	53.4839	-113.5227	45145.0
6	Edmonton	East Castledowns	53.6072	-113.5183	41912.0
7	Edmonton	The Meadows	53.4768	-113.3662	37150.0
8	Edmonton	Ellerslie	53.4154	-113.4917	35649.0
9	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413	35049.0
10	Edmonton	West Londonderry, East Calder	53.5923	-113.5168	32242.0
11	Edmonton	West Lake District	53.5966	-113.4882	30715.0
12	Edmonton	Southgate, North Riverbend	53.4839	-113.5227	30171.0
13	Edmonton	West Mill Woods	53.4816	-113.4623	29838.0
14	Edmonton	Southwest Edmonton	53.4967	-113.6162	28187.0
15	Edmonton	NorthDowntown Fringe, East Downtown Fringe	53.555	-113.4822	26290.0

Fig. 4: Snapshot of Top 30 Neighborhood population in Edmonton (All > 10,000) - Full list in the notebook

Lastly, a data frame with just Calgary and Edmonton Borough was created as shown in figure 5.

26	Calgary	Symons Valley	51.1793	-114.1333	15578.0
27	Calgary	Elbow Park, Britannia, Parkhill, Mission	51.0171	-114.0812	14637.0
28	Calgary	Inglewood, Burnsland, Chinatown, East Victoria...	51.0415	-114.0599	14370.0
29	Calgary	Brentwood, Collingwood, Nose Hill	51.0917	-114.1127	14214.0
30	Calgary	Connaught, West Victoria Park	51.0426	-114.0791	13879.0
31	Calgary	City Centre, Calgary Tower	51.0472	-114.0802	13056.0
32	Calgary	Northeast Calgary	51.1494	-114.0019	12483.0
33	Calgary	Highfield, Burns Industrial	50.9857	-114.0631	9372.0
34	Edmonton	West Jasper Place, West Edmonton Mall	53.5157	-113.6339	69914.0
35	Edmonton	East Mill Woods	53.4681	-113.4339	51562.0
36	Edmonton	Heritage Valley	53.4129	-113.4957	50904.0
37	Edmonton	Kaskitayo, Aspen Gardens	53.4822	-113.5269	47109.0
38	Edmonton	Horse Hill, East Lake District	53.6026	-113.3837	45947.0
39	Edmonton	Southgate, North Riverbend	53.4839	-113.5227	45145.0
40	Edmonton	East Castledowns	53.6072	-113.5183	41912.0

Fig. 5: Snapshot of Calgary_edmonton data frame - Full list in the notebook

3. Exploratory Data Analysis

Using the Foursquare API, the top 100 venues for each neighborhood were extracted. As shown in figure 6 below, most of the neighborhood are not up to 100 venues except North Downtown in Edmonton. This is understandable as there is a large concentration of restaurants/fast food/attractions/specialized venues in downtown Edmonton since that is where most of the office population is during the week. The other neighborhoods are fairly busy as well. “Inglewood” for instance has 56 venues reported and “The meadows” has 20 venues reported.

[34]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
North Downtown	100	100	100	100	100	100
Inglewood, Burnsland, Chinatown, East Victoria Park, Saddledome	56	56	56	56	56	56
Connaught, West Victoria Park	43	43	43	43	43	43
Oak Ridge, Haysboro, Kingsland, Kelvin Grove, Windsor Park	37	37	37	37	37	37
City Centre, Calgary Tower	27	27	27	27	27	27
Bridgeland, Greenview, Zoo, YYC	26	26	26	26	26	26
West Lake District	21	21	21	21	21	21
Rosscarrock, Westgate, Wildwood, Shaganappi, Sunalta	20	20	20	20	20	20
West Northwest Industrial, Winterburn	20	20	20	20	20	20
The Meadows	20	20	20	20	20	20
Millrise, Somerset, Bridlewood, Evergreen	16	16	16	16	16	16

Fig. 6: Snapshot of Neighborhood with the highest number of venues

Now, since we are interested in opening a restaurant, it is a good idea to find all venues that are restaurants in the neighborhoods. Figure 7 below shows the top 20 neighborhood with the highest number of restaurants. Just like the figure above with North downtown having highest number of venues, it has the highest number of restaurants as well and “Tuscany” has the lowest number of restaurant.

As shown in figure 3 and 4 above, top 30 neighborhood has population over 10,000 and are suitable to accommodate a restaurant. Since our main targets area are populated area with few competition, opening a restaurant in Edmonton North downtown does not meet the criteria. Likewise, we don’t want to open in a neighborhood with too few restaurants. In this analysis, I will consider a neighborhood with one of the lowest restaurant to venue ratios.



Fig. 7: Top 20 neighborhoods grouped by the number of restaurants

Comparing restaurant and total venues for the top 10 neighborhood (figure 8) shows “**Inglewood, Burnslan, Chinatown, East Victoria Park and Saddledome**” has the lowest restaurant to venue ratio. This means the neighborhood is busy (venues count) but with fewer restaurants (restaurants count).

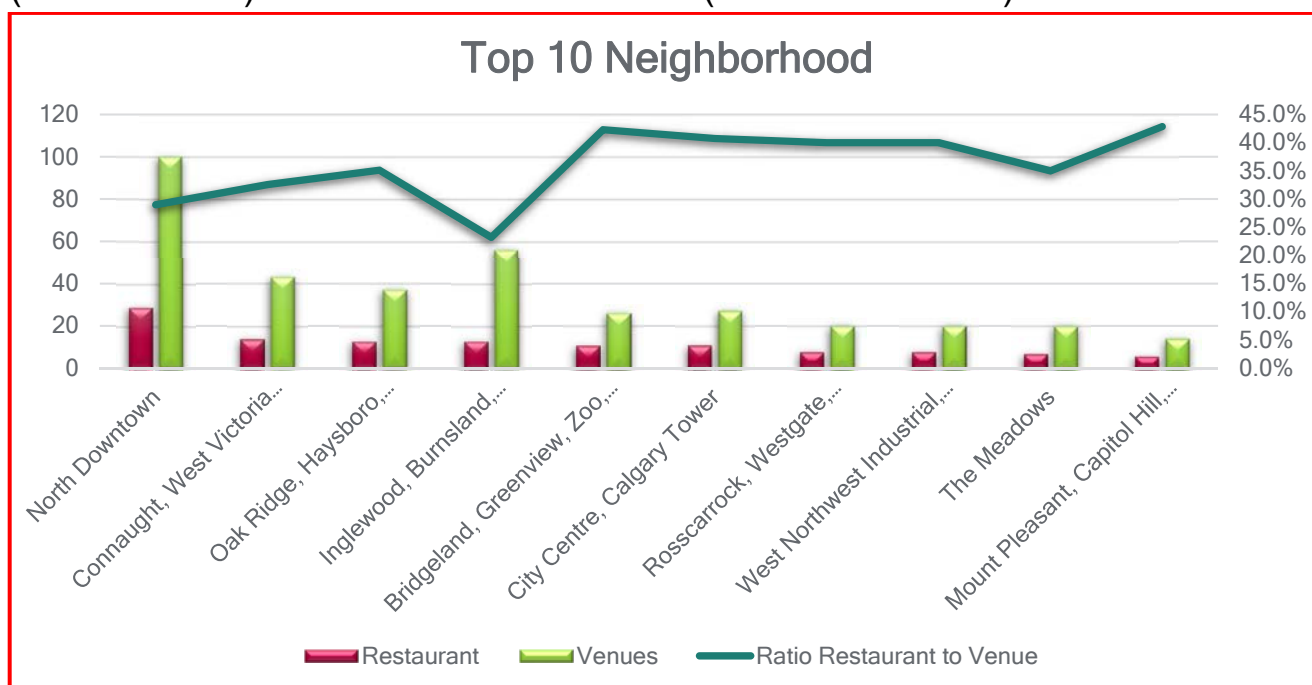


Fig. 8: Top 10 neighborhoods with restaurant to venue ratios

To also determine the most common restaurants in the neighborhood, I looked at the top 5 restaurants type in the neighborhood (figure 9)

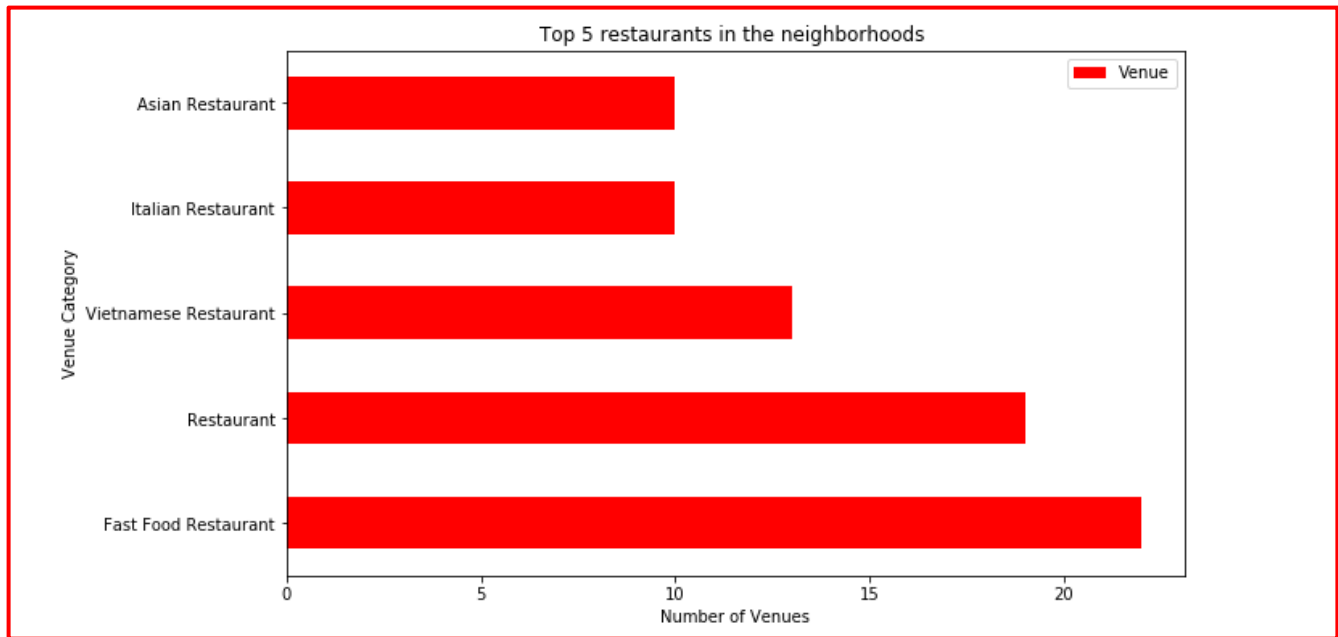


Fig. 9: Top 5 restaurants type in the neighborhood

The chart above shows that fast food restaurants have the greatest number of traffic. 2nd is “restaurant” which is not clearly defined in our data set. I will assume these are not specialized restaurant as our interest is in specialized restaurant. 3rd on the list, Vietnamese seems is the most popular when it comes to specialized restaurants. I performed on hot coding on the datasets, group the resulting data frame by neighborhoods and compute the mean of occurrence for each category as shown in figure 10 below.

[49]:	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Braeside, Cedarbrae, Woodbine	Pharmacy	Hockey Rink	Pizza Place	Convenience Store	Ice Cream Shop	Pub	Gym	Coffee Shop	Donut Shop	Flower Shop
1	Brentwood, Collingwood, Nose Hill	Electronics Store	Yoga Studio	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fast Food Restaurant	Falafel Restaurant
2	Bridgeland, Greenview, Zoo, YYC	Bank	Scenic Lookout	Indian Restaurant	Falafel Restaurant	Sandwich Place	Seafood Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Poke Place	Clothing Store
3	Central Beverly	Construction & Landscaping	Smoke Shop	Grocery Store	Eastern European Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fast Food Restaurant
4	Central Bonnie Doon	American Restaurant	Trail	Liquor Store	Cosmetics Shop	Water Park	Arts & Crafts Store	Electronics Store	Forest	Food Truck	Food Court

Fig. 10: Ranking of the top 10 venues in some neighborhood.

Also, I visualize the top 5 venues for all of the neighborhood. As shown below for few neighborhood including our target neighborhood of “*Inglewood, Burnsland, Chinatown, East Victoria Park and Saddledome*” (figure 11), it is obvious no restaurants in its top 5 venues. This supports the fact that the probability for competition will be very minimal if a new restaurant come into play in the neighborhood.

```
----Highfield, Burns Industrial----
      venue  freq
0   American Restaurant  0.5
1   Athletics & Sports   0.5
2   New American Restaurant  0.0
3   Moroccan Restaurant  0.0
4           Motel        0.0
```

```
----Horse Hill, East Lake District----
      venue  freq
0   Pharmacy  0.33
1   Department Store  0.17
2   Convenience Store  0.17
3   Food & Drink Shop  0.17
4   Coffee Shop  0.17
```

```
----Inglewood, Burnsland, Chinatown, East Victoria Park, Saddledome----
      venue  freq
0   Coffee Shop  0.09
1   Pub          0.07
2   Hotel        0.05
3   Theater      0.05
4   Performing Arts Venue  0.05
```

```
----Kaskitayo, Aspen Gardens----
      venue  freq
0   Lake      1.0
1   American Restaurant  0.0
2   Music Venue  0.0
3   Moroccan Restaurant  0.0
4   Motel      0.0
```

```
----Kensington, Westmont, Parkdale, University----
      venue  freq
0   Park      0.25
1   Carpet Store  0.25
2   Hobby Shop  0.25
3   Furniture / Home Store  0.25
4   American Restaurant  0.00
```

Fig. 11: Top 5 venues per neighborhood highlighting chosen neighborhood of Inglewood.

4. Results and Discussion

From the graph in figure 7, it is evident that North Downtown in Edmonton has the highest number of restaurants count and since our strategy is to reduce competition by looking for a neighborhood with fewer restaurant count, I will suggest the neighborhood of “*Inglewood, Burnsland, Chinatown, East Victoria Park and Saddledome*”.

As shown in figure 8, it has the lowest restaurant to venue ratios amongst the top 10 neighborhood. This ensures we open a restaurant in a fairly populated with little competition on our hand. To buttress above point, figure 11 shows restaurants is not among the top 5 venues in “Inglewood”.

Figure 9 also shown that Vietnamese restaurants are popular aside from fast food and non-specialized restaurants. In chosen neighborhood, Italian is the most popular restaurant and Vietnamese restaurant is not amongst the top 10 venues (Figure 12). Since Vietnamese restaurant is popular as well, it will be good to open one in a busy neighborhood with less competition.

	Borough	Neighborhood	Latitude	Longitude	Population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Calgary	Sandstone, MacEwan Glen, Beddington, Harvest H...	51.127	-114.0787	80792.0	0.0	Pharmacy	Bank	Liquor Store	Italian Restaurant	Coffee Shop	Grocery Store	Pizza Place	Diner	Eastern European Restaurant	Food & Drink Shop
1	Calgary	Martindale, Taradale, Falconridge, Saddle Ridge	51.0999	-113.9422	77605.0	0.0	Arts & Crafts Store	Dog Run	Moving Target	Yoga Studio	Electronics Store	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market
2	Calgary	Discovery Ridge, Signal Hill, West Springs, Ch...	51.0566	-114.1815	71251.0	0.0	Pizza Place	Vietnamese Restaurant	Gas Station	Bar	Convenience Store	Department Store	Eastern European Restaurant	Food Court	Food & Drink Shop	Deli / Bodega
3	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.9023	-113.9873	68438.0	0.0	Pub	Dry Cleaner	Business Service	Bus Stop	Gas Station	Intersection	Bar	Thrift / Vintage Store	Forest	Asian Restaurant
4	Calgary	Millrise, Somerset, Bridlewood, Evergreen	50.9093	-114.0721	61344.0	0.0	Pub	Bank	Sporting Goods Shop	American Restaurant	Bookstore	Health & Beauty Service	Grocery Store	Gas Station	Light Rail Station	Multiplex
5	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320	59641.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	Calgary	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	51.1147	-114.1796	59025.0	0.0	Pizza Place	Miscellaneous Shop	Pub	Yoga Studio	Dry Cleaner	Food & Drink Shop	Flower Shop	Flea Market	Fast Food Restaurant	Falafel Restaurant
24	Calgary	Mount Pleasant, Capitol Hill, Banff Trail	51.0696	-114.0862	21086.0	0.0	Coffee Shop	American Restaurant	Sushi Restaurant	Comic Shop	Pizza Place	Pub	Chinese Restaurant	Fast Food Restaurant	Sandwich Place	Mediterranean Restaurant
25	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.9878	-114.0001	20406.0	0.0	Pizza Place	Convenience Store	Diner	Clothing Store	Dessert Shop	Electronics Store	Food Truck	Food Court	Food & Drink Shop	Flower Shop
26	Calgary	Symons Valley	51.1793	-114.1333	15578.0	0.0	Home Service	Convenience Store	Eastern European Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fast Food Restaurant	Falafel Restaurant
27	Calgary	Elbow Park, Britannia, Parkhill, Mission	51.0171	-114.0812	14637.0	0.0	Construction & Landscaping	Coffee Shop	Japanese Restaurant	Yoga Studio	Eastern European Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market
28	Calgary	Inglewood, Burnsland, Chinatown, East Victoria...	51.0415	-114.0599	14370.0	0.0	Coffee Shop	Pub	Theater	Performing Arts Venue	Hotel	Italian Restaurant	Restaurant	Deli / Bodega	New American Restaurant	Cocktail Bar
29	Calgary	Brentwood, Collingwood, Nose Hill	51.0917	-114.1127	14214.0	0.0	Electronics Store	Yoga Studio	French Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fast Food Restaurant	Falafel Restaurant
30	Calgary	Connaught, West Victoria Park	51.0426	-114.0791	13879.0	0.0	Bar	Coffee Shop	Mediterranean Restaurant	French Restaurant	Restaurant	Pub	Brewery	Middle Eastern Restaurant	Park	Pharmacy

Fig. 12: Top 10 common venues in each neighborhood highlighting chosen neighborhood of “Inglewood” showing Vietnamese restaurant not popular.

5. Conclusion

This project recommend a neighborhood to open a specialized restaurant for a client who recently moved to Alberta. I considered all the borough in Alberta and decided to explore the two most populated boroughs. I considered top 20 neighborhoods with the highest number of restaurants. I identified a neighborhood despite having a high venue count has a low restaurant count. In addition, I identified that Vietnamese restaurants are popular in the city but not amongst the top 10 venues in chosen neighborhood. Therefore, I am recommending my client to open a Vietnamese restaurant in the *Inglewood, Burnsland, Chinatown, East Victoria Park and Saddledome* neighborhood. I believe the neighborhood will provide the least competition and improve the success of the new restaurant. Finally, while it's beyond the scope of my project, I think data on population demographic, income per household and spending power data in each neighborhood will have a significant impact on this project.