

The Battle of the Neighbourhoods – (Week 2)

Table of contents

1. Introduction
 - 1.1 Background
 - 1.2 Business Problem
 - 1.3 Interest
2. Source of Data
3. Methodology
 - 3.1 Analytical Approach
 - 3.2 Exploratory Data Analysis
4. Results
5. Examining the Clusters
6. Discussions
7. Conclusion

1. Introduction

1.1 Background

The city of Calgary is one of the largest municipalities Canada located in the province of Alberta. The city had a population of 1,285,711 in 2019, making it Alberta's largest city and Canada's third-largest municipality.

Calgary's economy includes activity in the energy, financial services, film and television, transportation and logistics, technology, manufacturing, aerospace, health and wellness, retail, and tourism sectors.

The Calgary Metropolitan Area (CMA) is home to Canada's second-highest number of corporate head offices among the country's 800 largest corporations.

With a thriving population and rich economy like this, there is no doubt a restaurant might be a good business venture in the city of Calgary.

1.2 Business Problem

Finding the right location to open a new restaurant can be quite challenging because the investor has to be confident that the proposed restaurant isn't located in a neighbourhood already congested with restaurants where competition may be stiff. Also, the restaurant investor needs to be sure there's a market demand for its proposed cuisine/meals.

In this Capstone project, we will be focussing on using location analytics and machine learning algorithms such as clustering to provide answer these business questions.

1.3 Interest/Target Audience

The outcome of this project benefits an investor who is looking to setup a new restaurant business or expand an existing restaurant value chain.

2. Source of Data

For this analysis, I will be using the "List of neighbourhoods in Calgary" data scraped from Wikipedia (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary).

From the scraped data, there are total 257 neighbourhoods are in Calgary. This data will be trimmed down to two features ("Name" and "Sector") to remove irrelevant data for this analysis.

We will find latitude and longitude of each neighbourhood and cluster them according the restaurants present in each neighbourhood fetched from foursquare location data. Then we will make decision examining each cluster of neighbourhoods.

Sample records from Wikipedia.

	Name[9]	Quadrant	Sector[10]	Ward[11]	Type[10]	2012 PopulationRank	Population(2012) [9]	Population(2011) [9]	% change	Dwellings(2012) [9]	Area(km2) [10]	Populationdensity
0	Abbeydale	NE/SE	Northeast	10	Residential	82	5917.0	5700.0	3.8	2023.0	1.7	3480.6
1	Acadia	SE	South	9	Residential	27	10705.0	10615.0	0.8	5053.0	3.9	2744.9
2	Albert Park/Radisson Heights	SE	East	10	Residential	75	6234.0	6217.0	0.3	2709.0	2.5	2493.6
3	Altadore	SW	Centre	11	Residential	39	9116.0	8907.0	2.3	4486.0	2.9	3143.4
4	Alyth/Bonnybrook	SE	Centre	9	Industrial	208	16.0	17.0	-5.9	14.0	3.8	4.2

The dataframe will be enriched by writing a function to append "Calgary" to each neighbourhood to enhance the chances of looking-up the coordinates of each neighbourhood.

```
calgary_df.head()
```

	Neighborhood	Location
1	Acadia	South, Calgary
2	Albert Park/Radisson Heights	East, Calgary
3	Altadore	Centre, Calgary
4	Alyth/Bonnybrook	Centre, Calgary
5	Applewood Park	East, Calgary

Further enrichment of the dataframe will be performed to include the coordinates (latitude and longitude) for each neighbourhood using the geopy library.

```
[14]: calgary_df.head()
```

```
[14]:
```

	Neighborhood	Location	Latitude	Longitude
1	Acadia	South, Calgary	50.968655	-114.055587
2	Albert Park/Radisson Heights	East, Calgary	51.044845	-113.990195
3	Altadore	Centre, Calgary	51.015104	-114.100756
4	Alyth/Bonnybrook	Centre, Calgary	51.016669	-114.024294
5	Applewood Park	East, Calgary	51.044658	-113.928931

The Calgary demographic data will then be used as an input to the foursquare location data to fetch top 100 restaurants nearby to each neighbourhood within 3,000 meters radius.

```
[47]: calgary_restaurants_unique[calgary_restaurants_unique['Venue Category'].str.contains('Restaurant')].head()
```

```
[47]:
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Abbeydale	51.058836	-113.929413	A&W Canada	51.068291	-113.933571	Fast Food Restaurant
3	Abbeydale	51.058836	-113.929413	Song Huong Vietnamese Restaurant	51.038606	-113.942208	Vietnamese Restaurant
6	Abbeydale	51.058836	-113.929413	McDonald's	51.075787	-113.958094	Fast Food Restaurant
9	Abbeydale	51.058836	-113.929413	KFC	51.064316	-113.957155	Fast Food Restaurant
10	Abbeydale	51.058836	-113.929413	Barrio Fiesta	51.052695	-113.935544	Filipino Restaurant

3. Methodology

The objective of this project is to find the optimal location to open a new restaurant in the city of Calgary.

In this project we will direct our efforts on detecting areas of Calgary that have low restaurant density. We will limit our analysis to area ~3km around city center.

In the first step we scrape the Calgary neighbourhood data from Wikipedia collected the required data.

In the second step we have collected the required data: location and type (category) of every restaurant within 3km radius. We have also filtered restaurants (according to Foursquare categorization).

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighbourhoods /

addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

3.1 Analytical Approach

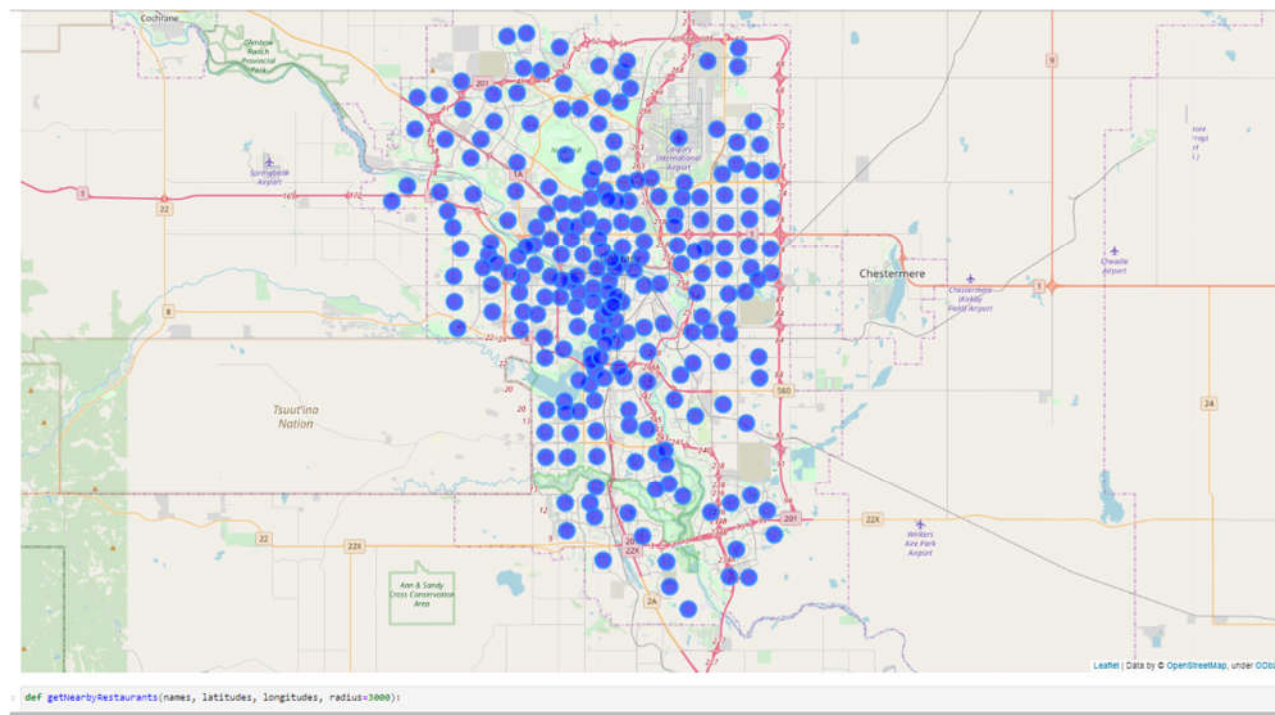
Calgary city has total of 257 neighbourhoods and our goal in this project is to cluster these neighbourhoods based on the similarities between each neighbourhood and their restaurants.

3.2 Exploratory Data Analysis

Calgary Neighbourhood data:

1. We first scrape the neighbourhood information from the Wikipedia and load that into a data frame dropping columns not needed for this analysis.
2. Then we perform data wrangling to convert the data into analysis ready form.
3. Then we use geopy library to get the latitude and longitude information for each neighbourhood
4. Then we use this data to get the 100 nearby restaurants within 3000 meters using foursquare location data.
5. Calculate and perform unique restaurant categories in each neighbourhood and find the top 10 most common restaurants.
6. Then we will cluster all the restaurants in all the neighbourhoods using k-means clustering algorithm.

Calgary city neighbourhood visualization:



Total number of venues in each Neighbourhood:

There are around 49 distinct types of restaurants across all neighbourhoods and total of 860 restaurants.

Out[33]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Abbeydale	15	15	15	15	15	15
Acadia	69	69	69	69	69	69
Albert Park/Radisson Heights	55	55	55	55	55	55
Altadore	58	58	58	58	58	58
Alyth/Bonnybrook	19	19	19	19	19	19
Arbour Lake	32	32	32	32	32	32
Aspen Woods	18	18	18	18	18	18
Auburn Bay	9	9	9	9	9	9
Aurora Business Park	31	31	31	31	31	31
Banff Trail	21	21	21	21	21	21
Bankview	33	33	33	33	33	33
Bayview	12	12	12	12	12	12
Beddington Heights	9	9	9	9	9	9

Find 10 most common restaurants in each neighbourhood:

Top 10 Venues in each neighborhood

In [39]: neighborhoods_venues_sorted.head(10)

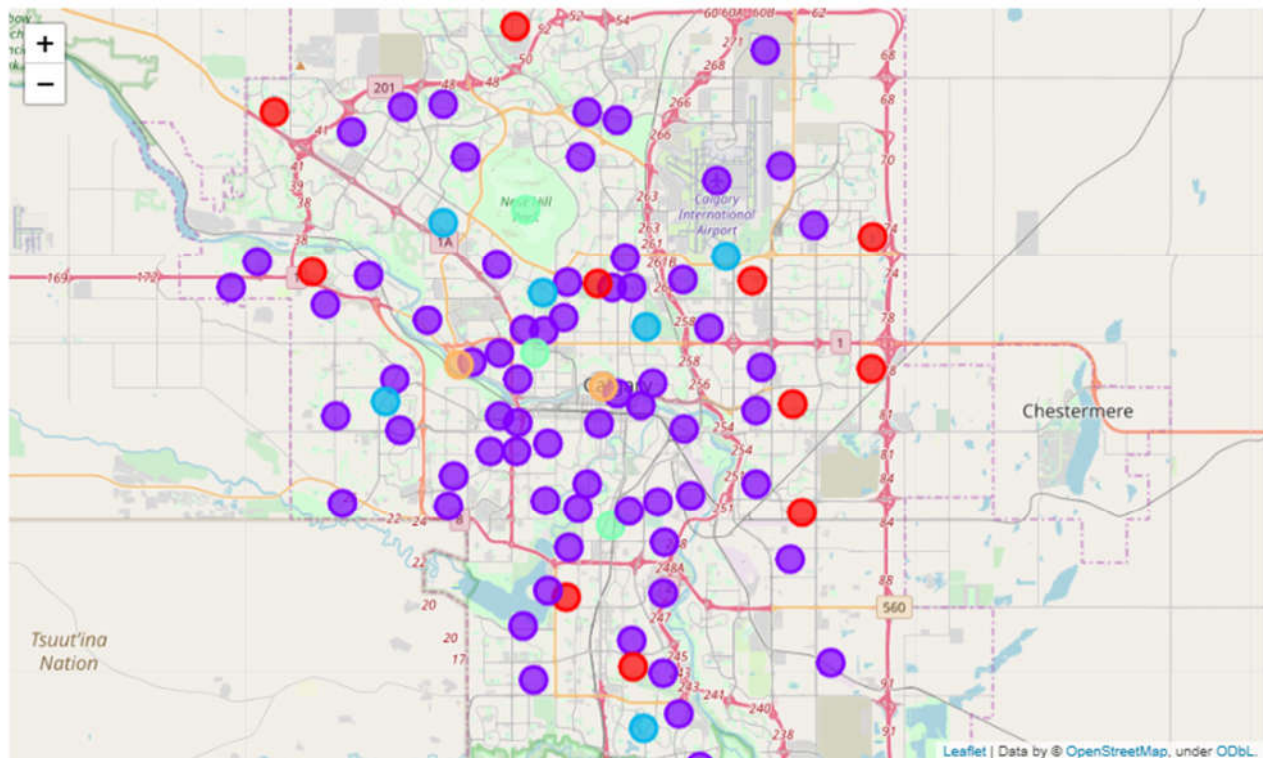
Out[39]:

	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	Abbeydale	Fast Food Restaurant	Vietnamese Restaurant	Italian Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Filipino Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant
1	Acadia	Fast Food Restaurant	Vietnamese Restaurant	Sushi Restaurant	American Restaurant	Restaurant	Italian Restaurant	Mexican Restaurant	Asian Restaurant	Chinese Restaurant	Greek Restaurant
2	Albert Park/Radisson Heights	Vietnamese Restaurant	Fast Food Restaurant	Indian Restaurant	Chinese Restaurant	Restaurant	Mexican Restaurant	Asian Restaurant	American Restaurant	Italian Restaurant	Falafel Restaurant
3	Altadore	Vietnamese Restaurant	Restaurant	Mexican Restaurant	Fast Food Restaurant	French Restaurant	American Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant
4	Alyth/Bonnybrook	Fast Food Restaurant	Restaurant	Vietnamese Restaurant	American Restaurant	Chinese Restaurant	Eastern European Restaurant	Mediterranean Restaurant	French Restaurant	Japanese Restaurant	Empanada Restaurant
5	Arbour Lake	Fast Food Restaurant	Vietnamese Restaurant	Chinese Restaurant	Japanese Restaurant	Greek Restaurant	Mexican Restaurant	Sushi Restaurant	Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant
6	Aspen Woods	Restaurant	Vietnamese Restaurant	American Restaurant	Asian Restaurant	Sushi Restaurant	Japanese Restaurant	Indian Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Fast Food Restaurant
7	Auburn Bay	Sushi Restaurant	Portuguese Restaurant	Italian Restaurant	Restaurant	Seafood Restaurant	Japanese Restaurant	Asian Restaurant	American Restaurant	Brazilian Restaurant	Fast Food Restaurant
8	Aurora Business Park	Fast Food Restaurant	Vietnamese Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Restaurant	Japanese Restaurant	Indian Restaurant	Hong Kong Restaurant	Mediterranean Restaurant
9	Banff Trail	Fast Food Restaurant	American Restaurant	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Restaurant	Asian Restaurant	Greek Restaurant	Indian Restaurant	Mediterranean Restaurant

4. Results

From the foursquare location data, we got 49 restaurants across all neighbourhoods and found most common restaurants. Performing the k-means clustering on the most common restaurants for each neighbourhood will partition the data set into 6 clusters. The 6 clusters are partitioned based on similar type of restaurants that belong to each neighbourhoods.

The folium representation of the clustered data is shown below.



5. Examining the Clusters

We can examine samples from each cluster and determine the discriminating types of restaurants that distinguish each cluster.

Cluster 1:

```
calgary_merged.loc[calgary_merged['Cluster Labels'] == 0, calgary_merged.columns[[0] + list(range(5, calgary_merged.shape[1]))]]
```

	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	Abbeyle	Fast Food Restaurant	Vietnamese Restaurant	Italian Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Filipino Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant
36	Chinook Park	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Vietnamese Restaurant	Falafel Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant
44	Coral Springs	Fast Food Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant
66	Eastfield	Fast Food Restaurant	Mediterranean Restaurant	Sushi Restaurant	Chinese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant
73	Evanston	Fast Food Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Sushi Restaurant	Greek Restaurant	Restaurant	Japanese Restaurant	Falafel Restaurant	Indonesian Restaurant
80	Forest Heights	American Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant
93	Greenwood/Greenbriar	Fast Food Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant
103	Horizon	Vietnamese Restaurant	Fast Food Restaurant	American Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant
160	Queens Park Village	Fast Food Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant
184	Rocky Ridge	Fast Food Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant

Cluster 2:

```
In [98]: calgary_merged.loc[calgary_merged['Cluster Labels'] == 1, calgary_merged.columns[[0] + list(range(5, calgary_merged.shape[1]))]]
```

	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
1	Acadia	Fast Food Restaurant	Vietnamese Restaurant	Sushi Restaurant	American Restaurant	Restaurant	Italian Restaurant	Mexican Restaurant	Asian Restaurant	Chinese Restaurant	Greek Restaurant
2	Albert Park/Radisson Heights	Vietnamese Restaurant	Fast Food Restaurant	Indian Restaurant	Chinese Restaurant	Restaurant	Mexican Restaurant	Asian Restaurant	American Restaurant	Italian Restaurant	Falafel Restaurant
3	Altadore	Vietnamese Restaurant	Restaurant	Mexican Restaurant	Fast Food Restaurant	French Restaurant	American Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant
4	Alyth/Bonnybrook	Fast Food Restaurant	Restaurant	Vietnamese Restaurant	American Restaurant	Chinese Restaurant	Eastern European Restaurant	Mediterranean Restaurant	French Restaurant	Japanese Restaurant	Empanada Restaurant
6	Arbour Lake	Fast Food Restaurant	Vietnamese Restaurant	Chinese Restaurant	Japanese Restaurant	Greek Restaurant	Mexican Restaurant	Sushi Restaurant	Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant
7	Aspen Woods	Restaurant	Vietnamese Restaurant	American Restaurant	Asian Restaurant	Sushi Restaurant	Japanese Restaurant	Indian Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Fast Food Restaurant

Cluster 3:

```
calgary_merged.loc[calgary_merged['Cluster Labels'] == 2, calgary_merged.columns[[0] + list(range(5, calgary_merged.shape[1]))]]
```

	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
42	Collingwood	Chinese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant
52	Dalhousie	Asian Restaurant	Chinese Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant
112	Lake Bonavista	Chinese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant
155	Pegasus	Chinese Restaurant	French Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant
228	Strathcona Park	Chinese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant
254	Winston Heights/Mountview	Chinese Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant

Cluster 4:

```
calgary_merged.loc[calgary_merged['Cluster Labels'] == 3, calgary_merged.columns[[0] + list(range(5, calgary_merged.shape[1]))]]
```

	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
105	Hounsfield Heights/Briar Hill	Japanese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
119	Manchester	Japanese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
145	Nose Hill Park	Japanese Restaurant	Falafel Restaurant	Italian Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant

Cluster 5:

```
calgary_merged.loc[calgary_merged['Cluster Labels'] == 4, calgary_merged.columns[[0] + list(range(5, calgary_merged.shape[1]))]]
```

	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
67	Eau Claire	Italian Restaurant	Vietnamese Restaurant	Falafel Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant
158	Point Mckay	Italian Restaurant	Vietnamese Restaurant	Falafel Restaurant	Indonesian Restaurant	Indian Restaurant	Hotpot Restaurant	Hong Kong Restaurant	Greek Restaurant	Gluten-free Restaurant	French Restaurant

6. Discussions

1. From the data visualization of the clusters we can say that there is an opportunity to open more “Fast Food Restaurant” close to the Nose Hill Park axis and around the Point Mckay neighbourhood since there are none in that neighborhood from the result shown in the cluster.
2. It can also be suggested that opening a restaurant around the Acadia may not be the best idea, since that neighborhood is already densely populated with restaurants.

7. Conclusion

This analysis is performed on limited data. This may be right or may be wrong. But if a richer source of data such as data obtained from google places api is been utilized for this analysis, we are likely to have a scope to come up with better results.

If there are lot of restaurants probably there is lot of demand. The western axis of Calgary and downtown area is saturated with restaurant business this may be due to high demand and opportunities that presents itself.

The northern part of Calgary also has good number of restaurants but not as many as required, so this can be explored further.

As per the neighbourhood or restaurant type mentioned like Indian Restaurant further analysis can be carried out. A venue with lowest risk and competition can be identified.