Predicting an ideal city to migrate, by exploring its neighbourhoods in Canada

Anu Varshini

1. Introduction

1.1 Background

Canada, the largest country of North America and the second largest country in the world has only 35 million inhabitants ,translating to less than 4 persons per square kilometre! In total, roughly 90 percent of the Canadian population consists of immigrants or descendants of those people.

It's no wonder that Canada is the most coveted immigration destination for people around the world. Canada's reputation of being an immigration-friendly nation has also made it easier to move here and explore opportunities, education and quality of life.

Many foreign nationals look to Canada as a land of opportunity, hoping to secure a comfortable life for themselves and their families. However, without proper research one might end up in a place that isn't aligned with their hopes and needs.

1.2 Problem

This project aims to select the most suitable city to live in terms of safety, cost and quality of living, industrial location, and explore the neighbourhoods of that city using the Foursquare API to find the most common venues in each neighbourhood and finally cluster the neighbourhoods using the k-means clustering. It aims to aid migrants, in their choice of location based on amenities that the neighbourhood offers.

2. Data Description

2.1 Data Source

Source 1: List of companies in Canada via Wikipedia

Canada has 162 cities as per Stats Canada's Census 2016. In order to narrow down, lets choose only the top 5 cities that has the maximum number of industrial headquarters located in that city. The dataset contains the following columns:

- Name - Name of the Company
- Industry - Type of the industry
- Sector - Type of the sector
- Headquarters - Name of the city the headquarters is located
- Founded - The year in which the company was founded
- Notes - Additional information about the company

	Name	Industry	Sector	Headquarters	Founded	Notes
0	1-800-GOT-JUNK?	Industrials	Waste & disposal services	Vancouver	1989	Junk removal
1	Norda Stelo	Industrials	Construction & Materials	Quebec City	1963	Integrated projects
2	3Way International Logistics	Industrials	Business support services	Mississauga	2001	Freight forwarding
3	A Buck or Two	Consumer services	Specialty retailers	Vaughan	1988	Retail
4	A&W	Consumer services	Restaurants & bars	North Vancouver	1956	Fast food chain

Figure 1: List of companies in Canada

Source 2: Crime by region-Canada via Wikipedia

Safety is the prime factor to be considered while migrating. For each of the top 5 cities, Violent crime severity index by census metropolitan area is added and sorted by the lowest. The dataset contains the following columns:

- City - Name of the Census Metropolitan area
- Years column from 2009-2016

CMAs in Canada – Violent Crime Severity Index, by year[13][14][15][16][17][18][19][20]

	City	2016	2015	2014	2013	2012	2011	2010	2009
0	Abbotsford–Mission	82.3	90.4	81.1	70.7	79.7	72.4	89.8	118.8
1	Barrie	46.3	43.8	42.3	38.6	46.1	49.2	50.1	53.9
2	Brantford	88.4	70.0	73.5	73.9	67.6	84.5	92.5	91.5
3	Calgary	61.3	72.1	63.0	62.0	61.2	72.1	82.1	84.8
4	Edmonton	102.5	103.9	93.3	89.7	95.8	105.9	106.0	118.7

Figure 2: Crime Severity Index per Metropolitan area

Source 3: Quality of Life in Canada via Numbeo

Quality of Life Index (higher is better) is an estimation of overall quality of life by using an empirical formula which takes into account purchasing power index (higher is better), pollution index (lower is better), house price to income ratio (lower is better), cost of living index (lower is better), safety index (higher is better), health care index (higher is better), traffic commute time index (lower is better) and climate index (higher is better).

Though the dataset contains multiple indicator columns, lets take into account only the quality of life index.

	City	Quality of Life Index
0	Victoria	187.51
1	Quebec City	179.68
2	Calgary	177.05
3	Vancouver	175.97
4	Ottawa	175.32

Figure 3: Quality of Life Index per city

Source 4: List of Postal codes of Canada via Wikipedia

This Wikipedia page contains all the postal codes of the cities along with its neighbourhood. It also provides latitude and longitude data. The dataset contains the following columns.

- Postal Code - Postal code of the Borough
- Borough - Name of the Borough
- Neighborhood - Its neighbourhood
- Latitude - Latitude of the neighbourhood
- Longitude - Longitude of the neighbourhood

	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

Figure 4: List of Postal Codes of Canada

Source 5: Location data via Foursquare API

In order to retrieve information about different venues in different neighbourhoods, we can use the *Foursquare API*. Foursquare data is very comprehensive and it powers location data for Apple, Uber etc. It returns a list of recommended venues near the current location that is specified. It supports both userless and user authentication. Here we will be using a userless authentication where we have to specify a valid Client ID and Secret in the query string of each request.

The request parameters are

- Lat, Long - Latitude and Longitude of the place to be queried
- Radius - Radius to search within, in meters
- Limit - Number of results to return
- Client ID - unique access token
- Client Secret - unique access token

The responses of the API which we require are the venue's name, its category name and its lat and long values. Based on this we will perform k-means clustering to group similar neighbourhoods.


```
{'reasons': {'count': 0,
  'items': [{'summary': 'This spot is popular',
    'type': 'general',
    'reasonName': 'globalInteractionReason'}]},
 'venue': {'id': '4fbb978fe4b0450d1d488e52',
  'name': 'River Walk',
  'location': {'lat': 51.05141324870493,
   'lng': -114.05996249390982,
   'labeledLatLngs': [{'label': 'display',
     'lat': 51.05141324870493,
     'lng': -114.05996249390982}],
   'distance': 289,
   'cc': 'CA',
   'country': 'Canada',
   'formattedAddress': ['Canada']},
  'categories': [{'id': '4bf58dd8d48988d165941735',
    'name': 'Scenic Lookout',
    'pluralName': 'Scenic Lookouts',
    'shortName': 'Scenic Lookout',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/sceniclookout_',
     'suffix': '.png'},
    'primary': True}],
  'photos': {'count': 0, 'groups': []}},
 'referralId': 'e-0-4fbb978fe4b0450d1d488e52-0'}
```


Figure 5: Data retrieved from Foursquare API

2.2 Data Cleaning

The data downloaded from Sources 1-3 are combined together into a dataframe. The count of company headquarters are calculated for each city. The values obtained are added into a column ‘ Industrial Count’.

Taking all the crime data from the years 2009 to 2016(latest available) wouldn’t be advisable as it is very outdated and can affect our analysis. So lets consider the mean of only the years 2015 & 2016. These values are added into a column ‘Crime Index’. All the columns except City, Industrial Count, Crime Index and Quality of Life Index are dropped.

2.3 Data Visualization

In order to narrow down our search for the best city, let’s take into account of only the top 5 cities that has the highest industrial count.

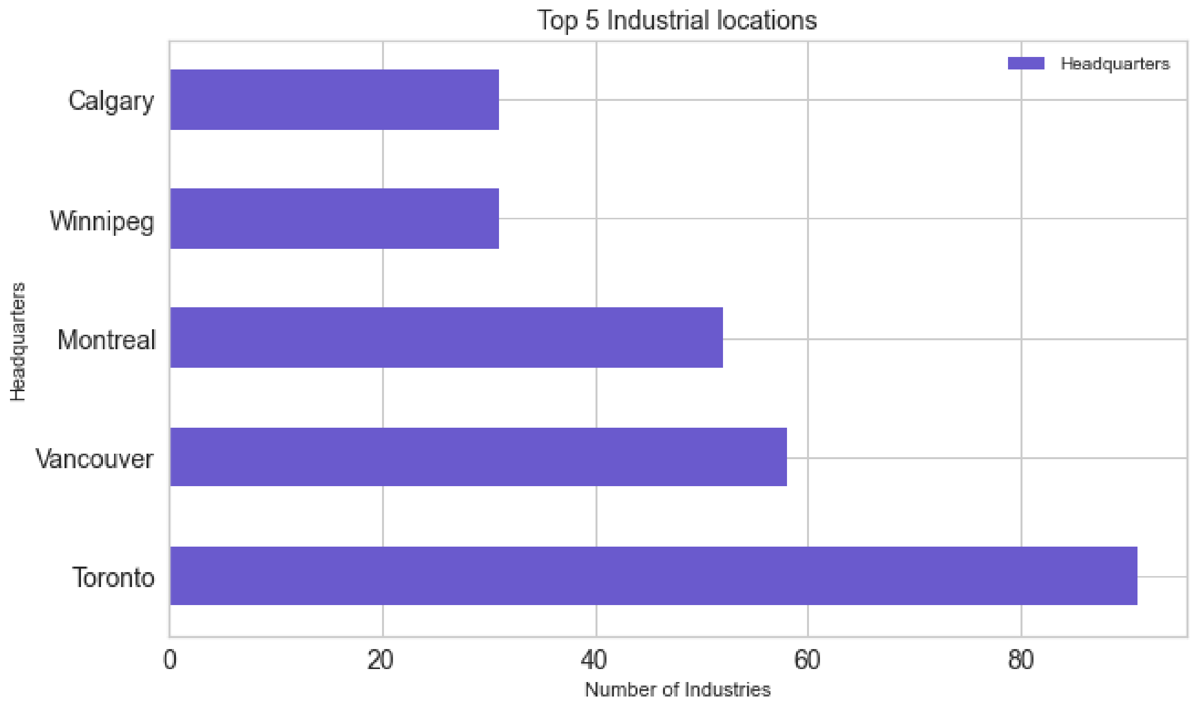

Figure 6: Top 5 Industrial Locations

From the bar plot, Toronto has the highest count of industries followed by Vancouver and Montreal. Both Calgary and Winnipeg have equal number of industries located within them.

Now lets sort the dataframe such that it has the lowest crime index and highest quality of life index.

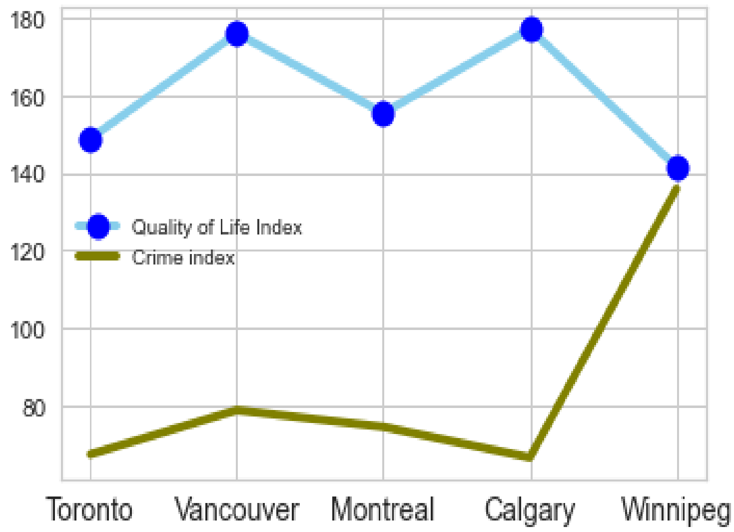

Figure 7: Quality of life index and Crime index per City

Clearly Calgary fits our criteria followed by Vancouver, Toranto, Montreal and Winnipeg.

3.Methodology

List of all the postal codes and other information about the boroughs of the province Alberta is scraped from the web using the pandas library. All the values of the boroughs that are not assigned and contains missing data are removed. The resulting dataframe is filtered to contain only the boroughs of Calgary.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
1	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
2	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786
3	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.0809	-114.1616
4	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.9878	-114.0001

Figure 8: Postal codes of Calgary Borough

3.1 Data Exploration

Since our data already contains the latitude and longitude data we can go ahead and visualize the geographic details of Calgary and its neighbourhood using the folium library. Here is the snippet of the code,

```
map_calgary = folium.Map(location=[latitude, longitude], zoom_start=11)
# adding calgary lat and long to the map
folium.CircleMarker(
    [51.0534234, -114.0625892],
    radius=10,
    popup='Calgary',
    fill=True,
    color='red',
    fill_color='red',
    fill_opacity=0.4
    ).add_to(map_calgary)
# add markers to map
for lat, lng, label in zip(calgary_df['Latitude'], calgary_df['Longitude'], calgary_df['Neighborhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='darkgreen',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_calgary)

map_calgary
```

Figure 9: Code Snippet to display Calgary map

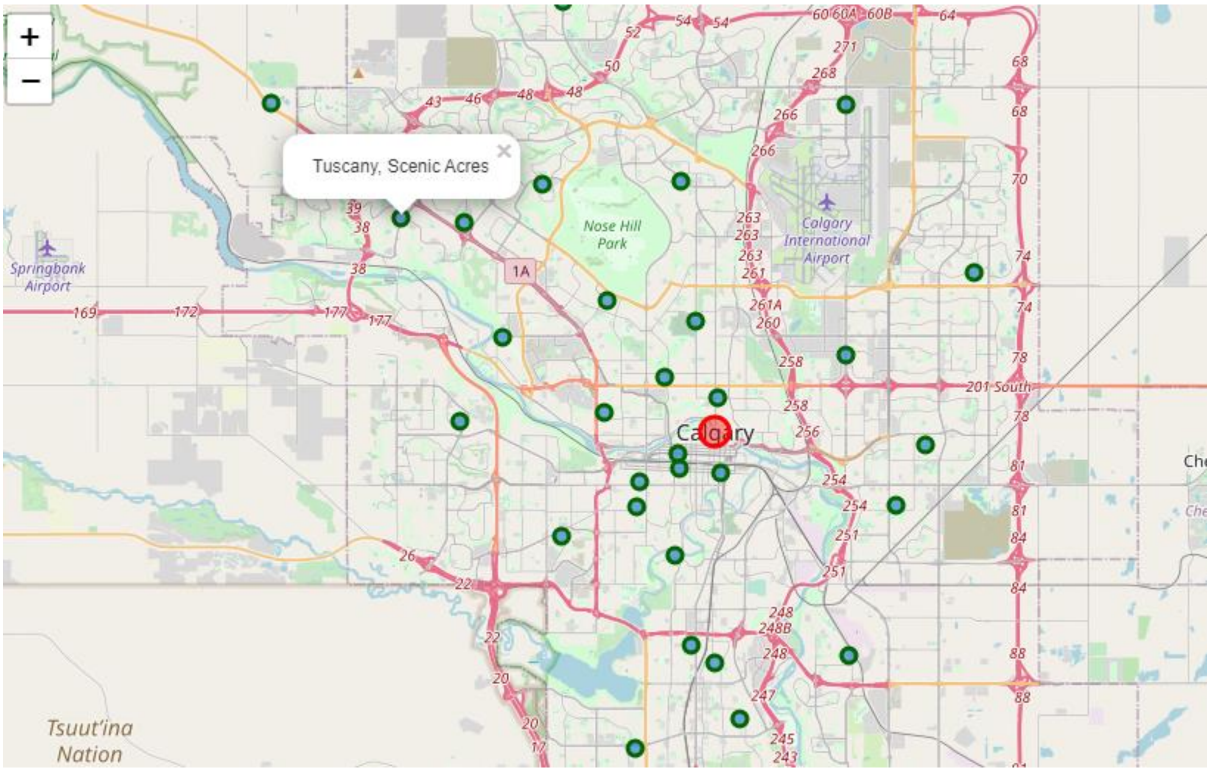

Figure 10: Map of Calgary along with its neighborhoods

Next we will use the FourSquare API to get the list of the venues in each neighbourhood of Calgary within 1000 meter radius and set the limit to 100 venues (popular spots) per neighbourhood given their latitude and longitude data.

Below is the dataframe obtained from the JSON file that was returned by Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Petro-Canada	51.128068	-114.138057	Gas Station
1	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Edgemont City	51.126473	-114.138997	Asian Restaurant
2	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Friends Cappuccino Bar & Bake Shop	51.126370	-114.138676	Café
3	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Mac's	51.128309	-114.137902	Convenience Store
4	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	Bonasera Pizza And Sports Bar	51.029893	-113.982543	Bar
5	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	7-Eleven	51.029839	-113.982060	Convenience Store
6	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	Hempisphere	51.031206	-113.981848	Smoke Shop
7	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	Foggy Gorilla Vaping Co.	51.030038	-113.972642	Smoke Shop
8	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600	Starbucks	51.084185	-114.156905	Coffee Shop
9	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600	Dale Hodges Park Lookout	51.080653	-114.166324	Scenic Lookout

Figure 11: Dataframe displaying API data

The name of the venue, the category it belongs to, and the geographical data such as venue latitude and longitude are added to the dataframe. Let's use a word cloud to visualize the most common venues in these neighbourhoods.

Figure 12: Word Cloud displaying most common venues

```
print('There are {} uniques categories.'.format(len(calgary_venues['Venue Category'].unique())))
```

```
There are 115 uniques categories.
```

Figure 13: Code snippet to identify unique venues

We can see that there are 115 unique venues in the neighborhood and from the word cloud we can infer that the most popular ones are Restaurants, Coffee shops, Convenience Stores, Bars, Pizza Places and Gas station.

3.2 Machine Learning

Many machine learning algorithms cannot handle categorical data very well. One-Hot Encoding is popular technique for treating categorical variables. It simply creates additional features based on the number of unique values in the categorical feature. Every unique value in the category will be added as a feature. Dummy variables are created for each venue category.

Now let's group rows by neighborhood and by mean of frequency of occurrence of each category.

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bakery	Bank	Bar	Bistro	Board Shop	Boat or Ferry	Bookstore	Bowling Alley	Brazilian Restaurant	Breakfas Spot
0	Braeside, Cedarbrae, Woodbine	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.00000	0.
1	Brentwood, Collingwood, Nose Hill	0.0	0.0	0.0	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.00000	0.
2	Bridgeland, Greenview, Zoo, YYC	0.0	0.0	0.0	0.045455	0.0	0.000000	0.045455	0.000000	0.0	0.0	0.0	0.0	0.0	0.00000	0.
3	City Centre, Calgary Tower	0.0	0.0	0.0	0.000000	0.0	0.071429	0.000000	0.071429	0.0	0.0	0.0	0.0	0.0	0.00000	0.
4	Connaught, West Victoria Park	0.0	0.0	0.0	0.000000	0.0	0.023810	0.000000	0.071429	0.0	0.0	0.0	0.0	0.0	0.02381	0.

Figure 14: One hot encoded and grouped by neighborhood

Here we will use the most popular clustering algorithm-Kmeans clustering. Clustering allows us to find groups of similar neighbourhoods, neighbourhoods that are more related to each other than to neighbourhoods in other groups. Lets find the optimal value of k, i.e., the number of clusters using the elbow method. The elbow method is a useful graphical tool to estimate the optimal number of clusters k for a given task. Intuitively, we can say that, if k increases, the within-cluster SSE ("distortion") will decrease. This is because the samples will be closer to the centroids they are assigned to.

```
a=calgary_grouped.drop('Neighborhood',1)
model=KMeans()
visualizer=KElbowVisualizer(model,k=(1,11))
visualizer.fit(a)
visualizer.poof()
```

Figure 15: Code snippet to visualize elbow method

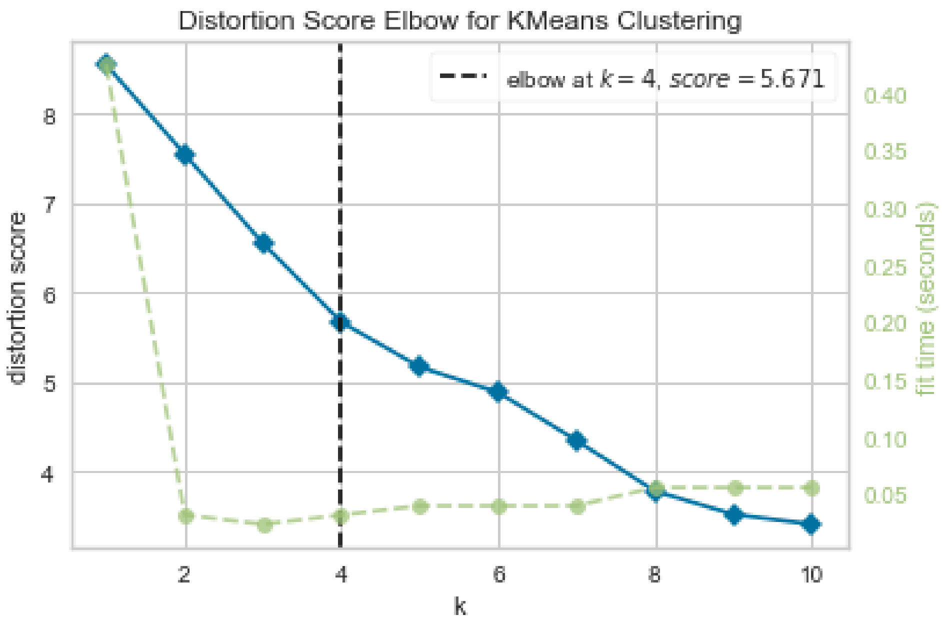

Figure 16: Finding optimal k via elbow method

Inferring from the previous graph, the optimal number of clusters should be 4.

After fitting the one-hot encoded data to the k-means model, all the NaN values are dropped inorder to prevent skewing.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	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
0	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.964320	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	1.0	Convenience Store	Asian Restaurant	Gas Station	Café	Yoga Studio	Dry Cleaner	Construction & Landscaping	Cosme Sl
2	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	1.0	Smoke Shop	Convenience Store	Bar	Yoga Studio	Electronics Store	Comic Shop	Construction & Landscaping	Cosme Sl
3	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600	0.0	Scenic Lookout	Bank	Food Court	Coffee Shop	Dry Cleaner	Comic Shop	Construction & Landscaping	Convenier St
4	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100	1.0	Clothing Store	Convenience Store	Pizza Place	Diner	Yoga Studio	Coffee Shop	Comic Shop	Construct Landscap

Figure 17: Dataframe after clustering into similar neighbourhoods

4.Results

4.1 Cluster Analysis

We have a total of 4 clusters (0,1,2,3). Before we analyze them one by one let's check the total count of neighbourhoods in each cluster.

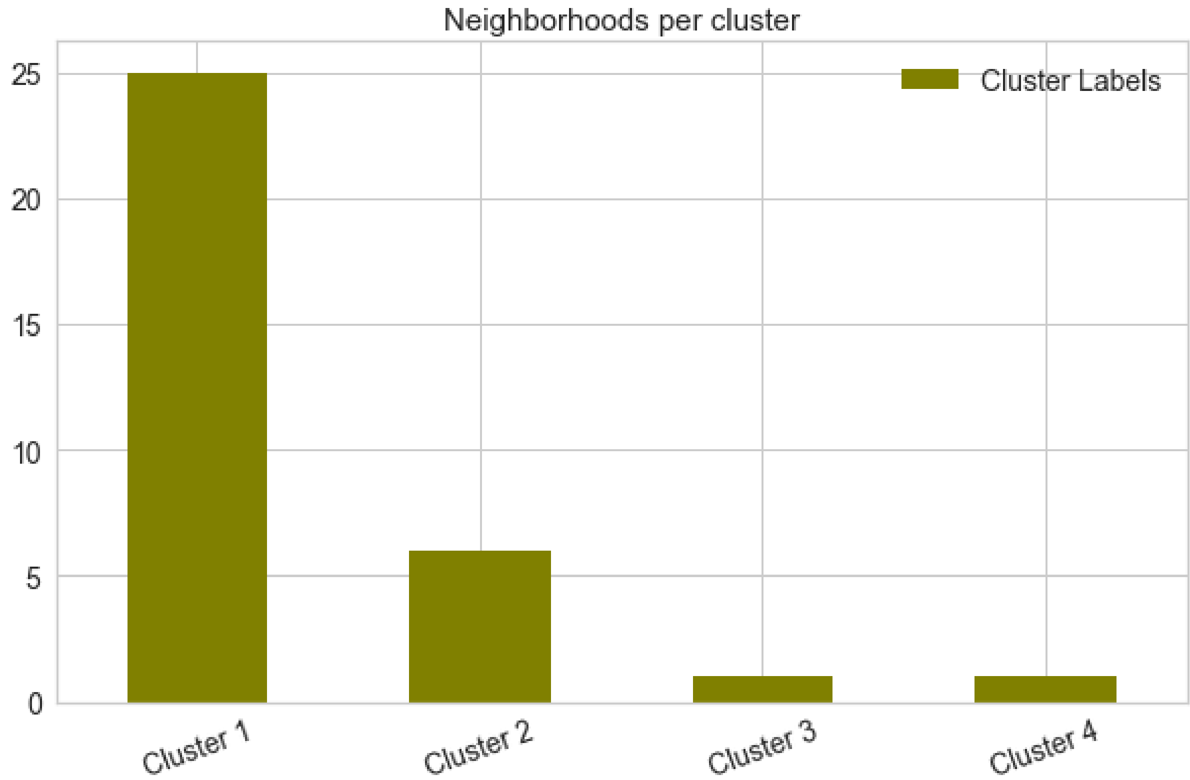

Figure 18: Neighborhoods per cluster

From the bar graph that was made using Matplotlib ,we can compare the number of Neighbourhoods per Cluster. We see that Cluster 1 has the highest neighbourhoods (25) followed by cluster 2 (6). The Cluster 3 and 4 has only one neighbourhood each.

Looking into the neighbourhoods of the Cluster 1

	Borough	Neighborhood	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
3	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	0.0	Scenic Lookout	Bank	Food Court	Coffee Shop	Dry Cleaner	Comic Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop
5	Calgary	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	0.0	Pub	Sandwich Place	Indian Restaurant	Mexican Restaurant	Sports Bar	Candy Store	Gas Station	Bookstore	Pizza Place
6	Calgary	Bridgeland, Greenview, Zoo, YYC	0.0	Fast Food Restaurant	Gym / Fitness Center	Seafood Restaurant	Chinese Restaurant	Middle Eastern Restaurant	Convenience Store	Indian Restaurant	Dim Sum Restaurant	Noodle House
7	Calgary	Lakeview, Glendale, Killarney, Glamorgan	0.0	Coffee Shop	Business Service	Wine Shop	Arts & Crafts Store	Falafel Restaurant	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store
8	Calgary	Inglewood, Burnsland, Chinatown, East Victoria...	0.0	Coffee Shop	Pub	Hotel	Restaurant	Theater	Performing Arts Venue	Cocktail Bar	Deli / Bodega	Steakhouse
9	Calgary	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	0.0	Pizza Place	Pub	Boat or Ferry	Yoga Studio	Donut Shop	Comic Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop
10	Calgary	Highfield, Burns Industrial	0.0	American Restaurant	Karaoke Bar	Bowling Alley	Electronics Store	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store
12	Calgary	Queensland, Lake Bonavista, Willow Park, Acadia	0.0	Chinese Restaurant	Child Care Service	Pizza Place	Insurance Office	Yoga Studio	Donut Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop
27	Calgary	South Calgary (Altadore / Bankview / Richmond)	0.0	Convenience Store	Pizza Place	Liquor Store	Coffee Shop	Yoga Studio	Donut Shop	Comic Shop	Construction & Landscaping	Cosmetics Shop
28	Calgary	Oak Ridge, Haysboro, Kingsland, Kelvin Grove, ...	0.0	Burger Joint	Sushi Restaurant	Coffee Shop	Restaurant	Steakhouse	Music Store	Pizza Place	Breakfast Spot	Bridal Shop
29	Calgary	Braeside, Cedarbrae, Woodbine	0.0	Pharmacy	Gym	Pub	Gas Station	Coffee Shop	Convenience Store	Hockey Rink	Pizza Place	Ice Cream Shop
30	Calgary	Midnapore, Sundance	0.0	Hardware Store	Mobile Phone Shop	Board Shop	Yoga Studio	Dry Cleaner	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega
31	Calgary	Rundle, Whitehorn, Monterey Park	0.0	Hotel	Coffee Shop	Grocery Store	Gastropub	BBQ Joint	Sports Bar	Discount Store	Diner	Dim Sum Restaurant
32	Calgary	Millrise, Somerset, Bridlewood, Evergreen	0.0	Sporting Goods Shop	Bank	Pub	American Restaurant	Gas Station	Light Rail Station	Convenience Store	Coffee Shop	Multiplex
33	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East	0.0	Bar	Park	Bus Stop	Pub	Dry Cleaner	Asian Restaurant	Cosmetics Shop	Deli / Bodega	Donut Shop

Figure 19: Neighborhoods of Cluster 1

Upon closely examining the neighbourhoods we can see that the most common venues in these neighbourhoods are Coffee shops, Restaurants, Parks and scenic outlooks , Bars and Pubs.

The second cluster consists of 6 Neighbourhoods and the most common venues can be easily identified as the Convenience stores and Vietnamese Restaurants.

	Borough	Neighborhood	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
1	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	1.0	Convenience Store	Asian Restaurant	Gas Station	Café	Yoga Studio	Dry Cleaner	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
2	Calgary	Forest Lawn, Dover, Erin Woods	1.0	Smoke Shop	Convenience Store	Bar	Yoga Studio	Electronics Store	Comic Shop	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store
4	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	1.0	Clothing Store	Convenience Store	Pizza Place	Diner	Yoga Studio	Coffee Shop	Comic Shop	Construction & Landscaping	Cosmetics Shop	Deli / Bodega
11	Calgary	Discovery Ridge, Signal Hill, West Springs, Ch...	1.0	Vietnamese Restaurant	Convenience Store	Gas Station	Bar	Pizza Place	Yoga Studio	Donut Shop	Comic Shop	Construction & Landscaping	Cosmetics Shop
14	Calgary	Thorncliffe, Tuxedo Park	1.0	Vietnamese Restaurant	Convenience Store	Bar	Fast Food Restaurant	Yoga Studio	Dry Cleaner	Comic Shop	Construction & Landscaping	Cosmetics Shop	Deli / Bodega
23	Calgary	Symons Valley	1.0	Convenience Store	Yoga Studio	Coffee Shop	Comic Shop	Construction & Landscaping	Cosmetics Shop	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner

Figure 20: Neighborhoods of Cluster 2

Looking into the neighbourhoods of cluster 3 and 4, we can see that these clusters have only one neighbourhood each. This is because of the unique venues in each neighbourhoods, hence they couldn't be clustered into similar neighbourhoods.

	Borough	Neighborhood	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
16	Calgary	Brentwood, Collingwood, Nose Hill	2.0	Electronics Store	Coffee Shop	Comic Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dim Sum Restaurant	Diner

Figure 21: Neighborhoods of Cluster 3

The third cluster consists of Venues such as Electronics store, Coffee shops and Comic shops.

	Borough	Neighborhood	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
25	Calgary	Northwest Calgary	3.0	Flea Market	Yoga Studio	Dry Cleaner	Comic Shop	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store	Dim Sum Restaurant

Figure 22: Neighborhoods of Cluster 4

The Fourth cluster consists of Venues such as Flea Market, Dry Cleaner and Yoga Studios.

4.2 Cluster Visualization

Now lets visualize these clustered neighbourhoods on the map using the folium library. Each cluster is being colour coded for the ease of representation. Its evident that majority of the clusters are marked red, which belongs to the first cluster. Purple markings represent the second cluster. The Blue and Beige markings represent the third and fourth cluster respectively.

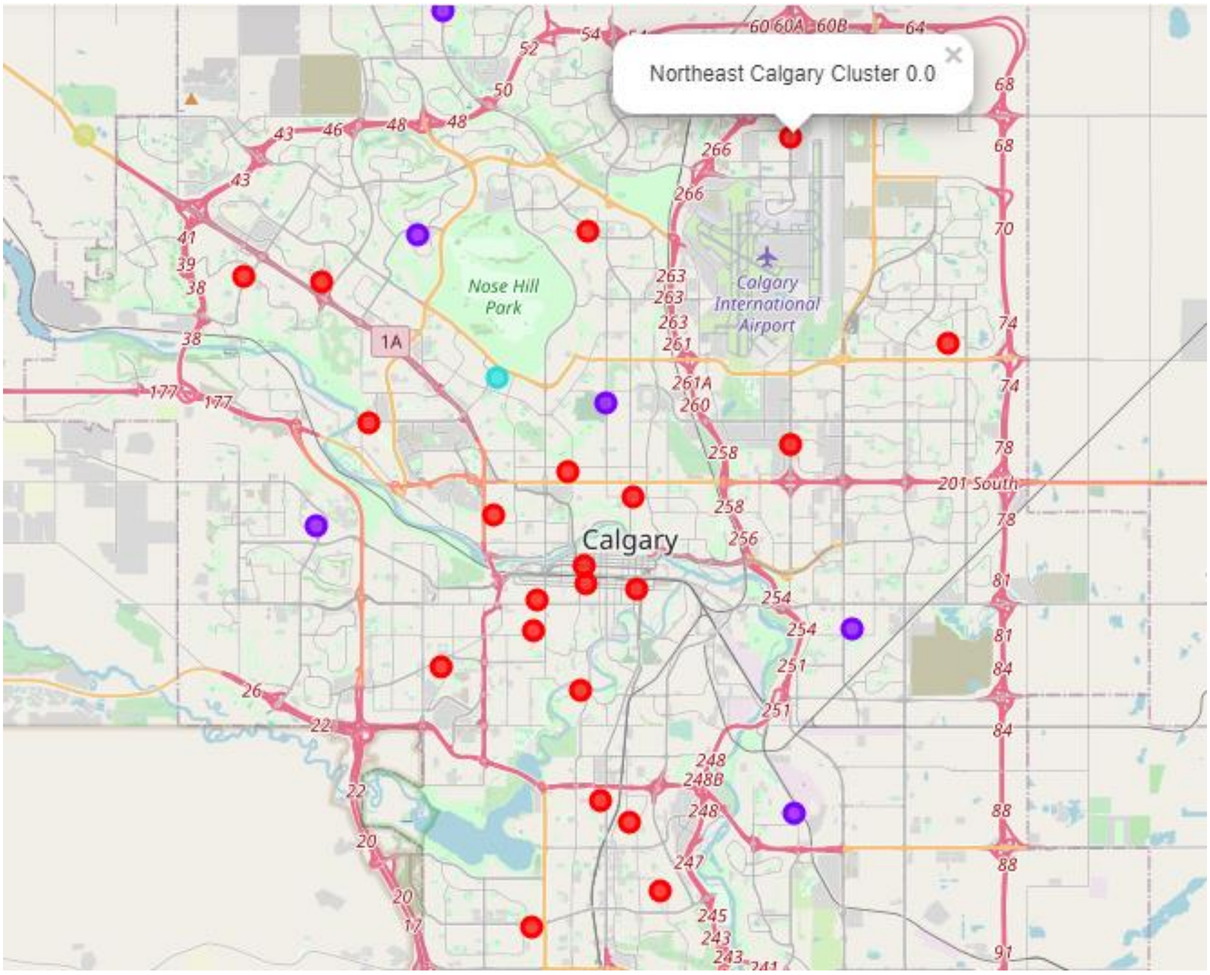

Figure 23: Map of Calgary with clustered neighborhoods

5.Discussion

After exploring the neighbourhoods of Calgary, we find that the neighbourhoods in the first cluster would be best suitable to migrate as it has a wide variety of venues nearby ranging from restaurants of different cuisines to parks and scenic outlooks. It also consists of necessary amenities such as banks, child care, convenience stores, pharmacy and gym. These neighbourhoods would be the ideal place for relocating as it also contains good transport network such as Bus stops, Light rail stations and plenty of gas stations. Also the neighbourhoods offer a great night life as Bars, Pubs are located in large numbers. Thus Cluster 1 would be our Housing Cluster.

The Neighbourhoods in the second cluster consists mostly of commercial venues such as convenience stores and clothing stores and can be conveniently named as the Commercial Cluster. These neighbourhoods are also hotspots for Vietnamese Cuisine. Thus these neighbourhoods would be ideal in order to either set up a store or a Vietnamese restaurant.

The third and fourth cluster consisting of one neighbourhood each doesn’t offer much as it has only a few venues such as Electronic stores and Flea markets located within them. This explains the population sparsity of Canada. As these neighbourhoods are not well occupied, they wouldn’t be ideal for migrating.

6.Conclusion

The aim of the project is to help the migrants to identify the an ideal place in terms of safety, job opportunity , quality of life and choosing a neighbourhood that would be the most suitable for them. Through our analysis we found that Calgary is the best city that fits our criteria. In terms of job opportunity, it is one of the top 5 cities to house large number of industries. Also it has low crime rate and high quality of life index.

With the help of Foursquare API , we then clustered the neighbourhoods of Calgary based on similar neighbourhoods using K-means Algorithm. From our results, the neighbourhoods in the first cluster would be the ideal place to migrate as has a well connected transport system and necessary amenities such as banks, pharmacy, convenience stores and child care. There are also lots of restaurants, bars and scenic outlooks to keep as entertained.

This project can further be improved by taking into account an additional city and compare both of them to select which would be more suitable to migrate or by using another machine learning algorithm such as DBSCAN to cluster the neighbourhoods.