# Final Report | Capstone Project – The Battle of Neighborhoods Finding a Better Place To Start a Vietnamese Restaurant Business in Calgary, Alberta, Canada

## 1. Introduction

The purpose of this Project is to help people in exploring better places to start a Vietnamese Restaurant Business around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Calgary City, Alberta, Canada.

Lots of people are migrating to various states of Canada and needed lots of research for good location and surrounding by their community for a better customer flow. This project is for those people who are looking for better start by doing research on the city of Calgary.

This Project aim to create an analysis of features for people migrating to Calgary City to search a best neighborhood as a comparative analysis between neighborhoods. The features include the most common venues ratings, asian restaurants rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.

## 2. Data Section

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T)

Initialzing data by scaping the table on the wikipedia page. Dataset consisting of Postal, Code, Borough, Neighborhood, Latitude, and Longitude.

**Foursquare API**

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 500 meter and limited to 100 venues

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood

2. Neighborhood Latitude

3. Neighborhood Longitude

4. Venue

5. Name of the venue e.g. the name of a store or restaurant

6. Venue Latitude

7. Venue Longitude

8. Venue Category

## Map of Calgary

```
In [6]:   Image('images/calgary_map.png')
```

Out[6]:
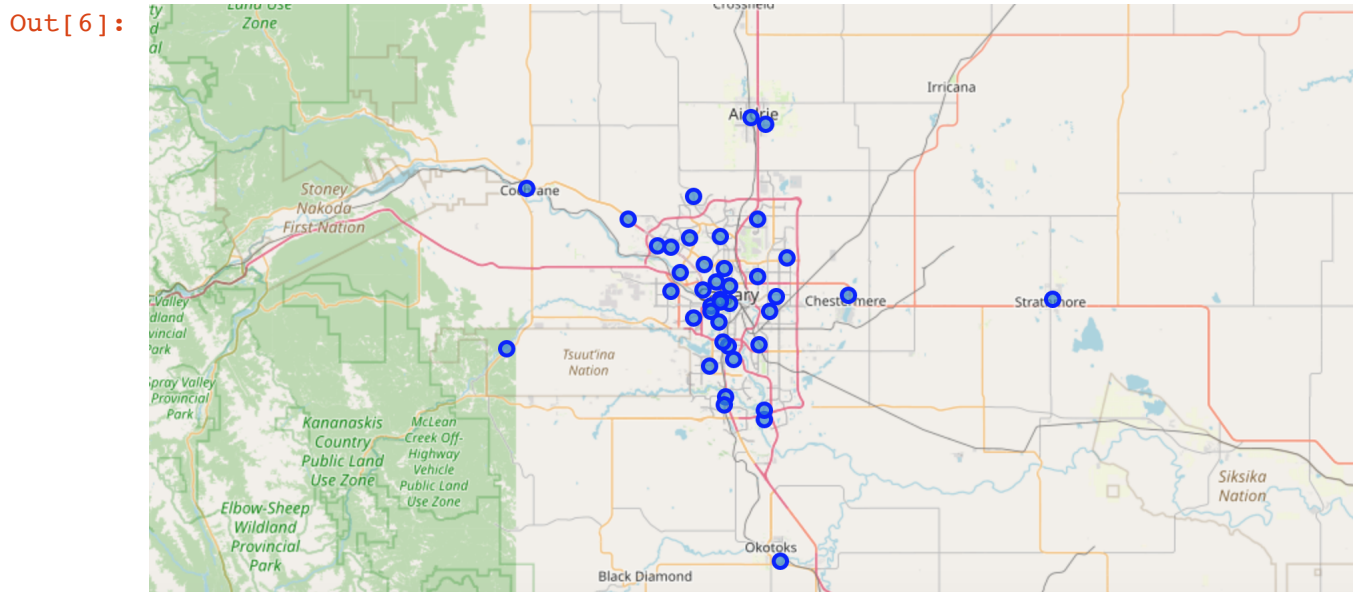
# 3. Methedology Section

**Clustering Approach:**

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

**Using K-Means Clustering Approach**

**Most Common venues near Neighborhood**

**Work Flow:**

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.
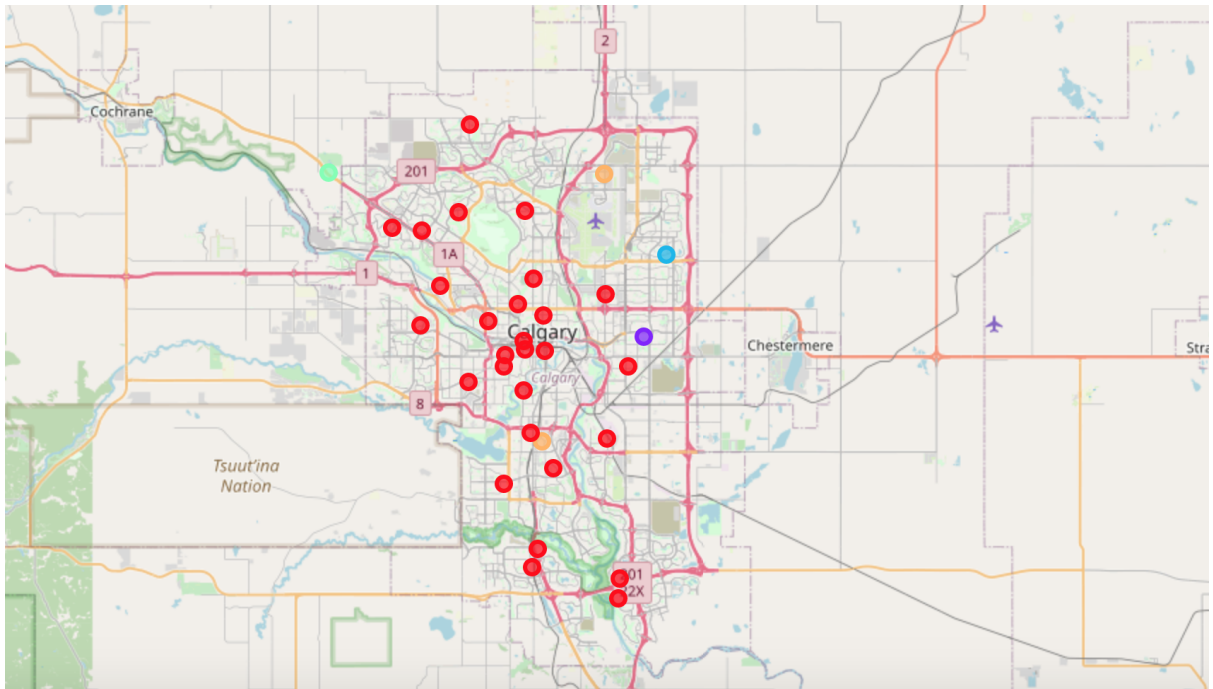
# 4. Result Section

**Map of Clusters in Calgary**

```
In [1]: from IPython.display import Image
```

```
In [5]:  Image('images/clustering_calgary.png')
```

Out[5]:



**Vietnamese Restaurant Venues**

**The location**

Calgary is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Alberta Province Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise.

# 5. Discussion Section

**Problem Which Tried to Solve:**

The major purpose of this project, is to suggest a better neighborhood in Calgary city for the Vietnamese people that plan to open a Vietnamese Restaurant there. After clustering section, we will look for the `1st Most Common Venue` going down to `10th Most Common Venue` for each cluster and try to find the Neighborhood has the highest rank for Vietnamese Restaurant.

After all the work, we will obtain those Neighborhood that has Vietnamese Restaurant as the highest rank in Common Venues.

# 6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 5(five) different clusters and for 33 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. After that we then look for the Neighborhood with the Highest Rank in Common Venues in `Vietnamese Restaurant`

At the end, we obtain 3 Neighborhood that has 1st Most and 2nd Most Common Venues are: `Thorncliffe, Tuxedo Park`, `Lakeview, Glendale, Killarney, Glamorgan`, `Discovery Ridge, Signal Hill, West Springs`

In [8]: `Image('images/1stmost_common_venues.png')`

Out[8]:

In [111]:    1 `calgary_merged[calgary_merged['1st Most Common Venue'] == 'Vietnamese Restaurant']`

Out[111]:

| index | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | T2K | Calgary | Thorncliffe, Tuxedo Park | 51.0857 | -114.0714 | 0 | Vietnamese Restaurant | Fast Food Restaurant | Convenience Store | Massage Studio | Home Service | Bar | Discount Store | Dry Cleaner |

In [7]: `Image('images/2ndmost_common_venues.png')`

Out[7]:

In [112]:    1 `calgary_merged[calgary_merged['2nd Most Common Venue'] == 'Vietnamese Restaurant']`

Out[112]:

| index | 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 Mo Comm Ven |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | T3E | Calgary | Lakeview, Glendale, Killarney, Glamorgan | 51.0227 | -114.1342 | 0 | Wine Shop | Vietnamese Restaurant | Coffee Shop | Falafel Restaurant | Dance Studio | Deli / Bodega | Department Store | Dim Si Restaura |
| 11 | T3H | Calgary | Discovery Ridge, Signal Hill, West Springs, Ch... | 51.0566 | -114.1815 | 0 | Convenience Store | Vietnamese Restaurant | Gas Station | Bar | Electronics Store | Cosmetics Shop | Dance Studio | De Bode |

In [ ]:    1

# Future Works:

This project can be continued for making it more precise in terms to find the best location to open a Vietnamese Restaurant Business in Calgary City. Best means we can look on other factors such as is there a Asian or Vietnamese Community? Which nearby businesses are most likely to go out to dine in at the Vietnamese Restaurants(schools, tech companies, retail stores etc), are there any market share (competitive if nearby have a lot of Vietnamese Restaurant etc..)

**Libraries Which are Used to Develope the Project:**

`Pandas` : For creating and manipulating dataframes.

`Folium` : Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

`Scikit Learn` : For importing k-means clustering.

`JSON` : Library to handle JSON files.

`XML` : To separate data from presentation and XML stores data in plain text format.

`Geocoder` : To retrieve Location Data.

`Matplotlib` : Python Plotting Module.