The Best location for the setting up of a football stadium in the Calgary neighborhoods in Alberta, Canada

Submitted by Toju Mary Awe

1. INTRODUCTION/BUSINESS PROBLEM

With the growing number of youth and large young migrants in Canada, Toju Football Academy is looking at building a football stadium in the neighborhoods of Calgary. The aim of this is to attract as many young talents as possible into the sport while also contributing to the development of the game in the country. Data obtained from https://en.wikipedia.org/wiki/Demographics_of_Canada showed that 30% of the Canadian population is less than 25 years old. Calgary, the largest city in Alberta, has a population of about 1.1million people (see https://en.wikipedia.org/wiki/Demographics_of_Calgary#:~:text=Pop.&text=In%20the%202011%20Census%2C%20the,area%20(CMA)%20was%201%2C214%2C839). Toju Football Academy is therefore taking the advantage of this huge young population to set up a football stadium in Calgary. The aim of this project therefore, is to help the Academy suggest the best location to site the football stadium in the neighborhoods of Calgary.

2. DATA DESCRIPTION

As explained above, the aim for this project is to suggest the best possible location to set up a football stadium in the neighborhood of Calgary. The data for the purpose of this project would be sourced from:

a. Wikipedia. At https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T is a list of the neighborhoods in Calgary together with their latitude and longitude coordinates. This dataset would be filtered and cleaned to extract only the borough of Calgary.

b. The Forsquare **API** would be used to get the most common venues in the city of Calgary.

Using this dataset, we are going to set certain parameters so as to achieve our objectives. These parameters are:

a. Locations not already crowded with the presence of football stadium or any other sporting centers

b. Locations where there is no football stadium.

3. METHODOLOGY

For this project, the latitude and longitude coordinates of Calgary was the data used and our source was Wikipedia. The table below shows the format the data appeared after being passed into the Dataframe and cleaned.

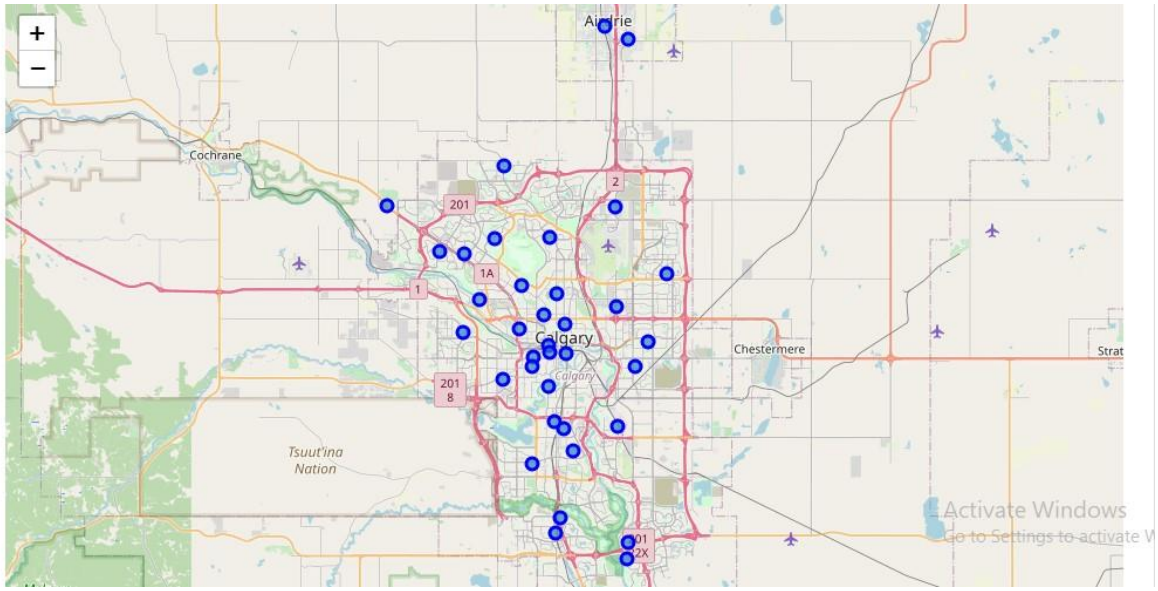
	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T1A	Medicine Hat	Central Medicine Hat	50.036460	-110.679250
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
3	T4A	Airdrie	East Airdrie	51.272450	-113.986980
4	T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413

This data was then filter to narrow down our analysis to only Calgary borough. The result is presented in the table below:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.964320
1	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158
2	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600
3	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600
4	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100

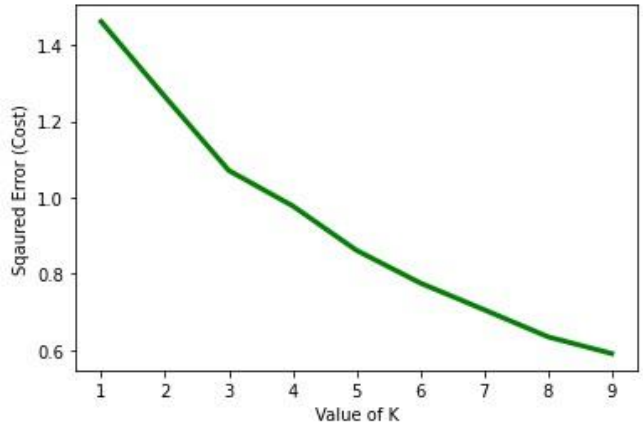
To visualize the map of Calgary neighborhoods, the folium library in python was used to create the map using the latitude and longitude coordinates given in the Calgary DataFrame above. Below is the map that was produced:

Out[7]:



The result of this cleaned dataset was passed into the Foursquare API to obtain the venues in each neighborhood within the radius of 1500 meters and the limit of 100. The result obtained using the Foursquare API was converted into a DataFrame and grouped by the number of neighborhoods. The result of this shows there are 191 unique categories of venues in the Calgary neighborhood.

Using the elbow method of calculating the optimal value of K, 3 was the fit value for the Kmeans algorithm. Therefore, the Kmeans cluster algorithm was used to cluster the Calgary neighborhood into 3.



Analysing the graph suggest the value of 3 as the best fit for k

4. RESULTS

The result of our clustering shows majority of the neighborhoods belong to the second cluster (cluster '1') while the remaining neighborhoods were sparsely distributed among the other two clusters.

Out[27]:

	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
0	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.964320	1	Fast Food Restaurant	Vietnamese Restaurant	Pharmacy	Grocery Store	Italian Restaurant	Ice Cream Shop	Restaurant
1	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	0	Convenience Store	Gas Station	Chinese Restaurant	Gym / Fitness Center	Trail	General Entertainment	Women's Store
2	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	1	Grocery Store	Vietnamese Restaurant	Fast Food Restaurant	Convenience Store	Bank	Sandwich Place	Fried Chicken Joint
3	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600	1	Coffee Shop	Restaurant	Clothing Store	Park	Grocery Store	Gas Station	Pizza Place
4	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100	1	Pizza Place	Coffee Shop	Pharmacy	Construction & Landscaping	Indian Restaurant	Spa	Gas Station

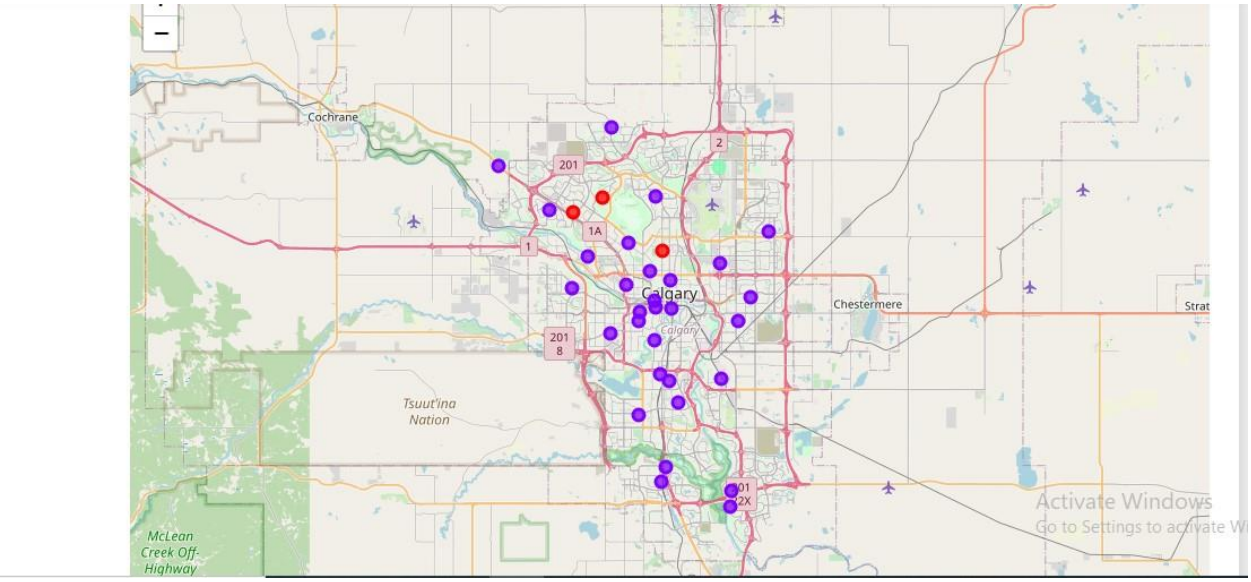
We can classify the clusters thus:

The first cluster (cluster ‘0’) – basically a relaxation center

The second cluster (cluster ‘1’) – for restaurants and café

The third cluster (cluster ‘2’) – more of a sports center

The result is projected on the map below:



5. DISCUSSION

The analysis of the result produced by the clustering of Calgary neighborhoods revealed that the third cluster has a football stadium and also a lot of sporting activities. For the second cluster,

there is also the presence of a football stadium, although the cluster reveals the preponderance of restaurants, shopping mall and the likes. Looking at the parameter we set at the beginning of this project, we may want to look away from second and the third cluster in setting our football stadium. In analyzing the first cluster, it is evident that there is no presence of any football stadium among the top ten venues in the cluster.

6. CONCLUSION

Comparing the parameters we set at the beginning of this project which are:

a. Locations not already crowded with the presence of football stadium or any other sporting centers
b. Locations where there is no football stadium.

According to the result of our analysis, the Northeast Calgary seems to be the center of sporting activities since there is a football stadium within the vicinity of the neighborhood. With the result we obtained from our clustering analysis, the best neighborhoods to site our football stadium are: Dalhousie, Edgemont, Hamptons, Hidden Valley, Hawkwood, Arbour Lake, Citadel, Ranchlands, Thorncliffe and Tuxedo Park, since these neighborhoods do not have much of sporting activities.