Italian Restaurant Opening

Capstone

TABLE OF CONTENTS

INTRODUCTION

Calgary is one of the cities where one of the popular Italian restaurants would like to be considering opening their one of the branches. Italian Restaurants CEO approached us to do vital aspect to consider physical location to settle in. Main purpose of this report is to make recommendation to a group of Executives including CEO on supporting opening of Italian Restaurants in Calgary. Foursquare services will be used to discover the main region of interest in Calgary and will provide meaningful data so that business can make a factful decision.

Data

The data necessary to accomplish the project are the following:

1. Neighborhoods in Calgary
    a. Postal Code
    b. Neighborhood

This data will be obtained from the following Wikipedia page.

2. Latitude and longitude for each Neighborhood

a. Postal Code
b. Neighborhood
c. Latitude
d. Longitude

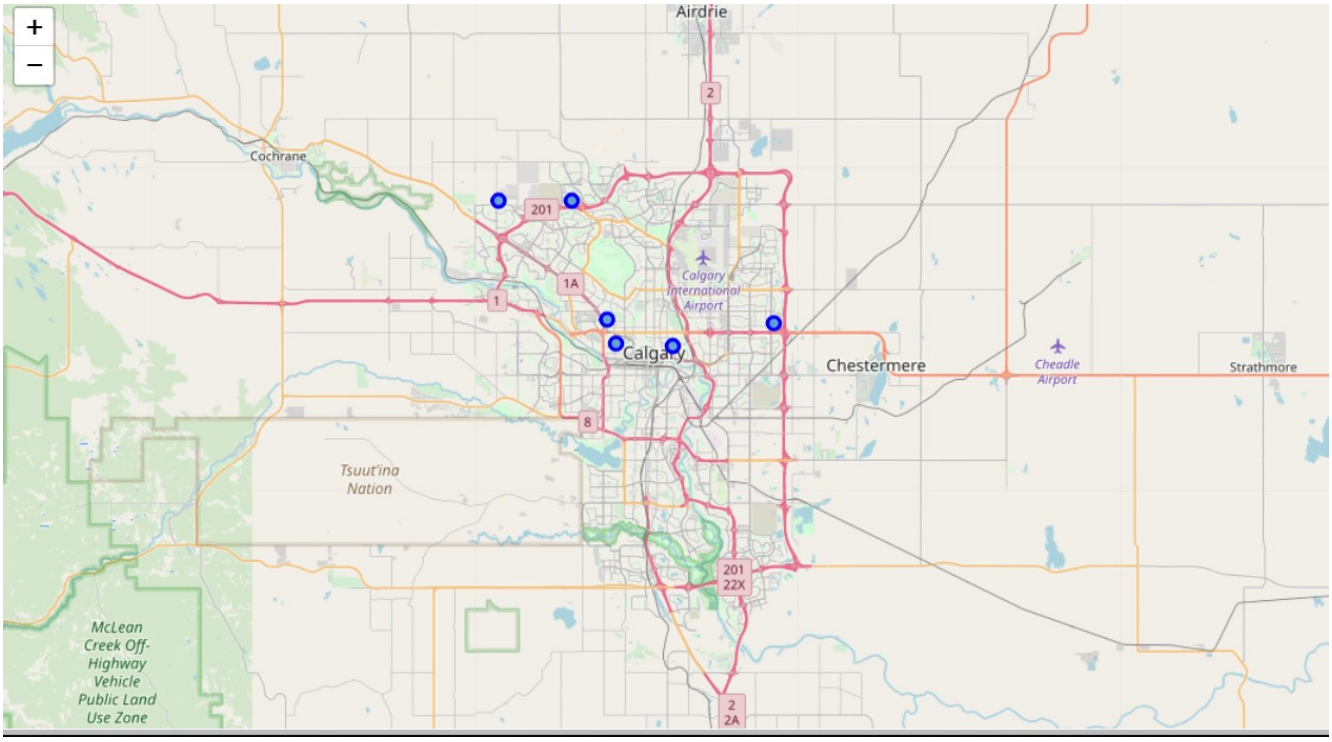
C and D items will be obtained using GeoPy (Python) library and using postal code as the argument. Foursquare will permit to get the main venues related to the listed postal codes, analyze and finally provide main point so that business can make a decision.

METHODOLOGY

We have used Canadian Postal Code as one of the sources to make decision, using Wikipedia page I have scrapped using Beautiful Soup and we obtained a data frame containing Postal Code and Neighborhood. In addition to that, using Nominatim from Geopy in Python, we found location coordinates using postal code as argument.

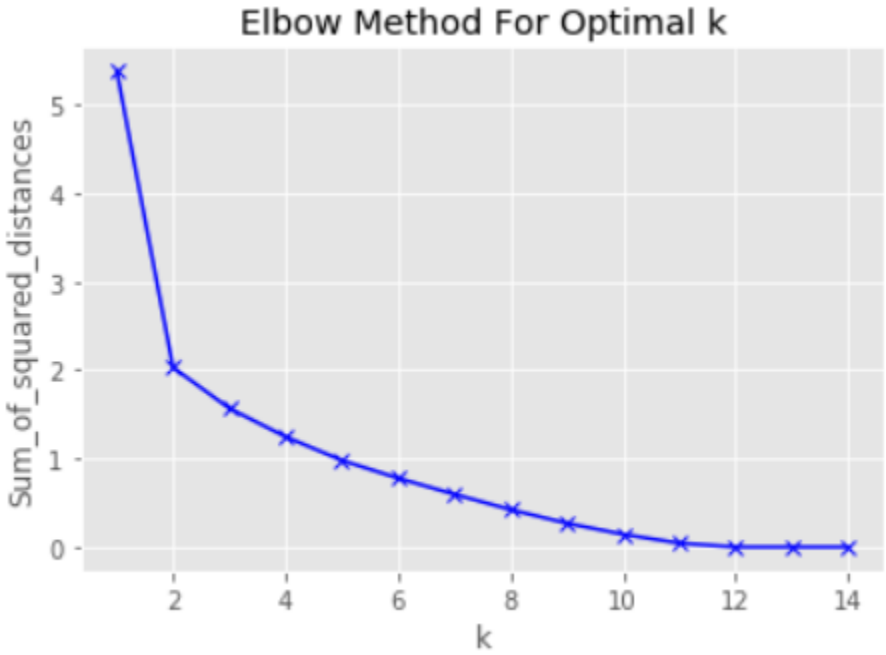
	PostalCode	Neighborhood	Latitude	Longitude
0	T1Y	Rundle, Whitehorn, Monterey Park	51.073658	-113.931857
1	T2A	Penbrooke Meadows, Marlborough	50.177444	-95.853973
2	T2B	Forest Lawn, Dover, Erin Woods	-33.723784	150.875606
3	T2C	Lynnwood Ridge, Ogden, Foothills Industrial, G...	-33.724066	150.875552
4	T2E	Bridgeland, Greenview, Zoo, YYC	51.057660	-114.041570
5	T2G	Inglewood, Burnsland, Chinatown, Victoria Park...	51.589034	-3.227128
6	T2J	Queensland Downs, Lake Bonavista, Willow Park,...	46.454554	4.117235
7	T2L	Brentwood, Collingwood, Nose Hill	49.457327	5.579155
8	T2M	Mount Pleasant, Capitol Hill, Banff Trail	51.075861	-114.113205
9	T2N	Kensington, Westmont, Parkdale, University	51.059908	-114.104076
10	T2P	City Centre, Calgary Tower	51.383267	0.516859
11	T2R	Connaught, Victoria Park	50.116367	8.727189
12	T2V	Oak Ridge, Haysboro, Kingsland, Kelvin Grove, ...	57.244347	22.605473
13	T3A	Dalhousie, Edgemont, Hamptons, Hidden Valley	-33.723398	150.874970
14	T3B	Montgomery, Bowness, Silver Springs, Greenwood	-33.723687	150.874921
15	T3C	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	-33.723973	150.874866
16	T3G	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	51.157034	-114.231181

In the map shown below we can see neighborhoods or groups of neighborhoods in Calgary.



We used Foursquare to get venues located in each neighborhood, within 500 m radius and limited to 100 venues per neighborhood.

To decide which neighborhood would be appropriated for a new Italian Restaurant we implemented k-means machine learning algorithm. In order to determine the best k-value for our k-means algorithm we used the elbow method.



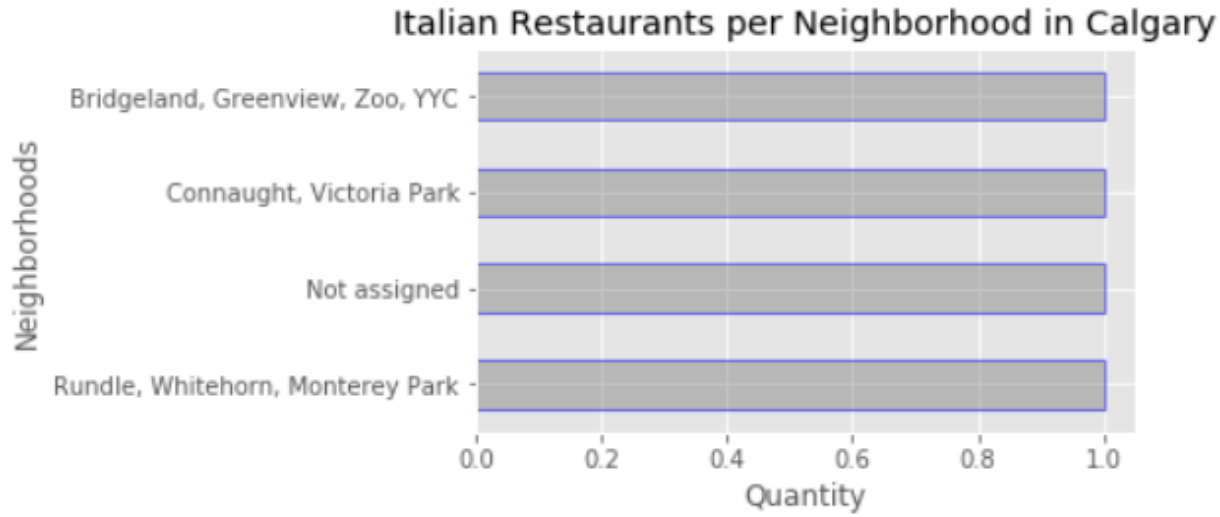
RESULTS

As a result of applying k-means algorithm and after analyzing its venues, below clusters were obtained.

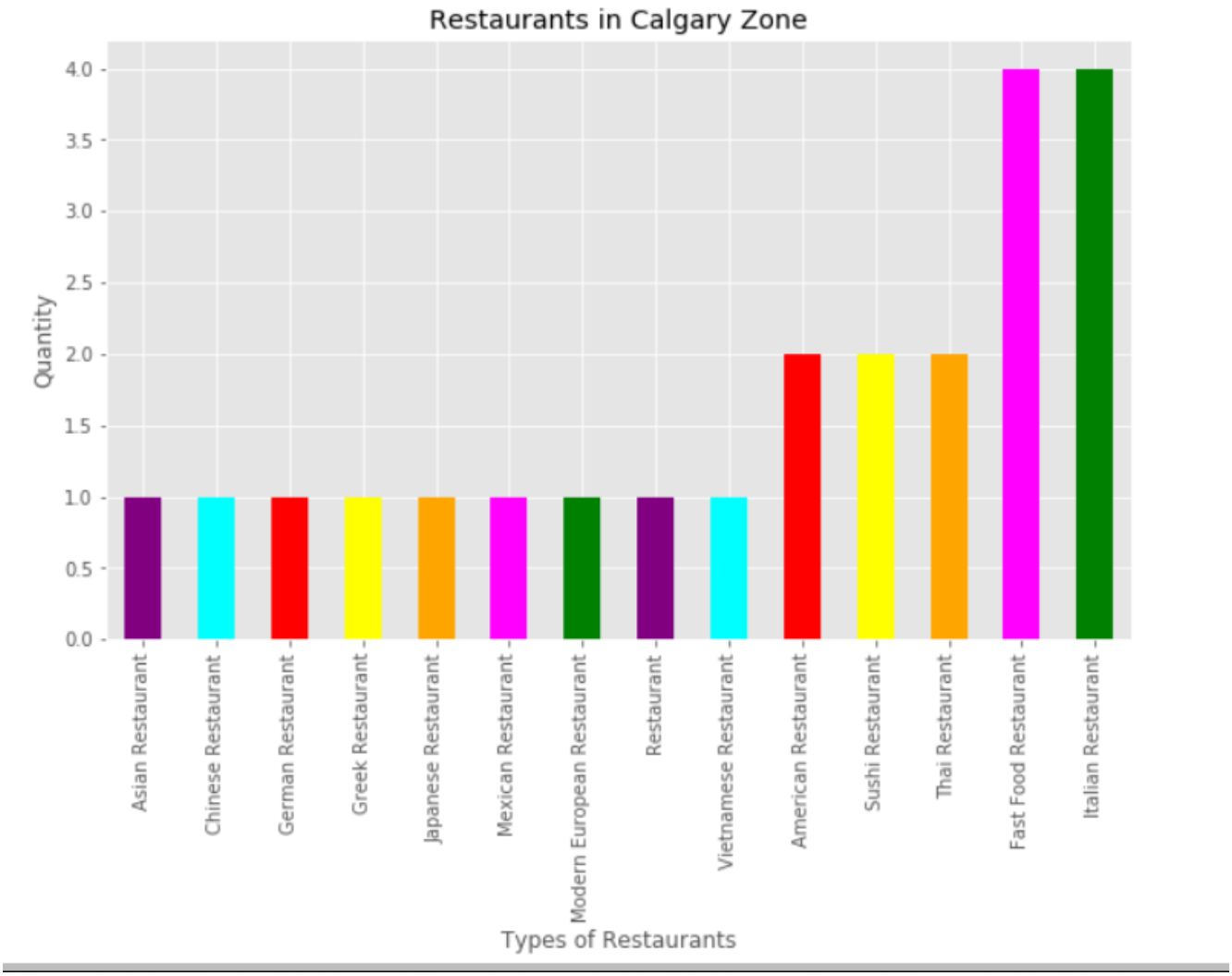
Cluster 1: 1 Neighborhood
Cluster 2 5 Neighborhoods
Cluster 3 4 Neighborhoods
Cluster 4 10 Neighborhoods
Cluster 5 4 Neighborhoods
Cluster 6 3 Neighborhoods

DISCUSSION

The following bar chart shows the amount of Italian Restaurants established in Calgary region and the neighborhood where they are located.



Additionally, below graph shows the variety of restaurants located in Calgary.



Based on limited data and its characteristics there are only 4 Italian Restaurants in the Calgary, which makes easier for investor to enter this market, on the other hand they are spread out between different clusters. In accordance with data we recommend considering neighborhood from Cluster 6 or 4.

CONCLUSION

We used k-means machine learning algorithm to construct a classifier and applied to a real word case scenario. Of course, methodology can be improved further to consider other data characteristics and elements.