Coursera Capstone- Battle of Neighborhoods

Recommending Neighborhoods to new home buyers in Calgary, AB

1.0 Introduction

Calgary sits in the sunny eastern foothills of Canada's Rocky Mountains. It is the major urban center for the entire southern half of the province of Alberta, and is surrounded by an area of profound beauty with an unspoiled, resource-rich natural environment.

In 2019, Calgary ranked the most livable city in North America and the 5th most livable city in the world by *the Economist Intelligence Unit.* Population density is low, quality of life is high and housing is abundant. Calgary is ranked amount the world's cleanest, healthiest and safest cities in numerous international studies, and is a global leader for overall quality of life.

A steady influx of new residents and historically strong economy made Calgary one of the most desirable places in the county to own property.

1.1 Problem Description

With 198 neighborhoods, newcomers to Calgary can be overwhelmed by the amount of options when choosing an area in which to purchase their home. A land area of 825.29 square kilometers means that choosing a neighborhood solely by price could mean driving hours to visit amenities you would like to frequent. Which would certainly influence quality of life.

1.2 Target Audience

Potential clients looking to buy a property in a neighborhood that closest fits their needs and budget. And real estate focused businesses looking to streamline their workflow for finding homes that cater to their clients.

1.3 Success Criteria

The success criteria of this project will be to deliver good recommendations of neighborhoods that closely match the clients needs for amenities while also falling below their ideal budget.

2.0 Data Description

We will be exploring, segmenting and clustering neighborhoods in the city of Calgary, and therefore require data which includes property values, neighborhoods, neighborhood locations. Luckily for us, the city of Calgary provides a dataset of Property Assessments
(Available here: https://data.calgary.ca/dataset/Property-Assessments/6zp6-pxei) .

We will clean up and process this data to create a dataset that is useful to us for this project. First by removing data that is irrelevant to this project, then by using median home values to group the neighborhoods. The resulting data is eventually structured like the example below.

1. Neighborhood Name
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Media Value of Homes in the Neighborhood

	Value	LATITUDE	LONGITUDE
Neighborhood			
ABBEYDALE	300000.0	51.058941	-113.929649
ACADIA	405500.0	50.972799	-114.058883
ALBERT PARK/RADISSON HEIGHTS	310000.0	51.041720	-113.995064
ALTADORE	775250.0	51.018423	-114.104126
APPLEWOOD PARK	347250.0	51.043497	-113.927941

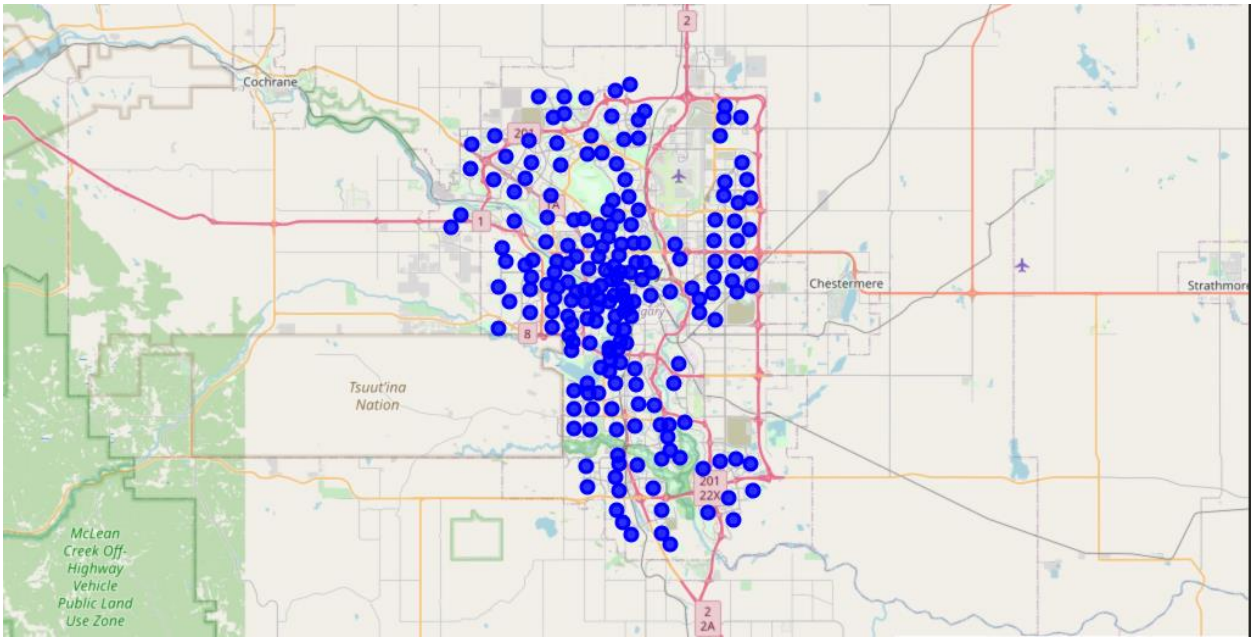
2.1 Data Features

We will be leveraging an API provided by Foursquare.com to explore the various types of venues and its categories available in each neighborhood. We will use the API to list venues nearby (Within 500m) of each respective neighborhood. The information provided by the API will be restructured like the example below to allow for further computation.

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ABBEYDALE	51.058941	-113.929649	Subway	51.059239	-113.934423	Sandwich Place
1	ABBEYDALE	51.058941	-113.929649	Mac's	51.059376	-113.934425	Convenience Store
2	ABBEYDALE	51.058941	-113.929649	roadside pub	51.059277	-113.934529	Wings Joint
3	ABBEYDALE	51.058941	-113.929649	Redbox	51.059108	-113.934845	Pizza Place
4	ACADIA	50.972799	-114.058883	Acadia Aquatic & Fitness Centre	50.973297	-114.059984	Gym / Fitness Center

3.0 Methodology



The dataset provided by the City of Calgary, contains property assessments, neighborhood names, and their respective latitude and longitude. By processing this dataset we are able to extract median home values for each neighborhood in addition to the entire neighborhood list and their locations.

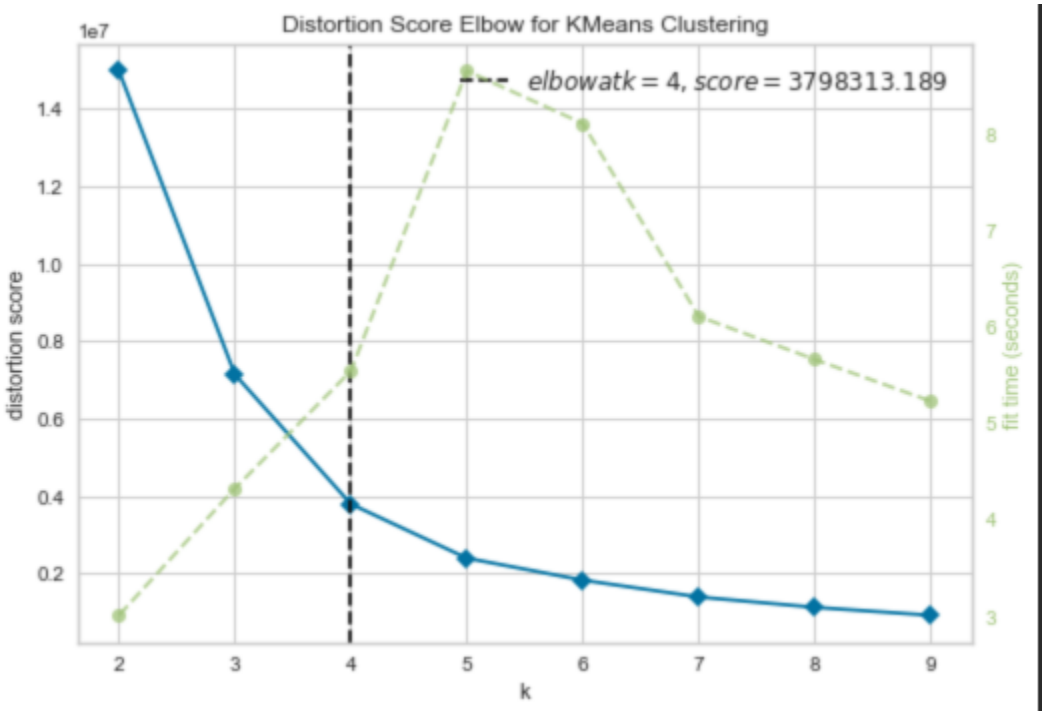
By extracting the neighborhood's latitude and longitude from the first dataset, we're able to input this information into Foursquare's API to obtain venue information in these neighborhoods.

We will also use the machine learning technique, K-Means to segment and cluster these neighborhoods so that we can group together to understand their similarities.

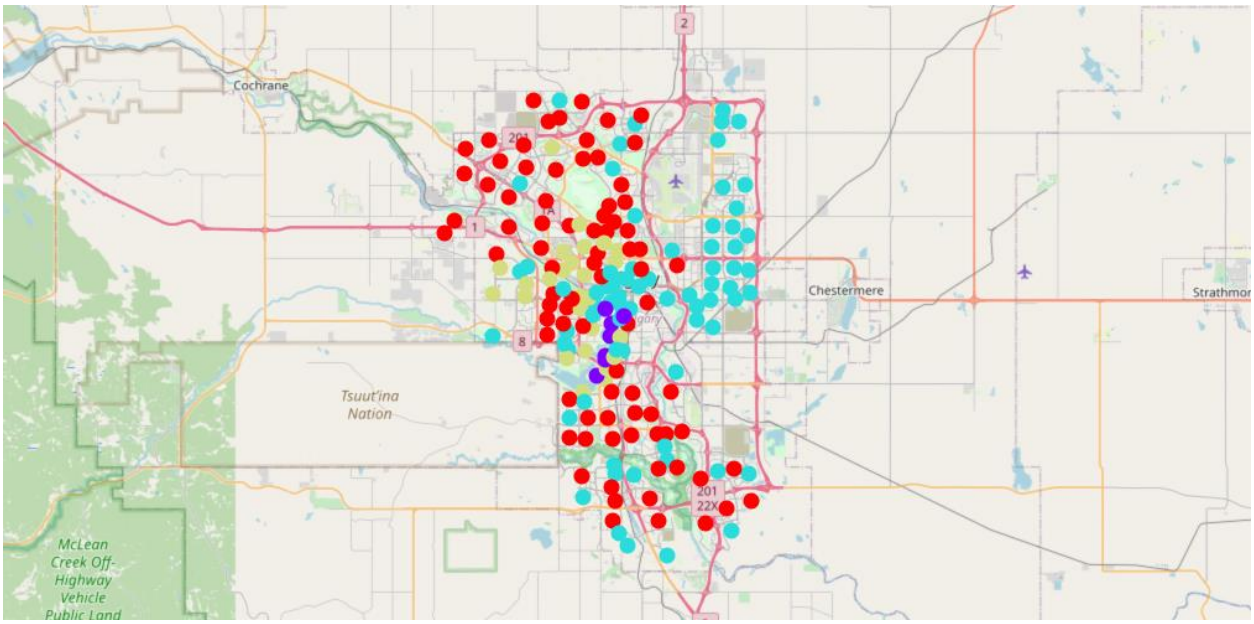
With all these methodologies, we will be able to come up with a best recommendation for clients looking to purchase a home based on their budgets and the amenities they would like nearby.

4.0 Result

First, we have used the elbow method to determine the ideal value of k in K-Means



With the result we know our best case is to use K-Means to cluster our neighborhoods into 4 clusters. Seen below.



5.0 Discussion

Our client has provided us with a budget of $400000 and has specified that Coffee Shops, and Pubs are their two most frequented venues so they'd prefer a neighborhood that has many of both.

By inputting the desired parameters provided by the client:

1. Budget – 400,000
2. Venues – Coffee Shop, Pub

We are provided with the following results:

Cluster	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	LATITUDE	LONGITUDE	Value
0	BOWNESS	Pub	Coffee Shop	Café	Stadium	Food & Drink Shop	Eastern European Restaurant	51.088532	-114.196535	383500.0
2	DOWNTOWN WEST END	Coffee Shop	Sandwich Place	Light Rail Station	Pub	Nightclub	Pharmacy	51.047856	-114.086675	199250.0
2	MILLRISE	Pub	Bank	Pizza Place	Coffee Shop	Golf Course	Grocery Store	50.916793	-114.078138	346500.0

With this data, we can recommend Bowness, Downtown West End, and Millrise as neighborhoods they should focus on for their home search.

We could also infer from these results that other Neighborhoods in Cluster 2 would be similar should they decide to expand their search outside of the 3 top Neighborhoods.

6.0 Conclusion

With that, we have concluded the best neighborhoods for our clients to pursue based on their needs are Bowness, Downtown West End, and Millrise. Should they need to broaden their search, cluster 2 would provide the closest fits to their needs.