Coursera Capstone Project Report

Introduction / Business Problem

Searching for an apartment can be a stressful endeavour; there are many variables to consider such as proximity to public transit, shopping centers, restaurants; safety, and price. Leveraging the data available from Foursquare, I will focus on find the lowest rent price for a 1-bedroom apartment comparable to other locations with similar venues in close proximity.

The result will be a clustered list of neighbourhoods with similar venue options in proximity and the median rental price of a 1-bedroom apartment – filtered by price to select the best option with all desired amenities.

Data

The dataset used is focused on 1-bedroom rental prices in the city of Calgary, Alberta, Canada neighbourhoods.

"geopy.geocoders" library will be used to augment the data to include address coordinates (latitude and longitude) for each neighbourhood.

The data will then be used as input to Foursquare to obtain the top 10 venues per neighbourhood.

Data samples are provided below:

	Neighborhood	Median Rent
0	Beltline	1213.0
1	Downtown Calgary	1208.0
2	Mission	1100.0
3	Skyview Ranch	1297.0
4	Varsity	1057.0
5	Bridgeland	1195.0
6	The Properties	1000.0

	Neighborhood	Median Rent
7	Crescent Heights	1088.0
8	Bankview	950.0
9	Hillhurst	1300.0
10	Sunalta	968.0
11	Thorncliffe	1129.0
12	Mount royal	1000.0
13	Bowness	925.0
14	Kingsland	1035.0
15	Capital Hill	1065.0
16	Altadore	1400.0
17	Acadia	1000.0
18	Sunnyside	925.0
19	Banff Trail	972.0
20	Inglewood	1340.0
21	Haysboro	1200.0
22	Windsor Park	1500.0
23	Forest Heights	875.0
24	West Hillhurst	1212.0
25	South Calgary	925.0

	Neighborhood	Median Rent
26	Auburn Bay	1325.0
27	Glenbrook	1049.0
28	Palliser	1156.0
29	Glenmorgan	1145.0
30	Killarney	695.0
31	Rosscarrock	1050.0
32	Radisson Heights	1131.0
33	Mayland Heights	1279.0
34	Applewood Park	1050.0
35	Lakeview	1045.0
36	Cedarbrae	1075.0
37	University Heights	1063.0
38	Quarry Park	1359.0
39	Spruce Cliff	1239.0
40	LIncoln Park	1150.0
41	Wildwood	1304

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beltline	51.040498	-114.072593	Milk Tiger Lounge	51.039357	-114.071403	Lounge
1	Beltline	51.040498	-114.072593	Central Memorial Park	51.041281	-114.070479	Park
2	Beltline	51.040498	-114.072593	Shelf Life Books	51.040737	-114.071526	Bookstore
3	Beltline	51.040498	-114.072593	Ox And Angela	51.037933	-114.073106	Latin American Restaurant
4	Beltline	51.040498	-114.072593	Yoga Passage	51.042799	-114.073123	Yoga Studio

Methodology

First data was collected which had the Calgary, AB neighbourhoods and median rent prices for a 1-bedroom apartment. This data was augmented using the geopy.geocoders library to include latitude and longitude information for each neighbourhood. A map was created to visualize the locations of neighbourhoods.
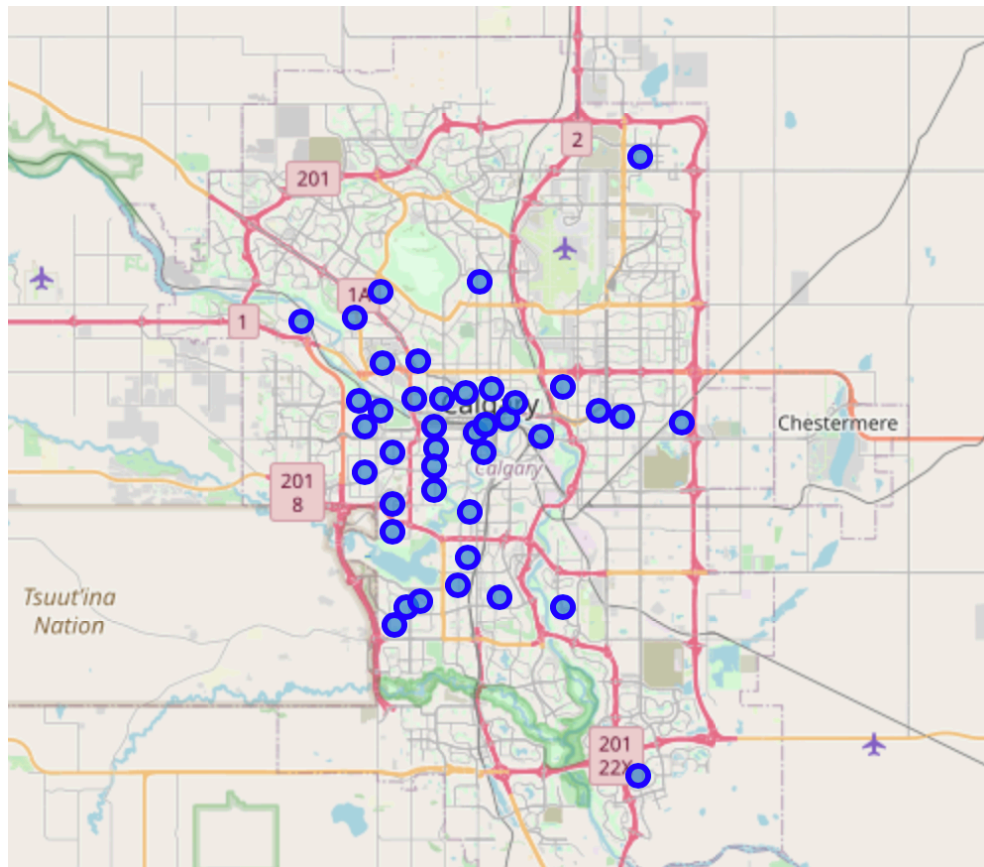
Next, Foursquare was used to extract the limit 100 closest venues to each neighbourhood frequently visited. The results were tallied up and compared by neighbourhood.

It was determined that an unsupervised machine learning algorithm would be best suited to find the top 5 clusters of neighbourhoods with similar active venues nearby.

K-means clustering was selected as the algorithm.

First the data had to be preprocessed by using one-hot encoding to separate the venues by neighbourhood.

The top 10 venues per neighbourhood were used to generate 5 clusters of similar areas.

A sample of the output is given below:

	Neighborhood	Median Rent	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	9th Most Common Venu
0	Beltline	1213.0	51.040498	-114.072593	0	Pub	Yoga Studio	Restaurant	Bar	Cocktail Bar	French Restaurant	Gym / Fitness Center	Lounge	Pizz Plac
1	Downtown Calgary	1208.0	51.046496	-114.050643	0	Coffee Shop	Music Venue	Hotel	BBQ Joint	Bakery	Shopping Plaza	Museum	Hostel	Ic Crear Sho
2	Mission	1100.0	51.031758	-114.066720	0	Coffee Shop	French Restaurant	Japanese Restaurant	Indian Restaurant	Gym / Fitness Center	Cosmetics Shop	Bakery	Breakfast Spot	Burge Joir
3	Skyview Ranch	1297.0	51.160534	-113.958135	2	Home Service	African Restaurant	Moroccan Restaurant	Music Store	Music Venue	New American Restaurant	Noodle House	Office	Othe Repa Sho
4	Varsity	1057.0	51.090998	-114.155759	0	Pizza Place	Gas Station	Yoga Studio	Café	Sandwich Place	Convenience Store	Shipping Store	Chinese Restaurant	Shoppin Ma

Results

The machine learning algorithm was a success at identifying the 5 clusters with most similarity in terms of common venues. Below is a mapping of the clusters colour coded:
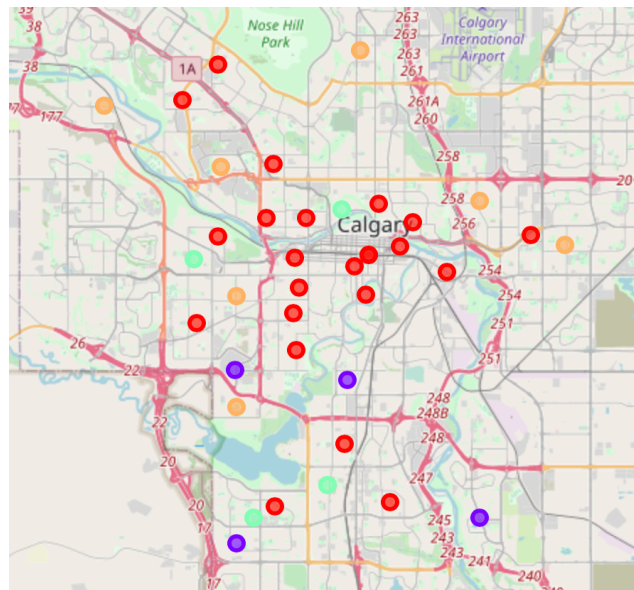
Finally, the clusters were filtered for the neighbourhood with the lowest median rent price. This would allow the user to select the neighbourhood with the lowest rent price but still enjoy the same top venues as the more expensive neighbourhoods.

Cluster Labels	Neighborhood	Median Rent	Latitude	Longitude
0	South Calgary	925.0	51.026135	-114.101907
1	Cedarbrae	1075.0	50.956318	-114.129323
2	Skyview Ranch	1297.0	51.160534	-113.958135
3	Sunnyside	925.0	51.057870	-114.078740
4	Killarney	695.0	51.031579	-114.129749

Discussion

The project was an overall success. A few recommendations would be to extend the analysis to include other features such as crime rates, neighbourhood populations, and transit proximity to help guide the user in making their final decision.

Conclusion

The project provides a useful data summary of the most similar neighbourhoods in Calgary in terms of top visited venues; however, it has the added benefit of providing median rent prices so that the user can get the best deal without sacrificing their favourite activities.