

The Battle of Alberta - Clustering Neighbourhoods in Calgary and Edmonton

Randy Wanye

April 19, 2020

1. Introduction

1.1 Background

The **Battle of Alberta** is used to refer to the rivalry between the two largest cities in the province of Alberta, Canada - Calgary and Edmonton. The term is mostly used in comparing sporting events between sports teams in these two cities. However, it is sometimes used in any comparison between these two cities. For more about the origins of the Battle of Alberta, see the [wikipedia page](#).

Edmonton is the administrative capital of the province, whilst Calgary is the largest city in the province and also recognised as a leader in the Canadian oil and gas industry. While public discourse focuses on their differences, these two cities are mostly similar whether it be about their demographics, the outdoor and urban adventures they offer, the food scene and urban planning.

1.2 Business Problem

The aim of this project is to determine how similar Caglary and Edmonton are, in terms of the experiences neighbourhoods in each city have to offer. This will not only contribute to the Battle of Alberta debate but can also help anyone looking to settle or emigrate to Alberta make a decision. To this end, recommendations will be provided to advise a person on a limited budget looking to settle in Alberta.

2. Data Description

To solve this problem, we will require three datasets. The three required sets of data are: neighbourhood data, geographical location data and real estate pricing data. Postal codes in the two cities will be used in obtaining the two sets of data.

Information about the various venues in a specific neighbourhood is also required. For this, the Foursquare API will be used to get this data.

2.1 Neighbourhood Data

The location data will be obtained by scraping postal codes information on [wikipedia](#) for Alberta. The information on wikipedia has postal codes by borough and neighbourhood names, this will help us limit ourselves to Calgary and Edmonton. The

wikipedia page also has latitude and longitudes for all postal codes. The neighbourhood data will consist of the following:

- Postal code
- Borough
- Neighbourhood
- Latitude
- Longitude

2.2 Real Estate Pricing Data

To get the real estate pricing data, the HousePriceHub [website](#) will be scrapped for house pricing in Calgary and Edmonton. HousePriceHub tracks the listing prices of houses in Canada, as well as provides insights on the real estate market in Canada. The real estate pricing data will consist of the following:

- Postal code
- Average Price
- Total listings

2.3 Foursquare API Data

In order to get specific neighbourhood information, the Foursquare API will be used. Foursquare is a geographical location data provider with information about venues and events within an area of interest. This location data includes venue names, locations, menus, photos and sometimes ratings. The neighbourhood information of Calgary and Edmonton will be connected to the Foursquare API and venues in each neighbourhood within a 500 meter radius will be obtained. The Foursquare API data will consist of the following:

- Neighbourhood
- Neighbourhood Latitude
- Neighbourhood Longitude
- Venue
- Venue Name
- Venue Latitude
- Venue Longitude
- Venue Category

These three data sets will assist us in determining the similarities between neighbourhoods in Calgary and Edmonton.

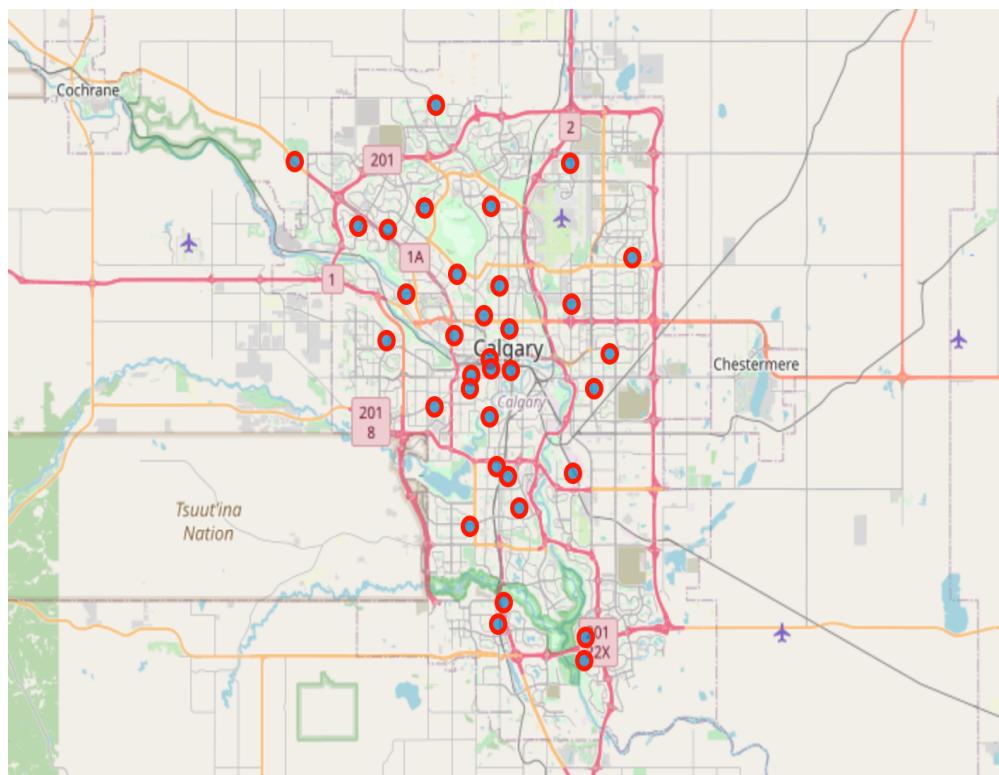
3. Methodology

The problem to be solved involves grouping various neighbourhoods in both cities based on their similarities. To this end, neighbourhoods will be explored, segmented and grouped to find similar neighbourhoods. This will be approached analytically by using

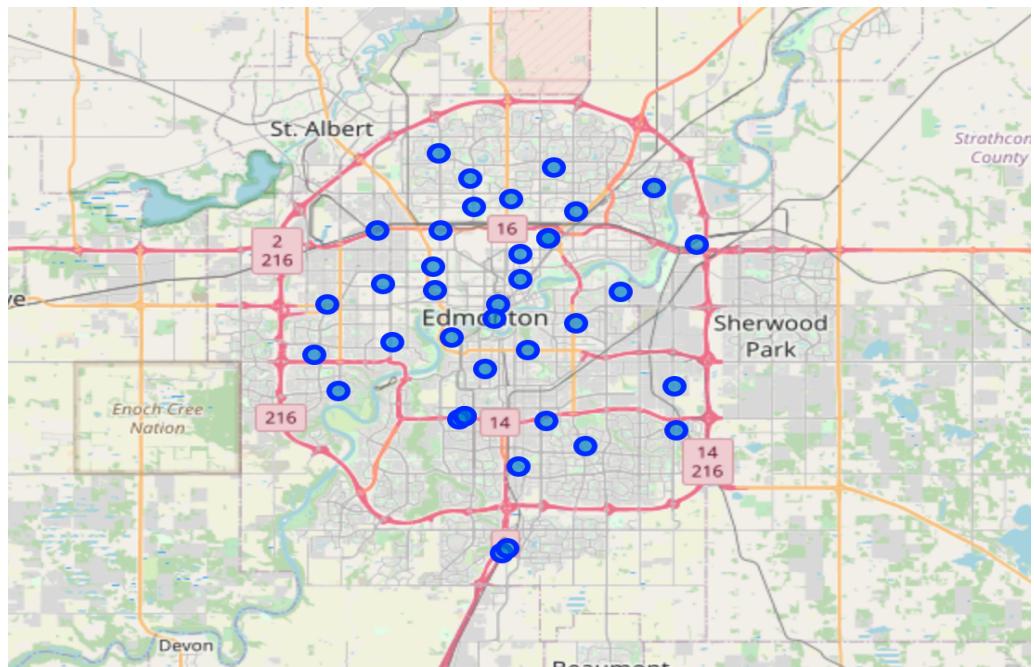
the KMeans clustering algorithm. All the neighbourhoods in Calgary and Edmonton will be explored to find the 10 most common venue categories in each neighbourhood and cluster them using KMeans clustering.

Finally, the most similar neighbourhoods will be selected and their real estate prices compared. A recommendation will be provided to advise a person looking to settle in Alberta with a limited budget on which city to choose.

Map of Calgary



Map of Edmonton

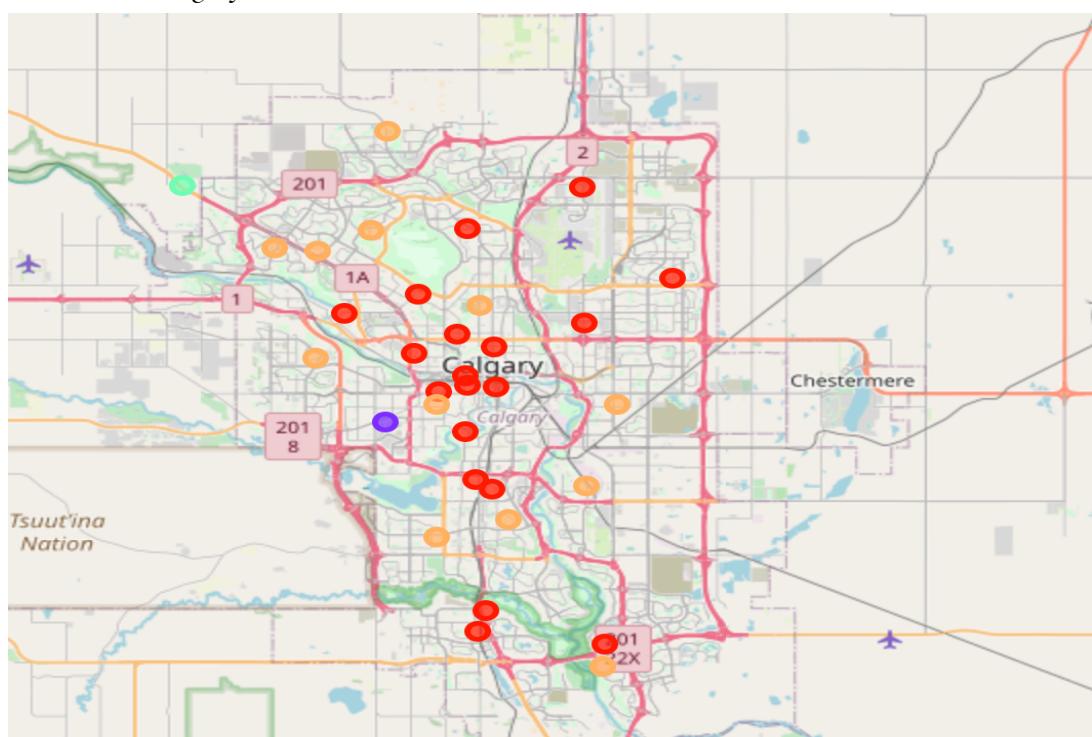


4. Results

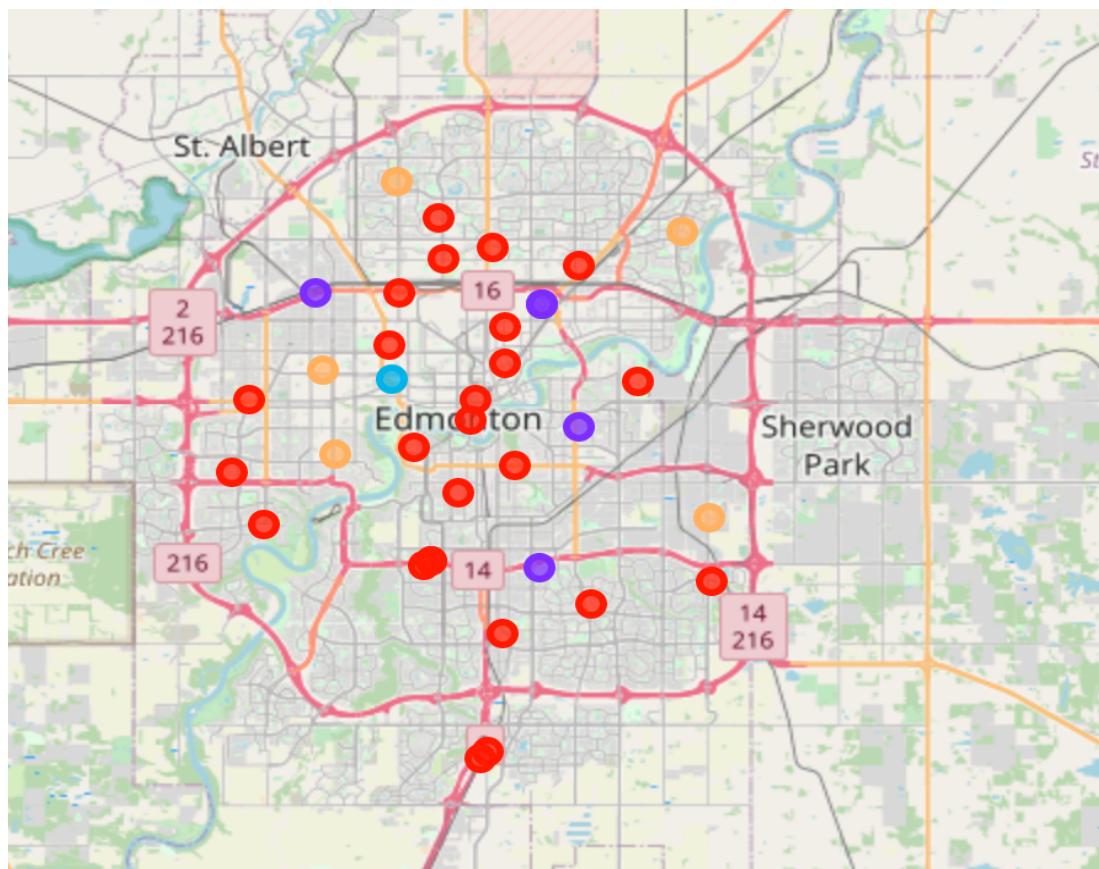
In total 69 neighbours in both cities were grouped into 5 clusters. Some neighbourhoods returned no venues and therefore could not be grouped.

4.1 Visualize Clusters

Clusters in Calgary



Clusters in Edmonton

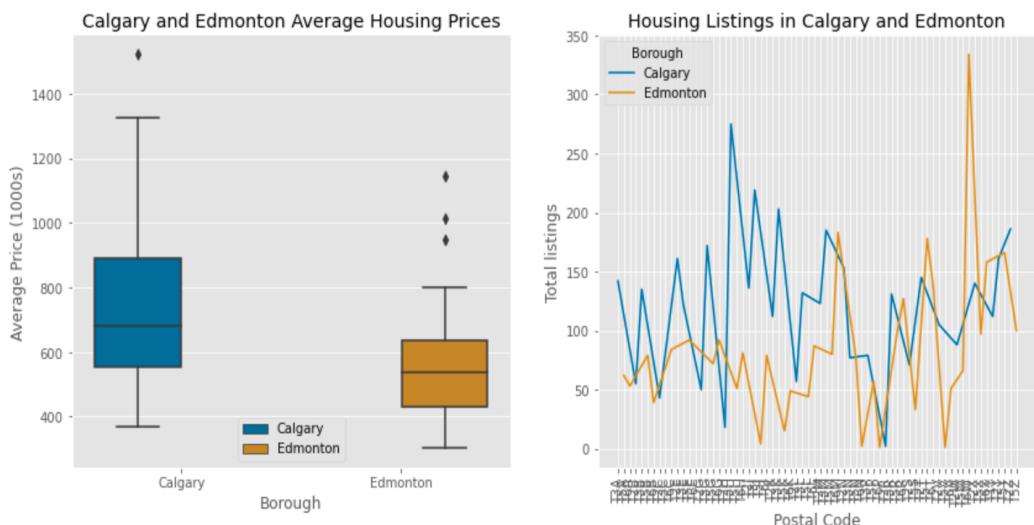


Summary of Neighbourhoods in Clusters

Cluster	Calgary	Edmonton
1	19	25
2	1	5
3	0	1
4	1	0
5	12	5

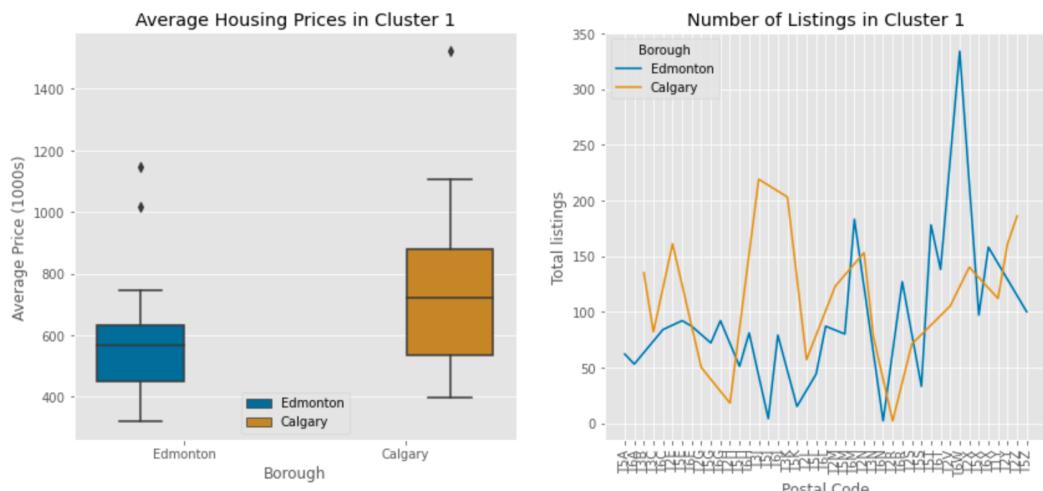
4.2 Compare Real Estate Prices

The figure below shows a comparison of real estate prices in all neighbourhoods of the two cities. The figure shows that the average housing prices in Calgary neighbourhoods are significantly higher than neighbourhoods in Edmonton.



We will now look at average housing prices in neighbourhoods that were in Clusters 1 and 5.

Cluster 1



Cluster 5



It can be observed that average housing prices of Calgary neighbourhood in Cluster 1 follows the earlier trend of being significantly higher than prices in Edmonton. In Cluster 5 where fewer neighbourhoods were grouped together, average housing prices in Calgary are still higher, however the difference is not so much.

5. Discussion

The neighbourhoods in Calgary and Edmonton were grouped into 5 different clusters. Together 69 neighbourhoods were clustered, the majority (44) of which were in Cluster 1 representing over 50% of the total neighbourhoods. Cluster 5 was the next largest group of neighbourhoods with 19. Clusters 2, 3 and 4 had very few neighbourhoods between the two cities, with clusters 3 and 4 having only one neighbourhood of Edmonton and Calgary respectively.

The results of the clustering suggests that the majority of venues in the neighbourhoods of Calgary and Edmonton are similar. The venues in these neighbourhoods tend to be places like restaurants, fast food outlets, pubs, liquor stores, movie theaters and parks/museums among others.

The real estate pricing data showed that average housing prices are significantly higher in Calgary compared to Edmonton. There are comparatively the same number of listings in both cities. One observation from the plot of the maps is that Calgary tends to be spread out whilst Edmonton is more compact.

Overall, the two cities are similar and offer the same type of experiences in their neighbourhoods. The recommendation for someone on a limited budget looking to move to Alberta will be to settle in Edmonton due to the lower average real estate pricing for a more affordable and less sprawl for decreased commute time within the city.

6. Conclusion

The purpose of this project was to explore and group neighbourhoods in Calgary and Edmonton in terms of the experiences they offer, and to provide recommendations to a person on a limited budget looking to settle in Alberta on which of the two cities to choose. Neighbourhoods in both cities were explored based on their postal codes.

Similar neighbourhoods were grouped together and real estate prices were compared. It was observed that neighbourhoods in both cities offer similar experiences, and that real estate prices in Calgary are significantly higher than prices in Edmonton.

Based on the findings, the recommendation to a person on a limited budget looking to settle in Alberta was to choose to move to Edmonton.