SEGMENTING AND CLUSTERING CALGARY NEIGHBOURHOODS BASED ON CURRENT REAL ESTATE AND LOCATION INFORMATION

By Moriah Rempel

Help Buyers find a home using Neighbourhood Clusters

■ Finding the perfect home can be a challenge as there are many different factors to consider when purchasing a home.

■ Clustering can be used to group Calgary neighbourhoods with similar real estate and venue traits

■ Buyers can be paired to these clusters to help them narrow down the neighbourhoods they are searching for homes in

Data

- Real Estate Data retrieved from realtor.ca using an API

	Neighbourhood	lng	lat	property_type	building_type	price	bathrooms	bedrooms	InteriorSize
0	Red Carpet	-113.941113	51.040125	Single Family	Mobile Home	$19,000	1	2 + 0	715 sqft
1	Red Carpet	-113.937916	51.038809	Single Family	Mobile Home	$29,000	1	2 + 0	960 sqft
2	Greenwood/Greenbriar	-114.214703	51.089611	Single Family	Mobile Home	$32,500	1	3 + 0	1249 sqft
3	Bowness	-114.216046	51.089684	Single Family	Mobile Home	$38,500	1	2 + 0	983 sqft
4	Arbour Lake	-114.217323	51.127602	Single Family	Mobile Home	$39,900	2	2 + 0	892 sqft

Data

- Neighbourhood list scraped from Wikipedia with coordinates found by averaging coordinates from neighbourhood listings in the real estate dataframe

	Neighbourhood	Latitude	Longitude
0	Abbeydale	51.059	-113.926
1	Acadia	50.9735	-114.061
2	Albert Park/Radisson Heights	51.0404	-113.993
3	Altadore	51.0168	-114.104
4	Applewood Park	51.0422	-113.93

Data

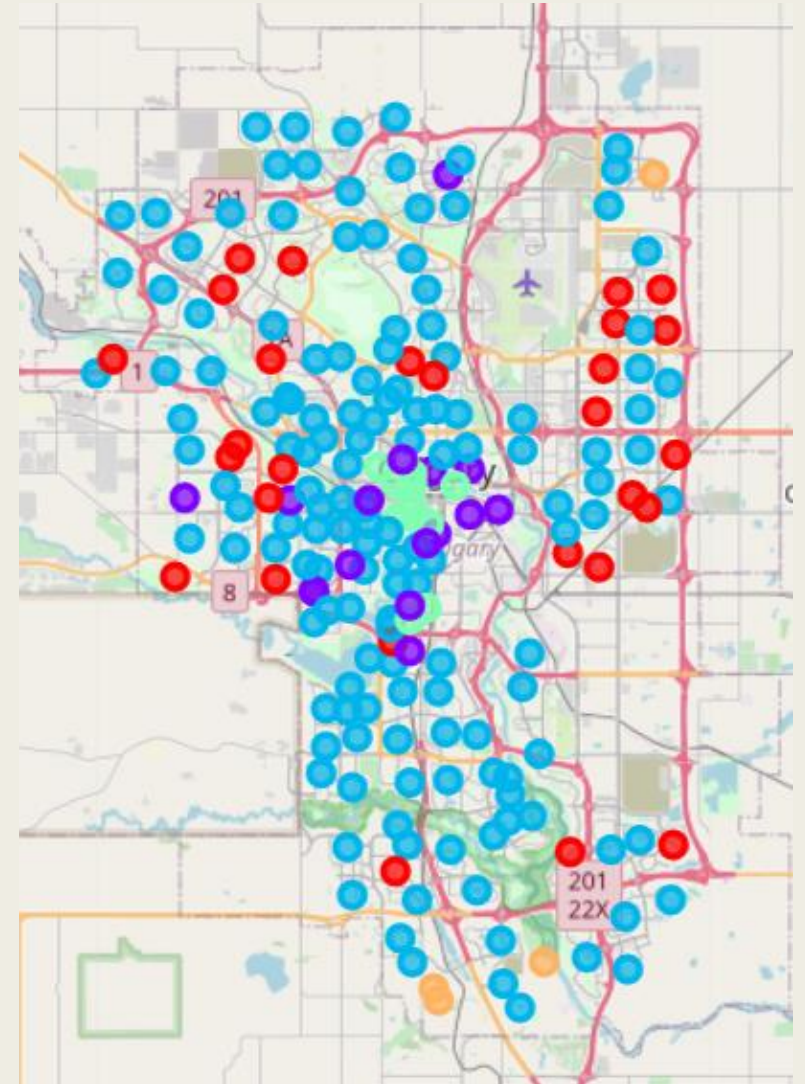- Neighbourhood venue data retrieved off Four Square using the API

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.059006	-113.925914	Subway	51.059239	-113.934423	Sandwich Place
1	Abbeydale	51.059006	-113.925914	Vanity Fitness	51.062982	-113.926238	Health & Beauty Service
2	Abbeydale	51.059006	-113.925914	Mac's	51.059376	-113.934425	Convenience Store
3	Abbeydale	51.059006	-113.925914	roadside pub	51.059277	-113.934529	Wings Joint
4	Abbeydale	51.059006	-113.925914	Magic Touch Stone Ltd	51.065379	-113.927304	Construction & Landscaping

K-means Clustering

- define a set number of clusters

- randomly distributing the set number of cluster centroids throughout the data

- data is assigned to the nearest cluster centroid

- the average of the data within the cluster is found and the centroid is moved to the average.

- Repeat assigning data and averaging until there is no change in the centroid location.
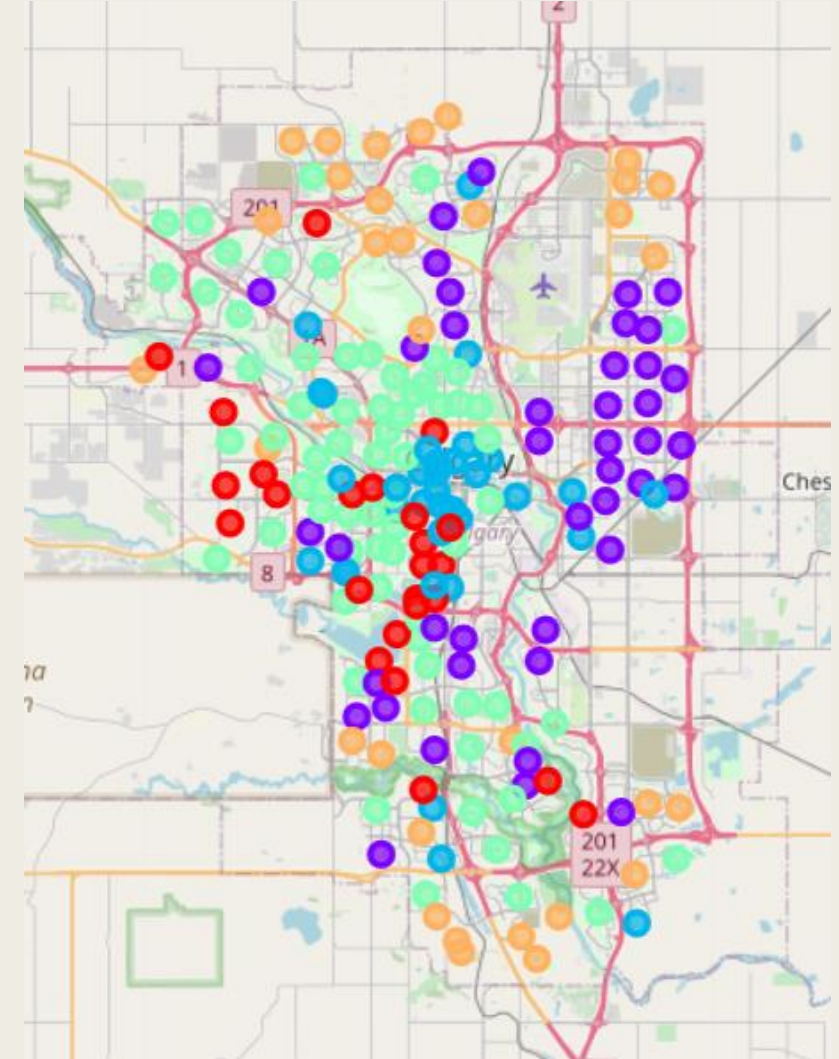
K-means Clustering Neighbourhoods using Venue Data

- K-means clustering of neighbourhoods performed using the venue data
  - *Top 5 venues*
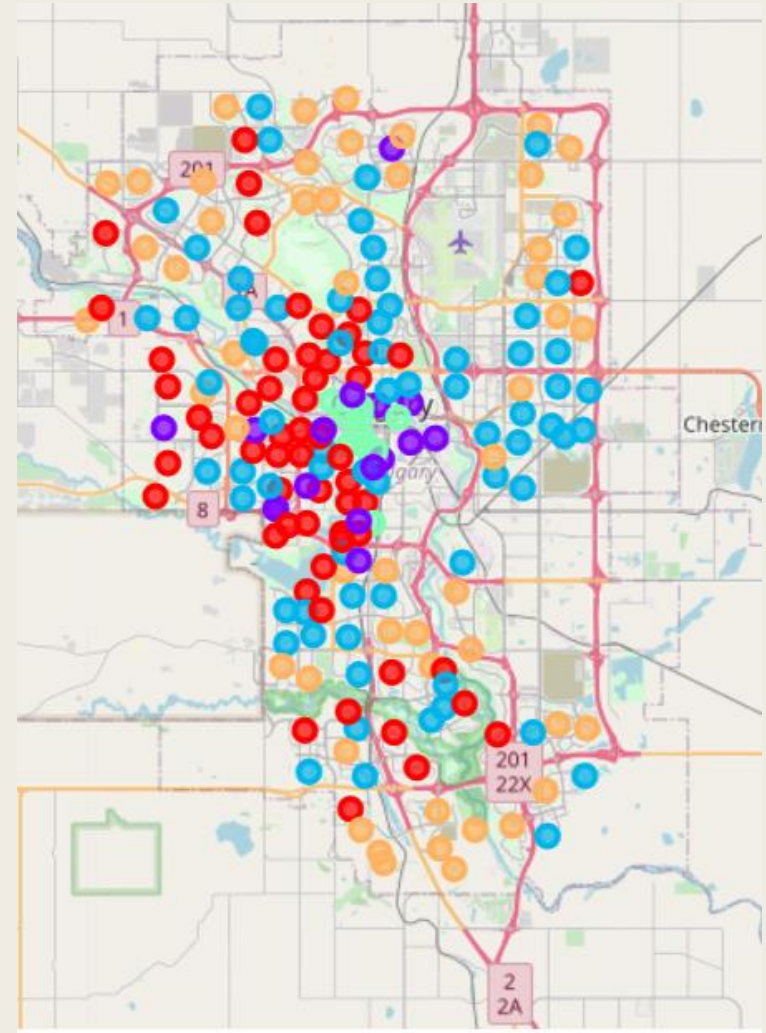 - *Number of venues*

K-means Clustering Neighbourhoods using Real Estate Data

- K-means clustering of neighbourhoods performed using the real estate data
  - *Building type (house, apartment, etc)*
  - *Price*
  - *Number of Bathrooms*
  - *Number of Bedrooms*
  - *Interior Size (sqft)*

K-means Clustering Neighbourhoods using both Venue and Real Estate Data

- K-means clustering of neighbourhoods performed using a combination of the venue and real estate data mentioned previously

Results

- 5 Clusters created
  - *Expensive and big houses, few venues*
  - *Less Expensive apartments and townhouses, good amount of venues*
  - *Less expensive, smaller houses, few venues*
  - *Mid-priced apartments and townhouses, in city core with lots of venues*
  - *Mid-priced, mid-sized houses, few venues*

Conclusion

- Neighbourhoods clustered into 5 groups based on real estate and venue data

- Future directions
  - *Split data by building type to find most and least expensive neighbhourhoods by comparing similar listings*
  - *Incorporate transit data*