

Evaluation of Calgary Communities and their Business Potentials Based on K-Means Method and Content Based Recommender System

By Siavash Fard 

June 2020



Outline

- Background
- Objective
- Data Sources
- Methodology
- Results and Discussion
- Conclusion
- References

Background

- Calgary is one of the fastest growing metropolitans in North America with the second youngest population in Canada
- Alberta families earned an average of \$72,700 in 2018, about \$11,000 more than the national average of \$61,400
- Calgary small business account for 95 percent of all businesses, the second highest number of small businesses per capita of the major cities in Canada
- Every year, in average, between 4 to 5 per cent of businesses decide to relocate

Objective

- Understanding Calgary communities and their demographics
- Understanding income distribution and property values for Calgary communities
- Clustering similar communities based on their demographics, income and property values
- Clustering communities based on their venues and frequencies
- Using recommender systems to suggest potential businesses for community of Montgomery

Data Sources

- Demographics data from the City of Calgary (<https://data.calgary.ca/>)
- Property assessments from the City of Calgary (<https://data.calgary.ca/>)
- Communities median income (great-news.ca)
- Communities venues and their ratings from Foursquare

Methodology – Data Wrangling

- Removing outliers and columns with missing values (e.g. SING_FAMILY_RATIO of 1.0 or 0.0)
- Removing irrelevant information (e.g. dog and cat counts, sector, converted dwellings, etc.)
- Converting counts to ratios:

$$\text{PRSCH_CHLD_RATIO} = \text{PRSCH_CHLD} / \text{RES_CNT}$$

$$\text{OWNSHIP_RATIO} = \text{OWNSHIP_CNT} / \text{DWELL_CNT}$$

$$\text{SING_FAMILY_RATIO} = \text{SING_FAMILY} / \text{DWELL_CNT}$$

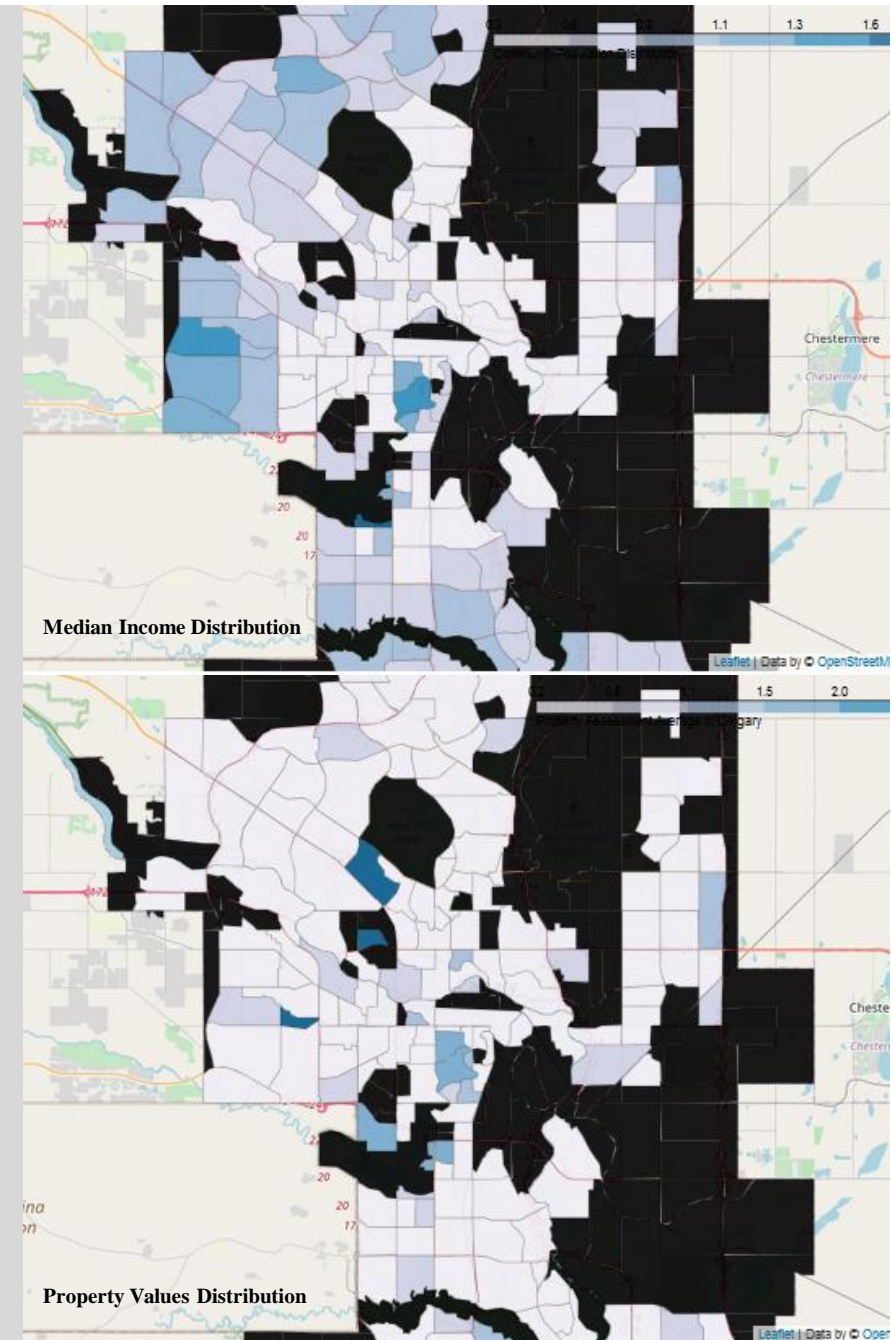
$$\text{RES_CNT_THOUSANDS_PER_SQ_KM} = \text{RES_CNT} / \text{AREA} * 10^3$$

Methodology – K-Means Method

- An iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**
- Tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible
- The less variation, the more homogeneous (similar) the data points are within the same cluster
- Communities clustering with K-Means have been implemented using sklearn

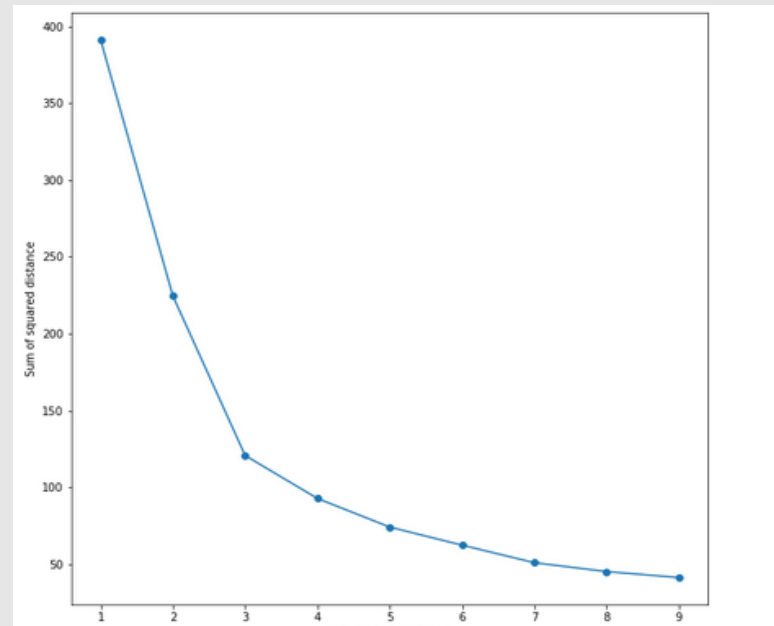
Results and Discussion

- Outer city communities have relatively higher income (except three inner city communities)
- Calgarians with higher income would prefer to reside in the western outer city communities
- Brentwood and University Heights communities have higher property values as they are closer to the University of Calgary



Results and Discussion – Elbow Method

- Based on the elbow method, cluster number of three (3) is recommended for K-Means modelling. However, cluster number of four (4) has been used to better differentiate between the communities.



Results and Discussion – Clustering

- Based on the K-Means method, Calgary communities were grouped into four (4) clusters

Clus_km	RES_CNT_THOUSA NDS_PER_SQ_KM	MEDIAN_INCOME_H UNDRED_THOUSAND	OWNSHIP_RATIO	SING_FAMILY_ RATIO	ASSESSED_V ALUE	PRSCH_CHLD_ RATIO
0	4.107248	0.589976	0.561534	0.490928	0.380450	0.079689
1	2.752917	0.672690	0.662595	0.612507	0.635986	0.066085
2	9.251569	0.347463	0.229917	0.045762	0.329398	0.020671
3	1.672295	0.820945	0.726011	0.709260	0.671622	0.055932

Results and Discussion – Clustering

- **Cluster 0:** Communities with relatively high population density, medium income level who own relatively more affordable properties; examples are South Calgary, Killarney and Crescent Heights
- **Cluster 1:** Communities with higher income and relatively higher property values and older demographics than Cluster 0. A large majority of Calgary communities fall into this cluster.
- **Cluster 2:** Communities with low income in highly populated communities. This cluster may be mostly younger adults with lower ownership ratio (most likely renting). Only a few communities fall into this category; examples are Mission, Beltline, Bankview and Lower Mount Royal.
- **Cluster 3:** Highest paid communities with lowest population density and high ownership ratios and highest property values with older demographics; examples are Upper Mount Royal, Elbow Park, Inglewood, Southview, Sprinbank Hill.

Results and Discussion – Community of Montgomery

- Foursquare API data was filtered to only include Cluster 3 communities to better evaluate the nearby venues
- There are approximately 40 communities in this cluster with total of 97 available ratings

	Neighborhood	Venue id	Venue Name	Venue Rating	Venue Category
0	MONTGOMERY	4c62047ceb82d13a969604d6	NOtaBLE	8.3	Restaurant
1	MONTGOMERY	4bb3ec87737d76b061e83a7c	Shouldice Athletic Park	8	Park
2	MONTGOMERY	55351231498ee5b605d43cb9	Tim Hortons	6.3	Coffee Shop
3	MONTGOMERY	4d2624b8d2668cfa2a14c5db	Subway	6.2	Sandwich Place
4	MONTGOMERY	4dcb5515c65bccd86744de6d	Pizza Hut	6.3	Pizza Place
5	MONTGOMERY	4dae25c55da3cca6f0a39f60	KFC	6.2	Fast Food Restaurant
6	MONTGOMERY	4bf4a947370e76b04796bd4a	7-Eleven	6	Convenience Store

Ratings available for the community of Montgomery

Results and Discussion – Community of Montgomery

- The community profile was calculated using the Venue Rating and the binarized category of each Venue Category
- Based on the content-based recommender system, Bayview, Wildwood and Britannia are the most similar communities to Montgomery

Neighborhoods	Weighted Average
BAYVIEW	0.084567
WILDWOOD	0.084567
BRITANNIA	0.053277
LAKE BONAVISTA	0.043793
EAGLE RIDGE	0.043693

Results and Discussion – Community of Montgomery

	Neighborhood	Venue id	Venue Name	Venue Rating	Venue Category
8	WILDWOOD	4bba279598c7ef3be3c43202	Edworthy Park	9.3	Park
9	WILDWOOD	4b3ffecbf964a520e2b325e3	Edworthy Dog Park	7.5	Dog Run
10	LAKE BONAVISTA	5325e880498eb79a55daadb7	The Lake House	7.4	Restaurant
11	LAKE BONAVISTA	4b7a26c3f964a520b0242fe3	Brewsters Lake Bonavista	7.1	Brewery
12	LAKE BONAVISTA	4d6d1a14cf7e41bd2cad8285	TD Canada Trust	6.5	Bank
13	LAKE BONAVISTA	4ba3cba4f964a520406038e3	Subway	6.4	Sandwich Place
14	LAKE BONAVISTA	4c9bdfe90e9bb1f7e7c3ce5f	Shoppers Drug Mart	6.2	Pharmacy
15	LAKE BONAVISTA	4bd4f1d429eb9c7460b592e1	Safeway Bonavista Shopping Plaza	6.1	Grocery Store
16	LAKE BONAVISTA	4b5b257af964a5206ce628e3	Lake Bonavista Promenade	5.6	Shopping Mall
24	BRITANNIA	54457979498e5391918f06c4	Village Ice Cream	8.5	Ice Cream Shop
25	BRITANNIA	5a679a8a59c42311fd6bec88	Monogram Coffee	8.3	Coffee Shop
26	BRITANNIA	4b771183f964a520007a2ee3	Sunterra Market	8	Food & Drink Shop
27	BRITANNIA	4ba12ae3f964a520419e37e3	Starbucks	7.1	Coffee Shop
28	BRITANNIA	4c08f9a6a1b32d7fbcd96f0	RBC Royal Bank	6.5	Bank
68	EAGLE RIDGE	4b0586e9f964a520c17422e3	Heritage Park Historical Village	8.5	History Museum
69	EAGLE RIDGE	4ba50faef964a5206cd738e3	Railroad Cafe	6.8	Sandwich Place
70	EAGLE RIDGE	4e8673f7d66a9b178e910fed	Petro-Canada	6.2	Gas Station
83	BAYVIEW	4b0586e9f964a520cb7422e3	South Glenmore Park	9.3	Park
84	BAYVIEW	4bc930f03740b71375705e65	Petro-Canada	6.3	Gas Station

Available Venue Rating and Name for top similar communities to Montgomery

Results and Discussion – Community of Montgomery

	NAME	Clus_ km	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	MONTGOMERY	3	1.0	Japanese Restaurant	Restaurant	Food & Drink Shop	Park	Coffee Shop
12	LAKE BONAVISTA	3	1.0	Park	Bank	Skating Rink	Shopping Mall	Brewery
25	INGLEWOOD	3	1.0	Gas Station	Diner	Cycle Studio	Pool	Liquor Store
65	VALLEY RIDGE	3	1.0	Restaurant	Greek Restaurant	Convenience Store	Yoga Studio	Flower Shop
117	GLENDALE	3	1.0	Chinese Restaurant	Pizza Place	Yoga Studio	Flower Shop	Garden Center
152	PATTERSON	3	1.0	Health & Beauty Service	Vietnamese Restaurant	Trail	Pizza Place	Convenience Store
155	CHARLESWOOD	3	1.0	Building	Pharmacy	Coffee Shop	Gas Station	Garden Center
159	BAYVIEW	3	1.0	Gas Station	Grocery Store	Other Great Outdoors	Park	Gastropub

Clustering of communities in Cluster 3 based on K-Means Method and their venue frequencies

Results and Discussion – Community of Montgomery

- Montgomery has relatively younger population with lower median income when compared to Cluster 3 communities.
- Assuming that we are not interested in businesses which already exist in the area, we can conclude that the following venues could be a match for Montgomery:
 - Coffee shop (local or high end)
 - Grocery store
 - Brewery
 - Gastropub

Conclusion

- Communities across Calgary were clustered and visualized based on their demographics and venue categories
- Calgary communities were clustered into four different groups based on their demographics, income level and average property assessments of communities
- A number of venues for one of the neighborhoods (Montgomery) has been recommended based on its clustered communities and using content-based recommender system.
- The same approach in this report can also be used to determine the similarity of two communities based on their demographics and venue category
- This report can be improved by taking the locations and demographics of venue reviewers on Foursquare to understand the customers behavior in different communities

References

- <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
- <https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/>
- <https://data.calgary.ca/Demographics/Census-by-Community-2019/rkfr-buzb>
- <https://data.calgary.ca/dataset/2020-Assessed-Property-Values/qwrb-nw8u>
- <https://data.calgary.ca/dataset/2020-Assessed-Property-Values/qwrb-nw8u>