Predicting Cluster Labels of Edmonton Banks based on Neighborhood Characteristics

Sean Hong

April 06, 2020

Business Problem: What are the neighborhood characteristics of establishments that are found near banks?

Hypothetically, banks would be found in areas where they can provide banking services to a good number of people enough to stay profitable.

- Do these establishments tend to cluster in the retail industry?
- In coffee or other refreshment services?
- How about other banks themselves?
- Given information on these types of establishments, can they be used to predict which cluster a particular bank branch belongs to?

Edmonton, Canada

Who are the major stakeholders? What can be gleaned from the analysis?

The beneficiaries are: *Edmonton's urban planning department*, *business investors*, and even the *banking companies* by providing useful insight into the businesses that cluster around banks, including but not limited to:

- Which banks are clustered around each other
- The types of establishments that are located near banks
- Relation between a specified bank cluster and the most common establishments that operate within said cluster

Data

To retrieve the banks' addresses in Edmonton, I referred to BankChart.ca for information from the site: https://bankchart.ca/catalogue/branches/3. Two variables are of particular interest:
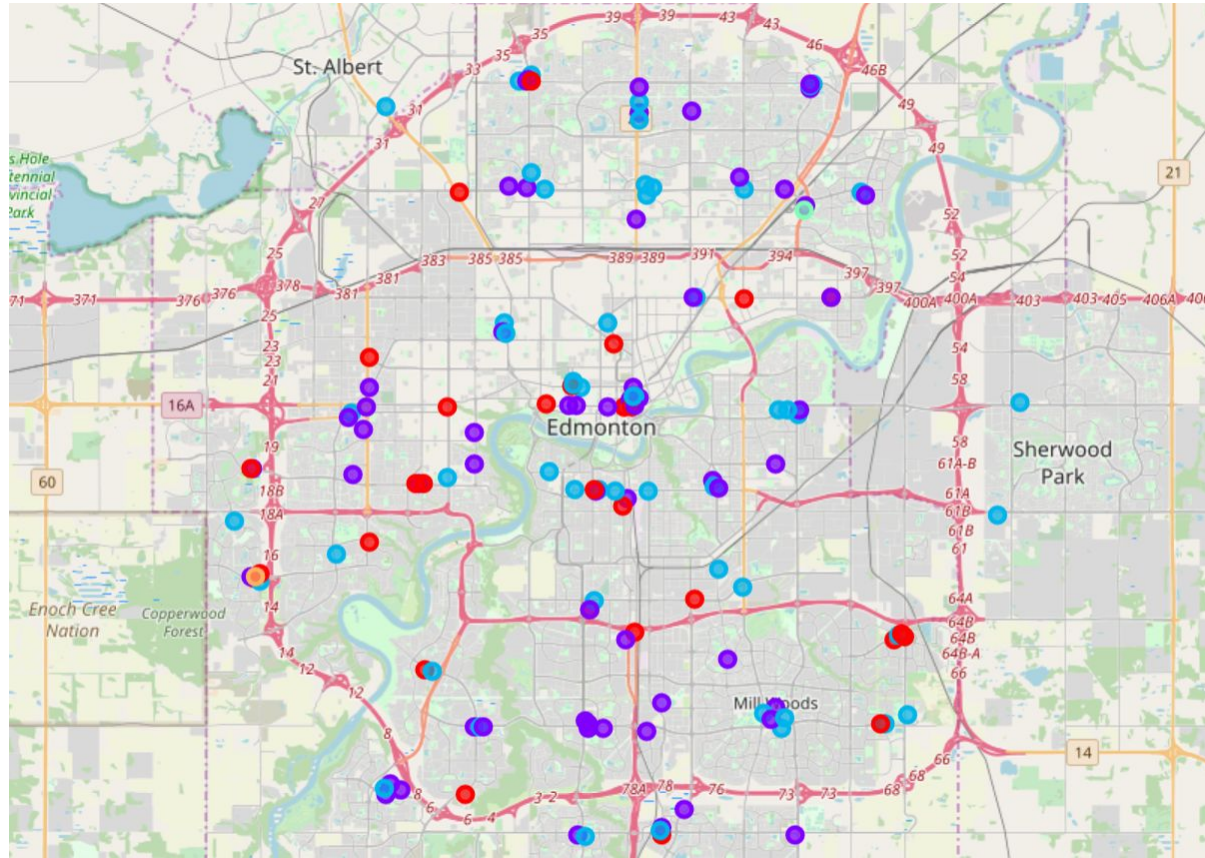
- Bank name
- Addresses of the bank branches

Once the relevant data was stored in a dataframe, the following were performed:

- The approximate locations of each branch's address were retrieved using Google Maps API reverse geocoding.
- Information on the venue category of the establishments were collected using Foursquare API.

	Bank Name	Address	long	lat
0	Bank of Montreal	208 KINGSWAY MALL, EDMONTON, AB T5G3A6: Edmonton	-113.505	53.5634
1	Bank of Montreal	2 HEBERT RD. UNIT 200, ST.ALBERT, AB T8N5T8: E...	-113.607	53.6223
2	Bank of Montreal	2304-24TH STREET NW, EDMONTON, AB T6T0G9: Edmo...	-113.378	53.4544
3	Bank of Montreal	17504 - 100 AVE NW, EDMONTON, AB T5S2S2: Edmonton	-113.623	53.5392
4	Bank of Montreal	300 222 BASELINE RD, SHERWOOD PARK, AB T8H1S8:...	-113.317	53.5421

The coordinates were retrieved using Google Maps API reverse geocoding.

A clustering analysis was performed to segment the bank into five clusters.

	Bank Name	Address	long	lat	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Bank of Montreal	208 KINGSWAY MALL, EDMONTON, AB T5G3A6: Edmonton	-113.505	53.5634	2	Fast Food Restaurant	Clothing Store	Coffee Shop
1	Bank of Montreal	2 HEBERT RD. UNIT 200, ST.ALBERT, AB T8N5T8: E...	-113.607	53.6223	2	Fast Food Restaurant	Grocery Store	Coffee Shop
2	Bank of Montreal	2304-24TH STREET NW, EDMONTON, AB T6T0G9: Edmo...	-113.378	53.4544	2	Liquor Store	Grocery Store	Pharmacy
3	Bank of Montreal	17504 - 100 AVE NW, EDMONTON, AB T5S2S2: Edmonton	-113.623	53.5392	2	Fast Food Restaurant	Coffee Shop	Hotel
4	Bank of Montreal	300 222 BASELINE RD, SHERWOOD PARK, AB T8H1S8:...	-113.317	53.5421	2	Fast Food Restaurant	Pizza Place	Gas Station

Foursquare data was used to form a '**neighborhood picture**' of the bank location, example:

"Bank of Montreal, located at *208 KINGSWAY MALL, EDMONTON, AB T5G3A6* with the neighbourhood description: "Fast Food Restaurant", Coffee Shop", and "Clothing Store", is found in Cluster 2."

Cluster Labels	index	Coffee Shop	Fast Food Restaurant	Pizza Place	Bank	Restaurant	Pharmacy	Sandwich Place	Grocery Store	Hotel
2	0	1	1	0	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0	1	0
2	2	0	0	0	0	0	1	0	1	0
2	3	1	1	0	0	0	0	0	0	1
2	4	0	1	1	0	0	0	0	0	0

Data used for classification analysis.

Results

	Establishment	Count	Percentage
0	Coffee Shop	67	0.151927
1	Fast Food Restaurant	51	0.115646
2	Bank	32	0.072562
3	Pizza Place	29	0.065760
4	Restaurant	25	0.056689
5	Sandwich Place	23	0.052154
6	Grocery Store	16	0.036281
7	Ice Cream Shop	15	0.034014
8	Pharmacy	14	0.031746
9	Hotel	11	0.024943
10	Italian Restaurant	9	0.020408

From the results, it was observed that 'Coffee Shop' is the most common type of establishment that is located close to Edmonton banks, at around 15%. This was followed by 'Fast Food Restaurant' and 'Bank', which came at 11% and 7% respectively. In fact, the top five venue categories (excluding bank) involved sit-down dining establishments and collectively account for almost half of establishments that are closest to banks.

Results

	F-1 Score	Jaccard Similarity Score
K-Nearest Neighbours	0.410	0.533
Logistic Regression	0.371	0.533

The logistic regression analysis that was performed resulted in a Jaccard Similarity Score of 0.533 and an F-1 score of 0.371. Whereas, the K-Nearest Neighbours resulted in a Jaccard Similarity Score of 0.533 and an F-1 score of 0.410. This suggests that the K-Nearest Neighbours is slightly better suited than logistic regression to predict the cluster labels given the data we have on the most common venues as provided by Foursquare.

Conclusion

The results of this project provides further insight into the methodology of assigning clusters following the theory of 'neighborhood characteristics'. Narrowing down the criteria for selecting which venue category to include (and exclude), and performing other classification techniques, are possible avenues for further refinement of the methodology.

As mentioned earlier, the main stakeholders that will benefit from this project are Edmonton's urban planning department, investors, and even the bank companies themselves. Inasmuch as they learn what types of establishments are located near banks, decisions on how land in Edmonton should be zoned, what businesses (e.g. coffee shops, clothing stores) are suited to be located nearby banks, and where the banks ideally should expand their footprint are all decisions that this study will help in facilitating.