

IDENTIFY RESIDENTIAL NEIGHBORHOODS IN TORONTO, CANADA

Capstone Project: The Battle of Neighborhoods

Maja Ferle

maja.ferle@coursera.com



Photo by [MURUCUTU](#) on [Unsplash](#)

February 2019

Table of contents

Introduction/Business problem	3
Data	4
Postal codes in Toronto.....	4
Geographical coordinates	4
Venue categories.....	5
Public schools	5
Population	5
Lists of best neighborhoods for families with children	6
Methodology	8
Exclude postal codes with few venue categories.....	8
Prepare the dataset for analysis.....	9
Perform clustering.....	9
Results	10
Cluster 0: Skating rink.....	11
Cluster 1: Residential.....	11
Cluster 2: Park	12
Cluster 3: Downtown.....	12
Cluster 4: Quick eats.....	13
Discussion	14
Compare with population counts.....	14
Best Toronto neighborhoods for families with children	15
Compare with school data	16
Conclusion	17
References.....	17
Acknowledgment.....	17

Introduction/Business problem

As part of Coursera IBM Data Science Professional Certificate Capstone project, we were asked to segment neighborhoods in Toronto. The main purpose of the exercise was to retrieve a list of neighborhoods in Toronto, find their geographical coordinates and use the coordinates as input to the Foursquare API, from where we can find the top venue categories in each neighborhood, for example restaurants, sporting venues, hotels, shops and many more. Using the frequency of venue categories in each neighborhood, we used the k-means clustering algorithm to cluster neighborhoods into segments with similar venue categories.

After completing the assignment and examining the resulting clusters, I noticed that some clusters probably represent residential areas, and other clusters represent more business/downtown areas. This opened the idea to explore the clusters further to validate this observation. As a practical outcome of this exploration I want to identify neighborhood clusters that are residential in nature and that would be suitable for families with children.

In addition to venue categories, I will introduce additional datasets about public schools and population counts to help me identify residential neighborhoods that are suitable for families with school age children.

Finally, I want to compare my results with a few blogs that I found online listing the best Toronto neighborhoods for families to determine how my analysis compares with the claims in these blogs.

The result of the analysis would be identification of clusters of neighborhoods that would be suitable for families with children. The benefit of knowing these results is that families who are looking for residence in Toronto might choose to live in neighborhoods that are family friendly but want to avoid the most popular neighborhoods that are mentioned frequently because those neighborhoods might become overcrowded and real estate prices might be high due to the demand. However, there may be similar neighborhoods that offer a comparable quality of life but are less known and thus might prove a better choice for a family with children to settle.

Data

This section introduces the datasets that will be used and their sources.

Postal codes in Toronto

My initial dataset is the [List of postal codes of Canada: M](#), representing postal codes in Toronto.

Postcode ↕	Borough ↕	Neighbourhood ↕
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue

Figure 1: Postal codes and neighborhoods in Toronto (sample)

Geographical coordinates

The previous dataset will be enriched with geographical coordinates using the [GeoPy](#) library from which I will add latitude and longitude for each postal code.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

Figure 2: Postal codes and neighborhoods in Toronto enriched with latitude and longitude information (sample)

Venue categories

Next, I will use the Foursquare API, using the latitude and longitude of each postal code to retrieve venues in a given radius around each location.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood, Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant

Figure 3: Sample Venue Categories returned by Foursquare API per neighborhood in Toronto

Public schools

I will enrich the previous dataset with information about how many Elementary and Secondary schools there are in each neighborhood by postal code. A list of public schools can be found on the [Ontario public school contact information](#) web site and this list can be transformed into a table with school counts.

	Postal Code	School Level	Number of Schools
0	M1B	Elementary	16
1	M1B	Secondary	1
2	M1C	Elementary	8
3	M1C	Secondary	2
4	M1E	Elementary	13
5	M1E	Secondary	9
6	M1G	Elementary	10
7	M1G	Secondary	2
8	M1H	Elementary	4
9	M1H	Secondary	1
10	M1J	Elementary	5

Figure 4: Number of elementary and secondary schools in Toronto by postal code (sample)

Population

Finally, I will enrich the dataset by adding population counts for each neighborhood by postal code as published on the Statistics Canada [Population and Dwelling Count Highlight Tables, 2016 Census](#) page.

	Geographic code	Population, 2016	Total private dwellings, 2016	Private dwellings occupied by usual residents, 2016
895	M1B	66108	20957	20230
896	M1C	35626	11588	11274
897	M1E	46943	17637	17161
898	M1G	29690	10116	9767
899	M1H	24383	9274	8985
900	M1J	36699	12797	12274
901	M1K	48434	18620	17930
902	M1L	35081	12884	12428
903	M1M	22913	8908	8623
904	M1N	22136	9535	9095
905	M1P	45571	17129	16540

Figure 5: Population in Toronto by postal code (sample)

Source: Statistics Canada. 2017. Population and dwelling counts, for Canada and forward sortation areas© as reported by the respondents, 2016 Census. Population and Dwelling Count Highlight Tables. 2016 Census. Statistics Canada Catalogue no. 98-402-X2016001. Ottawa. Released February 8, 2017.

The three values that are provided in this table represent:

- **Population, 2016:** number of Canadians whose usual place of residence is in that area, regardless of where they happened to be on census day
- **Total private dwellings, 2016:** total private dwellings and private dwellings occupied by usual residents in Canada
- **Private dwellings occupied by usual residents, 2016:** usual residents of Canada, not including tourists

For the analysis, we are looking at residential areas which means that we want to exclude tourists. Thus our column of interest is *Private dwellings occupied by usual residents, 2016*. We will divide this number by the total population count to derive a percentage of residents as compared to the total population. In residential areas where most of the population is not made up of tourists, the percentage should be higher as elsewhere.

Lists of best neighborhoods for families with children

The final part of my analysis will be an informal comparison between my resulting neighborhood segments to see how they match with neighborhoods that have been listed as the best Toronto neighborhoods for families with children on a few blogs that I found online.

Because the content of the blogs is unstructured data, I will have to manually match each neighborhood to its corresponding postal code. This will be feasible because the amount of data is relatively small.

Here is the unstructured data from 5 blogs that came up highest based on a Google search for "Toronto best neighborhoods for young families":

Top 10 Neighborhoods in Toronto for Families with Children	Best Neighborhoods in Toronto for Families	Top 6 Family Friendly Neighbourhoods in Toronto	Best Toronto Neighborhoods for Families	The 10 most family-friendly neighbourhoods in Toronto
Don Mills Deer Park Etobicoke West Mall Humber Valley Village Milliken North Toronto Riverdale Roncesvalles Runnymede The Beaches	Bedford Park Bloor West Village Riverdale Roncesvalles The Beaches	Bloor West Village Danforth Village Davisville Village Leslieville Rockcliffe-Smythe The Beaches	Allenby Bayview Village Bloor West Village Davisville Village Don Mills Humber Summit The Beaches	Allenby Bayview Village Bloor West Village Davisville Village Deer Park Don Mills Humber Summit The Beaches York Mills

Table 1: Lists of best Toronto neighborhoods for families with children (retrieved from a few blogs as listed in References)

Methodology

First, I collected all my data and stored it into dataframes. I scraped the Wikipedia page for a list of postal codes and added the latitude and longitude of each neighborhood.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

Figure 6: Postcodes with their geographical coordinates (sample)

Next, I used the Foursquare API to explore the neighborhoods. I passed the geographical coordinates of each neighborhood to the Foursquare API, which returned a list of venues in the neighborhood within a given radius and a limit of maximum 100 venues per neighborhood. In my case I decided to use 500 meters for the radius.

The resulting dataset became a list of all neighborhoods by postal code with added venues and venue categories.

	Postcode	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	M1E	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
4	M1E	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
5	M1E	43.763573	-79.188711	Marina Spa	43.766000	-79.191000	Spa
6	M1E	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
7	M1E	43.763573	-79.188711	Enterprise Rent-A-Car	43.764076	-79.193406	Rental Car Location

Figure 7: Neighborhoods by postal code with their venues and venue categories (sample)

Exclude postal codes with few venue categories

I counted the number of venue categories in each postal code and drew a histogram to get an understanding of how much data we have per postal code.

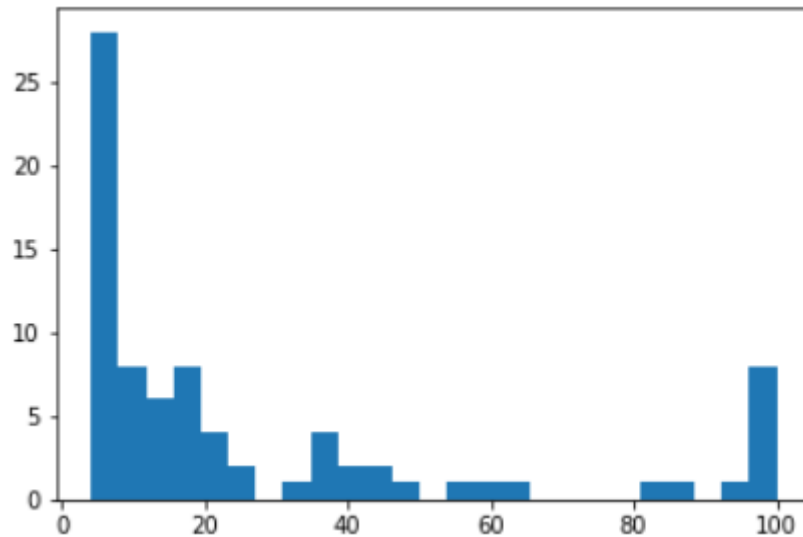


Figure 8: Number of venue categories per postal code

From the above diagram we can see that many postal codes have only a few venue categories (the leftmost bar on the diagram is much higher than the others). Because these postal codes contain too little data to make a meaningful analysis, I excluded them from the dataset. I decided to exclude all postal codes with fewer than 4 venue categories from the analysis.

Prepare the dataset for analysis

After excluding sparse data, I prepared the dataset for analysis. I used one hot encoding to pivot venue categories from rows to columns and calculated the mean of the frequency of occurrence of each category.

	Postcode	Accessories Store	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	...	Vegetarian / Vegan Restaurant	Video Game Store
0	M1E	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
1	M1H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
2	M1K	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
3	M1L	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
4	M1N	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
5	M1P	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
6	M1R	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
7	M1S	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
8	M1T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0
9	M1W	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0

Figure 9: Venue categories as columns (sample)

Perform clustering

Finally, I ran the k-means clustering algorithm on the above dataframe to derive clusters of neighborhoods by postal code using 5 as the number of clusters.

Results

First, I visualized the resulting clusters on a map as below:

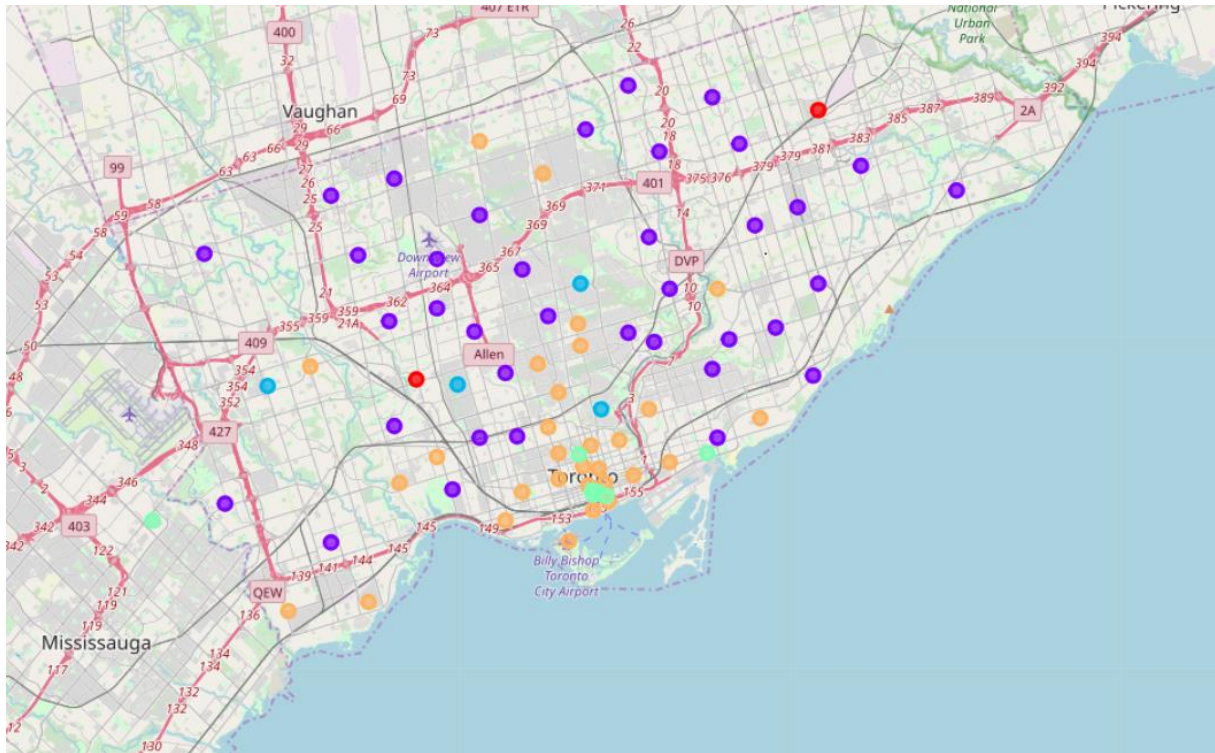


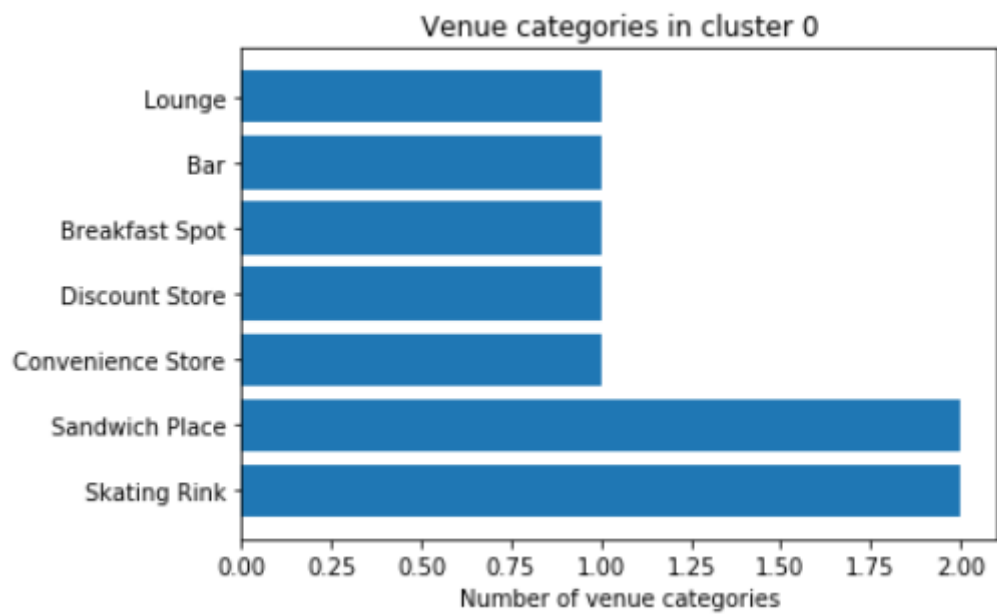
Figure 10: Resulting clusters displayed on a map of Toronto

The different colored dots represent these clusters:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4

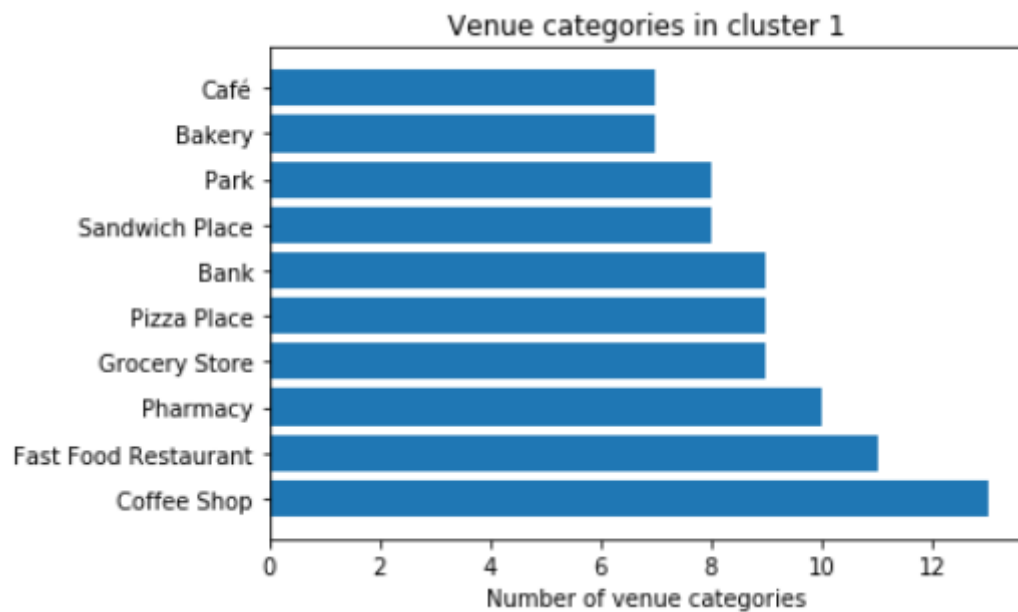
Then I explored each cluster to determine the common venue categories that define each cluster and I named the clusters accordingly.

Cluster 0: Skating rink



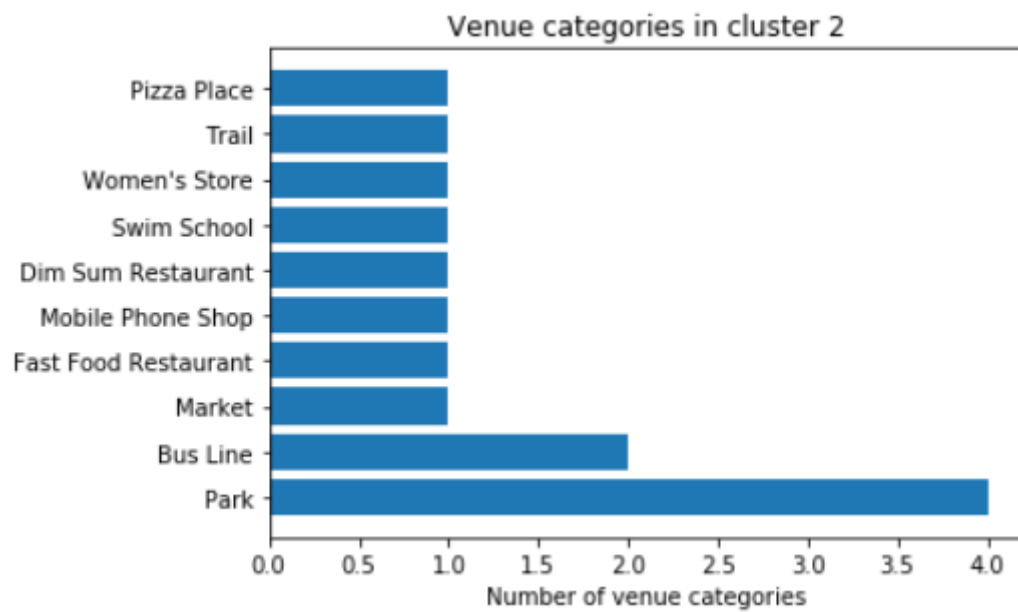
Venue categories in this cluster are skating rinks (no surprise, this is Canada!).

Cluster 1: Residential



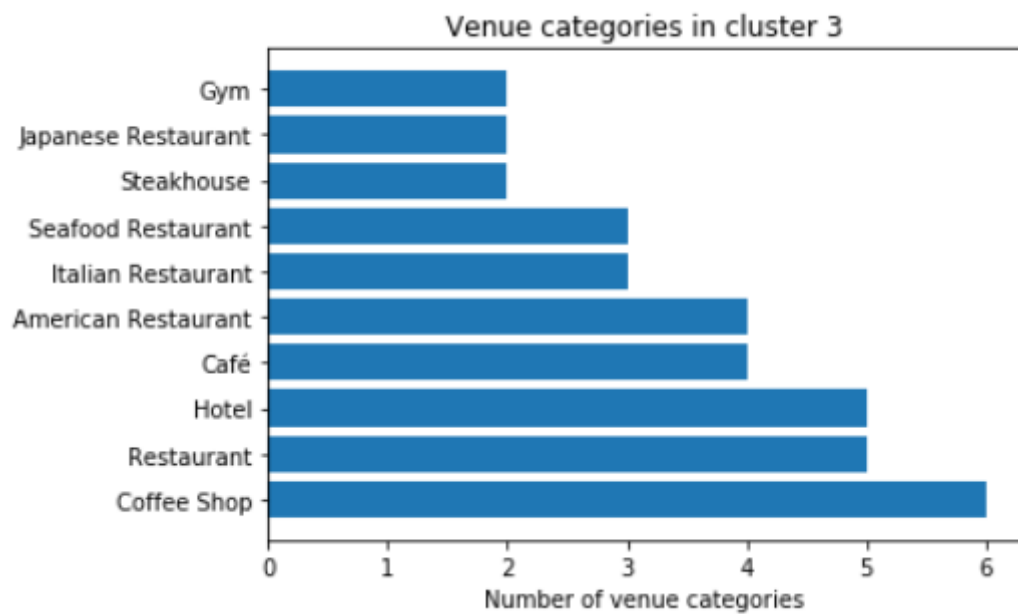
Venue categories in this cluster appear to be predominantly restaurants and shops, suggesting places that are found in residential areas.

Cluster 2: Park



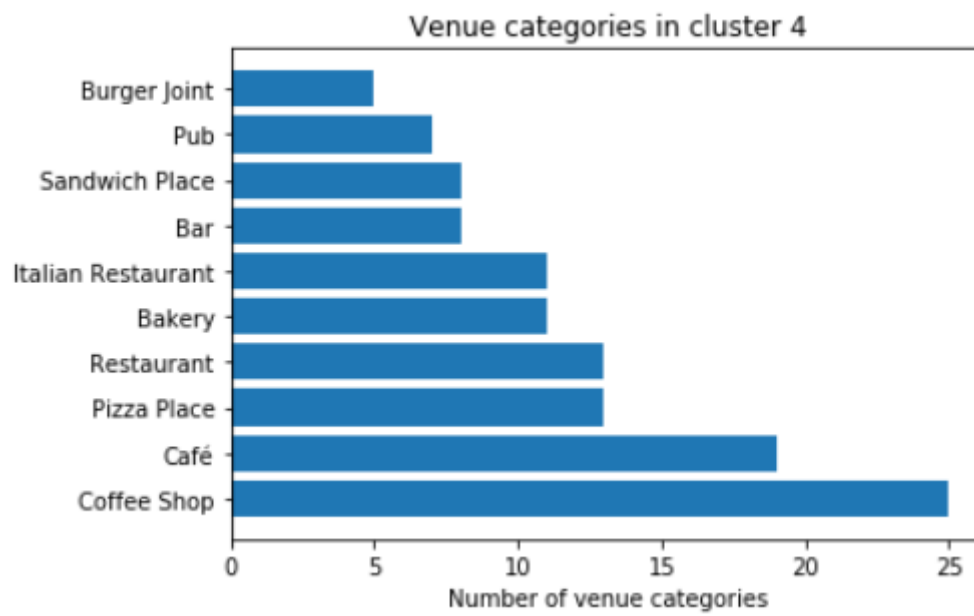
This is a cluster of neighborhoods with parks.

Cluster 3: Downtown



This cluster has mostly high-end restaurants and hotels with a couple of gyms which suggests it is made up of downtown neighborhoods.

Cluster 4: Quick eats



Venue categories in this cluster appear to be predominantly coffee shops and fast food type restaurants which all suggest places where one can find something quick to eat.

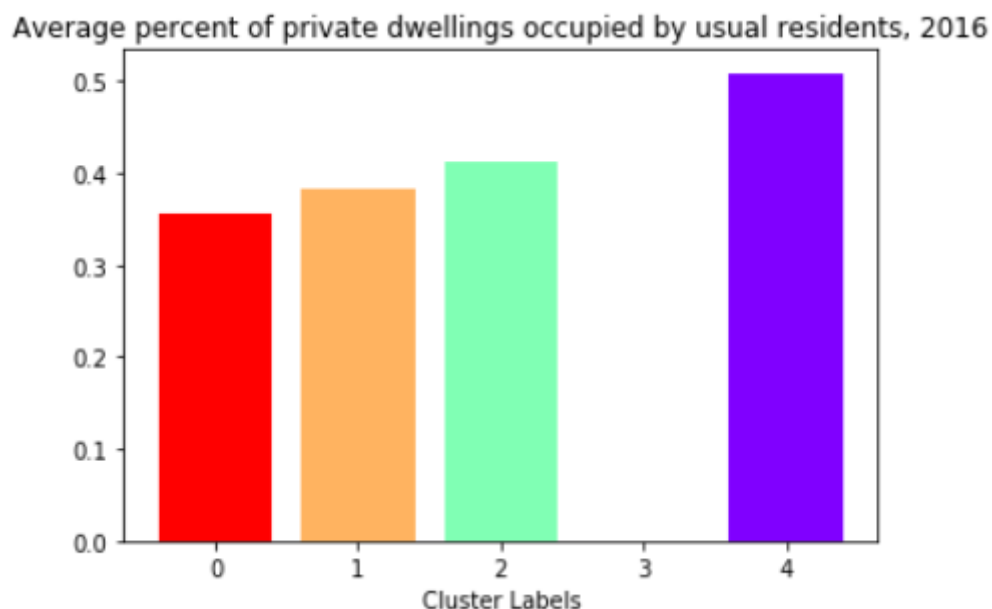
Discussion

Among the above clusters it would appear that the following clusters are best suited for families with children:

- **Cluster 1: Residential.** Venue categories in this cluster are predominantly shops with some interspersed coffee shops and restaurants as well as parks and sporting venues which is all suitable for families.
- **Cluster 4: Quick eats.** Venue categories in this cluster appear to be predominantly fast food restaurants, coffee shops, sandwich places, pizza places, grocery stores which all suggests places where one can find something quick to eat. These types of places are often located in shopping malls which suggest residential areas.

Compare with population counts

I calculated the average percent of private dwellings occupied by usual residents in each cluster.



We would expect that clusters 1 and 4 that we have identified as best suited for families with children would have a higher percent occupied than other neighborhoods, but in our case, this is only approximately so. Cluster 4 does have the highest percent occupied value, but cluster 1 is outranked by cluster 2 which according to our segmentation is mostly made up of parks where we would not expect many residents to live.

A possible explanation is that the neighborhoods with parks are suburban areas with individual family houses but without many venues such as restaurants and shops and for this reason they are not representative of the data that was used. Further analysis of this result is suggested.

Best Toronto neighborhoods for families with children

I have collected data from several blogs using the google search term "best toronto neighborhoods for families with children". I used the top search results and manually collected the names of suggested neighborhoods and stored them into an Excel file.

I read this Excel file into a dataframe and I joined it with my neighborhood data as a flag indicating that this particular neighborhood has been suggested as one of the best neighborhoods for families.

	Postal Code	TopFamilyFlag	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 Venue	10th Most Common Venue	Borough	Neighbourhood
0	M5N	1	4	Music Venue	Home Service	Garden	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Central Toronto	Roselawn
1	M2K	1	4	Japanese Restaurant	Café	Chinese Restaurant	Bank	NaN	NaN	NaN	NaN	NaN	NaN	North York	Bayview Village
2	M5M	1	4	Coffee Shop	Fast Food Restaurant	Italian Restaurant	Grocery Store	Thai Restaurant	Pharmacy	Pizza Place	Pub	Restaurant	Café	North York	Bedford Park, Lawrence Manor East
3	M6S	1	4	Coffee Shop	Pizza Place	Café	Italian Restaurant	Sushi Restaurant	Gourmet Shop	Bar	Restaurant	Pub	Dessert Shop	West Toronto	Runnymede, Swansea
4	M6H	1	4	Pharmacy	Bakery	Supermarket	Furniture / Home Store	Liquor Store	Fast Food Restaurant	Middle Eastern Restaurant	Discount Store	Music Venue	Park	West Toronto	Dovercourt Village, Dufferin

Figure 11: Neighborhood data with TopFamilyFlag added (sample)

Then I superimposed these flagged neighborhoods that have been suggested as the best neighborhoods for families on top of the clusters map:

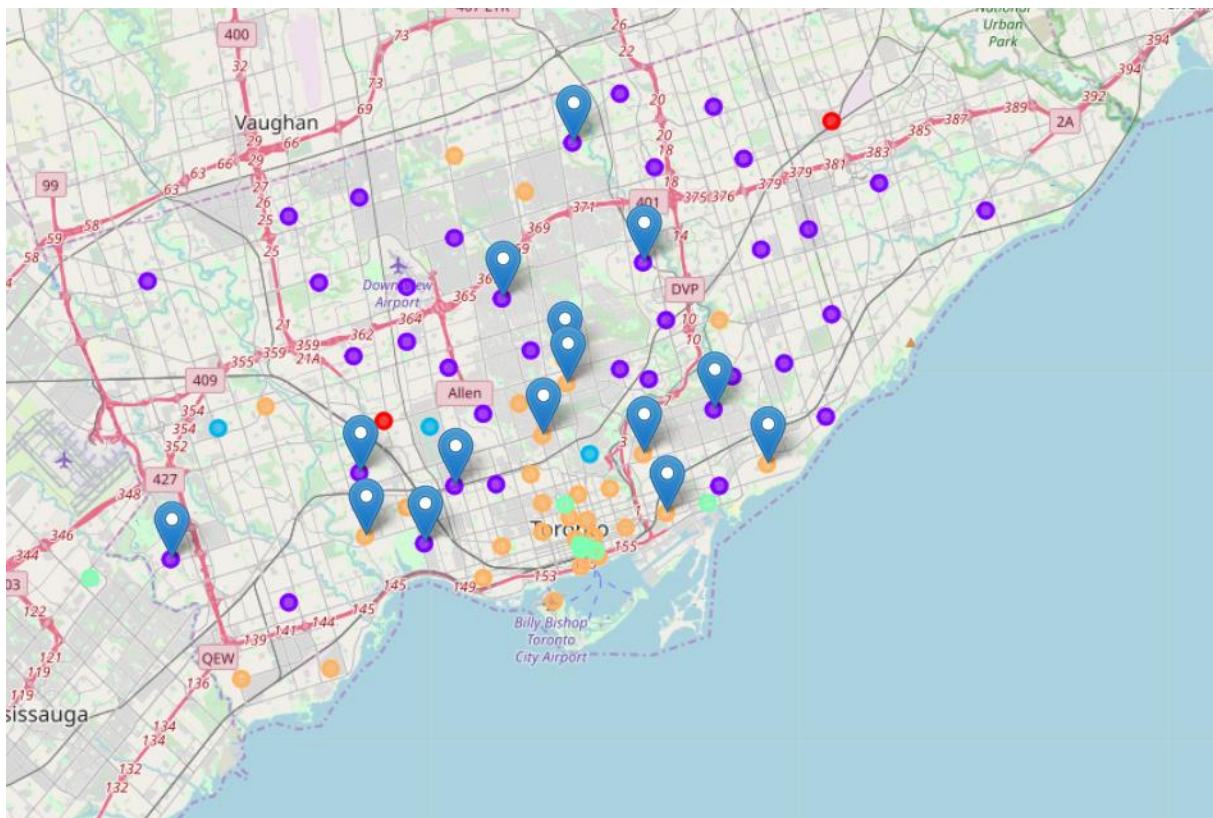


Figure 12: Flagged neighborhoods as markers on top of clusters

We can already see visually that all of the flagged neighborhoods correspond to either cluster 1 or cluster 4 which have both been identified by my analysis as residential neighborhoods suitable for families with children.

Checking counts, just to be sure we see that out of all flagged neighborhoods, 8 of them are indeed in cluster 1 and 7 of them in cluster 2.

	Postal Code	Neighbourhood
Cluster Labels		
1	8	8
4	7	7

Figure 13: Number of flagged neighborhoods per cluster

Compare with school data

Finally, I checked my assumption that neighborhoods that are best for families with children are those with a higher number of public elementary and secondary schools. It turns out that this is not the case, as can be seen on the below map where a choropleth map with number of schools per postal code is superimposed on the map with clusters and markers.

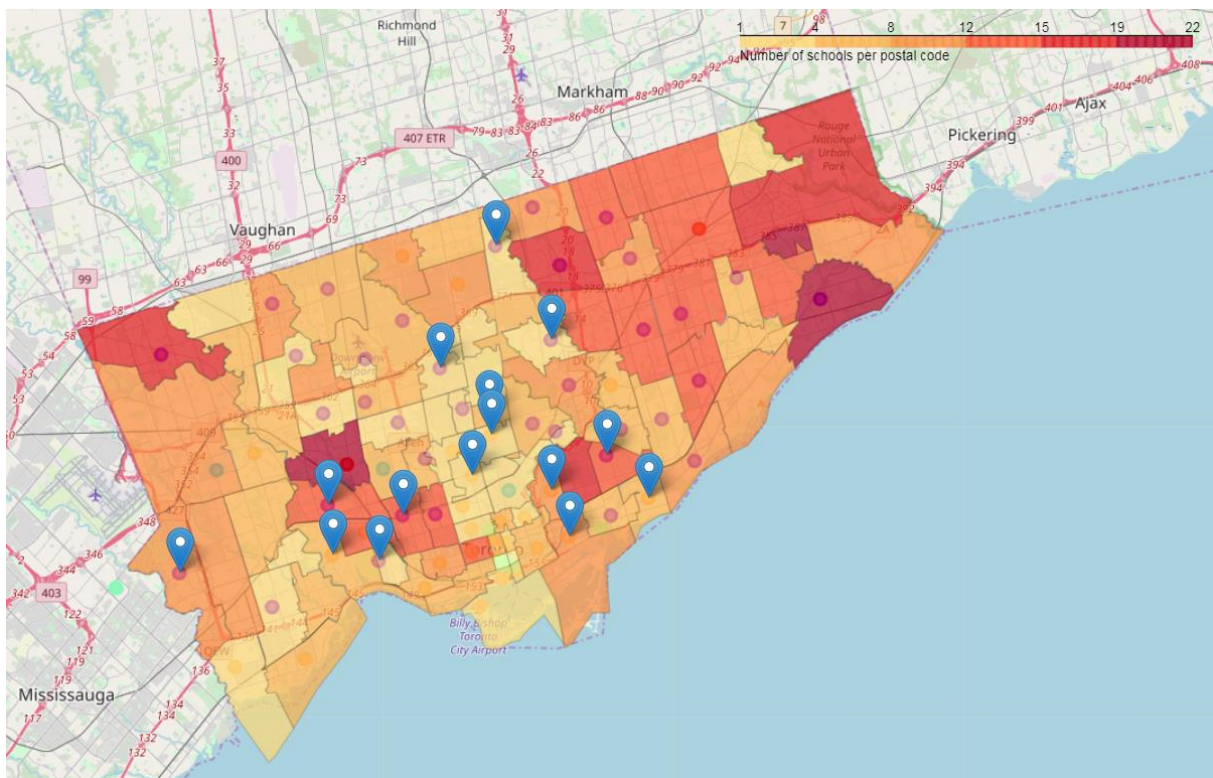


Figure 14: Number of schools as choropleth added to the previous map

A possible explanation might be that we are only looking at public schools. It could be that private schools are more desirable for families in residential areas but this would have to be examined further.

Conclusion

As my analysis has shown, the recommendation for families with children who are looking for a neighborhood to move to in Toronto would be to look for places in clusters 1 and 4, in addition to those that have been suggested on blogs.

Some suggestions for further analysis that came to mind while I was working on this project are:

- Perform additional analysis with school datasets, particularly look for private schools (only public schools were considered in my dataset)
- If possible, find and use additional datasets, with data about daycare centers, health facilities, crime rates, real estate listings, etc. as these would all be considered when a family is choosing a neighborhood to live in
- Since Foursquare provides crowd-sourced data, I found that data for Toronto is quite sparse (as compared to US cities, specifically New York which has much more data). In order to get more/better data, there might be other crowd-sourced data providers (TripAdvisor comes to mind, but this is just one example) that are more popular in Canada.

References

1. Coursera Capstone: Segmenting and Clustering Neighborhoods in Toronto, https://github.com/mferle/Coursera_Capstone/blob/master/Coursera_Capstone.ipynb
2. Ontario public school contact information, <https://www.ontario.ca/data/ontario-public-school-contact-information>
3. Population and Dwelling Count Highlight Tables, 2016 Census, <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&S=22&O=A> (accessed February 15, 2019)
4. Top 10 Neighborhoods in Toronto for Families With Children, <https://rankmyagent.com/realestate/top-10-neighborhoods-in-toronto-for-families-with-children/>
5. Best Neighborhoods in Toronto for Families, <https://highlevelmovers.ca/best-toronto-neighborhoods-for-families/>
6. Top 6 Family Friendly Neighbourhoods in Toronto, <https://www.zoocasa.com/blog/top-6-family-friendly-neighbourhoods-in-toronto/>
7. Best Toronto Neighborhoods for Families, <https://www.ryanroberts.ca/best-places-to-raise-a-family-in-toronto/>
8. The 10 most family-friendly Neighbourhoods in Toronto, <https://www.torontorentals.com/blog/the-5-most-kid-friendly-neighbourhoods-in-toronto>
9. Neighborhoods in Toronto, https://www.tripadvisor.com/NeighborhoodList-g155019-Toronto_Ontario.html
10. Foursquare API, <https://developer.foursquare.com/docs>

Acknowledgment

A big thanks to Coursera and IBM for offering the IBM Data Science Professional Certificate.