

The Battle of Neighbourhoods

Introduction

Toronto is one of the populated areas in Canada with an estimated population of over 6 million people. This brings opportunities for entrepreneurs to start or grow their business. Toronto is well known for its food

The objective of this project is to use Foursquare location data and clustering of venue information to determine what might be the 'best' neighbourhood in Toronto to open a restaurant. Pizza and Pasta are one of the most bought dishes in Toronto originating from Italy. Toronto is the fourth largest home to Italians with a population of over 500k, there are numerous opportunities to open a new Italian restaurant. Through this project, we will find the most suitable location for an entrepreneur to open a new Italian restaurant in Toronto, Canada.

Target Audience

Entrepreneurs or Business owners who want to open a new Italian Restaurant or grow their current business

Data

1. [Wikipedia](#)

Toronto - 103 FSAs [\[edit \]](#)

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0. However, the postal code M0R 8T0 is assigned to an [Amazon](#) warehouse in Mississauga, suggesting that Canada Post may have reserved the M0 FSA for high volume addresses.

Postal Code	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson
M6B	North York	Glencairn
M7B	Not assigned	Not assigned
M8B	Not assigned	Not assigned
M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale

2. [CSV File](#)

3. Venue Data using FourSquare API

Data Cleaning

1. Borough that are not assigned are dropped
2. If postal code appear twice, the rows will be combined
3. If there is a borough but neighbourhood is not assigned, neighbourhood will be the same as borough

Results after cleaning

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

After cleaning, we merge with geospatial data from CSV File base on Postal Code

Results after merging

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

After merging, venue data was taken from Foursquare API to get a list of all Venues in Toronto. We then merged the Foursquare data with the Neighbourhood data which gave us nearest Venue of each Neighbourhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Parkwoods	43.753259	-79.329656	Corrosion Service Company Limited	43.752432	-79.334661	Construction & Landscaping
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

Machine Learning

To analyse the data, we transformed the Venue Category into Numerical Data for Machine Learning algorithms. This technique is called **One hot encoding**. The results are as shown below

[illegible]

Next, we group rows by neighbourhood and by taking the mean of the frequency of occurrence of each category

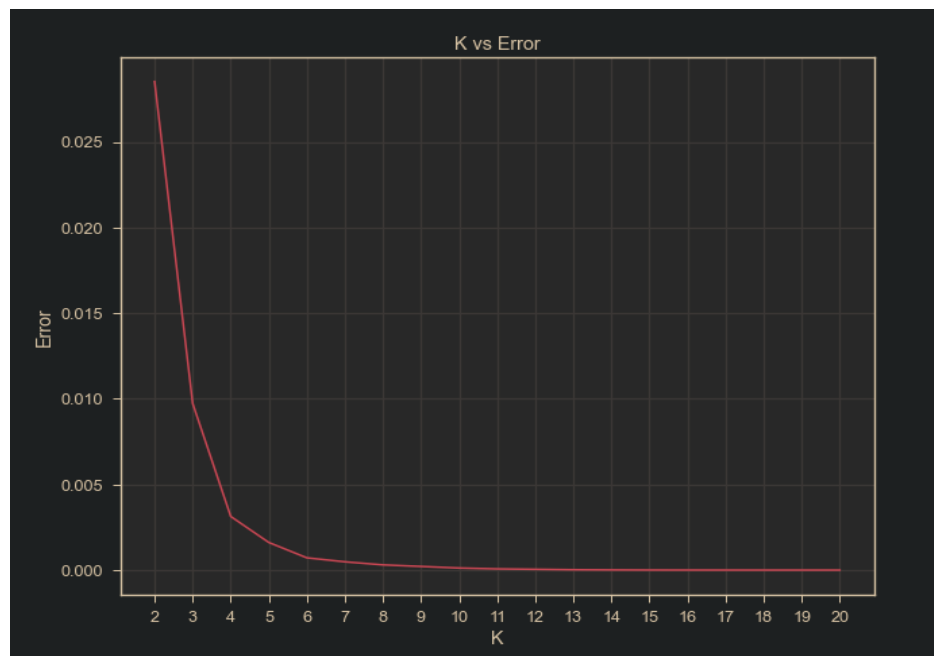
	Neighborhoods	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Train Station	Turkish Restaurant	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Warehouse Store	Wine Bar	Wings Joint	Women's Store	Yoga Studio
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.043478	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Next, a new dataframe was created based on neighbourhoods and italian restaurants

	Neighborhoods	Italian Restaurant
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.086957

Clustering

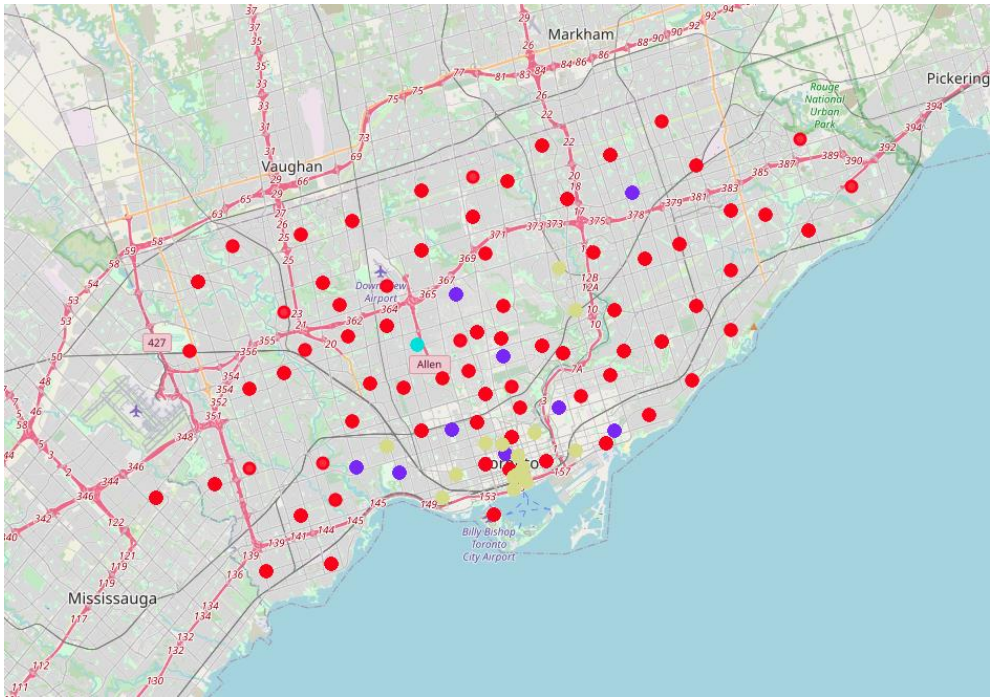
Now, we want to cluster the neighbourhoods base on neighbourhoods having similar averages of Italian restaurants in that neighbourhood. We use the K-means clustering method. For K-means clustering method, we need to get our optimum K using the Elbow Point method. The best K value is chosen at the point which line has the sharpest turn (K=4)



Now we use K-means clustering to group neighbourhoods into 4 clusters

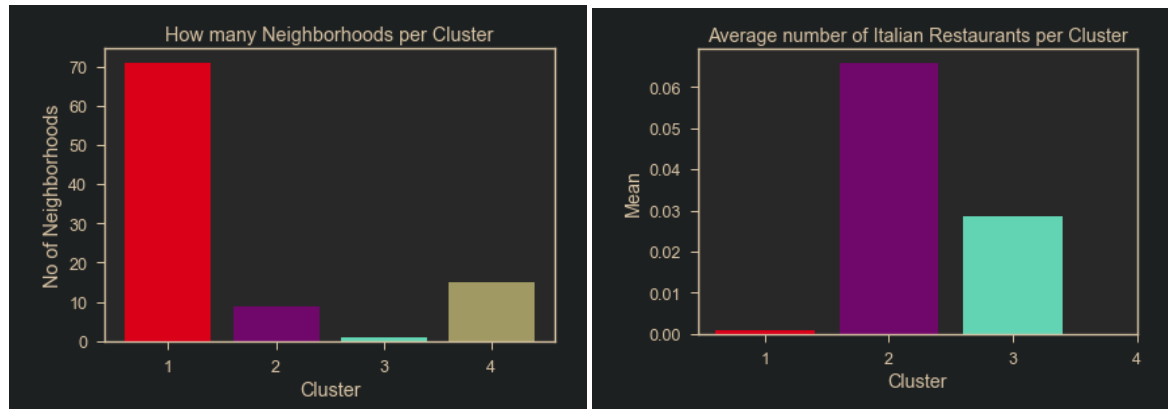
	Neighborhoods	Italian Restaurant	Cluster Labels
0	Agincourt	0.00	0
47	Leaside	0.00	0
46	Lawrence Park	0.00	0
46	Lawrence Park	0.00	0
46	Lawrence Park	0.00	0
...
36	High Park, The Junction South	0.04	3
35	Harbourfront East, Union Station, Toronto Islands	0.02	3
35	Harbourfront East, Union Station, Toronto Islands	0.02	3
35	Harbourfront East, Union Station, Toronto Islands	0.02	3
31	Garden District, Ryerson	0.03	3

Now we visualize the different clusters using folium



Analysis:

We have a total of 4 clusters. Let's check the total amount of neighbourhoods in each cluster and the average Italian Restaurants in that clusters



We see that cluster 3 has the least neighbourhoods and cluster 1 has the most (70). Then we compare average Italian Restaurants per cluster. Even though cluster 2 has about 10 neighbourhoods, it has the highest number of italian restaurants. Now let's analyze the clusters individually

Cluster 1 (Red)

	Borough	Neighbourhood	Italian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Parkwoods	0.0	0	43.753259	-79.529656	Brookbanks Park	43.751976	-79.332140	Park
1	North York	Parkwoods	0.0	0	43.753259	-79.529656	Corrosion Service Company Limited	43.752432	-79.334661	Construction & Landscaping
2	North York	Parkwoods	0.0	0	43.753259	-79.529656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	North York	Victoria Village	0.0	0	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
4	North York	Victoria Village	0.0	0	43.725882	-79.315572	The Frig	43.727051	-79.317418	French Restaurant
...
1002	Etobicoke	Mimico NW, The Queensway West, South of Bloo...	0.0	0	43.628841	-79.520999	Kingsway Boxing Club	43.627254	-79.526684	Gym
1003	Etobicoke	Mimico NW, The Queensway West, South of Bloo...	0.0	0	43.628841	-79.520999	Once Upon A Child	43.631075	-79.518290	Kids Store
1004	Etobicoke	Mimico NW, The Queensway West, South of Bloo...	0.0	0	43.628841	-79.520999	Koala Tan Tanning Salon & Sunless Spa	43.631370	-79.519006	Tanning Salon
1005	Etobicoke	Mimico NW, The Queensway West, South of Bloo...	0.0	0	43.628841	-79.520999	Islington Florist & Nursery	43.630156	-79.518718	Flower Shop
1006	Etobicoke	Mimico NW, The Queensway West, South of Bloo...	0.0	0	43.628841	-79.520999	Royal Canadian Legion #210	43.628855	-79.518903	Social Club

Cluster 1 has 71 unique neighbourhoods and only 1 Italian Restaurant in First Canadian Place, Underground city

Cluster 2 (Purple)

	Borough	Neighbourhood	Italian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Downtown Toronto	Central Bay Street	0.0625	1	43.657952	-79.387383	JOEY Eaton Centre	43.656094	-79.381878	New American Restaurant
1	Downtown Toronto	Central Bay Street	0.0625	1	43.657952	-79.387383	Panago	43.658258	-79.384313	Pizza Place
2	Downtown Toronto	Central Bay Street	0.0625	1	43.657952	-79.387383	The Queen and Beaver Public House	43.657472	-79.383524	Gastropub
3	Downtown Toronto	Central Bay Street	0.0625	1	43.657952	-79.387383	The Elm Tree Restaurant	43.657397	-79.383761	Modern European Restaurant
4	Downtown Toronto	Central Bay Street	0.0625	1	43.657952	-79.387383	KAKA	43.657457	-79.384192	Japanese Restaurant
...
253	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	0.1000	1	43.781638	-79.304302	Petro-Canada	43.779337	-79.307682	Gas Station
254	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	0.1000	1	43.781638	-79.304302	KFC	43.780400	-79.300700	Fast Food Restaurant
255	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	0.1000	1	43.781638	-79.304302	Popeyes Louisiana Kitchen	43.780476	-79.298460	Fried Chicken Joint
256	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	0.1000	1	43.781638	-79.304302	Little Caesars Pizza	43.780563	-79.298624	Pizza Place
257	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	0.1000	1	43.781638	-79.304302	Gusto Pizza	43.783607	-79.298983	Pizza Place

268 rows × 10 columns

Cluster 2 has 9 unique neighbourhoods but they have 17 Italian Restaurants

Cluster 3(Blue)

	Borough	Neighbourhood	Italian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude		Venue	Venue Latitude	Venue Longitude	Venue Category
0	Downtown Toronto	Queen's Park, Ontario Provincial Government	0.029412	3	43.662301	-79.389494		Como En Casa	43.665160	-79.384796	Mexican Restaurant
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	0.029412	3	43.662301	-79.389494		401 Games	43.663623	-79.384037	Hobby Shop
2	Downtown Toronto	Queen's Park, Ontario Provincial Government	0.029412	3	43.662301	-79.389494		Coach House Restaurant	43.664991	-79.384814	Diner
3	Downtown Toronto	Queen's Park, Ontario Provincial Government	0.029412	3	43.662301	-79.389494		Hart House Theatre	43.665571	-79.394616	Theater
4	Downtown Toronto	Queen's Park, Ontario Provincial Government	0.029412	3	43.662301	-79.389494		Flock Rotisserie + Greens	43.659167	-79.389475	Fried Chicken Joint
...
941	Downtown Toronto	St. James Town, Cabbagetown	0.041667	3	43.667967	-79.367675		TD Canada Trust	43.664655	-79.367887	Bank
942	Downtown Toronto	St. James Town, Cabbagetown	0.041667	3	43.667967	-79.367675		Thai Room - Carlton	43.664159	-79.368189	Thai Restaurant
943	Downtown Toronto	St. James Town, Cabbagetown	0.041667	3	43.667967	-79.367675		Winchester Park	43.666231	-79.371631	Playground
944	Downtown Toronto	St. James Town, Cabbagetown	0.041667	3	43.667967	-79.367675		LCBO	43.665586	-79.368531	Liquor Store
945	Downtown Toronto	St. James Town, Cabbagetown	0.041667	3	43.667967	-79.367675		Daniel et Daniel Event Creation & Catering	43.664384	-79.368328	Bakery

Cluster 3 has 15 unique neighbourhoods and 26 Italian Restaurants

Cluster 4(Turquoise)

	Borough	Neighbourhood	Italian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude		Venue	Venue Latitude	Venue Longitude	Venue Category
--	---------	---------------	--------------------	----------------	-----------------------	------------------------	--	-------	----------------	-----------------	----------------

Cluster 4 has 0 neighbourhoods and 0 Italian Restaurants

Discussion

Most of the Italian restaurants are located in cluster 3 represented by the blue circles. We know that Cluster 4 has no neighbourhoods and no Italian restaurants. Looking at all of the clusters, the most appropriate cluster to put a new Italian restaurant is in cluster 1 as it has 71 neighbourhoods but only 1 Italian restaurant, therefore eliminating any competition. Some drawbacks of this analysis are clustering based on data obtain from Foursquare API and also does not consider the population across the neighbourhoods which will affect the business of the new Italian restaurant

Conclusion

A business problem was came up and it was analysed by how a data scientist would analyse the problem. Different python libraries were used to consolidate data, visualize datasets, analyse datasets and eventually coming up with a business solution. We can now address more complex problems using data science.