

Battle of Toronto Neighbourhoods

M S Thiruchenduran

Jan 6, 2019

Table of Contents

<i>Introduction</i>	3
<i>Business Problem</i>	3
<i>Stakeholders</i>	3
<i>Data</i>	3
<i>Methodology</i>	3
<i>Solution Approach</i>	3
<i>Data Collection and Cleansing</i>	4
<i>Machine Learning Algorithm</i>	8
<i>Result</i>	8
<i>Discussion</i>	10
<i>Conclusion</i>	11

Introduction

Toronto is the largest city of Canada and most ethnically diverse city. Many immigrant cultures have brought their traditions, language and food. The cuisine of Toronto reflects its size and multicultural diversity. It provides ample opportunity for people to start their own business and one of the most common one is to open restaurant. The big question when coming to open a restaurant is the location. This report covers the details that would help to narrow down the location to open the restaurant.

Business Problem

There is lot of research one has to do before starting the restaurant like the idea, resources, budget, etc. We are considering aspiring entrepreneur who has unique idea to start an Indian Restaurant in Toronto. The most important problem is to identify the location to start the restaurant. Location is one of the key to restaurant success. In an ethnically diverse city like Toronto, this is even harder to narrow down the location to start the restaurant.

Stakeholders

Aspiring entrepreneur who wants to start an Indian Restaurant in the neighbourhood of Toronto.

Data

In order to identify the location best suited to start the Indian restaurant, we need to collect the required data that we will need in our analysis.

1. All neighbourhoods in the city of Toronto. This information is available in Wikipedia link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
2. Demographics information of Toronto neighbourhoods to understand the population of Indian and Sri Lankan (interested in Indian food) living in each neighbourhood. This information is available as csv file in Census 2016 Neighbourhood of Toronto.
https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv
3. Postal Code information for each neighbourhood. This information is available through Coursera assignments through the link http://cocl.us/Geospatial_data
4. Nearby venues in each neighbourhood. This information is available through Foursquare APIs. <https://foursquare.com/developers/apps>

Methodology

We will be collecting the data from the sources discussed in the data section and cleanse the data and get to a state which is good enough to perform any data analysis or apply machine learning algorithms and extract information from it.

Solution Approach

In order to identify the best location to start the Indian restaurant, following the solution approach. We will collect all the neighbourhoods in Toronto, by making use of the demographics information we will collect the top 50 neighbourhoods having high Indian and Sri

Lankan population. For these top 50 neighbourhoods we will explore nearby venues to identify the restaurants available currently and group them into 3 groups based on the most frequent restaurant type available now. Using this information, we will pick a group and recommend the neighbourhood to start the restaurant based on our findings.

Data Collection and Cleansing

The list of all neighbourhoods in Toronto is available in Wikipedia link discussed in the data section.



The screenshot shows a Wikipedia article titled "List of postal codes of Canada: M". The page is from the English Wikipedia. At the top, there are "Article" and "Talk" tabs. Below the title, it says "From Wikipedia, the free encyclopedia". The main content area starts with a paragraph about postal codes in Canada where the first letter is M. It mentions Canada Post's website and smartphone applications. A section titled "Toronto - FSAs" notes that there are no rural FSAs in Toronto, so no postal codes start with M0. Below this is a table with five rows of data:

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

We made use of BeautifulSoup library to web scrap the information from a table in the Wikipedia link. As seen from the table structure in the image above, the data is not in a state for data analysis directly. Following methods were applied to cleanse the data before applying any analysis.

1. All Boroughs that had value 'Not assigned' were removed.
2. All Neighbourhood that had value 'Not assigned' were assigned with the value from their respective Boroughs.
3. Neighbourhoods were grouped where the postal code is same.

At the end of cleansing we had neighbourhood data as below.

	PostalCode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West

To get the population information we made use of the census information from 2016. This information was available as csv file, so it was straight forward to read the details from the file. To keep it simple only immigration information was retained and all other data were removed from the file.

	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South	Alderwood	Annex	Banbury-Don Mills	... Willowdale	West
0	Immigration and citizenship	Immigrants by selected place of birth	Census Profile 98-316-X2016001	Total - Selected places of birth for the immigr...	1266005	19990	19990	3965	8275	13205	...	9395
1	Immigration and citizenship	Immigrants by selected place of birth	Census Profile 98-316-X2016001	Americas	212010	1635	1415	450	1630	1335	...	530
2	Immigration and citizenship	Immigrants by selected place of birth	Census Profile 98-316-X2016001	Brazil	7025	10	15	30	60	40	...	55
3	Immigration and citizenship	Immigrants by selected place of birth	Census Profile 98-316-X2016001	Colombia	8715	15	10	15	40	100	...	50
4	Immigration and citizenship	Immigrants by selected place of birth	Census Profile 98-316-X2016001	El Salvador	6955	10	30	20	10	10	...	0

Following cleaning operation were performed before applying data analysis.

1. Immigration information specific to India and Sri Lanka were retained.
2. Few of Neighbourhood names were changes to keep it in sync with the neighbourhood data extracted from Wikipedia.
3. The data were sorted based on the population in each neighbourhood and top 50 neighbourhood were extracted.

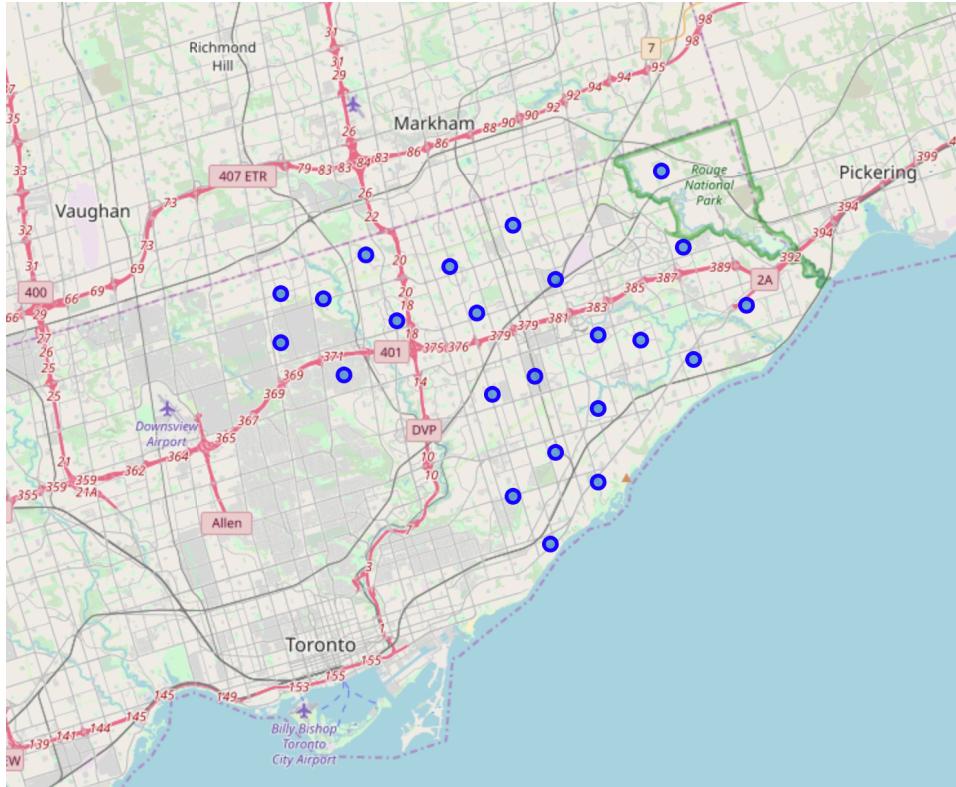
After applying the above cleansing methods, following data were extracted for further analysis.

	India	Sri Lanka	Population
	Woburn	6680	4405
	Rouge	2940	7385
	Malvern	3380	4860
	West Humber-Clairville	6525	585
	Mount Olive	4815	945
	L'Amoreaux	1360	2620
	Bendale	1760	1795
	Dorset Park	1325	2190
	Agincourt North	945	2210
	Eglinton East	985	2070
	Morningside	875	1570
	Flemington Park	1740	635
			2375

The top 50 highly populated Indian and Sri Lankan neighbourhood information extracted above were combined with Postal code and Borough details extracted before. The latitude and longitude information provided through csv file in coursera assignment was used to visualize this information in a map.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1J	Scarborough	Scarborough Village	43.773136	-79.239476
5	M1K	Scarborough	Ionview, Kennedy Park	43.744734	-79.239476
6	M1L	Scarborough	Clairlea	43.727929	-79.262029
7	M1M	Scarborough	Cliffcrest	43.711112	-79.284577
8	M1P	Scarborough	Dorset Park	43.716316	-79.239476
9	M1R	Scarborough	Wexford	43.692657	-79.264848
10	M1T	Scarborough	Tam O'Shanter	43.757410	-79.273304

The below map shows the city of Toronto with the blue marker showing the neighbourhoods having high Indian and Sri Lankan population that we would be analysing.



Foursquare location provider shares list of APIs that we will be using to understand more about the neighbourhood that we have picked for our analysis. Each of these neighbourhood would be explored to understand the top most cuisines currently available.

By making use of Foursquare API, we explored nearby venues to each of the neighbourhood and extract that data.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Rouge, Malvern	43.806686	-79.194353	Fratelli Village Pizzeria	43.784008	-79.169787	Italian Restaurant
Rouge, Malvern	43.806686	-79.194353	Mona's Roti	43.791613	-79.251015	Caribbean Restaurant
Rouge, Malvern	43.806686	-79.194353	Caribbean Wave	43.798558	-79.195777	Caribbean Restaurant
Rouge, Malvern	43.806686	-79.194353	Babu Catering & Take Out	43.791721	-79.251132	Sri Lankan Restaurant
Rouge, Malvern	43.806686	-79.194353	Lucky Lin's Restaurant	43.813613	-79.237943	Chinese Restaurant
Rouge, Malvern	43.806686	-79.194353	Silver Spoon Pak-Indian Restaurant	43.791824	-79.251340	Indian Restaurant
Rouge, Malvern	43.806686	-79.194353	Harvey's	43.800106	-79.198258	Fast Food Restaurant
Rouge, Malvern	43.806686	-79.194353	Wendy's	43.802008	-79.198080	Fast Food Restaurant
Rouge, Malvern	43.806686	-79.194353	Hakka Legend	43.776309	-79.234939	Chinese Restaurant
Rouge, Malvern	43.806686	-79.194353	La Sani Grill	43.776214	-79.234848	Indian Restaurant

Category specific to restaurant were filtered and by making use of the frequency of the category type, we extracted the top most restaurant types in each of the neighbourhood.

Neighborhood	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
Agincourt North, Milliken	Middle Eastern Restaurant	Chinese Restaurant	Indian Restaurant	Caribbean Restaurant	Mediterranean Restaurant	Falafel Restaurant	Sushi Restaurant	American Restaurant	Italian Restaurant	Restaurant
Bayview Village	Chinese Restaurant	Caribbean Restaurant	Indian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Hakka Restaurant	Asian Restaurant	Sushi Restaurant	Cantonese Restaurant	Vegetarian / Vegan Restaurant
Clairlea	Middle Eastern Restaurant	Chinese Restaurant	Indian Restaurant	Fast Food Restaurant	Thai Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant
Cliffcrest	Middle Eastern Restaurant	Thai Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	American Restaurant	Asian Restaurant	Chinese Restaurant	Ethiopian Restaurant	Filipino Restaurant	French Restaurant
Dorset Park	Chinese Restaurant	Japanese Restaurant	Greek Restaurant	Turkish Restaurant	Thai Restaurant	Sushi Restaurant	Indian Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant
Flemington Park	Chinese Restaurant	Caribbean Restaurant	Japanese Restaurant	Sushi Restaurant	Asian Restaurant	Korean Restaurant	Cantonese Restaurant	Vietnamese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
Henry Farm	Chinese Restaurant	Caribbean Restaurant	Middle Eastern Restaurant	Indian Restaurant	Sushi Restaurant	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant	Greek Restaurant	Asian Restaurant
Highland Creek	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Japanese Restaurant	Caribbean Restaurant	Chinese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Falafel Restaurant	Hakka Restaurant
Hillcrest Village	Chinese Restaurant	Caribbean Restaurant	Indian Restaurant	Sushi Restaurant	Vietnamese Restaurant	Hakka Restaurant	Greek Restaurant	Falafel Restaurant	Korean Restaurant	Fast Food Restaurant

Machine Learning Algorithm

At this stage we have the following information

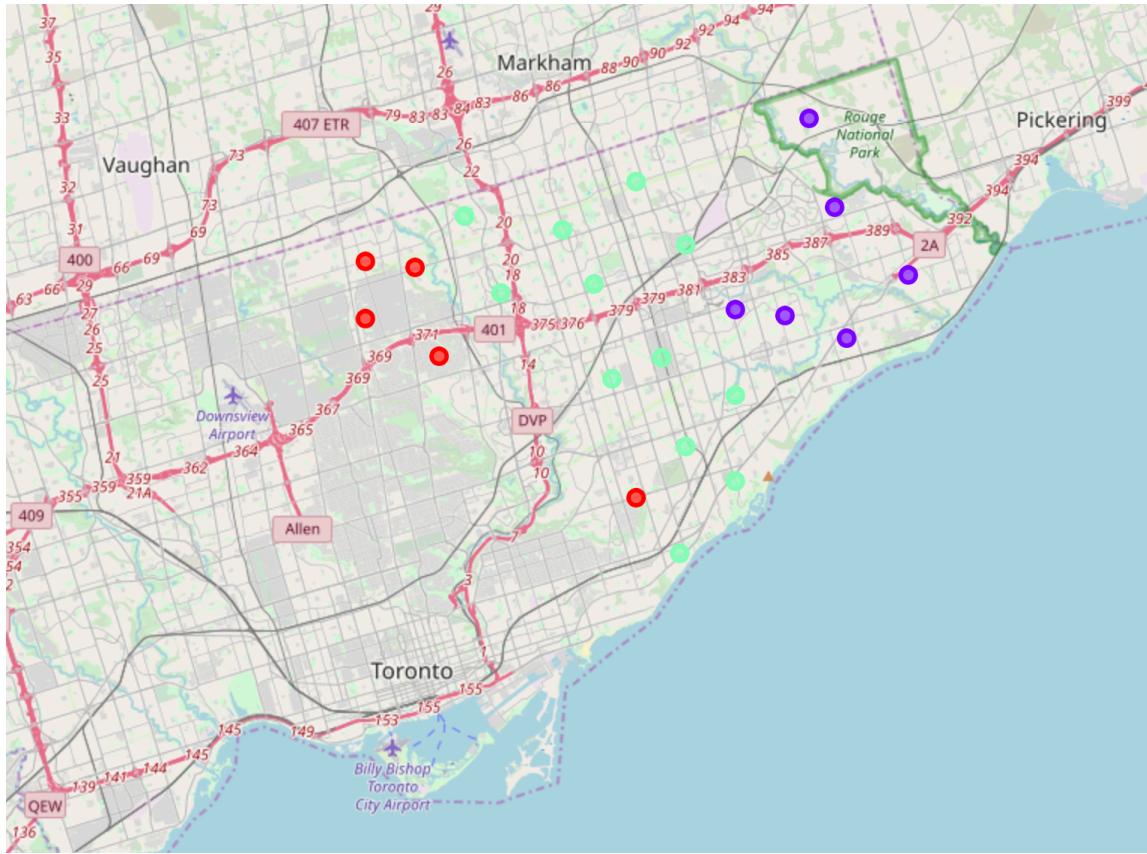
1. Top neighbourhood having high Indian and Sri Lankan population.
2. Top 10 most common restaurants for each of the above neighbourhoods.

To extract insights from these data, we need to segment these neighbourhood. One of the machine learning algorithms that is used to cluster a set of data is k-means algorithm. This is used to identify the similarity between each neighbourhood and group them into clusters. This cluster will help us to further narrow down the location for us to start the restaurant.

For k-means algorithm we need to specify the number of clusters that we need to group them into. We started with cluster size of 3 and applied the machine learning algorithm k-means to understand the cluster characteristic.

Result

The neighbourhoods were clustered in to three groups based on the similarity between them.



Cluster – 1

Cluster – 1 has neighbourhood like Cliffcrest, Humber Summit, etc which has moderate Indian and Sri Lankan population. Out of top 10 common restaurants currently available, none of them is Indian restaurant.

Neighborhood	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
Cliffcrest	Middle Eastern Restaurant	Thai Restaurant	Vietnamese Restaurant	Mediterranean Restaurant	American Restaurant	Asian Restaurant	Chinese Restaurant	Ethiopian Restaurant	Filipino Restaurant	French Restaurant
Humber Summit	Korean Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Japanese Restaurant	Ramen Restaurant	Doner Restaurant	French Restaurant
Mount Olive, Thistletown	Korean Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Japanese Restaurant	Seafood Restaurant	Ramen Restaurant	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant	Doner Restaurant
St. James Town	Korean Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Sushi Restaurant	Seafood Restaurant	Chinese Restaurant	Asian Restaurant	Ramen Restaurant	Caribbean Restaurant	French Restaurant
St. James Town	Korean Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Sushi Restaurant	Seafood Restaurant	Chinese Restaurant	Asian Restaurant	Ramen Restaurant	Caribbean Restaurant	French Restaurant

Cluster – 2

Cluster – 2 has neighbourhoods like Rouge, Malvern, Woburn, etc which has high Indian and Sri Lankan population. Indian Restaurants in these neighbourhoods forms one of the top 3 common restaurants.

Neighborhood	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
Highland Creek	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Japanese Restaurant	Caribbean Restaurant	Chinese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Falafel Restaurant	Hakka Restaurant
Morningside, West Hill	Indian Restaurant	Fast Food Restaurant	Caribbean Restaurant	Chinese Restaurant	Italian Restaurant	Hakka Restaurant	Mexican Restaurant	Vietnamese Restaurant	American Restaurant	Thai Restaurant
Rouge, Malvern	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant	Caribbean Restaurant	Italian Restaurant	Sri Lankan Restaurant	Mexican Restaurant	Falafel Restaurant	Hong Kong Restaurant	Hakka Restaurant
Scarborough Village	Fast Food Restaurant	Indian Restaurant	Chinese Restaurant	Caribbean Restaurant	Sushi Restaurant	Asian Restaurant	Hakka Restaurant	Malay Restaurant	Thai Restaurant	American Restaurant
Woburn	Fast Food Restaurant	Caribbean Restaurant	Indian Restaurant	Chinese Restaurant	Hakka Restaurant	Greek Restaurant	Italian Restaurant	Sushi Restaurant	Sri Lankan Restaurant	Seafood Restaurant
York University	Fast Food Restaurant	Hakka Restaurant	Indian Restaurant	Restaurant	Middle Eastern Restaurant	Afghan Restaurant	African Restaurant	Asian Restaurant	Greek Restaurant	Caribbean Restaurant

Cluster – 3

Cluster – 3 has neighbourhoods which has moderate Indian and Sri Lankan population. Indian Restaurants in these neighbourhoods forms one of the top 10 common restaurants and in few neighbourhood in top 5.

Neighborhood	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
Agincourt North, Milliken	Middle Eastern Restaurant	Chinese Restaurant	Indian Restaurant	Caribbean Restaurant	Mediterranean Restaurant	Falafel Restaurant	Sushi Restaurant	American Restaurant	Italian Restaurant	Restaurant
Bayview Village	Chinese Restaurant	Caribbean Restaurant	Indian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Hakka Restaurant	Asian Restaurant	Sushi Restaurant	Cantonese Restaurant	Vegetarian / Vegan Restaurant
Clairlea	Middle Eastern Restaurant	Chinese Restaurant	Indian Restaurant	Fast Food Restaurant	Thai Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant
Dorset Park	Chinese Restaurant	Japanese Restaurant	Greek Restaurant	Turkish Restaurant	Thai Restaurant	Sushi Restaurant	Indian Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant
Flemington Park	Chinese Restaurant	Caribbean Restaurant	Japanese Restaurant	Sushi Restaurant	Asian Restaurant	Korean Restaurant	Cantonese Restaurant	Vietnamese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
Henry Farm	Chinese Restaurant	Caribbean Restaurant	Middle Eastern Restaurant	Indian Restaurant	Sushi Restaurant	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant	Greek Restaurant	Asian Restaurant
Hillcrest Village	Chinese Restaurant	Caribbean Restaurant	Indian Restaurant	Sushi Restaurant	Vietnamese Restaurant	Hakka Restaurant	Greek Restaurant	Falafel Restaurant	Korean Restaurant	Fast Food Restaurant
Ionview, Kennedy Park	Chinese Restaurant	Indian Restaurant	Sushi Restaurant	Caribbean Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Japanese Restaurant	Korean Restaurant	Middle Eastern Restaurant	Hakka Restaurant
Tam O'Shanter	Middle Eastern Restaurant	Indian Restaurant	Chinese Restaurant	Caribbean Restaurant	Sushi Restaurant	Fast Food Restaurant	Italian Restaurant	Korean Restaurant	Asian Restaurant	Mediterranean Restaurant
Thorncliffe Park	Chinese Restaurant	Caribbean Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Cantonese Restaurant	Greek Restaurant	Italian Restaurant	Korean Restaurant	Restaurant	Hotpot Restaurant
Victoria Village	Chinese Restaurant	Japanese Restaurant	Caribbean Restaurant	Korean Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Asian Restaurant	Thai Restaurant	Cantonese Restaurant	Falafel Restaurant
Wexford	Vegetarian / Vegan Restaurant	French Restaurant	Turkish Restaurant	Thai Restaurant	Hungarian Restaurant	Sushi Restaurant	Indian Restaurant	Filipino Restaurant	Chinese Restaurant	Middle Eastern Restaurant

Discussion

1. Cluster-1 has moderate Indian and Sri Lankan population but there are no Indian restaurants available. There will be competition from other Asian and Middle Eastern restaurants but opening Indian restaurants in this area will pull crowd towards it. We would recommend these neighbourhoods as there is no Indian restaurants in it.
2. Cluster-2 has high Indian and Sri Lankan population as well as Indian restaurants. There will be heavy competition from other Indian restaurants. If the restaurant that is to be opened is up for the competition or already has earned name through its chain of restaurants, then these neighbourhoods is the one to look for.

3. Cluster-3 has moderate to high Indian and Sri Lankan population. Indian restaurants are currently available as one of the top 10 common restaurants but no so common in many. There will be competition from Chinese and Japanese restaurants as well apart from Indian restaurants. We would recommend these neighbourhood for those who can take up competition with other cuisines as well and as Indian Restaurants are not so common in some of the neighbourhoods, this will provide an opportunity.

Conclusion

With the public available of information like Wikipedia, Census data, Foursquare developer API, we were able to recommend options to narrow down on the location to start a restaurant. With integration of additional data like users review, restaurant ratings, etc, we could refine our recommendations and help user to decide the location.