Battle of Neighbourhoods

Scarborough, Toronto

Table of Contents

Problem Statement

Identifying a suitable neighbourhood in Toronto, to open a South Indian Coffee Shop.

Background of the Problem

A South Indian immigrant to Canada is our client and is planning to open a South Indian Coffee shop – Filter coffee, which is famous among the Tamils'. Though, the client knew the nuances of the coffee shop business, he is new to Canada and must know the demography of Canada. Toronto, being the financial capital of Canada, is one of the widely preferred city for new immigrants to set up a business. But, in order to choose the perfect neighbourhood in Toronto, the client has approached our Data Science team to come up with an analysis.

Objective

Data Science team has zeroed in on Scarborough, Toronto; because as per the 2016 statistics available on Wikipedia, Tamils are predominantly in Scarborough compared to the other Boroughs of Toronto. To further narrow down the options, our team has decided to use the Battle of Neighbourhoods approach for Scarborough using K-Means Clustering algorithm and FourSquare API - to understand the top 10 venues from each neighbourhood of Scarborough.

Data

Data collection is an important part for any Data science project. We will not be getting any readymade data to work with. Hence, with the limited data that we have obtained from Wikipedia, we will be proceeding with our analysis.

Wikipedia: *List of Postal Code for Toronto*

Below is the sample of the postal codes that we will be using for this analysis. We will be scrapping the below table data from Wikipedia and will be doing the necessary data wrangling and data cleansing activity and will use the final data for our further analysis.

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

Wikipedia: *Demography of Toronto*

Complete demography of Toronto has been provided in the above link. We have taken only the below section to select a Borough – Scarborough, upon which we will be applying the Battle of Neighbourhood analysis. We have taken Scarborough for analysis, because, as we can see from the below image that majority of Tamil populations are in Scarborough. Whereas, the Tamil population is very less in other boroughs which is evident from the below image. Apart from this, we will not be using the below data for any other purpose.


TORONTO & EAST YORK
- *Trinity-Spadina* (137,865): **1. English (60.2%)**, 2. Cantonese (4.7%), 3. Portuguese (4.5%), 4. Chinese, not otherwise specified (4.5%), 5. Mandarin (3.8%)
- *Toronto Centre* (123,430): **1. English (61.7%)**, 2. Spanish (2.5%), 3. Chinese, not otherwise specified (2.4%)
- *St. Paul's* (112,470): **1. English (69.4%)**, 2. Spanish (3.0%), 3. Tagalog (Pilipino, Filipino) (2.6%)
- *Beaches–East York* (103,625): **1. English (70.0%)**, 2. Bengali (3.7%), 3. Greek (2.0%)
- *Parkdale–High Park* (100,595): **1. English (63.0%)**, 2. Polish (4.3%), 3. Spanish (2.3%)
- *Toronto–Danforth* (100,420): **1. English (63.7%)**, 2. Cantonese (5.8%), 3. Greek (5.0%), 4. Chinese, not otherwise specified (3.6%)
- *Davenport* (98,425): 1. English (47.5%), **2. Portuguese (21.4%)**, 3. Italian (5.9%), 4. Spanish (5.6%)

NORTH YORK
- *Willowdale* (135,455): 1. English (35.1%), 2. Chinese, not otherwise specified (9.4%), 3. Persian (9.3%), 4. Korean (7.9%), 5. Mandarin (7.7%)
- *Don Valley West* (118,545): 1. English (54.4%), 2. Urdu (5.9%), 3. Persian (4.0%)
- *York Centre* (112,950): 1. English (40.5%), **2. Russian (11.9%)**, 3. Italian (7.6%), 4. Tagalog (Pilipino, Filipino) (7.4%) 5. Spanish (5.0%)
- *Eglinton–Lawrence* (108,730): **1. English (63.1%)**, 2. Tagalog (Pilipino, Filipino) (5.3%), 3. Italian (4.9%)
- *Don Valley East* (106,115): 1. English (42.1%), 2. Mandarin (6.5%), 3. Chinese, not otherwise specified (6.1%) 4. Persian (4.6%), 5. Arabic (3.5%), 6. Tagalog (Pilipino, Filipino) (3.2%), 7. Cantonese (3.2%)
- *York West* (103,395): 1. English (44.0%), 2. Italian (9.6%), 3. Spanish (8.5%), 4. Vietnamese (5.3%), 5. Urdu (3.0%)

SCARBOROUGH
- *Scarborough–Rouge River* (128,905): 1. English (37.4%), **2. Tamil (13.8%)**, 3. **Cantonese (13.1%)**, 4. Mandarin (4.6%), 5. Tagalog (Pilipino, Filipino) (4.3%), 6. Urdu (3.2%)
- *Scarborough–Agincourt* (107,465): 1. English (30.2%), **2. Chinese, not otherwise specified (15.8%)**, 3. **Cantonese (14.4%)**, **4. Mandarin (12.0%)**, 5. Tamil (5.0%)
- *Scarborough–Guildwood* (105,900): 1. English (57.6%), **2. Tamil (8.0%)**, 3. Gujarati (6.8%), 4 Tagalog (Pilipino, Filipino) (3.9%), 5. Urdu (3.2%)
- *Scarborough Centre* (105,880): 1. English (47.7%), **2. Tamil (8.5%)**, 3. Tagalog (Pilipino, Filipino) (6.1%) 4. Chinese, not otherwise specified (3.8%), 5. Cantonese (3.5%)
- *Scarborough Southwest* (103,270): 1. English (59.3%), 2. Bengali (6.1%), 3. Tagalog (Pilipino, Filipino) (5.0%), 4. Tamil (3.4%)

ETOBICOKE YORK
- *Etobicoke–Lakeshore* (119,120): 1. English (59.7%), 2. Polish (5.6%), 3. Ukrainian (3.5%)
- *York South–Weston* (111,710): 1. English (49.9%), 2. Portuguese (9.9%), 3. Spanish (8.2%), 4. Italian (7.6%), 5. Vietnamese (3.6%)
- *Etobicoke Centre* (109,250): 1. English (54.1%), 2. Italian (5.5%), 3. Ukrainian (4.1%), 4. Spanish (3.2%)
- *Etobicoke North* (105,620): 1. English (45.4%), **2. Panjabi (Punjabi) (10.1%)**, 3. Gujarati (4.8%), 4. Italian (3.8%), 5. Urdu (3.1%), 6. Somali (3.0%)

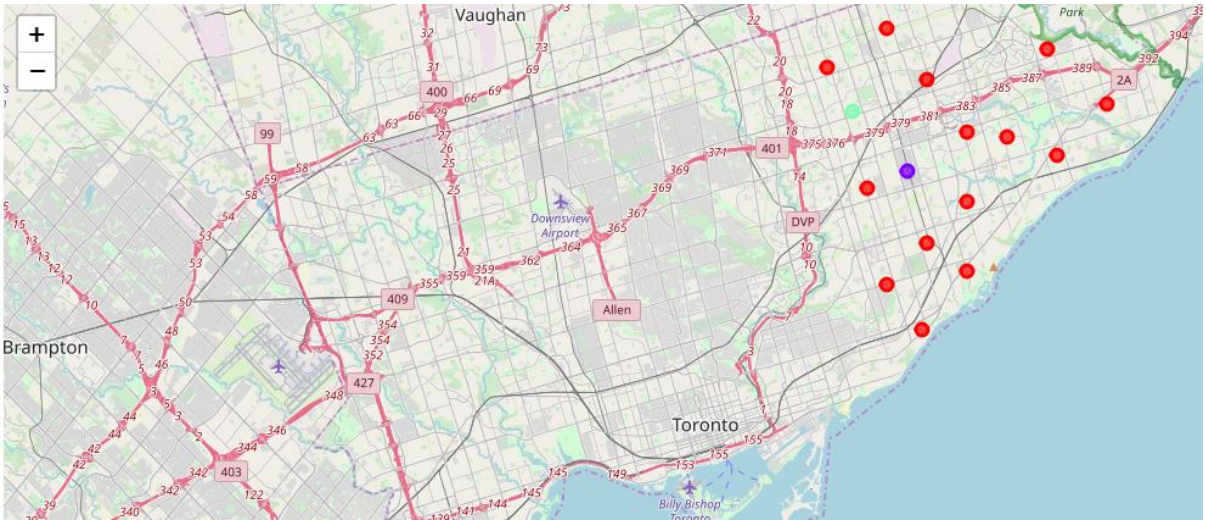
Coursera: *Geospatial Data*

We will be using the Geospatial data - Latitude and Longitude details of all the postal code of Toronto that we have received through the capstone project as well. A sample of the data is provided in the below image

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Methodology

We have cleansed the raw data and shaped it to a format that's easy to be analysed. Upon analysis we have extracted top 10 venues for each neighbourhood using FourSquare API. By applying Machine Learning algorithms (K-means clustering), we have come up with three clusters of neighbourhoods and their top 10 venues. Below is a map that depicts the clusters.



Result

Below are the results of each cluster with the top 10 venues of the neighbourhoods.

First Cluster

As per the K-Means clustering algorithm, first cluster is the red markers in the above map and most of the neighbourhoods (around 14) fall under this cluster. From the cluster result, we can infer that for neighbourhoods like Rogue, Malvern, Highland Creek, Rouge Hill, Port Union, Woburn, the Indian Restaurant venues comes in the 4th common venue of that neighbourhoods, showing clearly that more Indians are residing here.

Top 10 venues of first cluster

```
Scarborough_merged.loc[Scarborough_merged['Cluster Labels'] == 0, Scarborough_merged.columns[[2] + list(range(5, Scarborough_merged.shape[1]))]]
```

	Neighbourhood	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
0	Rouge, Malvern	0	Fast Food Restaurant	Vietnamese Restaurant	Italian Restaurant	Indian Restaurant	Hakka Restaurant	Grocery Store	General Entertainment	Fried Chicken Joint	Electronics Store	Discount Store
1	Highland Creek, Rouge Hill, Port Union	0	Bar	Vietnamese Restaurant	Coffee Shop	Indian Restaurant	Hakka Restaurant	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
2	Guildwood, Morningside, West Hill	0	Spa	Rental Car Location	Electronics Store	Pizza Place	Breakfast Spot	Medical Center	Mexican Restaurant	Vietnamese Restaurant	Coffee Shop	Grocery Store
3	Woburn	0	Coffee Shop	Korean Restaurant	Vietnamese Restaurant	Indian Restaurant	Hakka Restaurant	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
4	Cedarbrae	0	Caribbean Restaurant	Thai Restaurant	Athletics & Sports	Hakka Restaurant	Bakery	Bank	Fried Chicken Joint	College Stadium	Indian Restaurant	Grocery Store
5	Scarborough Village	0	Jewelry Store	Playground	Vietnamese Restaurant	Chinese Restaurant	Hakka Restaurant	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
6	East Birchmount Park, Ionview, Kennedy Park	0	Discount Store	Train Station	Coffee Shop	Bus Station	Department Store	Intersection	Indian Restaurant	Hakka Restaurant	Grocery Store	General Entertainment
7	Clairlea, Golden Mile,	0	Bakery	Bus Line	Soccer Field	Fast Food	Park	Metro Station	Intersection	Vietnamese	College	Hakka

Second Cluster

Second cluster is the purple marker in the above map and only one of the neighbourhood fall under this cluster. From the cluster result, we can infer that the Indian Restaurant venues comes as the 1st common venue of that neighbourhood. This shows that there are lots of Indians residing in this neighbourhood.

Top 10 venues of second cluster

```
Scarborough_merged.loc[Scarborough_merged['Cluster Labels'] == 1, Scarborough_merged.columns[[2] + list(range(5, Scarborough_merged.shape[1]))]]
```

	Neighbourhood	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
10	Dorset Park, Scarborough Town Centre, Wexford ...	1	Indian Restaurant	Pet Store	Latin American Restaurant	Light Rail Station	Vietnamese Restaurant	Chinese Restaurant	Department Store	Hakka Restaurant	Grocery Store	General Entertainment

Third Cluster

Third cluster is the cyan marker in the above map and only one of the neighbourhood fall under this cluster. From the cluster result, we can infer that the Indian Restaurant venue doesn't come in the top 10 most common venue of that neighbourhood and hence the chances of Tamils residing here is very less.

Top 10 venues of third cluster

```
Scarborough_merged.loc[Scarborough_merged['Cluster Labels'] == 2, Scarborough_merged.columns[[2] + list(range(5, Scarborough_merged.shape[1]))]]
```

	Neighbourhood	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
13	Clarks Corners, Sullivan, Tam O'Shanter	2	Pizza Place	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Fried Chicken Joint	Fast Food Restaurant	Pharmacy	Noodle House	Vietnamese Restaurant	Grocery Store

Discussion

We can observe from the clusters that Indian Restaurant tops the most common venue in 2nd Cluster, thus showing the higher probability of Indians residing in the neighbourhood - Dorset Park, Scarborough Town Centre, Wexford**.** Since, the more Indians are residing in this neighbourhood, we can **recommend** this neighbourhood to our client for opening a South Indian Coffee shop. This is also confirmed by the 8.5% of Tamils in Scarborough Centre Neighbourhood as per Wikipedia data.

Similarly, from the 1st cluster, couple of neighbourhoods can be **Highly recommended** to the client, since we can infer that there will be a considerable number of Indian population residing in these neighbourhoods. Hence, the Indian Restaurants are 4th most common venue in these neighbourhoods. This can be verified by the fact that Rogue and Guildwood neighbourhoods has a sum of 21% of Tamil population of Scarborough, as per Wikipedia data.

Alternatively, 3rd cluster contains most of the Asian type restaurants – Chinese, Thai, Vietnamese, confirming heavy Asian contingent residing here. Since, Indian restaurant didn't top the most common venue, the client will be less successful here. Or, he can open his coffee shop by thinking out of box in this neighbourhood.

Conclusion

With this limited data that we have collected, we were able to recommend a neighbourhood in Scarborough, Toronto to our client. Final decision has to be taken by our client based on the neighbourhood that we have recommended and their strategy.

This, Battle of Neighbourhoods capstone project can be a prototype for many larger projects, especially dealing with bigger cities and choosing between cities for any purpose. When the project gets larger, we will have to collect a lot more data and process them, analyse them and build a model that helps our client.