

Identifying the most ideal Toronto neighborhood to start an Indian restaurant

Amey Varangaonkar
June 16, 2019

1: Introduction / Business Problem

1.1 Discussion of the problem and the background

“Which Toronto neighborhood is the most ideal to start an Indian restaurant?”

My Canadian client, Mr Singh, owns a chain of Indian restaurants in various popular cities in Canada – including Montreal, and Vancouver. He now wishes to start an Indian restaurant in Toronto.

Mr Singh wanted me to consult him on which locality or neighborhood of Toronto is best suited to start his Indian restaurant. He had put forth the following criterion which needs to be satisfied in my work:

- The locality or the neighborhood ideally should be situated in Downtown Toronto. The second preference would be West Toronto.
- The neighborhood should not have any restaurant or eatery among its top 2 commonly visited places

The task here was to recommend the best possible location in Toronto which satisfies both the criteria above and allows Mr Singh to start a successful Indian restaurant in Toronto.

1.2 Target Audience

Toronto is one of the most ethnically diverse cities in Canada, with over 51% of the population belonging to a specific minority group. Per the survey conducted in 2016, approximately 12.6% of the population is South-Asian. Opening an Indian restaurant in Toronto could prove to be a profitable proposition for the client, given the popularity of the Indian cuisine not only among the south Asians, but all over the world.

2: Data

2.1 Description of the data

This project relied mainly on the publicly available data from Wikipedia as well as Foursquare.

2.1.1 Dataset 1: List of postal codes of Toronto along with the boroughs and neighborhoods

Since the focus of the project was the city of Toronto, I looked at procuring all the demographic information related to city, including all the boroughs and neighborhoods of Toronto along with their associated zip code.

To do this, I relied on the publicly available Wikipedia page for the same, titled List of postal codes of Canada: M (Link can found [here](#)). As can be seen from the description of the page:

"Postal codes beginning with M are located within the city of Toronto in the province of Ontario."

We scraped this data from the Wikipedia page with the help of Python's pandas and Wikipedia packages.

2.1.2 Dataset 2: Geographical co-ordinates of the neighborhoods

In order to plot the neighborhoods on the map, we also used the geographical co-ordinates of the neighborhoods of Toronto. Although this data could have been obtained using the Google Maps Geocoding API, given the unreliability of the package, I used the data from the following source: http://cocl.us/Geospatial_data.

2.1.3 Dataset 3: Data called from through Foursquare API

In this project, I leveraged the Foursquare API to obtain the geographical location data of various neighborhoods of Toronto. This data was used to explore the popular and commonly visited venues in each of the neighborhoods, which helped me to identify the best possible location for our client's Indian restaurant.

I identified the top 10 popular venues in each of the areas to satisfy the client's requirement, that the chosen locality should not have a restaurant in the top 2 commonly visited venues. Due to the API restrictions set by Foursquare, the search for the number of venues was limited only to 100.

2.1.4 Data Pre-processing

After scraping the initial data from Wikipedia, there were some improvements required before it could be used for analysis. Some modifications that were made to it were:

- Dropping all the rows from the derived table where boroughs were not assigned
- Combining different neighborhoods with the same postcode
- For neighborhoods with no name, assigning a borough name to it for the purposes of simplicity
- Concatenating the geospatial co-ordinates obtained from Dataset 2 to this dataframe.

The following is a sample of the final dataframe that was used for exploring the best neighborhood for the project:

| | Postcode | Borough | Neighbourhood | Postal Code | Latitude | Longitude |
|----|----------|-------------|---|-------------|-----------|------------|
| 0 | M1B | Scarborough | Malvern, Rouge | M1B | 43.806686 | -79.194353 |
| 1 | M1C | Scarborough | Highland Creek, Port Union, Rouge Hill | M1C | 43.784535 | -79.160497 |
| 2 | M1E | Scarborough | Guildwood, Morningside, West Hill | M1E | 43.763573 | -79.188711 |
| 3 | M1G | Scarborough | Woburn | M1G | 43.770992 | -79.216917 |
| 4 | M1H | Scarborough | Cedarbrae | M1H | 43.773136 | -79.239476 |
| 5 | M1J | Scarborough | Scarborough Village | M1J | 43.744734 | -79.239476 |
| 6 | M1K | Scarborough | East Birchmount Park, Ionview, Kennedy Park | M1K | 43.727929 | -79.262029 |
| 7 | M1L | Scarborough | Clairlea, Golden Mile, Oakridge | M1L | 43.711112 | -79.284577 |
| 8 | M1M | Scarborough | Cliffcrest, Cliffside, Scarborough Village West | M1M | 43.716316 | -79.239476 |
| 9 | M1N | Scarborough | Birch Cliff, Cliffside West | M1N | 43.692657 | -79.264848 |
| 10 | M1P | Scarborough | Dorset Park, Scarborough Town Centre, Wexford ... | M1P | 43.757410 | -79.273304 |
| 11 | M1R | Scarborough | Maryvale, Wexford | M1R | 43.750072 | -79.295849 |
| 12 | M1S | Scarborough | Agincourt | M1S | 43.794200 | -79.262029 |

Figure 1: Neighborhood data with geographical co-ordinates

3. Methodology

3.1 Exploratory Data Analysis

With the pre-processed data in place, the next step was to explore it to understand it better. The client's requirement was better understood when we explored the number of pin codes associated with each Toronto borough.

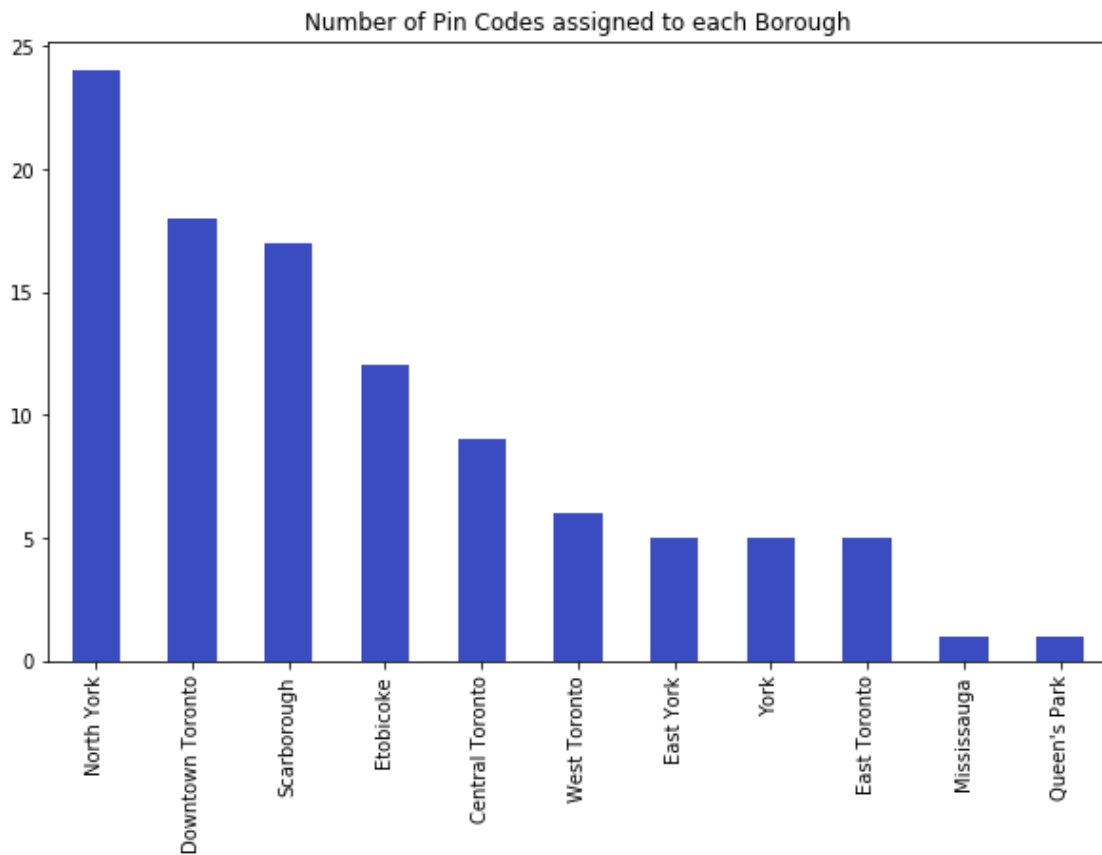


Figure 2: Pin codes assigned to each Toronto borough

Why did the client prefer Downtown Toronto over other, more populated boroughs such as North York? The answer to this question was easy - A quick Wikipedia search on Downtown Toronto gives the answer - with a density of 11725/ km², it is one of the most densely populated areas in North America.

The dataframe was thus narrowed down to include only the neighborhoods in Downtown Toronto, as that was the preference given by the client.

3.2 Analyzing the Toronto neighborhoods

For the analysis, the K-Means Clustering approach was used.

After using the one-hot encoding technique for all the columns in the source dataframe, we used the Foursquare API to extract the top 10 venues from each neighborhood, giving the following result:

| | 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 |
|----|---|-----------------------|-------------------------------|-----------------------|---------------------------|-----------------------|-----------------------|--------------------------|-----------------------|-----------------------|------------------------|
| 0 | Adelaide, King, Richmond | Coffee Shop | Café | Steakhouse | Bar | American Restaurant | Bakery | Cosmetics Shop | Restaurant | Hotel | Thai Restaurant |
| 1 | Bathurst Quay, CN Tower, Harbourfront West, Is... | Airport Terminal | Airport Lounge | Airport Service | Harbor / Marina | Boat or Ferry | Sculpture Garden | Bar | Plane | Boutique | Airport Food Court |
| 2 | Berczy Park | Coffee Shop | Cocktail Bar | Beer Bar | Bakery | Steakhouse | Seafood Restaurant | Cheese Shop | Italian Restaurant | Café | Farmers Market |
| 3 | Cabbagetown, St. James Town | Coffee Shop | Restaurant | Bakery | Park | Pub | Pizza Place | Café | Italian Restaurant | Indian Restaurant | Farmers Market |
| 4 | Central Bay Street | Coffee Shop | Café | Italian Restaurant | Middle Eastern Restaurant | Sandwich Place | Burger Joint | Sushi Restaurant | Bubble Tea Shop | Bar | Spa |
| 5 | Chinatown, Grange Park, Kensington Market | Café | Vegetarian / Vegan Restaurant | Coffee Shop | Bakery | Mexican Restaurant | Bar | Dumpling Restaurant | Vietnamese Restaurant | Chinese Restaurant | Cocktail Bar |
| 6 | Christie | Grocery Store | Café | Park | Diner | Baby Store | Coffee Shop | Convenience Store | Italian Restaurant | Restaurant | Nightclub |
| 7 | Church and Wellesley | Coffee Shop | Japanese Restaurant | Sushi Restaurant | Gay Bar | Restaurant | Men's Store | Mediterranean Restaurant | Hotel | Gym | Gastropub |
| 8 | Commerce Court, Victoria Hotel | Coffee Shop | Café | Hotel | Restaurant | American Restaurant | Seafood Restaurant | Bakery | Steakhouse | Italian Restaurant | Gastropub |
| 9 | Design Exchange, Toronto Dominion Centre | Coffee Shop | Café | Hotel | Restaurant | Italian Restaurant | Gastropub | Bakery | Deli / Bodega | American Restaurant | Burger Joint |
| 10 | First Canadian Place, Underground city | Coffee Shop | Café | Hotel | Restaurant | American Restaurant | Bakery | Bar | Steakhouse | Gastropub | Deli / Bodega |

Figure 3: Top 10 venues of each Toronto neighborhood

3.2.1 Clustering the neighborhoods

For clustering the elbow method was used to identify the most ideal number of clusters to be formed.

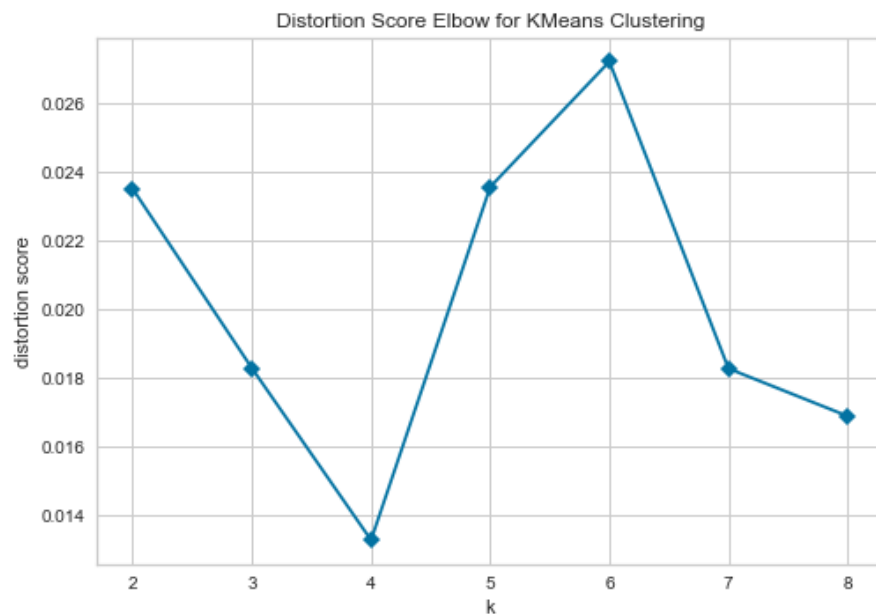


Figure 4: Identifying the best possible value for 'k'

Once the clustering was successfully done, we mapped the clusters on the map of Toronto to get the clusters of neighborhoods to explore.

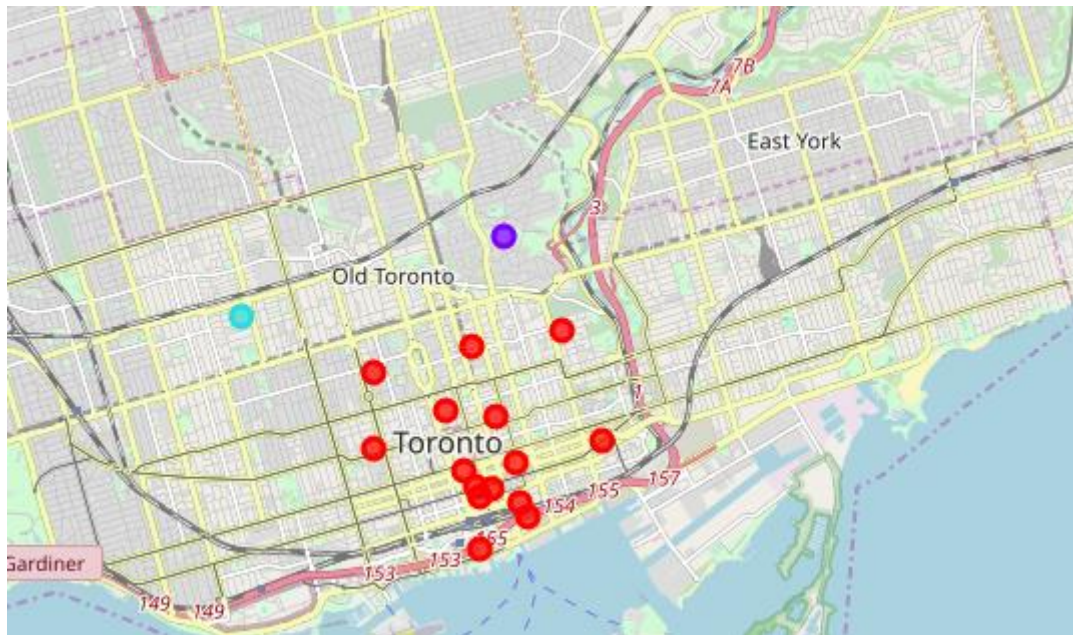


Figure 5: Clustered neighborhoods of Toronto

4. Results

The clustering algorithm efficiently clustered all the neighborhoods with similar top venues together. The segregation was done according to the following categories:

| Cluster | Top 2 common venues |
|------------|----------------------------------|
| 0 (Red) | Coffee shop, Café |
| 1 (Purple) | Park, playground |
| 2 (Teal) | Grocery store, Café |
| 3 (Yellow) | Airport terminal, airport lounge |

Table 1: Clustered neighborhoods and their top 2 venues

Clusters 0 and 2 were completely eliminated from the possibilities considering the neighborhoods in in had restaurants and eateries in abundance, and they did not meet the client's criteria of not having any eatery in the top 2 common venues.

Note: A café and a coffee shop were assumed to be eateries

The geographic locations of clusters 1 and 3 were considered, and the following factors were considered:

- Proximity with the other neighborhoods within Toronto
- Strategic positioning of the neighborhood

The clusters 1 and 3 were mapped to get the following results:



Figure 6: Cluster location comparison 1 vs. 3

Cluster 3 was situated in the Toronto city airport, which explained the lack of restaurants and other eateries as compared to the other neighborhoods. Also, it was isolated from other neighborhoods of Toronto. The neighborhood in cluster 1, Rosedale, Downtown Toronto can thus be shortlisted as the best possible location for the Indian restaurant.

However, all of the 6 neighborhoods from Clusters 1 and 3 satisfy the client's requirement and can thus be taken into consideration while making a decision.

| PostalCode | Borough | Neighborhood |
|------------|------------------|--------------|
| M4W | Downtown Toronto | Rosedale |

Figure 7: Shortlisted neighborhoods in Cluster 1

| Postcode | Borough | Neighbourhood |
|----------|------------------|-------------------|
| M5V | Downtown Toronto | Bathurst Quay |
| M5V | Downtown Toronto | CN Tower |
| M5V | Downtown Toronto | Harbourfront West |
| M5V | Downtown Toronto | Island airport |
| M5V | Downtown Toronto | King and Spadina |

Figure 8: Shortlisted neighborhoods in cluster 3

Since all of the 5 neighborhoods in Cluster 3 contained the same geographical co-ordinates, the suggested venues for all of them were the same.

5. Discussion

This analysis heavily relied on the data extracted from the FourSquare API. There is some scope for further improvement in the analysis by placing some additional filters such as:

- Population of the neighborhood
- Geographical coordinates of the neighborhoods as opposed to boroughs
- Average per-capita income of the population
- Consideration of the ratings of the restaurants in the neighborhood
- Crime rate of the neighborhoods
- Real-estate prices of the neighborhoods

In order to arrive at a more concrete conclusion, a thorough analysis on the above factors will be definitely helpful for the client to make a more informed decision.

6. Discussion

In this study, the neighborhoods of Toronto were explored and clustered to obtain the best and commonly visited venues in each of them, and accordingly the best possible neighborhood to start an Indian restaurant was determined. This was done in accordance with the preferences and conditions set by the client. However, this conclusion was subject to further refinement given a more thorough analysis of some other factors which could not be monitored in this study.