

Analyzing and Clustering Toronto Neighbourhoods

Sachchid Wast

26th May 2021

1. Introduction

1.1 Background

Toronto is considered as the most multicultural city of the world with nearly half of its population born outside of Canada. As a result, residents of this place get to enjoy a variety of cuisines just by walking through its neighbourhoods. Neighborhoods like little Italy, little Poland, little India offer a totally different cultural experience and a chance to try out new food without ever leaving the country.

1.2 Business Problem

We can safely assume that in an Indian neighbourhood there would be plenty of Indian restaurants. Consequently, there would be a lot of competition between the restaurants and a new restaurant would have to compete with the well-established places which is difficult to accomplish. Our problem is that we are looking to build a new Indian restaurant and looking for a suitable place for it. Therefore, we are analyzing the neighbourhoods of Toronto and look the places which are the best places to open an Indian restaurant.

2. Data

2.1 Data sources

We are using the Foursquare API which gives us the information about the venues in the neighbourhoods i.e. we get venue's latitude, longitude, name and category which is useful while clustering the neighbourhoods and the venues

We are using Wikipedia's Postal code data of Toronto to get all the boroughs and postal code of each neighbourhood which we helpful for plotting when combined with the geographical coordinates of Toronto.

To get the coordinates of each neighbourhood in Toronto, we have used Geopy library. As we have prepared the data from both Wikipedia and Geopy in last week, we going to import that data instead of going through all the steps.

2.2 Data cleaning

I web scraped the postal code and borough of each neighbourhood from [Wikipedia](#) using Beautiful Soup. I converted that data to a pandas data frame and removed all the postal code which didn't have any boroughs and neighbourhoods assigned to it.

I got the coordinates of each neighbourhood using Geopy and then finally merged the two data frames into one.

We cleaned the Foursquare data by converting it from JSON format to a pandas data frame and taking only the required columns which are discussed in the next section. The *venue* key in the JSON file contains all the information we require, and we access it just we would a pandas data frame.

2.3 Feature selection

In our neighbourhood dataset, we have the name of the neighbourhood, its borough and postal code along with its latitude and longitude. We have 15 boroughs and 103 neighbourhoods.

From the Foursquare API, we got the name of the neighbourhood its geographical coordinates its venues and the venue's geographical coordinates. Along with that we also got its name and category.

3. Methodology

In this project our goal is to find suitable neighbourhoods to open an Indian restaurant. We are going to find out areas which have low density of Indian restaurants.

Also, another way of predicting whether our restaurant would be successful or not is by observing the number of venues in that area. If there are a lot of venues in that neighborhood, it is safe to assume that people also go there. We have capped the number of venues to 100 per neighborhood.

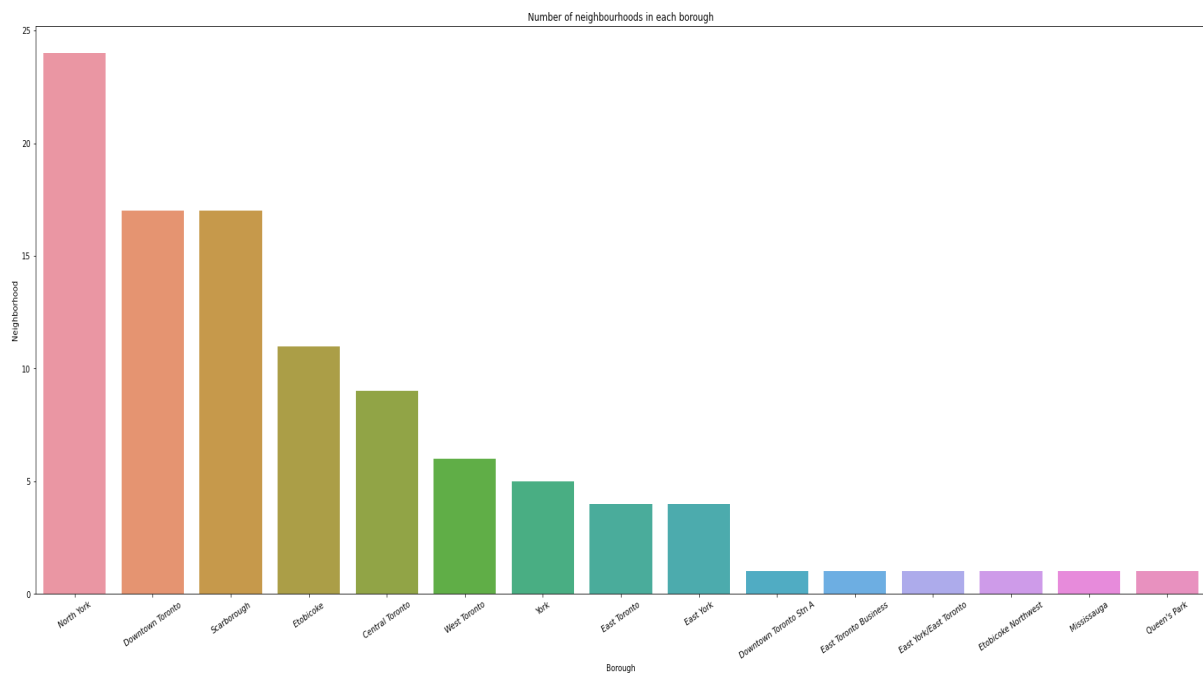
In the first step we have collected the required data: location and category of every venue and plotted them on the map and identified all the Indian restaurants and plotted them.

In the second step we explore the neighborhoods and look at the density of the venues and plot them using heatmaps.

In the third step we find the neighbourhoods which have similar venues and cluster them using KMeans clustering. This would help us identify which neighbourhoods are similar and which would be better for our restaurant.

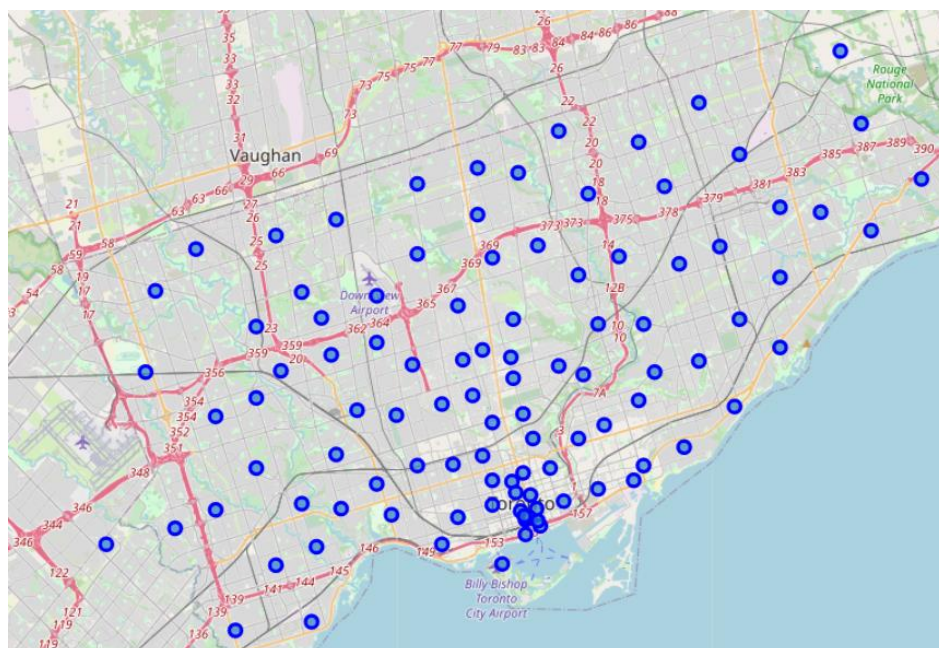
3.1 Exploratory Data Analysis

The following bar graph shows which borough has the highest number of neighbourhoods.



North York, Downtown Toronto and Scarborough are the top 3 boroughs with the greatest number of neighbourhoods.

In the following map, we have plotted all the neighbourhoods in Toronto, ON.



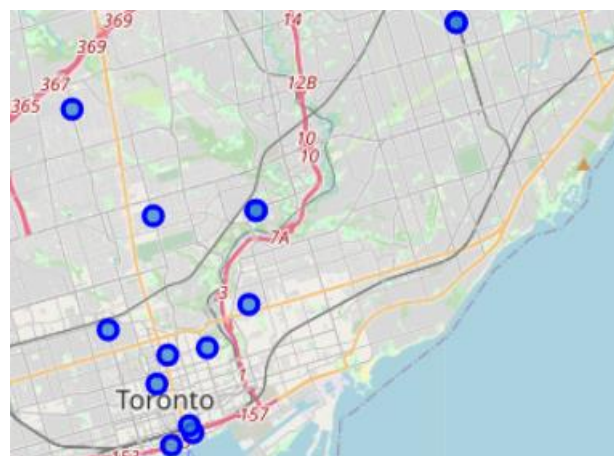
3.2 Analysis

We are assuming that if the neighbourhood has a higher number of venues as compared to another one, then it is a good place to open a restaurant since there would be more crowd than one with less venues. To get this data, we have used *groupby method* on the *neighbourhood* column and counted the number of venues in each neighbourhood by using the *count* method.

```
number_of_venues = trnt_venues.groupby('Neighborhood').count()
number_of_venues.sort_values(by = 'Venue', ascending = False).head(20)
```

Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
First Canadian Place, Underground city	100	100	100	100	100	100
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
Garden District, Ryerson	100	100	100	100	100	100
Toronto Dominion Centre, Design Exchange	100	100	100	100	100	100
Enclave of MSE	100	100	100	100	100	100
Richmond, Adelaide, King	92	92	92	92	92	92
St. James Town	82	82	82	82	82	82
Church and Wellesley	80	80	80	80	80	80
Fairview, Henry Farm, Oriole	69	69	69	69	69	69
Kensington Market, Chinatown, Grange Park	63	63	63	63	63	63
Central Bay Street	62	62	62	62	62	62
Berczy Park	59	59	59	59	59	59
St. James Town, Cabbagetown	46	46	46	46	46	46
Regent Park, Harbourfront	45	45	45	45	45	45
Little Portugal, Trinity	44	44	44	44	44	44
The Danforth West, Riverdale	41	41	41	41	41	41
Studio District	36	36	36	36	36	36
Davisville	35	35	35	35	35	35
Willowdale South	35	35	35	35	35	35

Then we found the Indian restaurants in Toronto and plotted them on the map.



As we can see there are not a lot of Indian restaurants in Toronto, meaning there isn't a lot of restaurants to go toe to toe against.

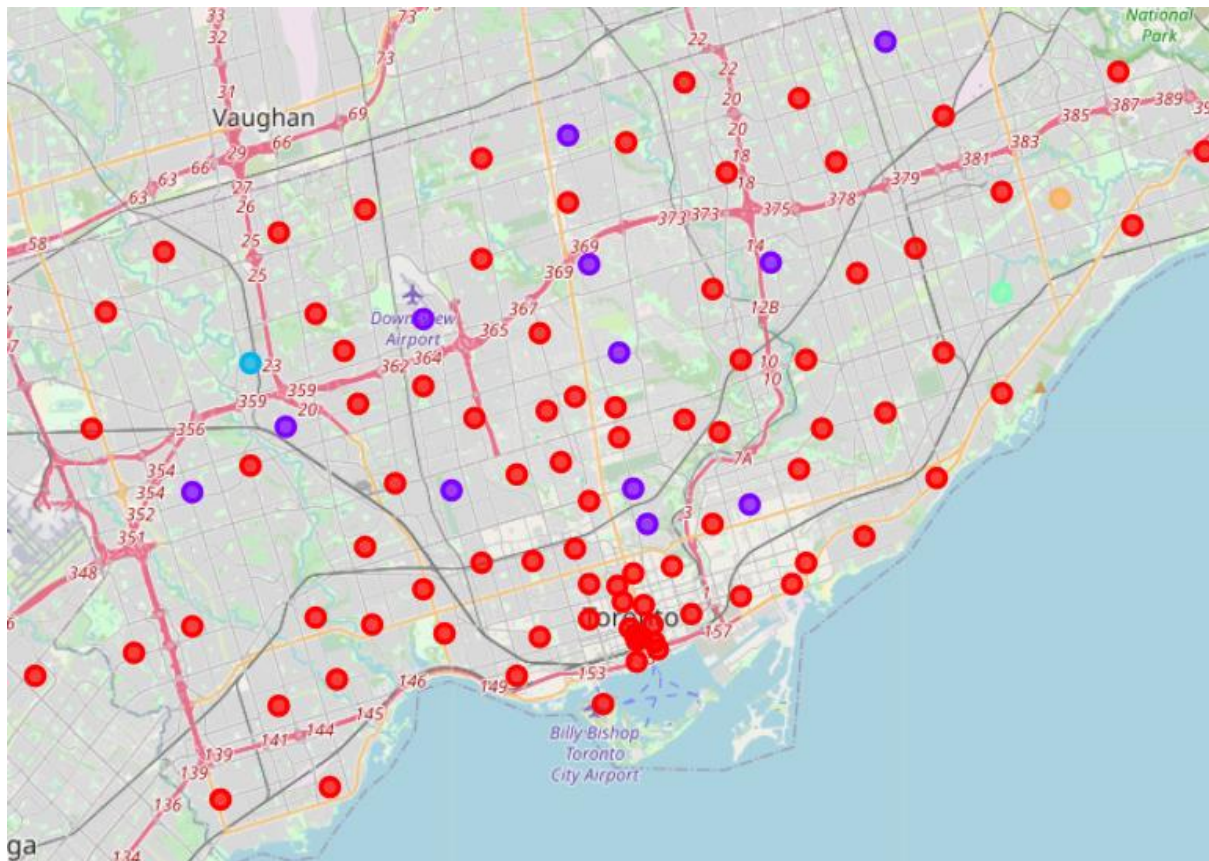
We then plot the neighbourhoods which have the highest density of venues by using Heatmaps, in which red indicates the highest density of venues and blues indicates the lowest number of venues.



From the above heatmap we can see that most of the venues located in Toronto are in Downtown Toronto and North York.

Since, there are a lot of places in **Downtown Toronto and North York**, it is *safe to assume that people visit these places* frequently and thus, are favourable places to open an Indian restaurant.

Last step is to cluster all the neighbourhoods according to the top 10 most common places found in those neighbourhoods. We set number of clusters to be 5 to find what differentiates the neighbourhoods. After running KMeans clustering, we found that cluster 0 i.e. the red cluster is the most dominant which consists mostly of restaurants, food places and parks. These are also the places where people usually go to. Thus, areas representing this cluster are good for an Indian restaurant. Following is the map where the clusters are plotted.



4. Results and Discussion

In our analysis we found that there are not a lot of Indian Restaurants in Toronto. So, we need not worry about the competing with other Indian restaurants. Therefore, we needed to find places which people frequent the most. I was planning to use the **trending** places in each neighbourhood instead of number of venues in a neighbourhood. I couldn't use the trending method because the Foursquare API was down at that moment.

So, by using a Heatmap we found out that areas around Downtown Toronto and North York have the greatest number of venues compared to other neighbourhoods. So, these are the areas where we can potentially open an Indian restaurant.

In the next step we clustered all the neighbourhoods based on the type of venue it had and found that cluster 0 i.e. the red cluster is the best cluster for us to consider as it has the most number of coffee shops, restaurants and delis/bodega. Therefore, are the high chances that if we open any type of restaurant in these areas, it will have a considerable amount of people coming in.

5. Conclusion

We did our analysis on Toronto's neighborhood to come up with the best place to open an Indian Restaurant. After plotting the Heatmap and clustering the neighborhoods for similarity in venues, we found out that Downtown Toronto and the areas surrounding it are the best places to build a Indian restaurant.