The Battle of Neighborhoods - Report

July 11, 2019

1 Battle of Neighborhoods (Week 2)

1.1 Introduction

1.1.1 Description of the Problem

A restaurant is a business which prepares and serves food and drink to customers. It is no surprise to learn that owning a restaurant comes with its fair share of responsibilities that require you to be on your feet. To be successful in the long run, a restaurant needs an accessible location, a theme or style that appeals to a broad range of customers, and a solid menu. Restaurants must stand out from the crowd and finalizing a space within the budget.

So it is evident that to survive in such competitive market it is very important to startegically plan. Often there's an issue of time constraint that keeps a restaurant owner preoccupied with intricacies. Whether it is sourcing the right ingredients, obtaining too many licences, finalizing a space within your budget to interviewing the right staff — restaurateurs have to consider everything beginning from a pin to piano before taking the right foot forward.

First move is very important, thereby choice of location is very important. So, HSB India (P) Ltd., plan to choose the correct location to start its first restaurant in Toronto, Canada.

1.1.2 Discussion of the Background - Prospects of a Indian Restaurant in Toronto, Canada.

Toronto, Canada's largest city is consistently rated among the world's top most liveable cities. Toronto boasts a multicultural population, diverse neighbourhoods, and world class cultural attractions, entertainment, dining and shopping.

Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. Its varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities, attract over 25 million tourists each year.

People have travelled through and inhabited the Toronto area and the diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. According to the 2011 National Household Survey (NHS), 1,264,395 non-Whites, or 20.2% of Canada's visible minority population, live in the city of Toronto; of this, approximately 70 percent are of Asian ancestry.

These indicates that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk and the Return on Investment will be reasonable.

Target Audience

1. Business personnel who wants to invest or open a ethnic restaurant. This analysis will be a comprehensive guide to start ot expand restaurants targeting the ethnica diversity in Toronto.
2. Any Freelancer who wants to start a new restaurant in Toronto.
3. New graduates, to find reasonable lunch/breakfast place close to office.
4. Budding Data Scientists, who wants to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it and, finally be able to tell a story out of it.

Success Criteria The success criteria of the project will be a good recommendation of Borough/Neighborhood choice to HSB India (P) Ltd., based on lack of such restaurants in that location and nearest suppliers of ingredients.

1.2 Data

Neighborhood has a total of 11 boroughs and 103 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 11 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.

Postal codes beginning with M are located within the City of Toronto in the province of Ontario. I used the BeautifulSoup website scraping library to extract the content on the Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) as XML and transfor data that is in the table of postal codes into a pandas dataframe.

I use the csv file that has the geographical coordinates of each postal code from http://cocl.us/Geospatial_data as a pandas dataframe. Perform a merge the Latitude and Longitude from the Geospatial dataframe with the Toronto neighbourhoods using inner join.

Then, sliced the dataframe to have only the boroughs that contain the word Toronto. The dataframe that is used for exploratory analysis and clustering is shown below.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

alt text

1.3 Methodology

I used the geopy library to get the latitude and longitude values of Toronto City which is used as input to the Foursquare API, to explore venues information in each neighborhood. The geograpical coordinate of Toronto City are 43.653963, -79.387207 is obtained with an instance of geocoder.

Next, I am going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

1.3.1 Exploring the Neighborhoods

The Foursquare API will be used to obtain the venues for the geographical location data. These will be used to explore the neighborhoods of Toronto accordingly.

1. Create the get request url (Foursquare ID and Secret are necessary)
   a. Number of Venues we will look for is 100
   b. Radius of Search Would be 0.5 km.
2. Create a json from the request object (using requests Module)
3. Create the lists containing the information
4. Create the dataframe from the list.

    The venues within the neighborhoods of Toronto City like restaurants and proximity to amenities would be correlated. Also, accessibility and ease of supplies would be considered as it relates to venues. I extracted the restaurant venues nearby Toronto City and the corresponding neighborhoods are . . .

```
The Beaches
The Danforth West, Riverdale
The Beaches West, India Bazaar
Studio District
Lawrence Park
Davisville North
North Toronto West
Davisville
Moore Park, Summerhill East
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
Rosedale
Cabbagetown, St. James Town
Church and Wellesley
Harbourfront, Regent Park
Ryerson, Garden District
St. James Town
Berczy Park
Central Bay Street
Adelaide, King, Richmond
Harbourfront East, Toronto Islands, Union Station
Design Exchange, Toronto Dominion Centre
Commerce Court, Victoria Hotel
Roselawn
Forest Hill North, Forest Hill West
The Annex, North Midtown, Yorkville
Harbord, University of Toronto
Chinatown, Grange Park, Kensington Market
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara
Stn A PO Boxes 25 The Esplanade
First Canadian Place, Underground city
Christie
Dovercourt Village, Dufferin
Little Portugal, Trinity
Brockton, Exhibition Place, Parkdale Village
High Park, The Junction South
Parkdale, Roncesvalles
Runnymede, Swansea
Business Reply Mail Processing Centre 969 Eastern
```

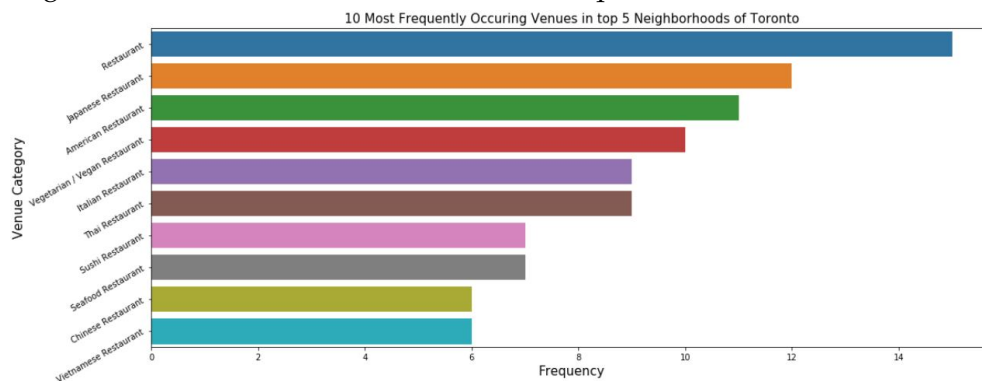I took the top 5 neighborhoods of Toronto City thost has the maximum number of restaurants.

	Neighborhood	Number of Restaurants
0	Chinatown, Grange Park, Kensington Market	35
1	Adelaide, King, Richmond	30
2	Church and Wellesley	29
3	Commerce Court, Victoria Hotel	27
4	St. James Town	26

I merged the venue details of those top 5 neighborhoods of Toronto City to keep the venue details and geographical coordinates in one dataframe.

Map Visualization   Created a map using the Folium library with new markers for those top 5 Toronto Neighborhoods.
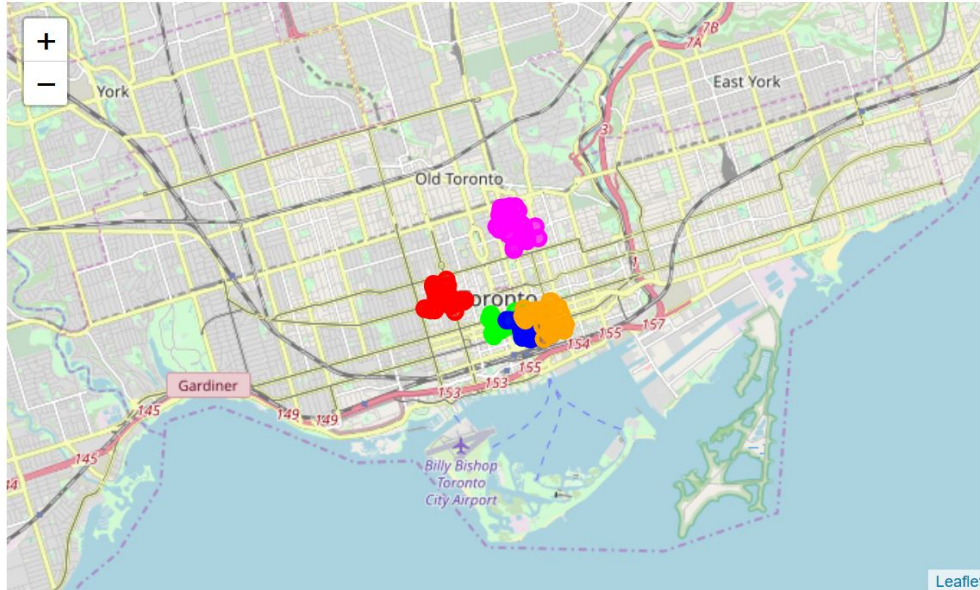

This visualization of Toronto City shows the density of restaurants in those top 5 neighborhoods and indicates the presence of more restaurants due to multi cultural population. I have seen that there are 37 unique venue categories curated from all the returned venues of the top 5 neighborhoods. Let's visualize the 10 most frequent restaurant venues in the top 5 neighborhoods.

Also, created a bar chart to show the top 5 neighborhoods with maximum number of restaurants.

Top-Most Neighborhoods of Toronto with max Number of Restaurants
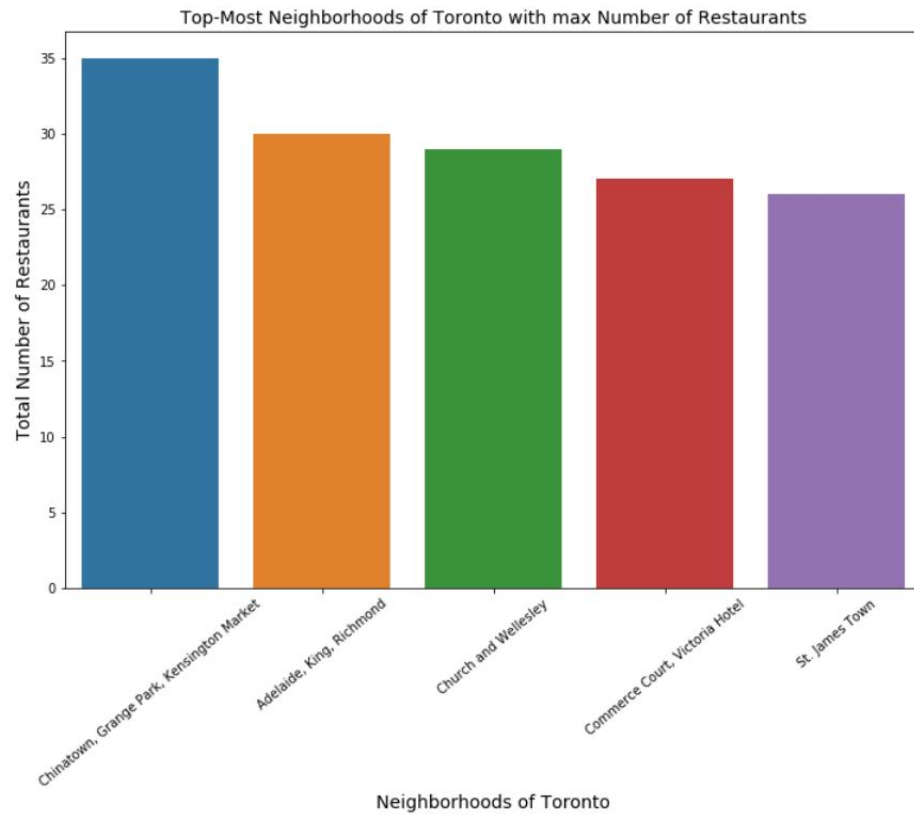
Next, I would like to explore the Indian Restaurants in those top 5 neighborhoods. I found 3 out of the top 5 neighborhoods has a Indian Restaurant.

Top-Most Neighborhoods of Toronto with max Number of Indian Restaurants
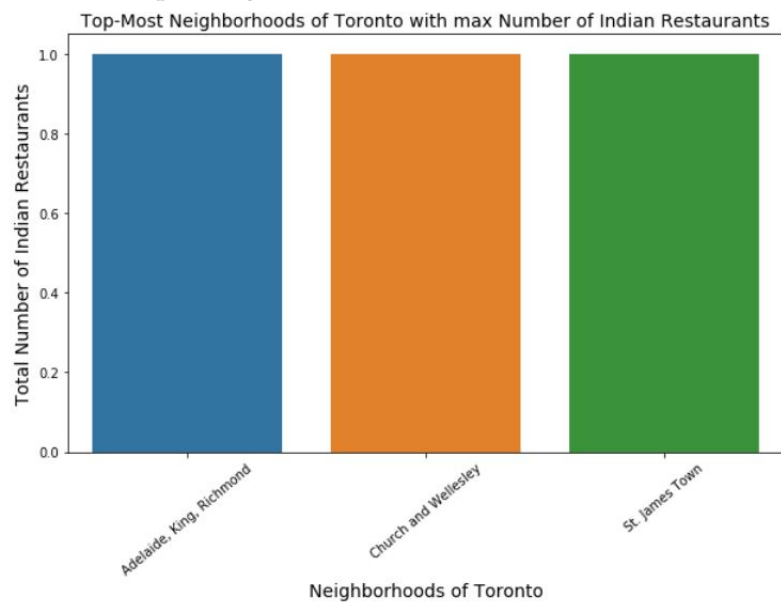
1.3.2 Analyze Top 5 Neighborhood

Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Let's print each neighborhood along with the top 5 most common venues.

—-Adelaide, King, Richmond—-

venue freq
0 American Restaurant 0.13
1 Thai Restaurant 0.13
2 Asian Restaurant 0.10
3 Sushi Restaurant 0.10
4 Restaurant 0.10
    —-Chinatown, Grange Park, Kensington Market—-
venue freq
0 Vegetarian / Vegan Restaurant 0.17
1 Vietnamese Restaurant 0.14
2 Chinese Restaurant 0.11
3 Dumpling Restaurant 0.11
4 Mexican Restaurant 0.11
    —-Church and Wellesley—-
venue freq
0 Japanese Restaurant 0.17
1 Sushi Restaurant 0.14
2 Restaurant 0.10
3 Fast Food Restaurant 0.07
4 Mediterranean Restaurant 0.07
    —-Commerce Court, Victoria Hotel—-
venue freq
0 American Restaurant 0.15
1 Restaurant 0.15
2 Seafood Restaurant 0.11
3 Italian Restaurant 0.07
4 Japanese Restaurant 0.07
    —-St. James Town—-
venue freq
0 Restaurant 0.19
1 Italian Restaurant 0.15
2 Japanese Restaurant 0.08
3 American Restaurant 0.08
4 Seafood Restaurant 0.08

1.3.3   Cluster Neighborhoods

K-means clustering is a type of unsupervised learning, which I used to to cluster the neighborhood into 5 clusters and I have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The results of the K-means clustering algorithm are:

The centroids of the K clusters, which can be used to label new data
Labels for the training data (each data point is assigned to a single cluster)

Finally, let's visualize the resulting clusters.
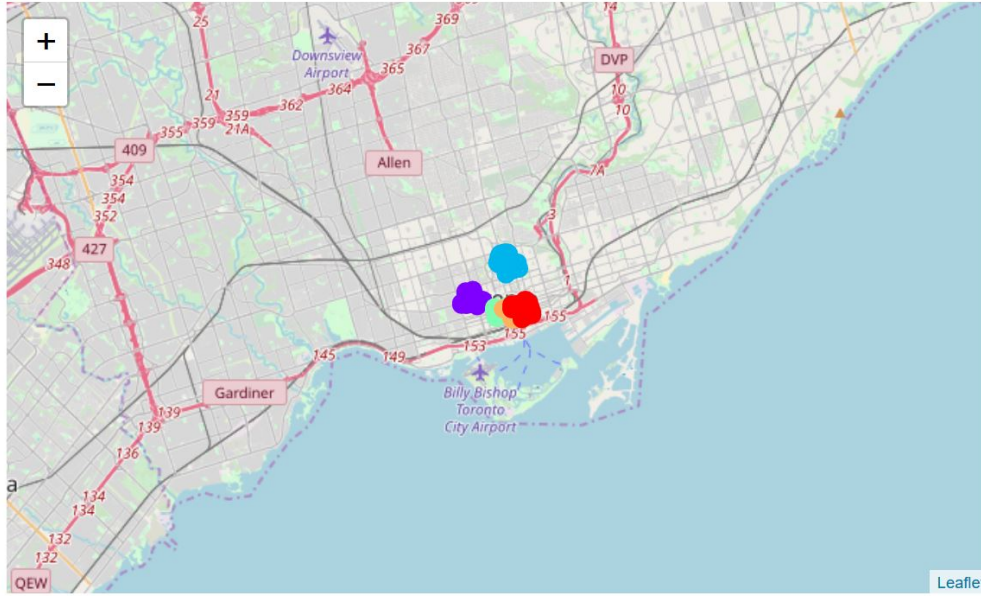
1.4 Results

Finally, I am analyzing the top 10 most common venues in the top 5 neighborhoods. For this, the clusters created using the K-Means algorithm will be of great help and let us examing each cluster.

The results of the exploratory data analysis and clustering are:

Japanese Restaurants tops the chart in the top 5 neighborhoods.

Indian Restaurant is one of the top 10 venue in St. James Town and presence in Church and Wellesley and Adelaide, King, Richmond neighborhoods.

Popular restaurants in each neighborhood:

1. St. James Town: Italian, American, Japanese
2. Chinatown, Grange Park, Kensington Market: Vietnamese, Chinese
3. Church and Wellesley: Japanese
4. Adelaide, King, Richmond: American, Thai
5. Commerce Court, Victoria Hotel: American, Thai, Japanese

As for restaurants, the Japanese Restaurants are very popular in the East Toronto area. Although, the clusters have variations, a very visible presence is the predominance of Asian Restaurants. It is very important to note that Clusters 1, 3 and 4 are the most suitable locations to open a brand Indian Restaurant.

1.5 Discussion

According to this analysis, it is very important to note that Clusters 1 and 3 are the most viable clusters to start a new Indian Restaurant. These 2 clusters do not have top restaurants that could rival their standards if they are opened.

Some drawbacks of this analysis are: the clustering is completely based on the most common venues obtained from Foursquare data. Since diverse population, accessible location, broad range of customers could all play a major role and thus, this analysis might definitely far from being conclusory. However, it definitely gives us some very important preliminary information on possibilities of opening restaurants around the important neighborhoods of Toronto.

1.6 Conclusion

In conclusion, I have got a small glimpse of how real life data-science projects look like. I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium leaflet map. This project would have had better results if there were more data in terms of accessibility data within the area, traffic access and allowance of more venues exploration with the Foursquare (limited venues for free calls).

Also, getting the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insight into the best location. Finally, all of the analysis were mostly concentrated on the possibilities of opening a restaurant targeting the diverse population.

1.6.1 Thank You !