
Capstone Project

The Battle of Neighborhoods – Report

- by Subhayan Sur (23-Sep-2019)

1. Introduction

1.1 Background

Every now and then we see new restaurants coming up in our neighborhoods. Some become successful, while others fail. Now if we have some insights about the location where the restaurant business has a high demand, has higher volume of customers, customer's meal and location preferences then we can successfully use these insights and find out where opening a new restaurant and its type will likely be a good choice.

To minimize the chances of failing in restaurant business, we should do proper research and planning before opening a restaurant. Consider the factors mentioned in this project when opening a new restaurant in any neighborhood so that you can save your precious - time, money and energy.

1.2 Problem

In this project we will attempt to find an optimal neighborhood for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a new restaurant in any of the neighborhoods of **Toronto**, Canada.

Since there are lots of restaurants in Toronto, this project aims to detect **neighborhoods that have - various types of restaurants as their most common venue**. By doing so we'll ensure that there is a demand for a restaurant in that neighborhood and our client can safely open one.

We will use our data science powers to generate the most promising neighborhood based on this criterion. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1.3 Interest

This project will be particularly interesting to those who are thinking about starting a new restaurant in a city for which they don't have much of an insight.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

Based on definition of our problem, factors that will influence our decision are number of existing restaurants in the neighborhood (any type of restaurant) and neighborhoods that have many restaurants as their most common venues.

Following data sources will be needed to extract/generate the required information:

1. The Borough and neighborhood information from Wikipedia
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

The columns were as follows:

- i) Postcode: The postal codes for various neighborhoods of Toronto.
 - ii) Borough: The names of the boroughs, 'not Assigned' – if they're not assigned to any borough.
 - iii) Neighborhood: The names of the neighborhoods, 'not Assigned' – if they do not have any assigned neighborhoods.
2. As Google Maps API is taking a long time to load and is failing hence we are using the geospatial data provided - http://cocl.us/Geospatial_data to perform the task of updating the coordinates and create clusters
 3. The number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
 4. We'll use Folium package in Python to create the map of Toronto using latitude and longitude values.

2.2 Data Cleaning

We performed the following data cleaning activities –

1. Updated the column names in the data frame with proper connotation.
2. Deleted the new lines from the data frame.
3. Deleted the records which doesn't have a borough assigned.
4. Removed spaces at the beginning of the strings.
5. Combined the neighborhoods with same Postal Code and Boroughs using a comma separator.

The dataset after performing cleaning and merging:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

3. Methodology

In this project we directed our efforts on detecting neighborhoods of Toronto that have high restaurant density, particularly those with which have restaurants as their most frequented venues.

In Part 1, We gathered Toronto's neighborhood information from Wikipedia sources and created a clean data frame.

Data set after Part 1, looks like below:

	PostalCode	Borough	Neighborhood
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

In Part 2, We prepared the data frames with Neighborhood coordinates, postal code, borough and neighborhood information that was used for clustering. We also used the geospatial data from the link.

Data set after Part 2, looks like below:

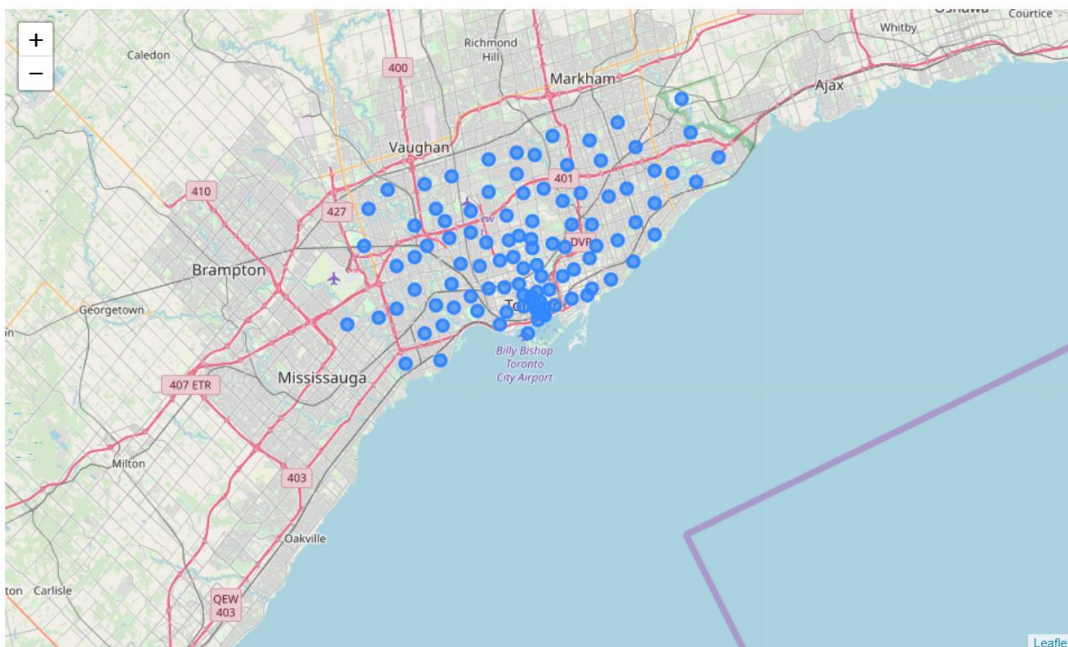
	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

In the final Part 3, we focused on most promising areas and within those areas create clusters of locations that meet our requirements established in discussion with stakeholders: we took into consideration locations with multiple restaurants in the vicinity, and we want locations that have various restaurants as their most common venue. We presented map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods where the conditions for opening a new restaurant will be the best.

Dataset after clustering in Part 3, looks like below:

	Borough	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	Scarborough	2.0	Fast Food Restaurant	Women's Store	Event Space	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Donut Shop	Dog Run
13	Scarborough	2.0	Pizza Place	Chinese Restaurant	Italian Restaurant	Noodle House	Bank	Breakfast Spot	Fried Chicken Joint	Fast Food Restaurant	Thai Restaurant	Dessert Shop
15	Scarborough	2.0	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Pizza Place	Sandwich Place	Bubble Tea Shop	Thrift / Vintage Store	Grocery Store	Pharmacy	Breakfast Spot
17	North York	2.0	Dog Run	Golf Course	Fast Food Restaurant	Mediterranean Restaurant	Pool	Discount Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
24	North York	2.0	Coffee Shop	Butcher	Home Service	Discount Store	Pharmacy	Pizza Place	Grocery Store	Airport Terminal	Falafel Restaurant	Ethiopian Restaurant
34	North York	2.0	Coffee Shop	Portuguese Restaurant	Pizza Place	Hockey Arena	Intersection	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
35	East York	2.0	Pizza Place	Fast Food Restaurant	Bank	Gym / Fitness Center	Pet Store	Pharmacy	Gastropub	Intersection	Athletics & Sports	Dessert Shop

Visualizing the clustered neighborhoods using the Folium library:



3.1. Summary of the clusters

During the 5 rounds of clustering we came across data segments that describe the most frequented places by the people based on most common venues. In each clusters the segmentations show a different grouping. However, from Cluster 2 and Cluster 3, it became apparent that Scarborough has the greatest number of restaurants as the 'most common venue' for people and visitors.

Cluster 2 - Segmentation based on Miscellaneous stores, restaurants, cafes etc

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] +
```

4]:

	Borough	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
2	Scarborough	1.0	Electronics Store	Mexican Restaurant	Breakfast Spot	Intersection	Medical Center	Rental Car Location
3	Scarborough	1.0	Coffee Shop	Convenience Store	Korean Restaurant	Women's Store	Department Store	Dessert Shop
4	Scarborough	1.0	Fried Chicken Joint	Bakery	Hakka Restaurant	Bank	Athletics & Sports	Thai Restaurant
5	Scarborough	1.0	Women's Store	Playground	Curling Ice	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

Cluster 3 - Segmentation based on Restaurants, Pizza Places

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] +
```

4]:

	Borough	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
0	Scarborough	2.0	Fast Food Restaurant	Women's Store	Event Space	Empanada Restaurant	Electronics Store	Eastern European Restaurant
13	Scarborough	2.0	Pizza Place	Chinese Restaurant	Italian Restaurant	Noodle House	Bank	Breakfast Spot
15	Scarborough	2.0	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Pizza Place	Sandwich Place	Bubble Tea Shop
17	North York	2.0	Dog Run	Golf Course	Fast Food Restaurant	Mediterranean Restaurant	Pool	Discount Store
24	North York	2.0	Coffee Shop	Butcher	Home Service	Discount Store	Pharmacy	Pizza Place
34	North York	2.0	Coffee Shop	Portuguese Restaurant	Pizza Place	Hockey Arena	Intersection	Deli / Bodega

4. Results

Our analysis shows that although there is a great number of restaurants in Toronto, yet the "1st Most Common Venue" for neighborhoods changes from Parks to Bus Stops to coffee shops to

various stores and restaurants. Highest concentration of restaurants was detected in Scarborough and North York.

From the segmentations in Cluster 2, Cluster 3 above, it is evident that Scarborough is the best place to open restaurants - as it has a wide variety of restaurants as the restaurants in this neighborhood accounts for the maximum number of "1st Most Common Venue". The various types of restaurants include - American Restaurant, Indian Restaurant, Greek Restaurant, Middle Eastern Restaurant, Fried Chicken Joint, Chinese Restaurant, Fast Food Restaurant, Pizza Place etc.

From cluster 2 and cluster 3 we can see that Scarborough has shops, bus stations, bus lines, lounges, food joints, pizza places, cafes all in one place. Hence it becomes as obvious choice to open a new restaurant as people can choose to take public transport to visit that place, do shopping and finally go to restaurant for a meal before heading back.

Which means our client can choose both to open a new restaurant that is unique to the area or a similar restaurant like the ones already existing.

5. Discussion

The aim of the project is to help people who want to open a new restaurant in a neighborhood borough of Toronto based on the most places for restaurants. If we look at cluster 2 and cluster 3, we can see that Scarborough has shops, bus stations, bus lines, lounges, food joints, pizza places, cafes all in one place. Hence it becomes as obvious choice to open a new restaurant as people can choose to take public transport to visit that place, do shopping and finally go to restaurant for a meal before heading back.

6. Conclusion

Purpose of this project was to identify Toronto neighborhoods in order to aid the stakeholder in narrowing down the search for optimal location for a new restaurant.

We first created the boroughs and neighborhoods from various data sources and added their location coordinates using geocode. Then we created a map to visualize Toronto's neighborhoods. Next we utilized the Foursquare API to explore all the neighborhoods and their most frequented venues. Then we analyzed each Neighborhood by taking the mean of the frequency of venue-occurrence for each category. Then we created a data frame with the top 10 - most common venues in each neighborhood. Then we used k-means clustering of those locations in order to create major neighborhoods of interest that were to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in the recommended neighborhood

(Scarborough), taking into consideration additional factors like attractiveness of restaurant location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every location etc.