

Capstone Project

Restaurant
Recommendation
in Toronto

Table of Content

1. Introduction/Business Problem
2. Data Description
3. Methodology
4. Results
5. Discussion
6. Conclusion

1. Introduction

Toronto is one of the largest city in Canada. It is also one of the most populous city. So, lot of analysis and investigation is required to open a restaurant in a particular place. There are lot of factors which need to be considered before opening a restaurant. Data science can be used to do such analysis. We can use Foursquare api and data analysis techniques to do analysis of neighbourhood of Toronto and decide in which place we can open the restaurant.

Target Audience –

Business personnel who wants to invest or open a restaurant.

2. Data Description

1. Toronto neighbourhood data from Wikipedia – We will use Wikipedia to get list of all the neighbourhood of Toronto city. Also we will get the latitude and longitude details of all the neighbourhood. 2. Foursquare Data – We will use Foursquare api to get the restaurant data of all the neighbourhood. We will get the data in json format and we will convert this data into dataframe.

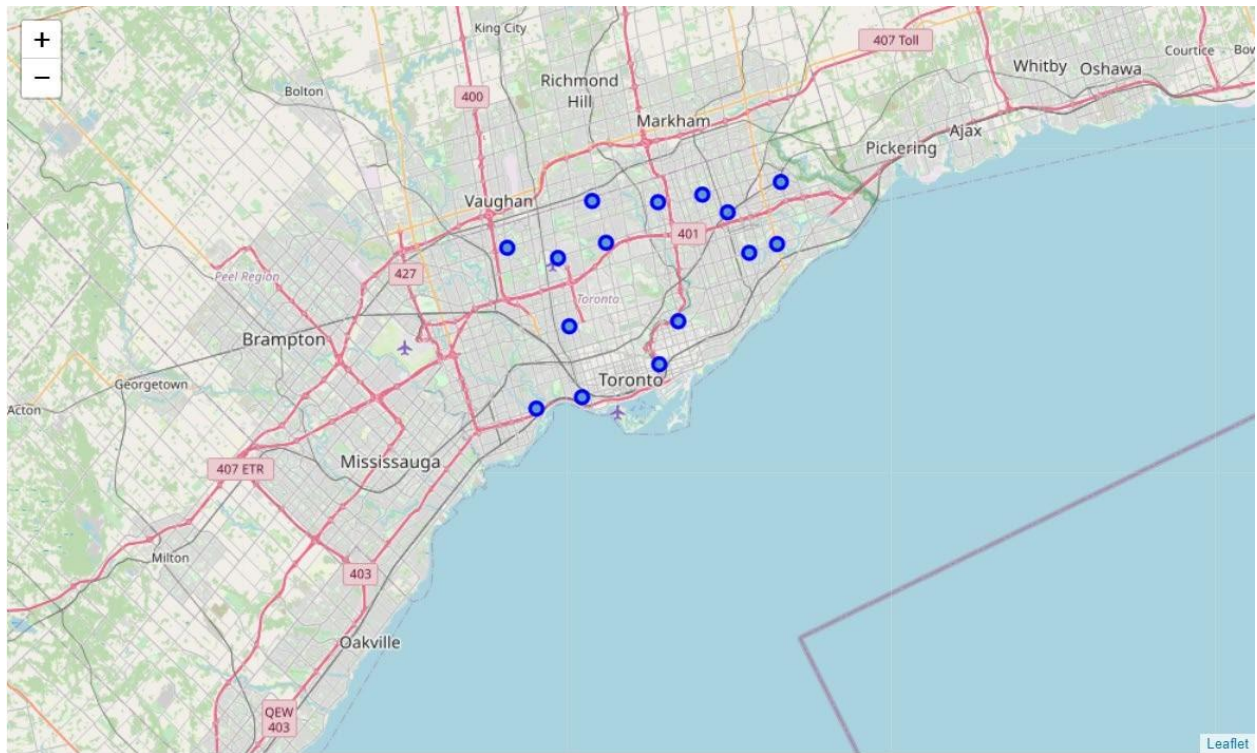
3. Methodology

Data scrapping from the Wikipedia page that contains the up-to-date population statistics of Toronto neighbourhoods has been used. Then we get the top 15 populated neighbourhoods from the all neighbourhoods in Toronto City.

Furthermore, we need to know the coordinates and locations of this neighbourhoods, and therefore the geopy package has been used for this. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighborhoods, and this is precisely what we have done for it in this project.

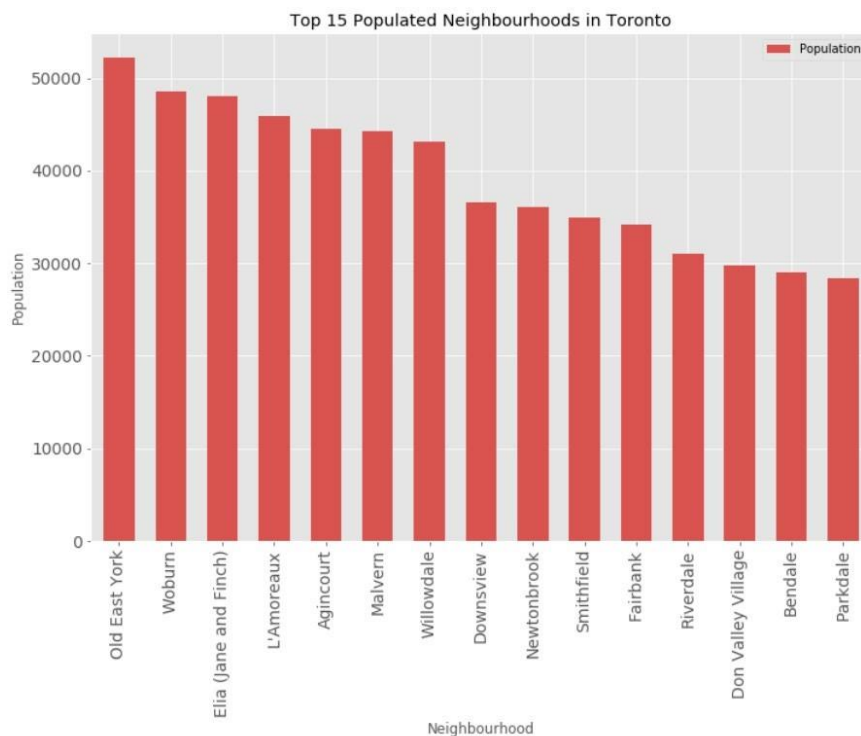
We will also use machine learnings techniques such as the K-Means Clustering to segment and cluster these neighbourhoods so that we can group them together to understand their similarities.

Top 15 Populated Neighbourhoods in Toronto



4. Results

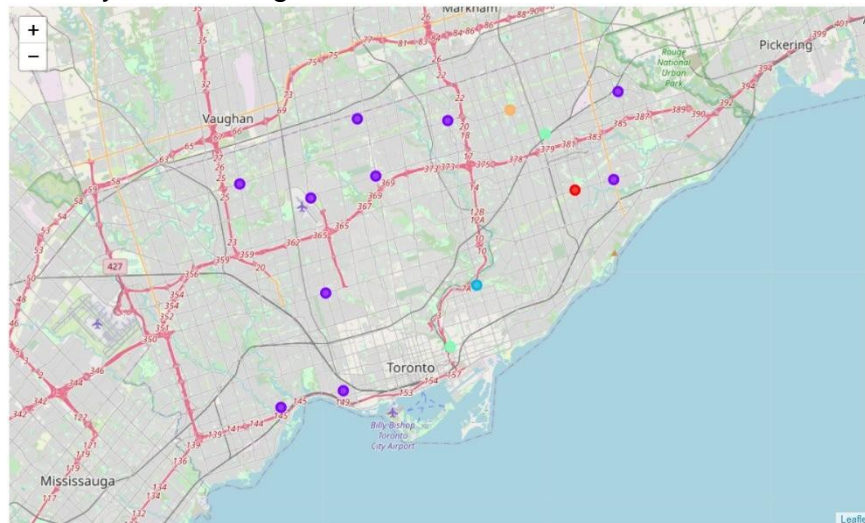
With the use of Bar Chart, we visualized the top 15 populated neighbourhoods and most populated neighbourhood was Old East York.



By Foursquare API, we were also able to leverage on the data to find out the top 10 common nearby venues and their categories in each of these neighborhoods.

	Neighbourhood	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	Agincourt	Chinese Restaurant	Asian Restaurant	Coffee Shop	Butcher	Korean Restaurant	Hong Kong Restaurant	Cantonese Restaurant	Shopping Mall	Vietnamese Restaurant	Food Court
1	Bendale	Dog Run	Grocery Store	Fast Food Restaurant	Tennis Court	Optical Shop	Discount Store	Yoga Studio	Chiropractor	Chocolate Shop	Coffee Shop
2	Don Valley Village	Sandwich Place	Bank	Coffee Shop	Park	Pharmacy	Pizza Place	Sushi Restaurant	Chinese Restaurant	Grocery Store	Farmers Market
3	Downsview	Coffee Shop	French Restaurant	Gas Station	Outdoor Supply Store	Park	Gym / Fitness Center	Gym Pool	Men's Store	Bus Station	Discount Store
4	Elia (Jane and Finch)	Grocery Store	Fast Food Restaurant	Pizza Place	Discount Store	Shopping Mall	Sandwich Place	Beer Store	Caribbean Restaurant	Coffee Shop	Liquor Store
5	Fairbank	Coffee Shop	Grocery Store	Food Truck	Park	Bakery	Bank	Pharmacy	Discount Store	Japanese Restaurant	Yoga Studio
6	L'Amoreaux	Shopping Mall	Coffee Shop	Athletics & Sports	Chinese Restaurant	Yoga Studio	Chocolate Shop	Convenience Store	Deli / Bodega	Department Store	Dessert Shop
7	Malvern	Fast Food Restaurant	Pizza Place	Pharmacy	Sandwich Place	Skating Rink	Bubble Tea Shop	Park	Convenience Store	Gym / Fitness Center	Grocery Store
8	Newtonbrook	Korean Restaurant	Coffee Shop	Middle Eastern Restaurant	Restaurant	Hardware Store	Vietnamese Restaurant	Discount Store	Fast Food Restaurant	Café	Cantonese Restaurant
9	Old East York	Park	Pastry Shop	Liquor Store	Pizza Place	Pub	Restaurant	Yoga Studio	Dessert Shop	Chiropractor	Chocolate Shop

Using K-Means Clustering technique, we divide neighbourhoods into 5 clusters. Within a cluster, we can say that all neighbourhoods have similar characteristics.



Cluster1

	Neighbourhood	Population	2nd Language	Latitude	Longitude	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
13	Bendale	28945	Tamil (3.7%)	43.75352	-79.255336	0	Dog Run	Grocery Store	Fast Food Restaurant	Tennis Court	Optical Shop	Discount Store	Yoga Studio

Cluster2

	Neighbourhood	Population	2nd Language	Latitude	Longitude	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
1	Woburn	48507	Gujarati (9.1%)	43.759824	-79.225291	1	Fast Food Restaurant	Coffee Shop	Discount Store	Bank	Beer Store	Hardware Store	Depa Store
2	Elia (Jane and Finch)	48003	Vietnamese (6.9%)	43.757253	-79.517697	1	Grocery Store	Fast Food Restaurant	Pizza Place	Discount Store	Shopping Mall	Sandwich Place	Beer
5	Malvern	44324	Tamil (12.2%)	43.809196	-79.221701	1	Fast Food Restaurant	Pizza Place	Pharmacy	Sandwich Place	Skating Rink	Bubble Tea Shop	Park

Cluster3

	Neighbourhood	Population	2nd Language	Latitude	Longitude	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
0	Old East York	52220	Greek (4.3%)	43.699971	-79.33252	2	Park	Pastry Shop	Liquor Store	Pizza Place	Pub	Restaurant	Yoga Studio

Cluster4

	Neighbourhood	Population	2nd Language	Latitude	Longitude	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
4	Agincourt	44577	Cantonese (19.3%)	43.785353	-79.278549	3	Chinese Restaurant	Asian Restaurant	Coffee Shop	Butcher	Korean Restaurant	Hong Kong Restaurant	Ca Re
11	Riverdale	31007	Cantonese (6.7%)	43.665470	-79.352594	3	Vietnamese Restaurant	Chinese Restaurant	Bakery	Coffee Shop	Grocery Store	Fast Food Restaurant	Trä

Cluster5

	Neighbourhood	Population	2nd Language	Latitude	Longitude	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
3	L'Amoreaux	45862	Unspecified Chinese (13.9%)	43.799003	-79.305967	4	Shopping Mall	Coffee Shop	Athletics & Sports	Chinese Restaurant	Yoga Studio	Chocolate Shop	Conver Store

5. Discussion

Based on above results from top 15 populated neighbourhoods, the second cluster (i.e. Cluster label=1) has a higher number of neighbourhoods with similar characteristics compare to other clusters.

Within the second cluster, we would like to recommend a neighbourhood with high demand for Fast Food. As you can see there are two neighbourhoods with Fast Food Restaurant as their 1st Most Common Venue namely *Woburn* and *Malvern*. Note that the neighbourhood **Woburn** looks to be **the best choice** as it is the most populated (i.e.48507) in the second cluster.

Another point to highlight is that the 2nd Language of Woburn is Gujarati. We can think that there is a higher number of Gujarati people in that neighbourhood. So that it would be better to concern about Gujarati Fast Foods to attract more customers. Since Woburn and Malvern are nearby neighbourhoods, we will also encourage to offer South-Indian fast food. (because the 2nd Language of Malvern is Tamil)

6. Conclusion

With all these information, we can conclude our recommendation as follows, Region: 2nd Cluster ,Best
Neighbourhood: Woburn