IBM Data Science - Capstone Project

Fast Food Restaurant

Recommendation

in Toronto

Table of Content

1. Introduction/Business Problem

2. Data Description

3. Methodology

4. Results

5. Discussion

6. Conclusion

1. Introduction

1.1 Background

Toronto is the financial capital of Canada, with lots of business opportunities and business friendly environment, it has no issue in attracting many different players into the market. However, that also means the market is highly competitive and as a well-developed city, the cost of doing business is also one of the highest in the country. And thus, any new business venture or expansion in the country needs to be reviewed carefully and strategically targeted so that the return on investment will be sustainably reasonable and more importantly the investment can be considerably less risker.

1.2 Business Problem

Let say if someone is looking to open a **Fast Food Restaurant** in Toronto. The **main problem/question** is where or which is the best neighbourhood to open the restaurant in Toronto city? In order to answer this problem, we have to consider may things such as there must be enough customers, preferences of residents in the relevant neighbourhood etc.

Toronto is one of the most multicultural and multiracial cities in the world. So, there are many ethnic origins live in the toronto city. Thus, we have to match their food preferences in order to make profits. In order to do that we will use clustering to find some similar neighbourhoods and choose the best neighbourhood among them.

1.3 Target Audience

This project's findings are targeted for anyone who wants to open a Fast Food Restaurant in Toronto city. The objective is to locate and recommend to the businessman which region of the neighborhoods in Toronto will be the best choice to start off their first fast food restaurant including online delivery services.

2. Data Description

As we need to explore, segment, and cluster the neighbourhoods in the city of Toronto, the Toronto neighbourhood data is key for this project. However, the neighborhood data is not readily available on the internet. Thus we need to scrap an existing Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format.

- Get Demographics of Toronto Neighbourhoods from Wikipedia page (https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods)

- Scape from the web page using **beautifulsoup** package and read into **pandas** dataframe

	Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
0	Toronto CMA Average	NaN	All	5113149	5903.63	866	9.0	40704	10.6	11.4	NaN	NaN	NaN
1	Agincourt	S	0377.01, 0377.02, 0377.03, 0377.04, 0378.02, 0...	44577	12.45	3580	4.6	25750	11.1	5.9	Cantonese (19.3%)	19.3% Cantonese	NaN
2	Alderwood	E	0211.00, 0212.00	11656	4.94	2360	-4.0	35239	8.8	8.5	Polish (6.2%)	06.2% Polish	NaN
3	Alexandra Park	OCoT	0039.00	4355	0.32	13609	0.0	19687	13.8	28.0	Cantonese (17.9%)	17.9% Cantonese	NaN
4	Allenby	OCoT	0140.00	2513	0.58	4333	-1.0	245592	5.2	3.4	Russian (1.4%)	01.4% Russian	NaN
5	Amesbury	NY	0280.00, 0281.01, 0281.02	17318	3.51	4934	1.1	27546	16.4	19.7	Spanish (6.1%)	06.1% Spanish	NaN
6	Armour Heights	NY	0298.00	4384	2.29	1914	2.0	116651	10.8	16.1	Russian (9.4%)	09.4% Russian	NaN

- Clean the dataframe by removing unwanted data

	Neighbourhood	Population	Land Area	Density	Average Income	2nd Language
1	Agincourt	44577	12.45	3580	25750	Cantonese (19.3%)
2	Alderwood	11656	4.94	2360	35239	Polish (6.2%)
3	Alexandra Park	4355	0.32	13609	19687	Cantonese (17.9%)
4	Allenby	2513	0.58	4333	245592	Russian (1.4%)
5	Amesbury	17318	3.51	4934	27546	Spanish (6.1%)

- Get location information using **GeoPy** package

	Neighbourhood	Latitude	Longitude
0	Old East York	43.699971	-79.332520
1	Woburn	43.759824	-79.225291
2	Elia (Jane and Finch)	43.757253	-79.517697
3	L'Amoreaux	43.799003	-79.305967
4	Agincourt	43.785353	-79.278549
5	Malvern	43.809196	-79.221701
6	Willowdale	43.761510	-79.410923
7	Downsview	43.749299	-79.462248
8	Newtonbrook	43.793886	-79.425679
9	Smithfield	43.631184	-79.485667
10	Fairbank	43.695689	-79.450310
11	Riverdale	43.665470	-79.352594
12	Don Valley Village	43.792673	-79.354722
13	Bendale	43.753520	-79.255336
14	Parkdale	43.640495	-79.436897

- Get venue information using **Foursquare API**

We need to leverage on features in a reliable location information provider namely [Foursquare](#) to explore the various types of venues and its categories available in each neighbourhood.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Old East York	43.699971	-79.33252	Mon K Patisserie	43.696922	-79.329520	Pastry Shop
1	Old East York	43.699971	-79.33252	LCBO	43.696728	-79.328875	Liquor Store
2	Old East York	43.699971	-79.33252	Remarks Bar & Grill	43.696726	-79.329219	Pub
3	Old East York	43.699971	-79.33252	Pizza Hut	43.696383	-79.328778	Pizza Place
4	Old East York	43.699971	-79.33252	Taylor Creek Park - West	43.701260	-79.331726	Park

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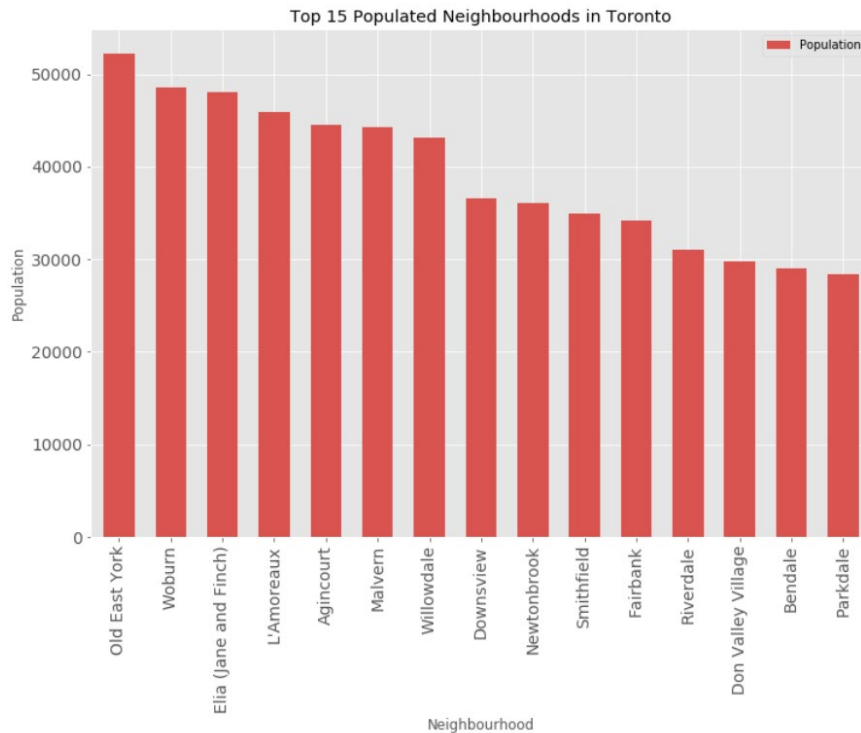
By Foursquare API, we were also able to leverage on the data to find out the top 10 common nearby venues and their categories in each of these neighborhoods.

	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.
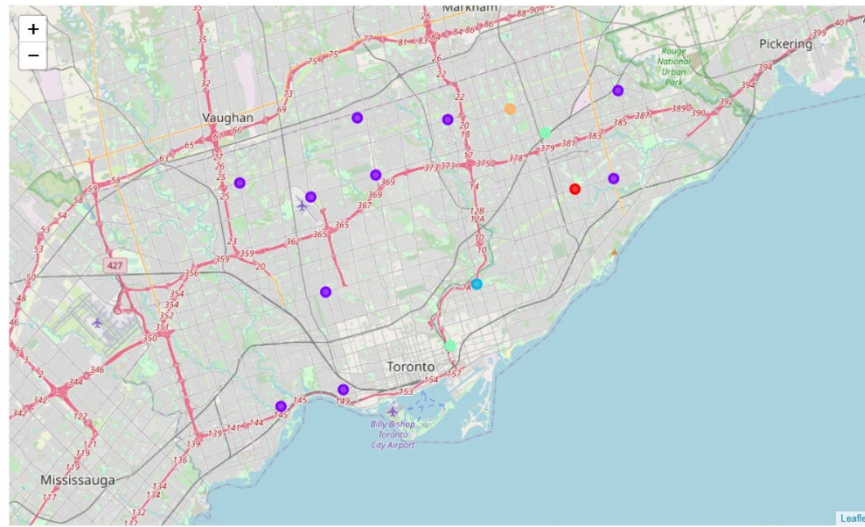
Cluster1

	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 Mos Common Venu
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 Co
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	7 C
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	Tra

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 Cor V
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
- Target customers: Mainly for Gujarati people
- Additional considerations: Indian Fast Food (Mainly Gujarati food and South-Indian food)