capstone project – battle of neighborhoods

Detecting promising neighborhoods for the next boba storefront

Jiaqi Chen

November 4, 2019

1. Introduction

1.1 Background

Toronto has a population of 2.9 million, being the most populous cities in Canada. As one of the most diverse places in the planet, Toronto has been attracting foods and drinks from different regions and countries. From 2016, the market of Taiwan's drink had been growing rapidly. Today, the number of bubble tea shops Toronto has exceeded 20 and they are busy establishing franchise chain.

One alluring thing about boba market is that the cost of opening a new storefront selling bubble tea- not much specialized equipment, low staffing needs- is much lower than a restaurant. But the gradually fierce war among boba shops making it difficult for players to survive for a long term.

The ones that will survive, He says, will either have built a strong enough brand that customers will pick them over other options, or they'll have enough shops to be a more convenient option than the competition.[1]

Why are so many new bubble tea shops opening in Toronto? BY NATALIA MANZOCCO

Therefore, making effective decision on the location of new storefronts to make shareholders one step ahead is bound to bring advantages.

1.2 Problem

The goal of this project is to detect promising neighborhoods for new boba storefronts. In this project, we only take peer competition and customer sources into consideration. Essential data include the number and location of existing boba shops, schools, shopping places in each neighborhood.

1.3 Interest

Shareholders and investors of boba brands targeting Toronto boba market can get profit from this report.

2. Data acquisition and cleansing

2.1 Data sources

1. names of Toronto neighborhoods will be obtained from **Wikipedia** by using **BeautifulSoup**
2. coordinate of neighborhoods will be obtained from **cousera** in case of the disconnection of **geocoder**

	PostalCode	Borough	Neighbourhood	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

figure 1 coordinate of neighborhoods on the basis of postal code

3. number and coordinate of existing boba storefronts will be obtained using **Foursquare API**

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Maryvale,Wexford	43.750071	-79.295849	Pho Metro	43.745365	-79.294462
1	Agincourt	43.794200	-79.262029	Health Oolong Tea	43.789042	-79.268513
2	Agincourt	43.794200	-79.262029	Real Fruit Bubble Tea 真果茶坊	43.797208	-79.271523
3	Agincourt North,L'Amoreaux East,Milliken,Steel...	43.815252	-79.284577	Go For Tea	43.814701	-79.292643
4	Agincourt North,L'Amoreaux East,Milliken,Steel...	43.815252	-79.284577	OneZo Tapioca 丸作食茶	43.815838	-79.293655

figure 2 coordinate of boba shops in each neighborhood

4. number and coordinate of universities and colleges and their location in every neighborhood will be obtained using **Foursquare API**

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Rouge,Malvern	43.806686	-79.194353	Evergreen College	43.802162	-79.199654
1	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	UTSC - Physics Labs	43.780079	-79.156193
2	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	st brendan catholic school	43.783052	-79.149267
3	Guildwood,Morningside,West Hill	43.763573	-79.188711	Z Cups Etc	43.767973	-79.188123
4	Guildwood,Morningside,West Hill	43.763573	-79.188711	Boys and Girls Club of East Scarborough	43.757549	-79.193691

figure 3 coordinate of schools in each neighborhood

5. number and coordinate of shopping places and their location in every neighborhood will be obtained using **Foursquare API**

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Guildwood,Morningside,West Hill	43.763573	-79.188711	Westhill Plaza	43.768039	-79.189845
1	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304	Midland Lawrence Plaza	43.756688	-79.265784
2	Clarks Corners,Sullivan,Tam O'Shanter	43.781638	-79.304302	Warden Plaza	43.785275	-79.310410
3	Agincourt North,L'Amoreaux East,Milliken,Steel...	43.815252	-79.284577	Midland Square	43.818752	-79.290249
4	Agincourt North,L'Amoreaux East,Milliken,Steel...	43.815252	-79.284577	Maxim ii	43.814862	-79.293021

figure 4 coordinate of shopping plazes in each neighborhood

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Rouge,Malvern	43.806686	-79.194353	Pleasant Corner	43.801164	-79.200254
1	Guildwood,Morningside,West Hill	43.763573	-79.188711	Kingston Square	43.769860	-79.187158
2	Guildwood,Morningside,West Hill	43.763573	-79.188711	Morningside Crossing	43.770599	-79.185541
3	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029	Midland Plaza	43.734730	-79.259544
4	Maryvale,Wexford	43.750071	-79.295849	Wexford Heights Plaza	43.746136	-79.293782

figure 5 coordinate of shopping malls in each neighborhood

The exploring radius of above venues of each neighborhood is set to be 1000.

2.1 Data cleansing

1. We're interested in schools because there are young people who are probably partial to novel goods like bubble tea, but students are more likely to stay at school, college, office,

institute and university, instead of other places like laboratories. So we will only consider venues that have any word in 'school', 'college', 'office', 'institute' and 'university'.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Rouge,Malvern	43.806686	-79.194353	Evergreen College	43.802162	-79.199654
2	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	st brendan catholic school	43.783052	-79.149267
5	Cedarbrae	43.773136	-79.239476	St. Richard Catholic School	43.768646	-79.240365
8	Cedarbrae	43.773136	-79.239476	Oxford College	43.777108	-79.250061
9	Cedarbrae	43.773136	-79.239476	UWIN Pro O/A the Canadian College for Higher S...	43.776911	-79.246111

figure 6 coordinate of schools after cleansing

2. Shopping places includes shopping malls and shopping plazas, so we should concatenate venues in these two categories.
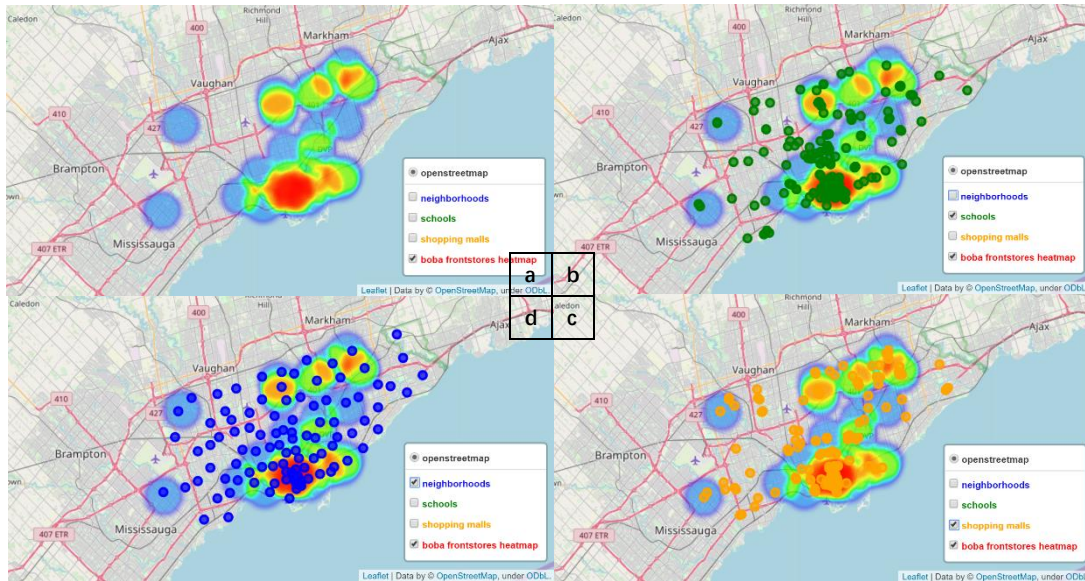

figure 7 map of Toronto **a**. heatmap of boba storefronts **b**. heatmap of boba storefronts and distribution of schools **c**. heatmap of boba storefronts and distribution of neighbourhoods **d**. heatmap of boba storefronts and distribution of shopping places.

3. Predictive Modeling

Analyzing Methodology

The goal of this project is to provide a list of promising neighborhoods for new boba storefronts. These neighborhoods should satisfy the primary criterium of low density of boba shops. Then we give preference to neighborhoods with as many as possible number of schools and shopping venues. To clearly define the scope of searching area, we will only take into consideration venues in radius of 1000 meters about the given coordinate of neighborhoods.

Above we have collected coordinate of each neighborhood in Toronto, location and number of boba shops in each neighborhood, location and number of school venues in each neighborhood, location and number of shopping venues in each neighborhood within 1km from neighborhood center.

In the next step, we will use machine learning methods to create clusters and try to find out ones that meet up with the expectation of promising candidates. After machine learning, we will also filter promising neighborhoods that have no more than 5 boba shops and more than 3 school venues.

In the final part, we will focus on visualizing the machine learning result and filter result on map. We will present all possible neighborhoods for new boba stores on map, to offer a 'neighborhood level' recommendation to a specific boba brand.

Language: python (anaconda jupyter notebook)
Data Scraping Libraries: Request, BeautifulSoup, etc.
Data preprocessing Libraries: pandas, numpy, etc.
Plotting Libraries: maplotlib, folium, etc.

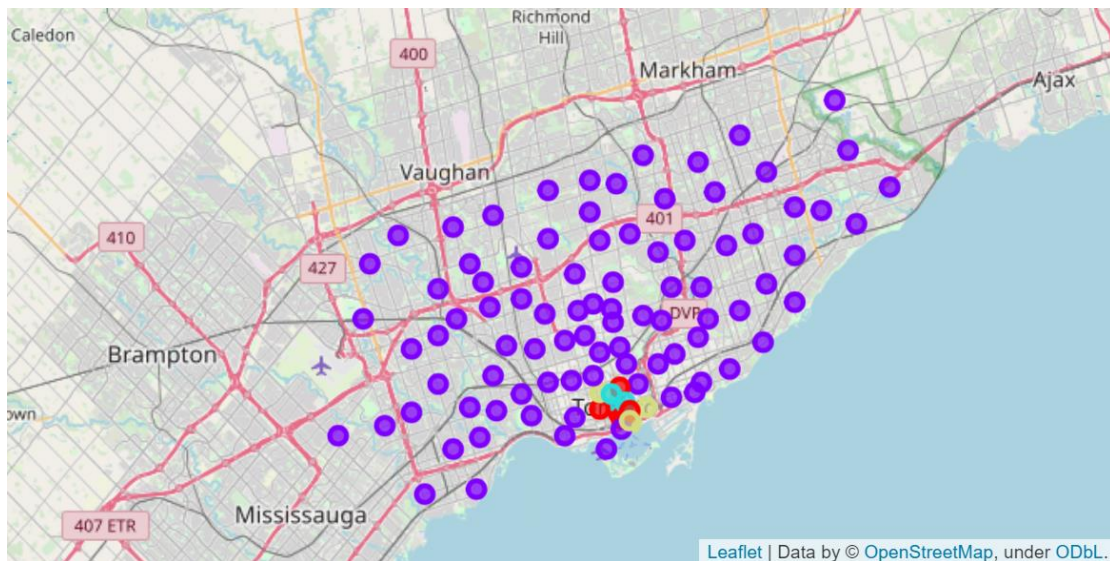**Performance of Model**

1. Results of k-means

figure 8 map of k-means resulting clusters. red: cluster0, purple: cluster1, blue: cluster2, grey: cluster3

There are 7 neighborhoods in cluster0, and the number of boba shops, schools, and shopping venues are close. There are 88 neighborhoods in cluster1, the biggest cluster. Most of them far from schools and shopping areas. This cluster supports the assumption that low density of schools and shopping hinders the development of boba market. Having 3 neighborhoods, cluster2 shows a distinguishing feature of extremely high number of boba shops, although there are also a large amount of schools and shopping places. The 4 neighborhoods in Cluster3 have less number of boba shops than cluster0 and cluster2, but higher number of schools and shopping places than cluster1.

After scrolling down the 4 clusters, we can regard cluster3 as the cluster that reaches the criterium of promising starting points for the next boba frontstore.
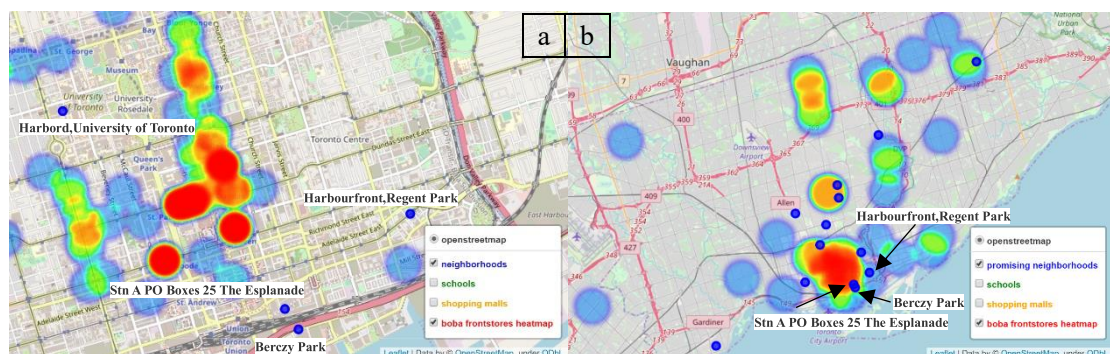
figure 9 a. map of cluster3 **b**. map of manual selection

Berczy Park, Harbourfront,Regent Park, Stn A PO Boxes 25 The Esplanade are selected by both approaches, and they all situate in the south of old Toronto near the seashore.

Conclusions

Purpose of this project was to detect promising neighborhoods for new boba storefront, which have low density of bubble tea shops but high of schools and shopping malls. At the beginning, we explore all bubble tea shops, schools, shopping places within 1km from each neighbor center. By plotting a heatmap, we roughly know the distribution of boba shops in Toronto. To achieve the goal, we try two methods, one is k-means and the other is manual selection. There are 4 neighborhoods in k-means results but 13 in manual selection. The coincidence of 3 neighborhoods somewhat proofs the work of machine learning.

Result of K-Means is a cluster including Berczy Park, Harbord, University of Toronto, Harbourfront, Regent Park, Stn A PO Boxes 25 The Esplanade. Result of manual selection is a cluster containing 13 neighborhoods. The common ones that they select are Berczy Park, Harbourfront, Regent Park, Stn A PO Boxes 25 The Esplanade. However, these results certainly do not imply these neighborhoods are ideal locations for a new boba storefront. Our goal in this analysis is detecting neighborhoods with as less as possible boba shops but as many as possible schools and shopping places within 1km.It is entirely possible that there is factors that stop boba brands from opening storefront in these neighborhoods, like low consumption level of people in the neighborhood, high rent, or stress of strong competitors. For further research, below the Top10 boba brands sorted by number of boba storefronts in Toronto is provided.

	boba brand
0	Real Fruit Bubble Tea
1	Tea Shop 168
2	Sharetea
3	Chatime
4	Presotea
5	CoCo Fresh Tea & Juice
6	Chatime Atealier
7	Chatime 日出茶太
8	The Alley
9	ZenQ

figure 10 rank of boba brands

Recommended neighborhoods should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no competition but also other factors taken into account and all other relevant conditions met.

Future directions

Final decision on optimal location will be made by stakeholders of boba shops based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (traffic, parking, sightseeing attractions),residence composition, real estate availability, prices, social and economic dynamics of every neighborhood etc.