Find the optimal location to open a bubble tea shop in North York, ON

Iris Li

July 27, 2020

1. Introduction

1.1 Background

Bubble tea is a Taiwanese drink contains milk and tea with tapioca. A bubble tea store usually sells this type of drink as well as other fruit tea. The bubble tea store is very popular in Asia and now it is introduced to America. In Toronto, although the product is still in the introductory stage, many new bubble tea shops opened and getting more popular in the recent years. There are many factors to determine whether a new bubble tea shops will thrive such as fund, employees and quality of the drinks, while geographical location is one of them that plays an essential role of the success. Thus, finding an optimal placement for a new shop can help the business gain customer base, brand recognition and more profit in the near future when it is opened.

1.2 Business Problem

This project aims to find an optimal place to open a brand-new bubble tea shop at North York district in Toronto. The goal is to find the best neighbourhood to open the shop by considering two essential factors. First, the market competition around each neighborhood will be examined to determine the market saturation and opportunity to gain customers. Second, I will find shopping malls as well as Asian restaurants in each neighborhood with which the bubble tea store can create marketing synergy and cooperation.

1.3 Interest

The target audience would be the people who plan to invest in a bubble tea shop in Toronto area especially North York. Investors who want to open other beverage business may also be interested in this analysis.

2. Data acquisition and cleaning

2.1 Data sources

In order to obtain the geographical information in North York, I use the data from Wikipedia and Foursquare.

The postal code, borough and neighborhood of Toronto are found in Wikipedia. The latitude and longitude of each neighborhood can be obtained from this CSV file. From these two datasets, I can get the postal code, borough, neighborhood name, and the related coordinate. For example, a neighborhood called Parkwoods belongs to the borough North York with postal code of M3A and coordinate of (43.753259, -79.329656). I also create a map to mark the venues around each neighborhood in North York using explore function in Foursquare API. By combining these three datasets and examining the surroundings with maximized cooperating business and minimized competitors, I am able to make a conclusion on the optimal location to open a bubble tea shop in North York, ON.

2.2 Data cleaning

All these datasets need to be cleaned and combined into one table. There is a table on the website of Wikipedia that will be used in the analysis. It contains postal code, borough, and neighbourhood of GTA area. To make it useable, I used the *Beautiful Soup* code to obtain the table from the website. I made three assumptions:

- Remove all the missing value in borough (not assigned value)
- Remove any duplicates

- If a cell has a value in borough but a not assigned in neighborhood, then the neighborhood will be the same as the borough.

I also read the CSV file with the latitude and longitude of Toronto into my code. And then use the *merge* function to combine these two table with columns of postal code, borough, neighbourhood, latitude and longitude. The first five rows of the table are showed in the table 1.

Table 1: Neighbourhoods in GTA area

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Because I only what to examine neighborhood in North York, I created a new table which only contain the borough with the word “North York”. This new table has 24 rows representing the 24 neighbourhoods in North York. The first five rows are showed in the table 2.

Table 2: Neighbourhoods in North York

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
3	M3B	North York	Don Mills	43.745906	-79.352188
4	M6B	North York	Glencairn	43.709577	-79.445073

In order to explore the venues around North York, I defined a method *getNearbyVenues* to get the nearby venues by creating and making the Foursquare API request URL. This table contains all the venues and 106 unique venue categories in each neighbourhood in North York (Table 3).

Table 3: Nearby venues in North York

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
5	Victoria Village	43.725882	-79.315572	The Frig	43.727051	-79.317418	French Restaurant
6	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
7	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	Roots	43.718214	-79.463893	Boutique
8	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	Kitchen Stuff Plus (Clearance Outlet)	43.719096	-79.462675	Furniture / Home Store
9	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	Lac Vien Vietnamese Restaurant	43.721259	-79.468472	Vietnamese Restaurant

2.3 Feature selection

After the cleaning, there are 24 neighbourhoods in North York with 248 venues and there are 106 unique venue categories. All the features can be used in the next step to analyze the competitors and cooperation around this area without any redundancy.

3. Methodology

There are two factors need to be considered when choosing the best location, which is competition and cooperation. As I mentioned above, bubble tea business is in the introductory stage in North America; however, market has been very competitive since the products that every shop sold are identical and coffee shops such as Starbucks will also be the direct competitor of a bubble tea store. Additionally, most of the bubble tea shops opened next to the shopping mall or Asian restaurants. It is a good strategy to increase the brand exposure and target the right customers. Therefore, in order to open a new milk tea shop with less competition and more brand exposure, I will look for a place close to shopping mall, movie theater, and all kinds of Asian restaurant. I defined a method to covert all the venue categories that contains those keywords (shopping, Thai, Chinese, etc.) into integer one or zero. If the category contains any of the keywords, the cooperation number will return 1; otherwise, it will return

0. Similarly, another method called *sameShopToInt* is designed to find the competitors in each neighborhood. If the keyword such as coffee, café, tea showed in the category, the competition number will return 1; otherwise, it will return 0. And then I extract all the neighborhood that only contains 0 competition number and 1 cooperation number. All the data is showed in Table 4:

Table 4: Competitors and cooperation business in each neighborhood

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	CompetitionNumber	CooperationNumber
9	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	Lac Vien Vietnamese Restaurant	43.721259	-79.468472	Vietnamese Restaurant	0	1
45	Don Mills	43.725900	-79.340923	Congee Star 帝王名粥	43.726586	-79.341833	Chinese Restaurant	0	1
61	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	Wakame Sushi	43.755382	-79.440945	Sushi Restaurant	0	1
67	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	Sheppard Plaza	43.755695	-79.439613	Shopping Mall	0	1
74	Fairview, Henry Farm, Oriole	43.778517	-79.346556	SilverCity	43.778681	-79.344085	Movie Theater	0	1
75	Fairview, Henry Farm, Oriole	43.778517	-79.346556	CF Fairview Mall	43.777994	-79.343665	Shopping Mall	0	1
132	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Manchu Wok	43.778225	-79.343302	Chinese Restaurant	0	1
150	Bayview Village	43.786947	-79.385975	Sun Star Chinese Cuisine 翠景小炒	43.787914	-79.381234	Chinese Restaurant	0	1
160	Downsview	43.739015	-79.506944	jane sheppard mall	43.740104	-79.512552	Shopping Mall	0	1
176	Bedford Park, Lawrence Manor East	43.733283	-79.419750	Satay on the Road	43.735310	-79.419783	Thai Restaurant	0	1
178	Bedford Park, Lawrence Manor East	43.733283	-79.419750	Sakura Garden	43.733398	-79.419491	Sushi Restaurant	0	1
208	Willowdale, Willowdale East	43.770120	-79.408493	Cineplex Cinemas	43.768625	-79.412613	Movie Theater	0	1
210	Willowdale, Willowdale East	43.770120	-79.408493	Empress Walk	43.768540	-79.412671	Shopping Mall	0	1
218	Willowdale, Willowdale East	43.770120	-79.408493	Aburi Room	43.769197	-79.414039	Sushi Restaurant	0	1
227	Willowdale, Willowdale East	43.770120	-79.408493	Pho 88 Vietnamese Cuisine	43.770456	-79.413064	Vietnamese Restaurant	0	1
233	Willowdale, Willowdale East	43.770120	-79.408493	PROJECT:FISH	43.769100	-79.414305	Sushi Restaurant	0	1

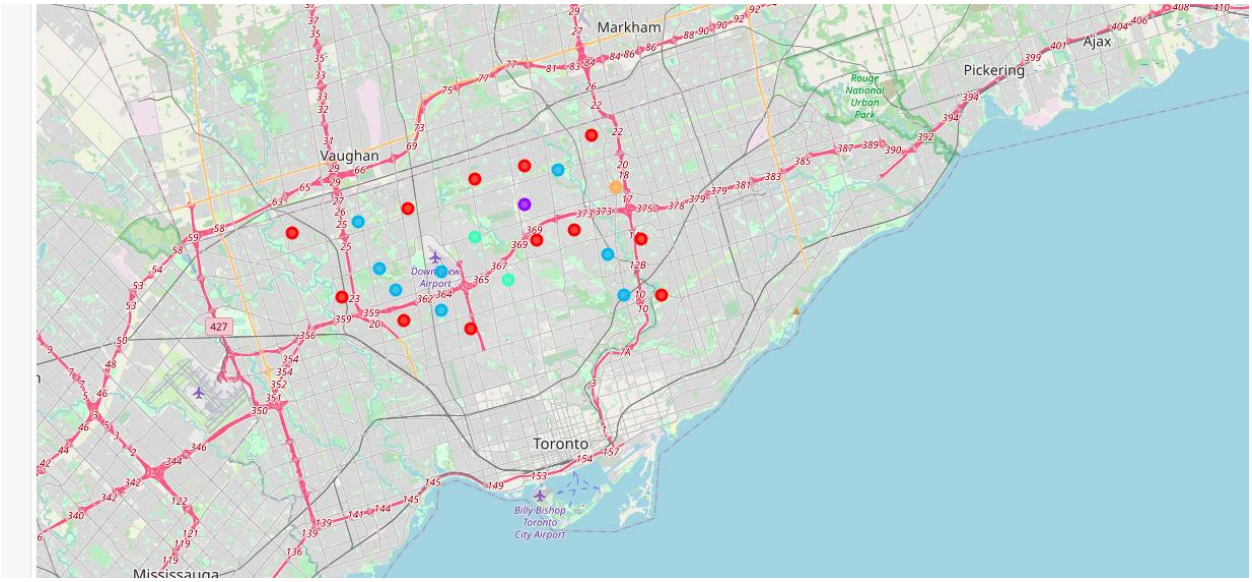
Table 4 only shows that the neighbourhoods that has cooperation businesses but does not demonstrate how many cooperation businesses in each neighbourhood. Thus, I summed up both the cooperation numbers and competition number for all the neighbourhoods in North York. The first ten rows of the table are showed in Table 5

Table 5: Summed up competitors and cooperation business in each neighborhood

	Cluster Labels	Postal Code	Borough	Neighbourhood	Latitude	Longitude	CompetitionNumber	CooperationNumber
0	0	M3A	North York	Parkwoods	43.753259	-79.329656	0	0
1	0	M4A	North York	Victoria Village	43.725882	-79.315572	0	0
2	2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	0	1
3	2	M3B	North York	Don Mills	43.745906	-79.352188	0	1
4	0	M6B	North York	Glencairn	43.709577	-79.445073	0	0
5	2	M3C	North York	Don Mills	43.725900	-79.340923	0	1
6	0	M2H	North York	Hillcrest Village	43.803762	-79.363452	0	0
7	3	M3H	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	0	2
8	3	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0	2
9	0	M3J	North York	Northwood Park, York University	43.767980	-79.487262	0	0

In this table, we can see different cooperation number in each neighbourhood. Our goal is to find a place with minimum competition number and maximum cooperation number; hence, I used K-means algorithm to cluster the neighborhoods. I run K-means into 5 clusters and add cluster labels to all the data and then create the map of clustered neighbourhoods using folium shows in Figure 1.

Figure 1: Map of clustered neighbourhoods



4. Result

The results of the clusters

Cluster #1

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
0	M3A	43.753259	-79.329656	0	0
1	M4A	43.725882	-79.315572	0	0
4	M6B	43.709577	-79.445073	0	0
6	M2H	43.803762	-79.363452	0	0
9	M3J	43.767980	-79.487262	0	0
12	M2L	43.757490	-79.374714	0	0
14	M6L	43.713756	-79.490074	0	0
15	M9L	43.756303	-79.565963	0	0
16	M2M	43.789053	-79.408493	0	0
19	M9M	43.724766	-79.532242	0	0
22	M2P	43.752758	-79.400049	0	0
23	M2R	43.782736	-79.442259	0	0

Cluster #2

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
20	M2N	43.77012	-79.408493	0	5

Cluster #3

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
2	M6A	43.718518	-79.464763	0	1
3	M3B	43.745906	-79.352188	0	1
5	M3C	43.725900	-79.340923	0	1
10	M2K	43.786947	-79.385975	0	1
11	M3K	43.737473	-79.464763	0	1
13	M3L	43.739015	-79.506944	0	1
17	M3M	43.728496	-79.495697	0	1
21	M3N	43.761631	-79.520999	0	1

Cluster #4

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
7	M3H	43.754328	-79.442259	0	2
18	M5M	43.733283	-79.419750	0	2

Cluster 5

	Postal Code	Latitude	Longitude	CompetitionNumber	CooperationNumber
8	M2J	43.778517	-79.346556	0	3

Since I already eliminated the competitors, these clusters show the different cooperation number of each neighborhood.

5. Discussion

Because the goal is to find a neighborhood where the competitors are minimized and cooperation businesses are maximized, Cluster 2 would be the best cluster to locate a brand-new bubble tea shop

with 0 coopetitors and 5 venues that would benefit the new shop. Therefore, the optimal location is M2N, the neighborhood of Willowdale and Willowdale East.

6. Conclusion

In this analysis, I labeled the neighbourhoods based on the characteristic of their venues. Because the goal is to open a bubble tea shop in the right place, the venues that examined are the competitors (Café, coffee shops and bubble tea shops) and the business that can be cooperated with (shopping mall, Asian restaurants). The fittest cluster is cluster 2 with 0 competitors and 5 business to cooperate with at M2N. This place will increase the possibility of the new bubble tea shop's success.