The Battle of the Neighborhoods

Xiaomeng Shang

November 19, 2020

Table of Contents

The Battle of the Neighborhoods

1. Introduction

1.1 Background

Toronto is Canada's largest city, the most populous city in Canada, and home to a diverse population of about 2.9 million people.

The city is ranked as one of the top destinations around the globe. It boasts world-class restaurants, cultural attractions as varied as the cultures themselves.

Moreover, Toronto is recognized for being Canada’s commercial capital and for its excellence in a number of sectors including life sciences, technology, and education. Thus, the outstanding opportunities attract investors all around the world.

1.2 Problem

A group of stakeholders have experience in running restaurant. They are attracted by culture diversity of Toronto and want to expand restaurant business.

They intend to open a Chinese restaurant in downtown Toronto, the main central business district of Toronto.

1.3 Interest

The location will make an impact on succeed of the restaurant. We particularly interested in:

1) areas with no Chinese restaurants;
2) areas which are not crowded with restaurants.

We are going to analysis location information of the restaurants in downtown Toronto and find an optimal location for stakeholders’ new restaurant.

2. Data acquisition and cleaning

2.1 Data sources

Information of Neighborhoods of Toronto can be found in a Wikipedia page here. A table in this page list postal code, borough and neighborhood name.

In week 3, the course provides a link of a csv document here, through which we can obtain the geographical coordinate conveniently.

Then, we can use Foursquare API to get venues' information in each neighborhood.

2.2 Data cleaning

1) Neighborhood information

We can use lxml package to scrape the table from Wikipedia page.

We only process with cells with valid values. Thus, we delete rows, which have Borough with "Not assigned" value. After checking Neighborhood column, we find that all the values are valid.

Then, we reset the index and get the data frame of Neighborhood information, including postal code, borough and neighborhood.

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

2) Geographical information

To get the data efficiently, we use the csv file to get the geographical information and load the data into a data frame.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Join the two data frames, we get a new data frame, which combines the neighborhood information and geographical information.

	PostalCode	Borough	Neighborhood	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

3) Venues' information

We can use Foursquare API to get venues' information in each neighborhood of downtown Toronto.

At first, we write a function to get the top 100 venues in a radius of 500 meters in every neighborhood.

Then, we get venues information. There are 1253 venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

2.3 Feature selection

2.3.1 General view of feature

As we care about the category of each venue. After checking, we found that there are 214 unique categories.

We should focus on category 'restaurant'. However, there are so many categories related to restaurant, in another words, they are specific restaurant categories, such as 'French Restaurant', 'Mexican Restaurant', 'Portuguese Restaurant', 'Italian Restaurant'.

Those categories should be taken into consideration as well. We examined every category and extracted categories, whose name contains ‘Restaurant’. We regard them as **competitor category**.

In total, there are 44 competitor categories.

```
['Restaurant', 'French Restaurant', 'Mexican Restaurant', 'Portuguese Restaurant', 'Italian Restaurant', 'Sushi Restaurant', 'C
hinese Restaurant', 'Ramen Restaurant', 'Thai Restaurant', 'New American Restaurant', 'Japanese Restaurant', 'Fast Food Restaur
ant', 'Middle Eastern Restaurant', 'Modern European Restaurant', 'Ethiopian Restaurant', 'Seafood Restaurant', 'Vietnamese Rest
aurant', 'American Restaurant', 'Latin American Restaurant', 'Vegetarian / Vegan Restaurant', 'German Restaurant', 'Comfort Foo
d Restaurant', 'Asian Restaurant', 'Moroccan Restaurant', 'Belgian Restaurant', 'Greek Restaurant', 'Eastern European Restauran
t', 'Falafel Restaurant', 'Indian Restaurant', 'Korean Restaurant', 'Colombian Restaurant', 'Mediterranean Restaurant', 'Brazil
ian Restaurant', 'Gluten-free Restaurant', 'Caribbean Restaurant', 'Dumpling Restaurant', 'Filipino Restaurant', 'Doner Restaur
ant', 'Dim Sum Restaurant', 'Molecular Gastronomy Restaurant', 'Taiwanese Restaurant', 'Sri Lankan Restaurant', 'Theme Restaura
nt', 'Afghan Restaurant']
```

Then, we extracted venues records from the 44 competitor categories and structured them into a data frame. There are 297 competitor venues, they are different styles restaurants.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
1	Regent Park, Harbourfront	43.654260	-79.360636	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restaurant
2	Regent Park, Harbourfront	43.654260	-79.360636	El Catrin	43.650601	-79.358920	Mexican Restaurant
3	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Nando's	43.661728	-79.386391	Portuguese Restaurant
4	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Mercatto	43.660391	-79.387664	Italian Restaurant

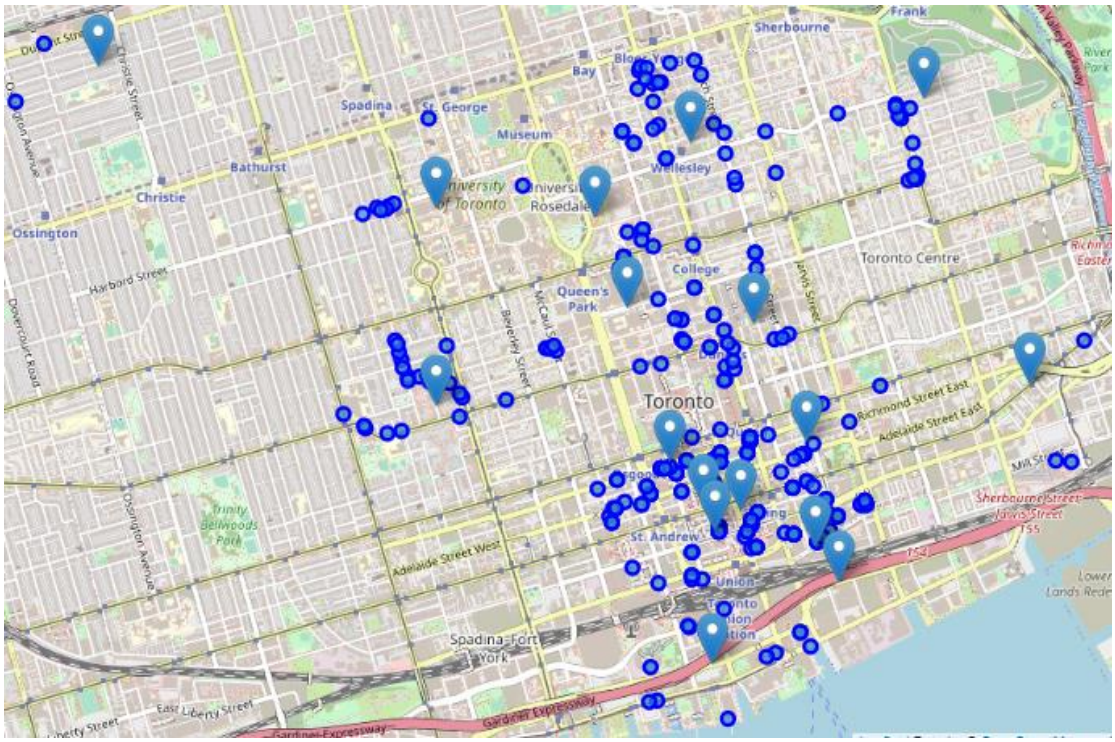
As we explored further, we discovered the top 10 popular competitor categories:

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue
Restaurant	44	44	44	44
Japanese Restaurant	31	31	31	31
Italian Restaurant	22	22	22	22
Seafood Restaurant	20	20	20	20
Sushi Restaurant	17	17	17	17
American Restaurant	16	16	16	16
Thai Restaurant	15	15	15	15
Vegetarian / Vegan Restaurant	14	14	14	14
Asian Restaurant	9	9	9	9
Mexican Restaurant	9	9	9	9

Obviously, Japanese and Italian restaurant are particularly popular.

2.3.2 Restaurants in each neighborhood

We drew a map to show how these restaurants distribute in each neighborhood.



In some neighborhoods, restaurants are crowed, such as Toronto Dominion Centre, Design Exchange, Commerce Court, Victoria Hotel, Richmond, Adelaide, King.

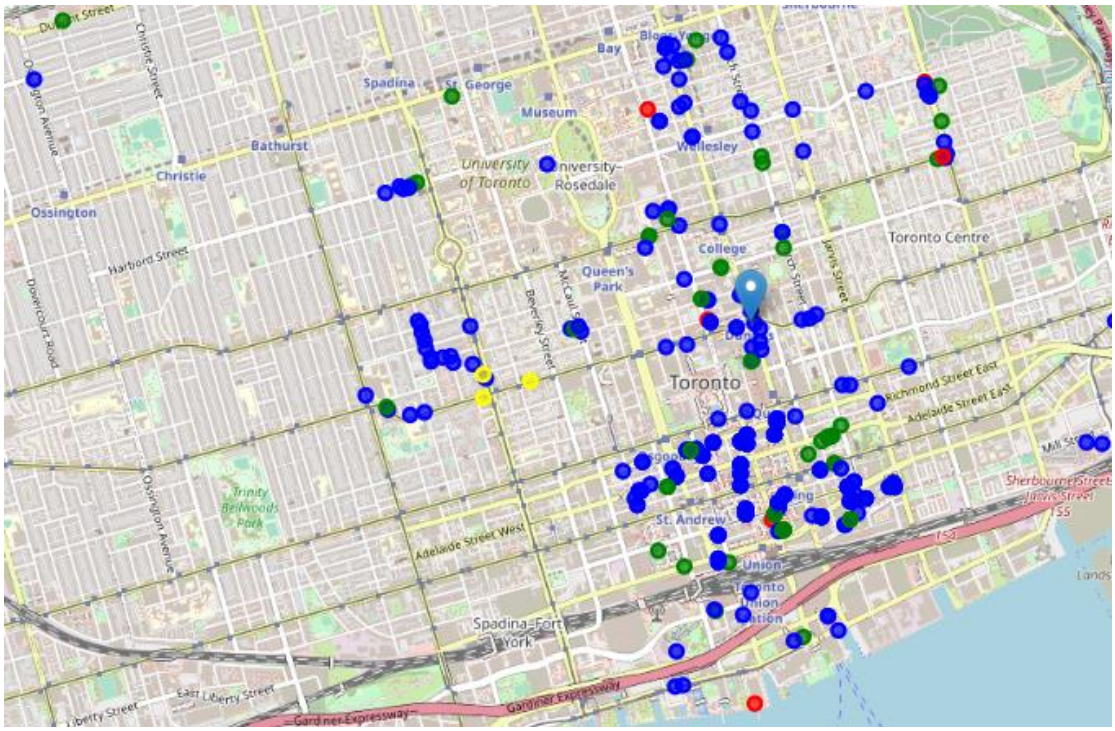
2.3.3 Important competitor categories

Except for ‘Chinese Restaurant’ category, we cared about **'Dumpling Restaurant'**, **'Dim Sum Restaurant’,** because they belong to Chinese food; we take **'Japanese Restaurant', 'Italian Restaurant'** into consideration as well, because they are two popular categories.

There are only 6 Chinese restaurants in Downtown Toronto. We still have an opportunity to enter the market.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
7	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Crown Princess Fine Dining 伯爵名宴	43.666455	-79.387698	Chinese Restaurant
23	Garden District, Ryerson	43.657162	-79.378937	GB Hand-Pulled Noodles	43.656434	-79.383783	Chinese Restaurant
114	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752	Pearl Harbourfront	43.638157	-79.380688	Chinese Restaurant
149	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576	Szechuan Express	43.646973	-79.379549	Chinese Restaurant
240	St. James Town, Cabbagetown	43.667967	-79.367675	China Gourmet	43.664180	-79.368359	Chinese Restaurant
241	St. James Town, Cabbagetown	43.667967	-79.367675	Tender Trap Restaurant	43.667724	-79.369485	Chinese Restaurant

Then, we create a new map and mark venues with different colors.



Red color refers to 'Chinese Restaurant', yellow color refers to 'Dumpling Restaurant' or 'Dim Sum Restaurant', green color refers to 'Japanese Restaurant' or 'Italian Restaurant', blue color refers to other categories.

3. Methodology

3.1 K-means algorithm

We will use K-means algorithm, one of the most popular “clustering” algorithms, to segment neighborhoods and to get clusters.

K-means algorithm can be explained as followings:

K-means stores *k* centroids that it uses to define clusters. A point is considered to be in a particular cluster if it is closer to that cluster's centroid than any other centroid.

K-Means finds the best centroids by alternating between (1) assigning data points to clusters based on the current centroids (2) choosing centroids (points which are the center of a cluster) based on the current assignment of data points to clusters.

We select venue categories as features, which are defined as “competitor category’, these restaurant categories contain 44 different types of restaurants.

3.2 Procedures of segment and cluster

At first, we apply one hot encoding method to features, the ‘venue category’.

Secondly, we calculate the mean of frequency of occurrence of each category.

After the two steps, we can transform distinct variable ‘venue category’ to numeric variable.

Thirdly, we conduct k-means process by using Python package and get clusters.

4. Results

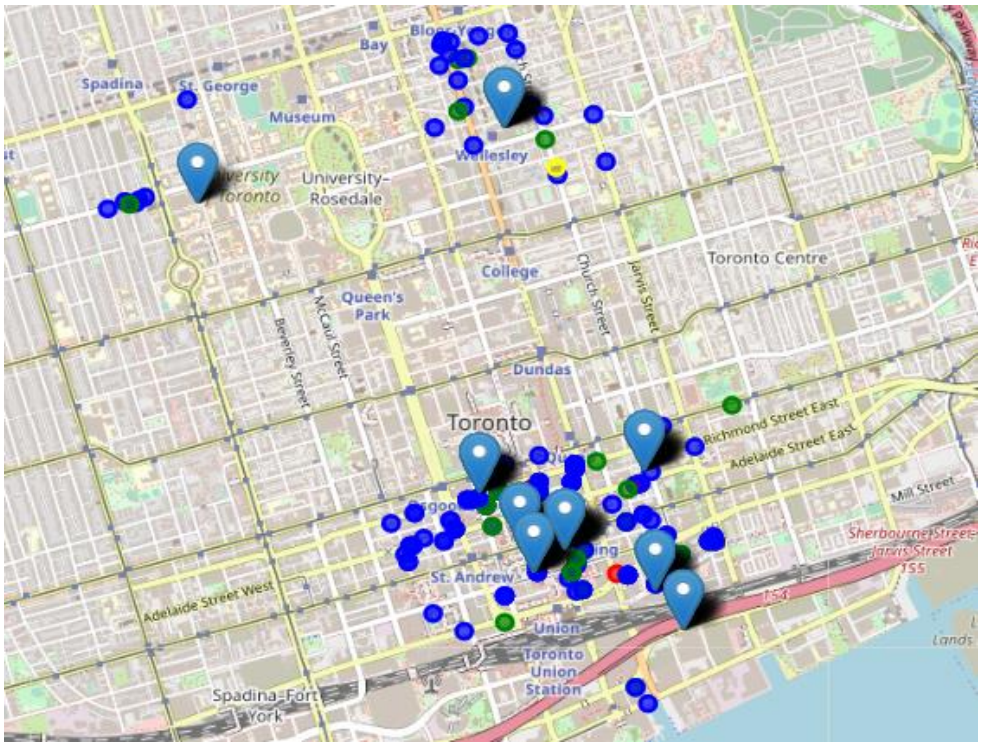
As we define k= 5, we segment all the neighborhoods into 5 clusters.

To display the result in a better way, we mark all the restaurants in the neighborhoods of important clusters with different colors.

Red color refers to 'Chinese Restaurant', yellow color refers to 'Dumpling Restaurant' or 'Dim Sum Restaurant', green color refers to 'Japanese Restaurant' or 'Italian Restaurant', blue color refers to other categories.

Cluster 1

	Neighborhood	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	Berczy Park	Seafood Restaurant	Restaurant	Comfort Food Restaurant	Thai Restaurant	Greek Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	French Restaurant
3	Church and Wellesley	Sushi Restaurant	Japanese Restaurant	Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Mexican Restaurant	American Restaurant	Caribbean Restaurant	Ethiopian Restaurant	Fast Food Restaurant
4	Commerce Court, Victoria Hotel	Restaurant	American Restaurant	Japanese Restaurant	Seafood Restaurant	Thai Restaurant	Asian Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	New American Restaurant	Fast Food Restaurant
5	First Canadian Place, Underground city	Japanese Restaurant	Restaurant	American Restaurant	Asian Restaurant	Seafood Restaurant	Thai Restaurant	Sushi Restaurant	Colombian Restaurant	New American Restaurant	Vegetarian / Vegan Restaurant
11	Richmond, Adelaide, King	Restaurant	Thai Restaurant	American Restaurant	Sushi Restaurant	Colombian Restaurant	New American Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	Modern European Restaurant	Fast Food Restaurant
12	St. James Town	Restaurant	American Restaurant	Japanese Restaurant	Seafood Restaurant	Moroccan Restaurant	Comfort Food Restaurant	New American Restaurant	Italian Restaurant	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant
14	Stn A PO Boxes	Italian Restaurant	Seafood Restaurant	Restaurant	Japanese Restaurant	Molecular Gastronomy Restaurant	American Restaurant	Thai Restaurant	Fast Food Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant
15	Toronto Dominion Centre, Design Exchange	Restaurant	American Restaurant	Japanese Restaurant	Seafood Restaurant	Asian Restaurant	Italian Restaurant	Chinese Restaurant	Vegetarian / Vegan Restaurant	New American Restaurant	French Restaurant
16	University of Toronto, Harbord	Japanese Restaurant	French Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Comfort Food Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Filipino Restaurant	Fast Food Restaurant



Cluster 2

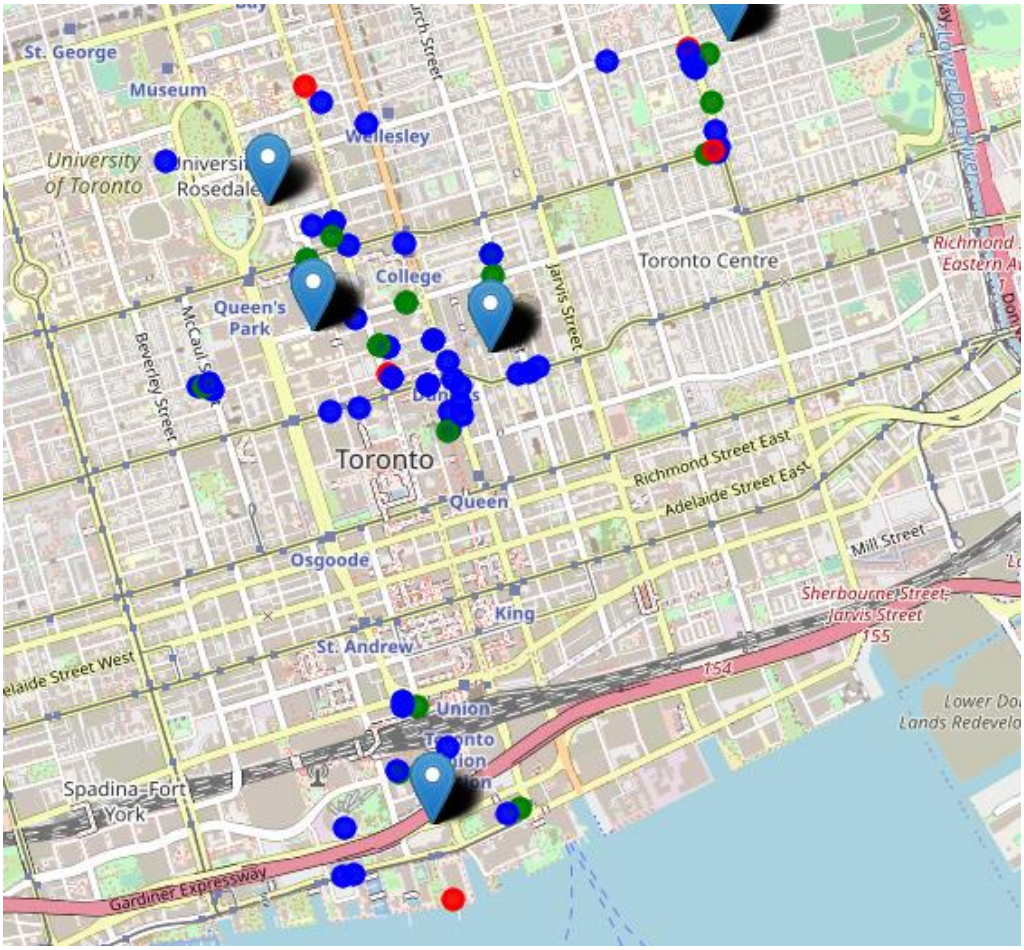
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Christie	Italian Restaurant	Restaurant	Vietnamese Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant

Cluster 3

	Neighborhood	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
10	Regent Park, Harbourfront	French Restaurant	Mexican Restaurant	Restaurant	Vietnamese Restaurant	Doner Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant

Cluster 4

	Neighborhood	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
1	Central Bay Street	Italian Restaurant	Thai Restaurant	Japanese Restaurant	Indian Restaurant	New American Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant	Middle Eastern Restaurant	Falafel Restaurant
6	Garden District, Ryerson	Japanese Restaurant	Italian Restaurant	Middle Eastern Restaurant	Fast Food Restaurant	Ramen Restaurant	Vietnamese Restaurant	New American Restaurant	Ethiopian Restaurant	Mexican Restaurant	Modern European Restaurant
7	Harbourfront East, Union Station, Toronto Islands	Restaurant	Italian Restaurant	Indian Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant	Seafood Restaurant	Japanese Restaurant	Chinese Restaurant	New American Restaurant
9	Queen's Park, Ontario Provincial Government	Italian Restaurant	Sushi Restaurant	Restaurant	Chinese Restaurant	Portuguese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Doner Restaurant	Filipino Restaurant	Fast Food Restaurant
13	St. James Town, Cabbagetown	Restaurant	Italian Restaurant	Chinese Restaurant	Indian Restaurant	Thai Restaurant	Taiwanese Restaurant	Sri Lankan Restaurant	Japanese Restaurant	Caribbean Restaurant	Dumpling Restaurant



Cluster 5

	Neighborhood	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
8	Kensington Market, Chinatown, Grange Park	Vegetarian / Vegan Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Comfort Food Restaurant	Japanese Restaurant	Filipino Restaurant	Dim Sum Restaurant	Doner Restaurant	Caribbean Restaurant

5. Discussion

Cluster 1 contains over 50 percentage of all neighborhoods and there are only 1 Chinese restaurants. Except for 'Church and Wellesleythe' and 'University of Toronto, Harbord', the restaurants of other neighborhoods are crowded.

Cluster 4 contains 30 percentage of all neighborhoods and there are 5 Chinese restaurants. Almost all the Chinese restaurant are located in these neighborhoods.

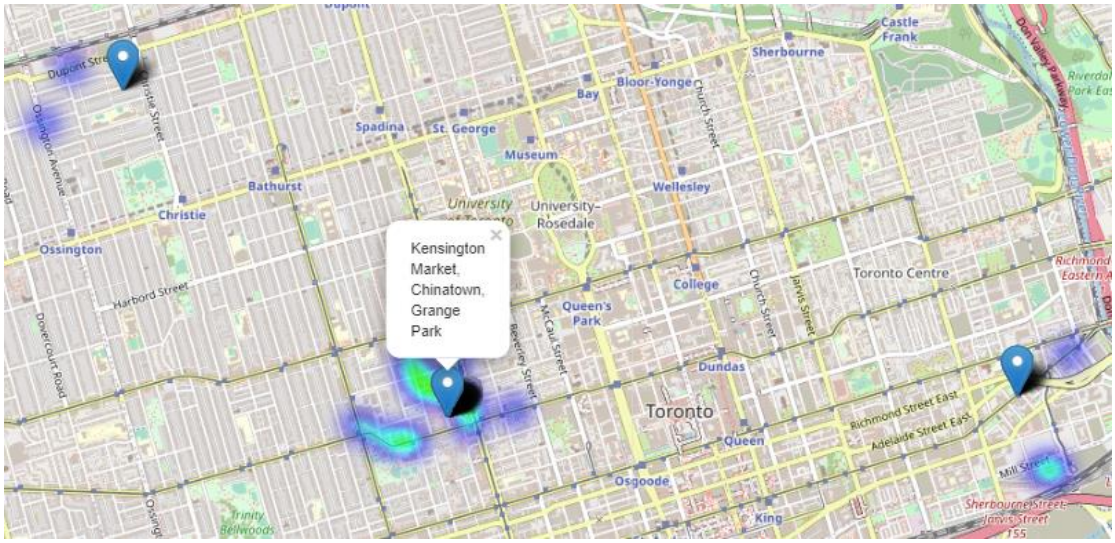
Except for stakeholders' interests, we will not consider neighborhoods of these two clusters. The reasons are as followings:

The restaurants are intensively located, especially **'Japanese Restaurant', 'Italian Restaurant',** thus the competition is severe.

For cluster 2,3,5, we create a heat map to present the intensity of restaurant.

The restaurants intensity of neighborhoods 'Kensington Market, Chinatown, Grange Park' is in a middle level.

Neighborhoods with a low intensity indicate that the demand for restaurant is not urgent.



6. Conclusion

Our stakeholders intend to open their first Chinese restaurant in Toronto and they chose the downtown Toronto, the main central business district of Toronto.

It is not easy to entry a new market, that's why they interest in area with no Chinese restaurants and area where restaurants are not crowded.

Based on the above analysis, we will recommend neighborhoods 'Kensington Market, Chinatown, Grange Park' for stakeholders.