Finding a better location for food business

Ziyun Zhou

Feb 02, 2020

A. Introduction:

a. Background

Food business is one of the most common business by its number of type and its store quantity. There are stores starting and stores closing every day. One of the most important parameters of their successes is location. However, the same location means different to different types of food business. Near a shopping center may be good for a fast food restaurant such as McDonald but may be not suitable for a high-end French restaurant. Finding a good location is more of finding a good location for a class or a type of food business rather than finding a location for all food businesses. Businesses of the same class do compete to each other. But it is riskier to open a business just by your own. That is why in many cases, same types of restaurants cluster together.

b. Problem

The problem we want to solve in my project is: if we are going to open a food business of one kind, where should we open inside a city? Our solution to the problem is finding clusters of locations for different types of food business. New York, Toronto and London are cities we are going to perform analyses. For simplicity, we are only going to distinguish food businesses into two classes: fast food business and non-fast food business. To better answer the above question, we are also going to find why this cluster of locations is good for one class and is there any difference among three cities? Both maps and tables will be used to analyze information collected and interpret findings and results.

c. Audience and interest

Our analysis will be valuable for food business owners, food business investors, and may be interested by investors who follow publicly listed food companies those have franchises in those three cities.

B. Data description

a. Data Sources

For each city, we will divide the data gathering and processing procedure into three steps by different data sources. Step 1: Get Borough, Neighborhood and Postal code of each city. For NYC, we get those data from[2] and for Toronto and London, we get them from Wikipedia; Step 2: Get geographical coordinates of neighborhoods. For all three cities, we get the data from pgeocode package from PyPL[1]; Step 3: Get nearby venue information for each geographical coordinate. We use Foursquare to get those data.

b. Data cleaning

As mentioned in Data Sources section, there are generally three steps to get our data. The Step 2 and 3 are quite similar for all three cities, however, the step 1 is a little different among cities. Here, we will explain this step in-detail for each city.

Step 1: Get Borough, Neighborhood and Postal code information for each city.

New York City

We find borough, neighborhood and postal code information from website[2] by using the pandas read-html function. Those data are quite clean. The only thing we need to adjust is the postal codes column since there are more than one postal code in the same neighborhood, but we only need one. Since those postal codes are separated by comma, we can split them and only take the first one. We then create a new column for postal code with only one postal code and we drop the old column. Now we have our table ready for the next step. We aim to create similar table for all three cities before going to the next step.

	Borough	Neighborhood	Postcode
0	Bronx	Central Bronx	10453
1	Bronx	Bronx Park and Fordham	10458
2	Bronx	High Bridge and Morrisania	10451
3	Bronx	Hunts Point and Mott Haven	10454
4	Bronx	Kingsbridge and Riverdale	10463

Toronto

We find borough, neighborhood and postal code information from wikipedia[3] by using the pandas read-html function. The initial table we get is in very good shape, however, there are several adjustments we need to make before entering to the next step.

Firstly, there are ‘not assigned’ cells in column ‘Borough’. We need to drop those rows since they cannot provide us enough information. Secondly, there are some ‘not assigned’ cells in

column ‘Neighborhood’ with a valuable cell under the column ‘Borough’. We replace those ‘not assigned’ with the value of ‘Borough’. Thirdly, there are different values of Neighborhood with the same Postcode. Those Neighborhood value should be merged together since we need unique postal code to get local information to avoid duplicate results.

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

London

We find borough, neighborhood and postal code information from wikipedia[4] by using the pandas read-html function. The initial table looks messy compare to initial tables of other two cities. London is a very big city. In order to simply our analysis, we only focus on the city of London rather than the entire London area. Then, we remove several columns since they provide irrelevant information. There are some ‘[#]’ characters behind some values of column ‘Borough’. Since this could mistakenly differentiate the same borough with those special characters, we need to remove them. Finally, some rows have more than one postal code. We will only use the first code if there are many.

	Postcode	Borough	Neighborhood
0	SE2	Bexley, Greenwich	Abbey Wood
1	W3	Ealing, Hammersmith and Fulham	Acton
2	EC3	City	Aldgate
3	WC2	Westminster	Aldwych
4	SE20	Bromley	Anerley

Step 2: Get geographical coordinates of neighborhoods. We will use the postal code from Step 1 to get coordinates. Each postal code has only one set of coordinates. Below is the result for City of London.

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	SE2	Bexley, Greenwich	Abbey Wood	51.4869	0.107500
1	W3	Ealing, Hammersmith and Fulham	Acton	51.5114	-0.265717
2	EC3	City	Aldgate	51.5085	-0.125700
3	WC2	Westminster	Aldwych	51.5142	-0.123382
4	SE20	Bromley	Anerley	51.4154	-0.056950

Sometimes there may be no coordinates for some boroughs. This could result 'Nan' in the value of coordinates, which could make some errors in Step 3. We will remove those rows if there is any.

Step 3: Get nearby venue information for each geographical coordinate. We set a limit of maximum 500 venues for each set of coordinates and a radius of 500 meters. Once the data of venue is ready, we group them by the value of neighborhood and get weight by each venue category. Below is the result for City of London. (The sum of weight of neighborhood by each venue category is 1)

	NEIGHBORHOOD	African Restaurant	Airport	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports
0	Abbey Wood	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.0
1	Acton	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.0
2	Aldgate	0.01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.02	0.02	0.02	0.0
3	Aldwych	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.01	0.00	0.0
4	Anerley	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.0

Then we get the 10 most common venue category for each neighborhood.

	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	Abbey Wood	Chinese Restaurant	Grocery Store	Indian Restaurant	Men's Store	Fish Market	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop
1	Acton	Pub	Park	Convenience Store	Mini Golf	Gas Station	Grocery Store	Bed & Breakfast	Train Station	Bakery	Japanese Restaurant
2	Aldgate	Theater	Hotel	French Restaurant	Pub	Plaza	Pizza Place	Wine Bar	Steakhouse	Café	Ice Cream Shop
3	Aldwych	Theater	Clothing Store	Coffee Shop	Bakery	Ice Cream Shop	Indian Restaurant	Dessert Shop	Cosmetics Shop	Museum	Wine Bar
4	Anerley	Fast Food Restaurant	Pub	Train Station	Coffee Shop	Pizza Place	Supermarket	Café	Hardware Store	Tunnel	Garden Center

Now we have our data ready to put into the model for analyzing.

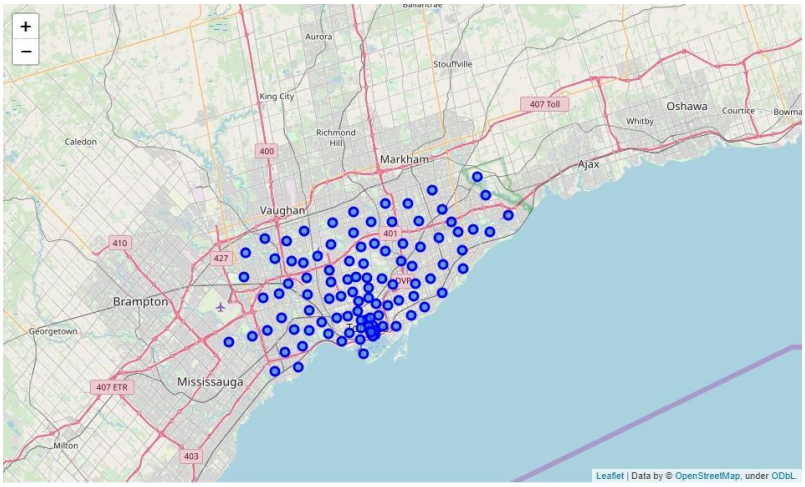
C. Methodology

Since all three cities are using the same method to analyze the problem, we only talk about Toronto in detail here.

At the end of step 2 of 'Data Cleaning', we have our data which contains ***Postcode, Borough, Neighborhood, Latitude and Longitude*** information of the city.

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.7545	-79.3300
1	M4A	North York	Victoria Village	43.7276	-79.3148
2	M5A	Downtown Toronto	Harbourfront	43.6555	-79.3626
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.7223	-79.4504
4	M7A	Downtown Toronto	Queen's Park	43.6641	-79.3889

As mentioned in section B, we then use folium library to visualize geographic details of Toronto and its boroughs and neighborhoods. We use Latitude and Longitude information to get a map as below.



Then we use the Foursquare API to explore the Neighborhoods. Here we use the limit as 100 venues and the radius 500 meters for each neighborhood from their given latitude and longitude data. Below shows 5 results or rows we get from the step.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.7545	-79.3300	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.7545	-79.3300	TTC stop - 44 Valley Woods	43.755402	-79.333741	Bus Stop
2	Parkwoods	43.7545	-79.3300	GreenWin pool	43.756232	-79.333842	Pool
3	Parkwoods	43.7545	-79.3300	Variety Store	43.751974	-79.333114	Food & Drink Shop
4	Victoria Village	43.7276	-79.3148	Victoria Village Arena	43.723481	-79.315635	Hockey Arena

We group the neighborhood together to see how many venues each neighborhood has to better understand our data limitation.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
Agincourt	5	5	5	5	5	5
Agincourt North, L'Amoreaux East, Milliken, Steeles East	2	2	2	2	2	2
Albion Gardens, Beaumond Heights, Humbergate, Jamestown, Mount Olive, Silverstone, South Steeles, Thistletown	12	12	12	12	12	12
Alderwood, Long Branch	9	9	9	9	9	9
Bathurst Manor, Downsview North, Wilson Heights	6	6	6	6	6	6
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	24	24	24	24	24	24
Berczy Park	90	90	90	90	90	90
Birch Cliff, Cliffside West	4	4	4	4	4	4
Bloordale Gardens, Eringate, Markland Wood, Old Burnhamthorpe	9	9	9	9	9	9
Brockton, Exhibition Place, Parkdale Village	38	38	38	38	38	38
Business Reply Mail Processing Centre 969 Eastern	15	15	15	15	15	15
CFB Toronto, Downsview East	4	4	4	4	4	4
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	57	57	57	57	57	57
Cabbagetown, St. James Town	44	44	44	44	44	44
Caledonia-Fairbanks	9	9	9	9	9	9
Cedarbrae	2	2	2	2	2	2
Central Bay Street	91	91	91	91	91	91
Chinatown, Grange Park, Kensington Market	80	80	80	80	80	80

From above table (here is only a part of it), we can see that ‘Adelaide, King, Richmond’, ‘Commerce Court, Victoria Hotel’, ‘Design Exchange, Toronto Dominion Centre’, ‘First Canadian Place, Underground city’, ‘Ryerson, Garden District’, ‘St. James Town’ and ‘Stn A PO Boxes 25 The Esplanade’ have reached the 100 limit of venues. On the other hand: there are neighborhood such as ‘Agincourt North, L'Amoreaux East, Milliken, Steeles East’, ‘Cedarbrae’ and more that are below 5 venues. Those results are in line with our analysis since the less the number of venues the lower the population density.

The next step is to group the data by the ‘Venue Category’ to get the weighted value for each neighborhood. This is done by the following procedure. We can take ‘Bank’ for example. During our search of venue information for each neighborhood, ‘Bank’ appears 100 times. Then each time is a weight of 0.01. A neighborhood who has three ‘Bank’s, has a total weight of 0.03 for ‘Bank’. We do the above procedure for each venue category and each neighborhood and we get the table as below.

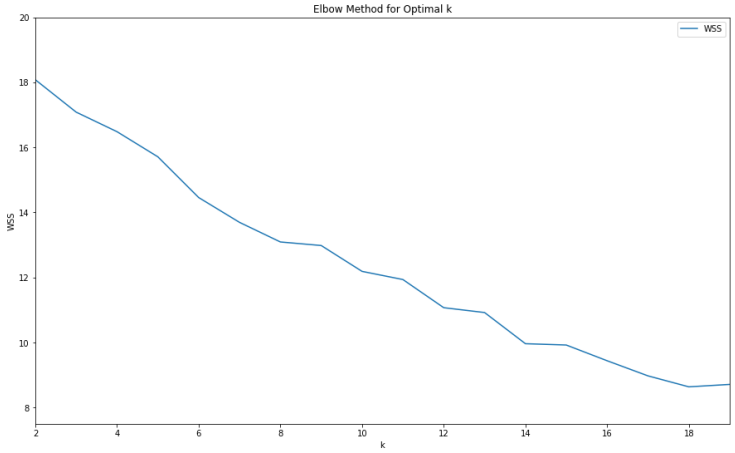
	NEIGHBORHOOD	Accessories Store	Afghan Restaurant	Airport	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership	Auto Garage	BBQ Joint	Baby Store	Badminton Court	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Basketball Court	Basketball Stadium	Beach Bar	Bed & Breakfast	Beer Bar	Beer Store
0	Adelaide, King, Richmond	0.0	0.0	0.0	0.03	0.01	0.0	0.03	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.03	0.01	0.03	0.0	0.0	0.0	0.0	0.0	0.01	0.000000
1	Agincourt	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.000000	0.0	0.0	0.0	0.0	0.2	0.0	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.000000
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.000000
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.083333
4	Alderwood, Long Branch	0.0	0.0	0.0	0.00	0.00	0.0	0.00	0.111111	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.000000

We can now get our most common 10 venue category for each neighborhood.

	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	Adelaide, King, Richmond	Coffee Shop	Café	Steakhouse	Restaurant	Bar	Gastropub	Gym	Asian Restaurant	Japanese Restaurant	American Restaurant
1	Agincourt	Breakfast Spot	Skating Rink	Badminton Court	Latin American Restaurant	Shanghai Restaurant	Yoga Studio	Falafel Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
2	Agincourt North, L'Amoreaux East, Milliken, St...	Sushi Restaurant	Pharmacy	Farmers Market	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space	Falafel Restaurant	Yoga Studio
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Liquor Store	Beer Store	Caribbean Restaurant	Coffee Shop	Fried Chicken Joint	Pizza Place	Hardware Store	Fast Food Restaurant	Pharmacy
4	Alderwood, Long Branch	Pizza Place	Athletics & Sports	Coffee Shop	Dance Studio	Pub	Gym	Convenience Store	Pharmacy	Sandwich Place	Farmers Market

Our goal is to find similar neighborhoods by looking their 10 most common venue categories. Now we can use unsupervised learning K-means algorithm to cluster neighborhoods. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

First, we need to select the K number, which determines how many clusters will be suitable for our data set. We use different K numbers to calculate each With-in-Sum-of-Square (WSS). Then we draw a graph of WSS versus K number and at the same time we create a table to see the decrease of WSS value as K increase.



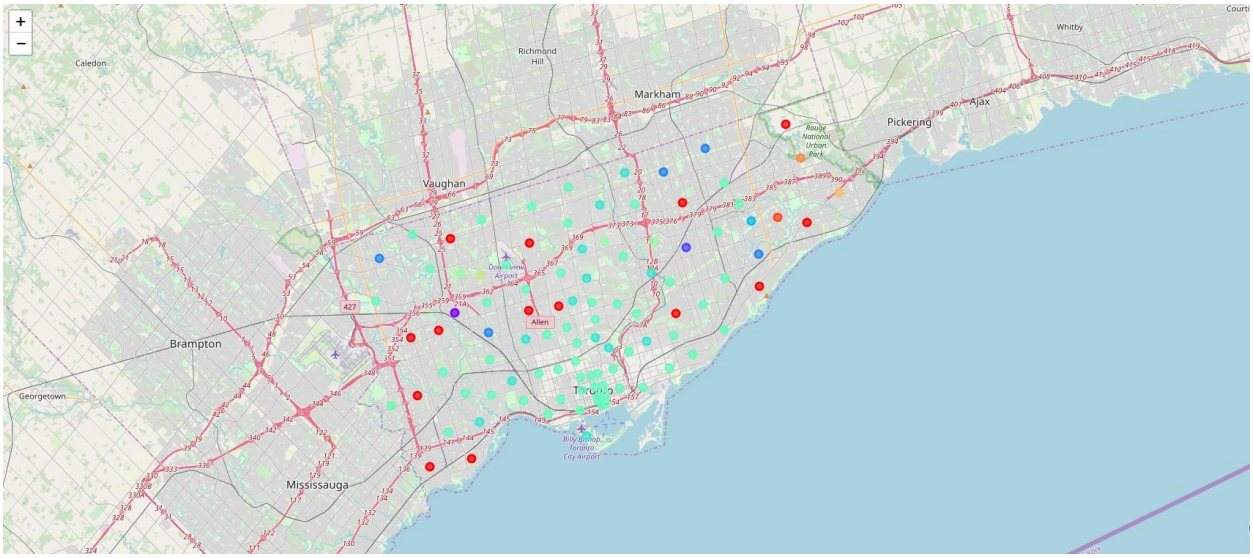
K	WSS	WSS Diff
2	18.079042	NaN
3	17.082677	-0.996365
4	16.481616	-0.601060
5	15.709200	-0.772416
6	14.456100	-1.253100
7	13.692914	-0.763186
8	13.090651	-0.602263
9	12.981026	-0.109625
10	12.185948	-0.795078
11	11.935022	-0.250926
12	11.069251	-0.865771
13	10.920317	-0.148933
14	9.962029	-0.958288
15	9.921269	-0.040761
16	9.438036	-0.483232
17	8.973160	-0.464877
18	8.635769	-0.337390
19	8.710062	0.074292

Using the Elbow Method, we can see the K number of 12 get us the best cluster.

Here is our merged table with cluster labels for each neighborhood.

	Postcode	Borough	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M3A	North York	Parkwoods	43.7545	-79.3300	7	Bus Stop	Food & Drink Shop	Pool	Park	Falafel Restaurant	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
1	M4A	North York	Victoria Village	43.7276	-79.3148	6	Hockey Arena	Intersection	Park	Portuguese Restaurant	Coffee Shop	French Restaurant	Pizza Place	Harbor / Marina	Dim Sum Restaurant	Fast Food Restaurant
2	M5A	Downtown Toronto	Harbourfront	43.6555	-79.3626	6	Coffee Shop	Restaurant	Yoga Studio	Breakfast Spot	Electronics Store	Thai Restaurant	Bakery	Pub	Italian Restaurant	Event Space
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.7223	-79.4504	6	Clothing Store	Coffee Shop	Cosmetics Shop	Electronics Store	Toy / Game Store	Men's Store	Sandwich Place	Jewelry Store	Food Court	Sushi Restaurant
4	M7A	Downtown Toronto	Queen's Park	43.6641	-79.3889	6	Gym	Pharmacy	Park	Burrito Place	Diner	College Cafeteria	Sushi Restaurant	Portuguese Restaurant	Coffee Shop	Ramen Restaurant

And we can see clusters with different colors on the map.



We are now able to see how many neighborhoods are in each cluster.

Cluster Labels	Postcode	Borough	Neighborhood	Latitude	Longitude	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	14	14	14	14	14	11	11	11	11	11	11	11	11	11	11
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13
6	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60
7	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Since there are less than 5 units in cluster 1, 2, 4 and 7-11, we only focus on cluster 0, 3, 5, 6.

Cluster 0 has its most restaurants Pizza Place or other kinds of fast food places with lots of coffee shops. It seems like places in Downtown suburb business center.

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	East York	0	Pizza Place	Fast Food Restaurant	Gym / Fitness Center	Breakfast Spot	Gastropub	Bank	Intersection	Bus Line	Pharmacy	Curling Ice
10	North York	0	Pizza Place	Mediterranean Restaurant	Gas Station	Latin American Restaurant	Fast Food Restaurant	Ice Cream Shop	Japanese Restaurant	Grocery Store	Event Space	Eastern European Restaurant
11	Etobicoke	0	Pizza Place	Construction & Landscaping	Tea Room	Sandwich Place	Coffee Shop	Chinese Restaurant	Wine Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant
18	Scarborough	0	Pizza Place	Electronics Store	Fast Food Restaurant	Coffee Shop	Liquor Store	Burger Joint	Sandwich Place	Food & Drink Shop	Breakfast Spot	Discount Store
28	North York	0	Pizza Place	Middle Eastern Restaurant	Coffee Shop	Fried Chicken Joint	Deli / Bodega	Mediterranean Restaurant	Fish & Chips Shop	Field	Fish Market	Fast Food Restaurant
51	Scarborough	0	Ice Cream Shop	Pizza Place	Bistro	Sandwich Place	Bank	Coffee Shop	Discount Store	Pharmacy	Cuban Restaurant	Cosmetics Shop
60	North York	0	Pizza Place	Grocery Store	Discount Store	Mobile Phone Shop	Liquor Store	Caribbean Restaurant	Gas Station	Fast Food Restaurant	Fried Chicken Joint	Sandwich Place
62	Central Toronto	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
70	Etobicoke	0	Pizza Place	Flea Market	Sandwich Place	Middle Eastern Restaurant	Ice Cream Shop	Supermarket	Chinese Restaurant	Coffee Shop	Yoga Studio	Electronics Store
76	Etobicoke	0	Pharmacy	Supermarket	Bus Line	Chinese Restaurant	Beer Store	Bank	Sandwich Place	Gas Station	Shopping Mall	Pizza Place

Cluster 3 is very similar to Cluster 0 with lots of fast food restaurant. However, it also has some full service restaurants such as Sushi, Chinese Restaurant. At the meantime, it also has many grocery stores, and other neighborhood retails. It seems like neighborhood center.

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
32	Scarborough	3	Spa	Grocery Store	Yoga Studio	Falafel Restaurant	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space	Farmers Market
56	York	3	Fast Food Restaurant	Playground	Coffee Shop	Event Space	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Falafel Restaurant
84	Scarborough	3	Sushi Restaurant	Pharmacy	Farmers Market	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space	Falafel Restaurant	Yoga Studio
88	Etobicoke	3	Grocery Store	Liquor Store	Beer Store	Caribbean Restaurant	Coffee Shop	Fried Chicken Joint	Pizza Place	Hardware Store	Fast Food Restaurant	Pharmacy
89	Scarborough	3	Fast Food Restaurant	Chinese Restaurant	Pizza Place	Other Great Outdoors	Coffee Shop	Grocery Store	Sandwich Place	Pharmacy	Event Space	Falafel Restaurant

Cluster 5 has lots of Parks. There are not many restaurants in this cluster; however, bakery and breakfast restaurants are very popular here.

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	North York	5	Gym	River	Park	Yoga Studio	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
21	York	5	Park	Sporting Goods Shop	Mexican Restaurant	Bakery	Beer Store	Fast Food Restaurant	Gym	Market	Fish & Chips Shop	Field
27	North York	5	Residential Building (Apartment / Condo)	Park	Yoga Studio	Event Space	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Falafel Restaurant
35	East York	5	Park	Convenience Store	Rental Car Location	Greek Restaurant	Intersection	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant
36	Downtown Toronto	5	Music Venue	Café	Harbor / Marina	Park	Event Space	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant
39	North York	5	Flower Shop	Gas Station	Trail	Park	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Yoga Studio
61	Central Toronto	5	Photography Studio	Park	Lawyer	Yoga Studio	Falafel Restaurant	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
66	North York	5	Convenience Store	Bank	Park	Falafel Restaurant	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space	Farmers Market
69	West Toronto	5	Bed & Breakfast	Park	Residential Building (Apartment / Condo)	Yoga Studio	Event Space	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Falafel Restaurant
73	Central Toronto	5	Playground	Garden	Park	Gym Pool	Event Space	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant

Cluster 6 has all kinds of restaurants with many Coffee shops. It also has lots of entertainment facilities. This cluster seems like Downtown area and centers of districts.

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	North York	6	Hockey Arena	Intersection	Park	Portuguese Restaurant	Coffee Shop	French Restaurant	Pizza Place	Harbor / Marina	Dim Sum Restaurant	Fast Food Restaurant
2	Downtown Toronto	6	Coffee Shop	Restaurant	Yoga Studio	Breakfast Spot	Electronics Store	Thai Restaurant	Bakery	Pub	Italian Restaurant	Event Space
3	North York	6	Clothing Store	Coffee Shop	Cosmetics Shop	Electronics Store	Toy / Game Store	Men's Store	Sandwich Place	Jewelry Store	Food Court	Sushi Restaurant
4	Downtown Toronto	6	Gym	Pharmacy	Park	Burrito Place	Diner	College Cafeteria	Sushi Restaurant	Portuguese Restaurant	Coffee Shop	Ramen Restaurant
5	Queen's Park	6	Gym	Pharmacy	Park	Burrito Place	Diner	College Cafeteria	Sushi Restaurant	Portuguese Restaurant	Coffee Shop	Ramen Restaurant
7	North York	6	Furniture / Home Store	Pool	Golf Driving Range	Park	Basketball Court	Yoga Studio	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
9	Downtown Toronto	6	Coffee Shop	Clothing Store	Japanese Restaurant	Café	Cosmetics Shop	Bakery	Pizza Place	Lingerie Store	Theater	Plaza
14	East York	6	Convenience Store	Spa	Asian Restaurant	Video Store	Beer Store	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
15	Downtown Toronto	6	Coffee Shop	Café	Restaurant	Bakery	Seafood Restaurant	Italian Restaurant	Breakfast Spot	Cosmetics Shop	Beer Bar	Hotel
16	York	6	Hockey Arena	Trail	Tennis Court	Playground	Field	Park	Deli / Bodega	Grocery Store	Donut Shop	Dumpling Restaurant

Now we have summarized each cluster characters using information of their common venues. We can now answer our question: if we are going to open a food business of one kind, where should we open inside a city?

D. Results

Toronto

As we have already summarized different clusters, we can differentiate clusters by their common restaurant categories. From above analysis, clusters 0 and 3 have better places for fast food restaurants. Cluster 5 has better places for bakery and breakfast restaurants. Cluster 6 welcomes all kinds of restaurants.

New York City

For New York City, we focus on cluster 0, 1 and 5.

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Central Bronx	0	Pizza Place	Chinese Restaurant	Bakery	Grocery Store	Donut Shop	Supermarket	Bank	Bus Station	Metro Station	Wings Joint
3	Hunts Point and Mott Haven	0	Pizza Place	Donut Shop	Chinese Restaurant	Mobile Phone Shop	Playground	Mexican Restaurant	Metro Station	Fruit & Vegetable Store	Discount Store	Diner
9	Borough Park	0	Chinese Restaurant	Convenience Store	Pizza Place	Bakery	Pharmacy	Bank	Grocery Store	Italian Restaurant	Bagel Shop	Asian Restaurant
10	Canarsie and Flatlands	0	Bakery	Italian Restaurant	Discount Store	Pizza Place	Martial Arts Dojo	Women's Store	Café	Cosmetics Shop	Restaurant	Liquor Store
11	Southern Brooklyn	0	Pizza Place	Bakery	Lounge	Arts & Crafts Store	Food & Drink Shop	Men's Store	Baseball Field	Bar	Donut Shop	Fish Market
14	East New York and New Lots	0	Pizza Place	Discount Store	Supermarket	Deli / Bodega	Jewelry Store	Sandwich Place	Grocery Store	Bus Line	Train Station	Mediterranean Restaurant
34	Rockaways	0	Food & Drink Shop	Beach	Pharmacy	Deli / Bodega	Chinese Restaurant	Donut Shop	Bus Stop	Metro Station	Fried Chicken Joint	Supermarket

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Bronx Park and Fordham	1	Deli / Bodega	Sandwich Place	Spanish Restaurant	Pizza Place	Coffee Shop	Plaza	Shoe Store	Chinese Restaurant	Fast Food Restaurant	Supplement Shop
2	High Bridge and Morrisania	1	Ice Cream Shop	Southern / Soul Food Restaurant	Sandwich Place	Mexican Restaurant	Fast Food Restaurant	Pharmacy	Shopping Mall	Sporting Goods Shop	Donut Shop	Food Court
4	Kingsbridge and Riverdale	1	Supermarket	Sandwich Place	Pizza Place	Spanish Restaurant	Mexican Restaurant	Donut Shop	Latin American Restaurant	Steakhouse	Flower Shop	Breakfast Spot
7	Central Brooklyn	1	Park	Pizza Place	Pool	Burger Joint	Food	Mobile Phone Shop	Donut Shop	Discount Store	Performing Arts Venue	Clothing Store
13	Flatbush	1	Caribbean Restaurant	Bank	Pharmacy	Bakery	Mobile Phone Shop	Liquor Store	Basketball Court	Supermarket	Fried Chicken Joint	Fish Market

	Neighborhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Southeast Bronx	5	Fast Food Restaurant	Donut Shop	Liquor Store	Sandwich Place	Supermarket	Restaurant	Bus Line	Gym / Fitness Center	Medical Center	Park
8	Southwest Brooklyn	5	Italian Restaurant	Cosmetics Shop	Spa	Bagel Shop	Middle Eastern Restaurant	Bar	Mobile Phone Shop	Greek Restaurant	Pharmacy	Grocery Store
12	Northwest Brooklyn	5	Bakery	Yoga Studio	Deli / Bodega	Pizza Place	Park	Gym	Asian Restaurant	Mexican Restaurant	Bagel Shop	Burger Joint
15	Greenpoint	5	Bar	Pizza Place	Wine Bar	Coffee Shop	Mexican Restaurant	Yoga Studio	Dive Bar	Burger Joint	Café	South American Restaurant
18	Central Harlem	5	Park	Coffee Shop	Café	Southern / Soul Food Restaurant	Mexican Restaurant	Seafood Restaurant	Restaurant	Bakery	Grocery Store	Plaza
19	Chelsea and Clinton	5	Gym / Fitness Center	Coffee Shop	Salad Place	Pizza Place	Spa	Bakery	Sandwich Place	Deli / Bodega	Restaurant	Martial Arts Dojo
21	Gramercy Park and Murray Hill	5	Bar	Italian Restaurant	Pizza Place	Bagel Shop	Thrift / Vintage Store	Comedy Club	Cocktail Bar	Mexican Restaurant	Grocery Store	American Restaurant
22	Greenwich Village and Soho	5	Italian Restaurant	Clothing Store	Café	Boutique	Coffee Shop	American Restaurant	Art Gallery	French Restaurant	Sporting Goods Shop	Sushi Restaurant
23	Lower Manhattan	5	Sandwich Place	Coffee Shop	Gym / Fitness Center	Hotel	Italian Restaurant	Park	Spa	Cocktail Bar	American Restaurant	French Restaurant
24	Lower East Side	5	Mexican Restaurant	Bar	Cocktail Bar	Coffee Shop	Chinese Restaurant	Art Gallery	Ice Cream Shop	Asian Restaurant	American Restaurant	Vegetarian / Vegan Restaurant

Cluster 0 has many restaurants, especially Chinese restaurants with supermarket and grocery stores. Cluster 1 has many fast food restaurants with other neighborhood retails. Cluster 5 has all kinds of restaurant, bar, retail stores and other shops. As a result, for restaurants with full services other than Chinese restaurants, Cluster 5 neighborhoods are better choices compared to other 2 clusters. Fast food restaurants can choose either Cluster 1 or 5 neighborhoods. For Chinese restaurants, Cluster 0 and 5 could be good choices.

London

For London, we focus on cluster 0, 1 and 3.

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Ealing, Hammersmith and Fulham	0	Park	Pub	French Restaurant	Convenience Store	Grocery Store	Bakery	Gas Station	Train Station	Discount Store	Mini Golf
12	Islington	0	Pub	Park	Café	Cocktail Bar	Gastropub	Grocery Store	Bakery	Trail	Boutique	Organic Grocery
27	Tower Hamlets	0	Pub	Grocery Store	Burger Joint	Rental Car Location	Locksmith	Light Rail Station	Bar	Coffee Shop	Park	Discount Store
33	Tower Hamlets	0	Pub	Grocery Store	Burger Joint	Rental Car Location	Locksmith	Light Rail Station	Bar	Coffee Shop	Park	Discount Store
44	Islington	0	Pub	Park	Café	Cocktail Bar	Gastropub	Grocery Store	Bakery	Trail	Boutique	Organic Grocery
53	Lewisham	0	Pub	Hotel	Café	Park	Film Studio	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
60	Barnet	0	Pub	Asian Restaurant	Café	Outdoor Supply Store	Park	Clothing Store	Ice Cream Shop	Dessert Shop	Dim Sum Restaurant	Food Court
73	Islington	0	Pub	Park	Café	Cocktail Bar	Gastropub	Grocery Store	Bakery	Trail	Boutique	Organic Grocery
79	Kensington and Chelsea	0	Pub	Italian Restaurant	Café	Pizza Place	Garden	Restaurant	Cupcake Shop	Speakeasy	Farmers Market	French Restaurant
81	Southwark	0	Pub	Vegetarian / Vegan Restaurant	Bus Stop	Café	Furniture / Home Store	Fish & Chips Shop	Gastropub	Thai Restaurant	Grocery Store	Convenience Store

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Bromley	1	Fast Food Restaurant	Pub	Supermarket	Pizza Place	Train Station	Platform	Café	Convenience Store	Garden Center	Grocery Store
6	Islington	1	Coffee Shop	Grocery Store	Pub	Café	Bar	Italian Restaurant	Pizza Place	Gastropub	Lounge	Bike Rental / Bike Share
11	Richmond upon Thames	1	Pub	Farmers Market	Grocery Store	French Restaurant	Train Station	Food & Drink Shop	Café	Gastropub	Park	Gym / Fitness Center
13	Wandsworth	1	Pub	Cocktail Bar	Café	Indian Restaurant	Coffee Shop	Grocery Store	Supermarket	Pizza Place	Portuguese Restaurant	French Restaurant
20	Tower Hamlets	1	Café	Vegetarian / Vegan Restaurant	Pizza Place	Coffee Shop	Cocktail Bar	Bakery	Pub	Art Gallery	Vietnamese Restaurant	Brewery
28	Haringey	1	Indian Restaurant	Café	Convenience Store	Fast Food Restaurant	Light Rail Station	Bar	Bakery	Train Station	Restaurant	Auto Garage
32	Lewisham	1	Convenience Store	Coffee Shop	Grocery Store	Beer Store	Café	Gastropub	Fish & Chips Shop	Furniture / Home Store	Brewery	Bus Stop
35	Brent	1	Coffee Shop	Pub	Café	Pizza Place	Middle Eastern Restaurant	Indian Restaurant	Bar	Korean Restaurant	Theater	Gastropub
38	Southwark	1	Grocery Store	Bakery	Café	Coffee Shop	Turkish Restaurant	Gym / Fitness Center	Beer Bar	Bus Stop	Gastropub	Pub
39	Tower Hamlets	1	Café	Vegetarian / Vegan Restaurant	Pizza Place	Coffee Shop	Cocktail Bar	Bakery	Pub	Art Gallery	Vietnamese Restaurant	Brewery

	Borough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	City	3	Theater	Hotel	French Restaurant	Pub	Bakery	Pizza Place	Ice Cream Shop	Café	Plaza	Steakhouse
3	Westminster	3	Theater	Coffee Shop	Ice Cream Shop	Bakery	Dessert Shop	Hotel	Clothing Store	Cosmetics Shop	Museum	Indian Restaurant
5	Islington	3	Coffee Shop	Pub	Italian Restaurant	Sandwich Place	Wine Bar	French Restaurant	Burrito Place	Beer Bar	Restaurant	Café
8	Wandsworth	3	Fish & Chips Shop	Music Store	Garden	Tennis Court	Zoo Exhibit	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant	Film Studio
9	Southwark	3	Pub	Coffee Shop	Italian Restaurant	Hotel	Wine Bar	Portuguese Restaurant	Restaurant	Seafood Restaurant	Food Truck	Modern European Restaurant
10	City	3	Coffee Shop	Pub	Italian Restaurant	Sandwich Place	Wine Bar	French Restaurant	Burrito Place	Beer Bar	Restaurant	Café
14	Westminster	3	Hotel	Pub	Chinese Restaurant	Greek Restaurant	Café	Coffee Shop	Garden	Ice Cream Shop	English Restaurant	Gym / Fitness Center
15	Ealing	3	Coffee Shop	Pub	Bakery	Café	Bookstore	Pizza Place	Bus Stop	Burger Joint	Ice Cream Shop	Grocery Store
16	Westminster	3	Hotel	Pub	Italian Restaurant	Café	Sandwich Place	Bakery	Grocery Store	Coffee Shop	Turkish Restaurant	Deli / Bodega
18	Camden	3	Coffee Shop	Hotel	Gym / Fitness Center	Café	Italian Restaurant	Pub	Hotel Bar	Department Store	Restaurant	Farmers Market

Cluster 0 has many parks with some restaurants (mostly fast food) and pubs nearby. Cluster 1 has many fast food restaurants with pubs, coffee shops, farmers market and convenient stores. Cluster 3 has all kinds of restaurants, supermarkets and retail stores with lots of entertainments. As a result, for restaurants with full services, Cluster 3 neighborhoods are better choices compared to other 2 clusters. Fast food restaurants can choose either Cluster 1 or 3 neighborhoods. Since Cluster 0 has only a small amount of restaurants, it may be risky for a new comer.

Three cities comparison

Toronto and NYC are very similar. London seems a little different. There are pubs in all neighborhoods of London. This is not the same for Toronto and NYC. It may be worth to dig if opening a pub or investing a pub in London could be a good choice.

E. Discussion

As I mentioned in the beginning, we only focus to differentiate the food businesses by two classes: fast food business and non-fast food business. This approach has its limitation since sometimes it is difficult to clearly define a restaurant as fast food or non-fast food. As we look into our data further, we could see it provides us with more detailed venue information. More approaches can be tried in clustering and classification studies.

In this project, I used the K-means algorithm as my main tool of this clustering study. I set the optimal k value by using the Elbow method. However, the Elbow figures are not clear to recognize. At the meantime, only one coordinates were used for each neighborhood. This may also bias the result of common venues searching. For more detailed and accurate guidance, the data set can be expanded, and the details of the neighborhood or street can also be drilled.

Beyond our model limitation, our target problem could be better analyzed if we could add in more features from other sources, such as population, racing, local traffic and so on.

F. Conclusion

Since it is always important but time consuming to find a right location to start a food business, our analysis could help to give business owners and investors a quick guide to start searching in those three cities.

Not only finding a location, our analysis could also help find some unique characteristics for each city, such as the 'Pub' in London as we mentioned at the end of section 'Results'. Such findings could help our audiences know the city better and further make a better decision.

G. References

1. Geocode package: https://pypi.org/project/pgeocode/
2. New York City Borough, Neighborhood and Postal Code:

https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

3. Toronto Borough, Neighborhood and Postal Code:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

4. London Borough, Neighborhood and Postal Code:

https://en.wikipedia.org/wiki/List_of_areas_of_London