

Coursera IBM data science – Capstone Project

Clustering Toronto neighborhood based on available restaurant

Swarup Barua

October 2019



Introduction

Toronto is the most popular city in Canada (the largest country in North America). Different ethnic neighborhood throughout the city focuses on specific cuisines such as Italian cuisine in Little Italy, Chinese in Chinatown. Numerous world cuisines are available throughout the city including Portuguese, Japanese, Indian, Greek and Caribbean.

For this capstone project, I am going to analyze all available restaurants in Toronto which may help to an entrepreneur to decide which cuisine restaurant will be best to open in a specific location in Toronto. Analyzing the available restaurant categories in different neighborhoods would give us an idea of the distribution of the restaurant types across Toronto.

Business Problem

The objective of this capstone project is to determine which places in Toronto has which type of restaurant more. By using machine learning algorithm such as clustering this project aims to find which cuisine restaurants dominate in which part of Toronto.

Data

To solve this problem, I am going use the below data sources.

- List of neighborhoods in Toronto, Canada.
- Latitude and longitude of these neighborhoods.
- Venues related to food category around different neighborhoods.

Extracting the data

- Web scrapping of Toronto neighborhood data from Wikipedia.
- Fetch location details (latitude and longitude) of these neighborhoods using google geocoder package.
- Use Foursquare API to fetch food venue details for neighborhoods in Toronto.

Data Cleaning

After extracting the data from the above sources data needs to be cleaned for visualization and clustering.

- In the dataframe, few of the records have Borough as 'Not assigned'. These data need to remove first.
- Once above data are removed, neighborhoods having the same borough has been merged and location of the borough are added to the dataframe using a dataset of location details of Toronto.
- Filter the dataframe which are related to Toronto Borough.

Methodology

- Use foursquare API to get all the restaurant details and location for each neighborhood in our dataframe using category id '4d4b7105d754a06374d81259'.

- There are a lot of different categories restaurants are available in Toronto. In this project we are going to analyze for Chinese Restaurant, Italian Restaurant, Mexican Restaurant, New American Restaurant, Fast Food Restaurant, Sushi Restaurant, Japanese Restaurant, Thai Restaurant, Seafood Restaurant, Indian Restaurant, French Restaurant, Asian Restaurant and Greek Restaurant. Below is the resultant dataframe after filtering the dataframe with the defined restaurants.

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Downtown Toronto	Harbourfront, Regent Park	43.65426	-79.380636	El Catrin	43.650601	-79.358920	Mexican Restaurant
5	Downtown Toronto	Harbourfront, Regent Park	43.65426	-79.380636	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restaurant
8	Downtown Toronto	Harbourfront, Regent Park	43.65426	-79.380636	Cocina Economica	43.654959	-79.365657	Mexican Restaurant
10	Downtown Toronto	Harbourfront, Regent Park	43.65426	-79.380636	Souvlaki Express	43.655584	-79.364438	Greek Restaurant
13	Downtown Toronto	Harbourfront, Regent Park	43.65426	-79.380636	Archeo	43.650667	-79.359431	Italian Restaurant

- Use Onehot Encoding technique to find similar neighborhoods based on restaurant categories. The result of the dataframe is the neighborhoods related restaurant categories where venue categories represented by 1 in the column of matching venue category.

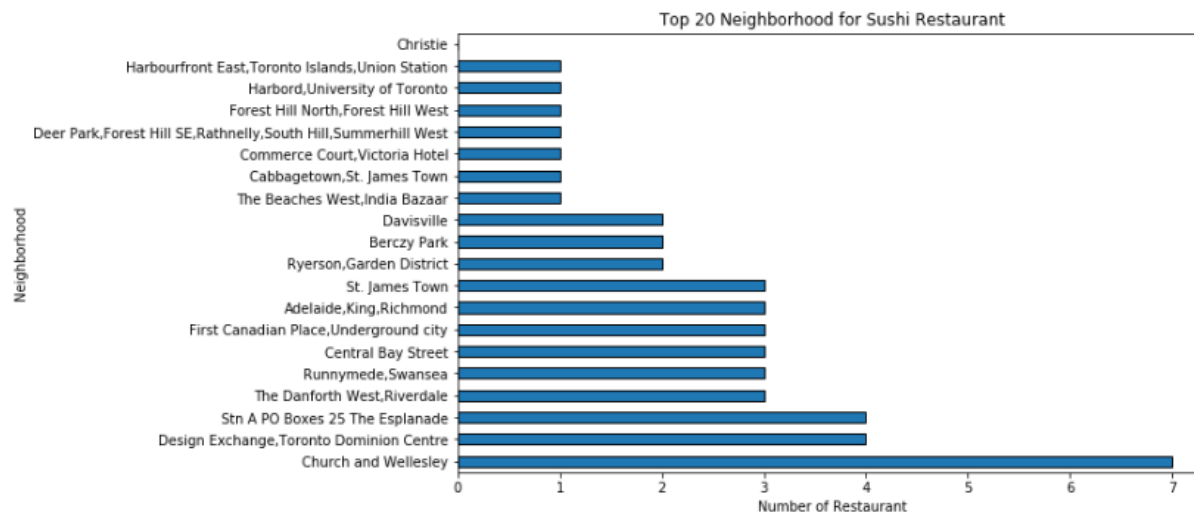
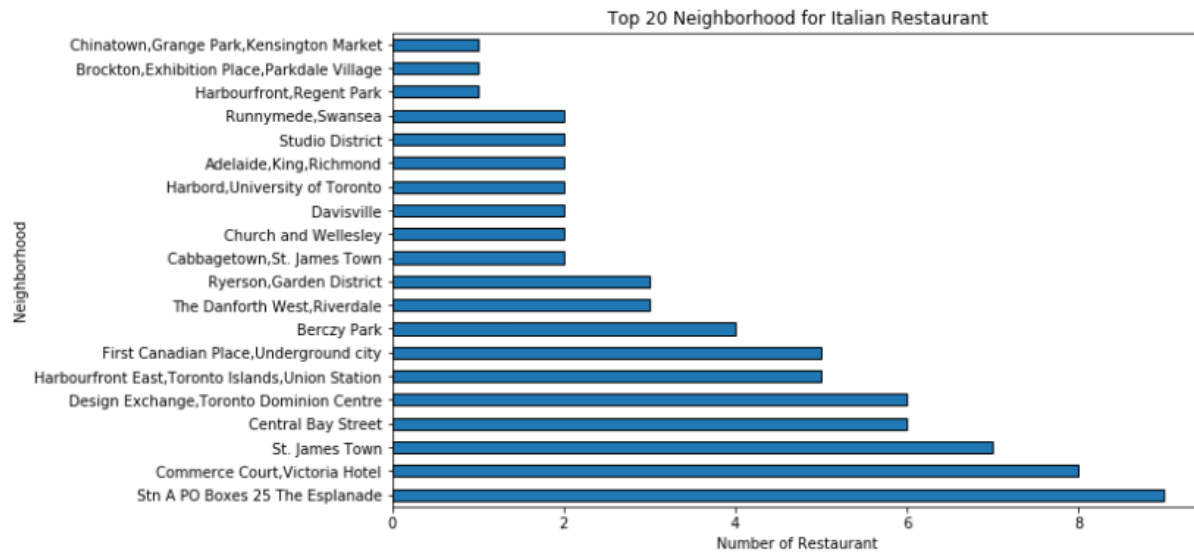
	Neighborhood	Asian Restaurant	Chinese Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	New American Restaurant	Seafood Restaurant	Re
4	Harbourfront, Regent Park	0	0	0	0	0	0	0	0	1	0	0	
5	Harbourfront, Regent Park	0	0	0	1	0	0	0	0	0	0	0	
8	Harbourfront, Regent Park	0	0	0	0	0	0	0	0	1	0	0	
10	Harbourfront, Regent Park	0	0	0	0	1	0	0	0	0	0	0	
13	Harbourfront, Regent Park	0	0	0	0	0	0	1	0	0	0	0	
14	Harbourfront, Regent Park	1	0	0	0	0	0	0	0	0	0	0	
15	Harbourfront, Regent Park	0	0	0	0	0	0	0	0	0	0	1	
17	Harbourfront, Regent Park	0	0	0	0	0	0	0	0	0	0	0	
18	Harbourfront, Regent Park	0	1	0	0	0	0	0	0	0	0	0	
20	Harbourfront, Regent Park	0	0	0	0	0	0	0	1	0	0	0	

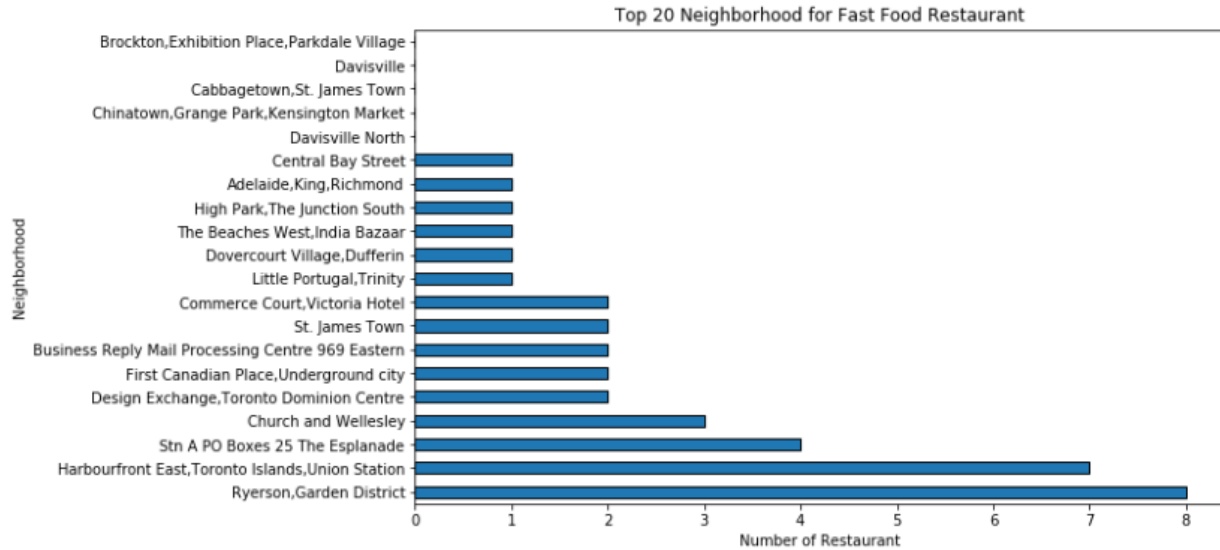
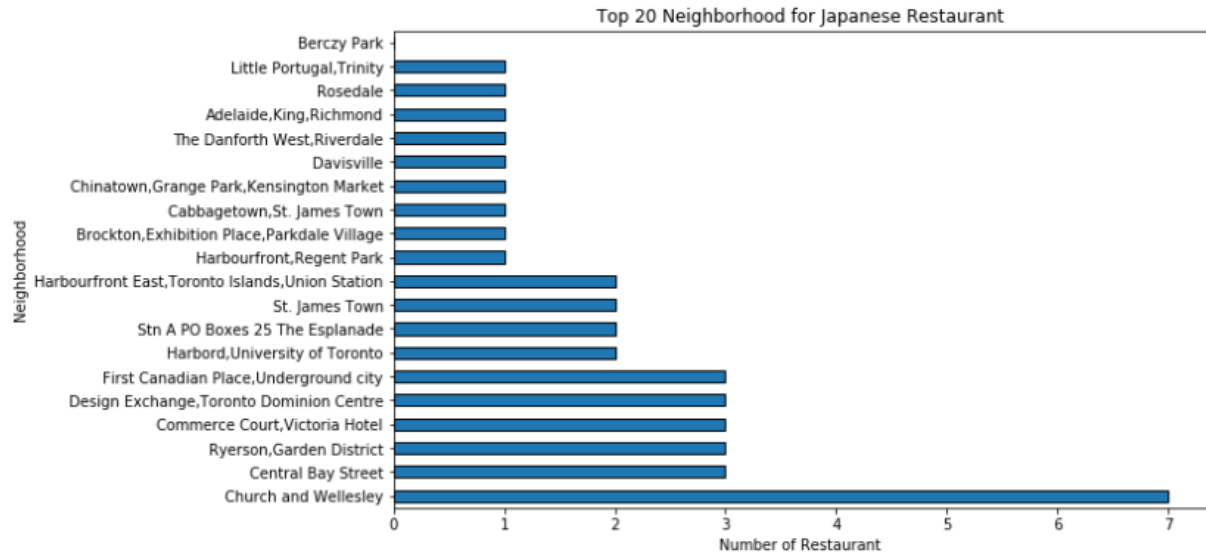
- Also get the count of restaurant category in each neighborhood through grouping the data based on neighborhood and then summing the restaurant count.

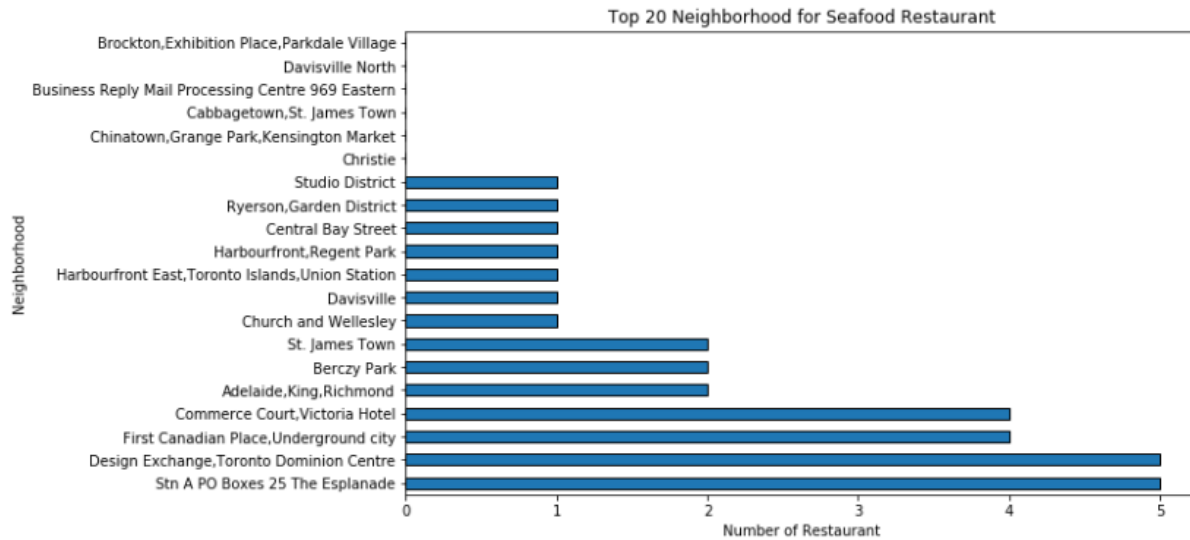
	Neighborhood	Asian Restaurant	Chinese Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	New American Restaurant	Seafood Restaurant	Re
	Adelaide, King, Richmond	6	1	1	0	1	1	2	1	2	1	2	
	Berczy Park	0	0	0	2	2	1	4	0	1	0	2	
	Brockton, Exhibition Place, Parkdale Village	0	0	0	0	0	0	1	1	0	0	0	
	Business Reply Mail Processing Centre 909 Eastern	0	0	2	0	0	0	0	0	0	0	0	
	Cabbagetown, St. James Town	0	2	0	0	0	1	2	1	0	0	0	

Data Visualization

To visualize the data of this project, I have plotted bar plot for top five restaurant categories (Italian, Sushi, Japanese, Fast Food and Seafood restaurant) with top twenty neighborhoods where these restaurant categories are having the maximum counts.







Result and discussion

From the analysis on the data I have arrived to a solution where I have focused on mainly two findings as below.

- Firstly, I have found the top five restaurant categories which are dominating others for each neighborhood. After my analysis below is the dataframe I have generated which shows the top five restaurant categories in a neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide, King, Richmond	Asian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	Mexican Restaurant
1	Berczy Park	Italian Restaurant	Sushi Restaurant	Seafood Restaurant	Greek Restaurant	French Restaurant
2	Brockton, Exhibition Place, Parkdale Village	Japanese Restaurant	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Fast Food Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	New American Restaurant
4	Cabbagetown, St. James Town	Italian Restaurant	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant

- Secondly, I have found the least five restaurant categories which are very less or not present around each neighborhood. After my analysis below is the dataframe I have generated which shows the least five restaurant categories in a neighborhood.

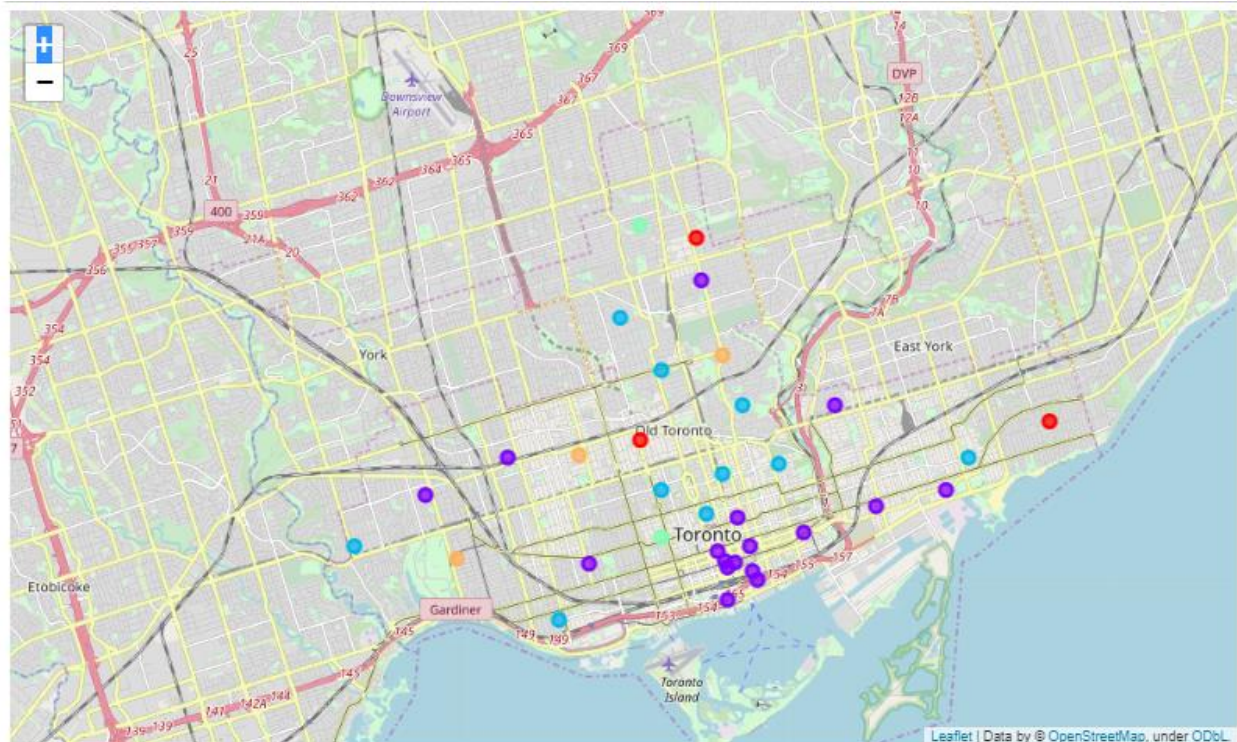
	Neighborhood	1st least Common Venue	2nd least Common Venue	3rd least Common Venue	4th Least Common Venue	5th Least Common Venue
0	Adelaide, King, Richmond	French Restaurant	Chinese Restaurant	Fast Food Restaurant	Greek Restaurant	Indian Restaurant
1	Berczy Park	Asian Restaurant	Chinese Restaurant	Fast Food Restaurant	Japanese Restaurant	New American Restaurant
2	Brockton, Exhibition Place, Parkdale Village	Asian Restaurant	Chinese Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Asian Restaurant	Chinese Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant
4	Cabbagetown, St. James Town	Asian Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant	Mexican Restaurant

Clustering

To find out the diversity of food culture in Toronto we have used machine learning unsupervised algorithm clustering which fits our model very well. K means clustering algorithm is used to cluster the neighborhood based on the available restaurant categories. Here I have used 5 different clusters. Below is the dataframe where each neighborhood has its cluster defined.

	PostalCode	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
22	M6R	West Toronto	Parkdale,Roncesvalles	43.648980	-79.456325	4	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	New American Restaurant
6	M6G	Downtown Toronto	Christie	43.669542	-79.422564	4	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	New American Restaurant
26	M4T	Central Toronto	Moore Park,Summerhill East	43.689574	-79.383160	4	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant	New American Restaurant
27	M5T	Downtown Toronto	Chinatown,Grange Park,Kensington Market	43.653206	-79.400049	3	Chinese Restaurant	Mexican Restaurant	Thai Restaurant	Japanese Restaurant	Italian Restaurant
20	M4R	Central Toronto	North Toronto West	43.715383	-79.405678	3	Mexican Restaurant	Italian Restaurant	Chinese Restaurant	Thai Restaurant	Sushi Restaurant
28	M4V	Central Toronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,....	43.686412	-79.400049	2	Sushi Restaurant	Thai Restaurant	Seafood Restaurant	New American Restaurant	Mexican Restaurant
24	M5S	Downtown Toronto	Harbord,University of Toronto	43.662696	-79.400049	2	Japanese Restaurant	Italian Restaurant	Sushi Restaurant	French Restaurant	Chinese Restaurant
14	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	2	Sushi Restaurant	Italian Restaurant	Fast Food Restaurant	Thai Restaurant	Seafood Restaurant
13	M6K	West Toronto	Brookton,Exhibition Place,Parkdale Village	43.636847	-79.428191	2	Japanese Restaurant	Italian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant
25	M6S	West Toronto	Runnymede,Swansea	43.651571	-79.484450	2	Sushi Restaurant	Italian Restaurant	French Restaurant	Thai Restaurant	Seafood Restaurant

To visualize the clusters better I have plotted different clusters on Toronto map with different colors.



Conclusion

From the above visualization i can see the similar neighborhoods based on available restaurant types and their distribution. Most of the places in Downtown Toronto are falling under cluster one. Also it is identified that most of the restaurants are from Downtown Toronto.

In this analysis we have listed top restaurant categories for each neighborhood. Based on this list someone can identify which neighborhood will be appropriate to find a particular type of restaurant.

The purpose of this project was to identify venues in Toronto where a particular type of restaurant is most available and which one is least available. Based on this data people interested in different cuisines can find their place of interest. Stakeholder how are trying to find a place to open a restaurant can find a place where a particular type of restaurant is least available.

Future direction

This project is developed to understand the most and least restaurant categories in Toronto neighborhoods. But this is not the optimum solution. In future different machine learning algorithm, techniques and foursquare API data can be used more extensively to investigate and determine the best possible neighborhoods to open a new restaurant of a specific categories which may help different interested stakeholders.