

## Introduction:

For someone who is going to start a restaurant in Toronto, it is really important to pick a right position for his business. In this project, we will use the venue information provided by the Foursquare to help the restaurant owner to decide the best place to start his business. The k-means cluster approach is used to classify the suburbs so that the stakeholder can start the restaurant in a suburb similar to the well-developed suburbs with many restaurants.

## Data:

The data used in this project are mainly two parts: the location data of Toronto, and the Foursquare data. The location data of Toronto ([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:M)) consisting of the postcode, borough, neighborhoods of the suburbs in Toronto. It is used along with the geospatial data to determine the exact latitudes and longitudes of the suburbs.

The Foursquare data is provided by the Foursquare. It will provide the information of the venues around certain locations. We will use these venues to determine the similarities of the suburbs.

## Methodology:

In this project, K-Means clustering is applied to solve the problem. The intuition behind this methodology is quite straightforward: find out the similar places where a lot of restaurant locate, which obviously means that these places are good to run a restaurant. And then among these places, we will find out the suitable places for a new restaurant based on the frequency of existing restaurants. If there are already too many restaurants in that place, running a new restaurant will be really hard facing the stress from others. The best place would be somewhere with enough existing restaurants to attract people but not too many restaurants to pose a great pressure.

About K-Means clustering: k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. [from Wikipedia]

Regarding this problem, we will use the frequency of the venues of a place as the feature. For example, below shows the venues of 5 top frequency of place M2J:

venue	freq
Clothing Store	0.12
Fast Food Restaurant	0.09
Coffee Shop	0.08
Restaurant	0.05
Toy / Game Store	0.03

Then the places will be clustered according to the frequencies of the venues inside them. After the clusters are found, we can inspect each cluster to find out the best cluster to run a restaurant, and inside the cluster, we can also find the best place to start a new restaurant.

Result:

The suburbs of Toronto are divided into 10 clusters. See the map below for detailed information about the cluster:

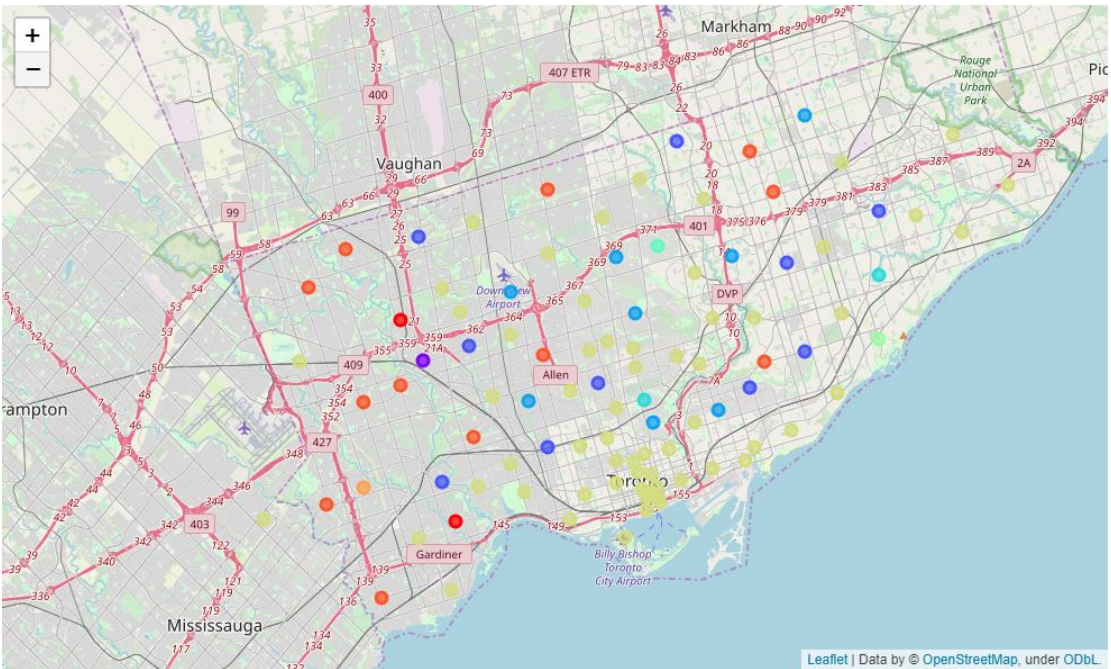


Figure - -1 Clustered map of Toronto

There are two clusters which are good for starting a new restaurant: cluster 7 and cluster 9. We can see the distribution of venues inside these two clusters with the following table:

	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
7	Scarborough	7	Intersection	Spa	Electronics Store	Pizza Place	Breakfast Spot	Tech Startup	Medical Center	Rental Car Location	Mexican Restaurant
8	North York	7	Restaurant	Coffee Shop	Ramen Restaurant	Japanese Restaurant	Café	Sandwich Place	Sushi Restaurant	Bubble Tea Shop	Ice Cream Shop
9	Etobicoke	7	Gym	Coffee Shop	Café	Fried Chicken Joint	Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Seafood Restaurant
13	East Toronto	7	Health Food Store	Trail	Neighborhood	Pub	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
14	Scarborough	7	Fast Food Restaurant	Print Shop	Dessert Shop	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
17	East York	7	Coffee Shop	Sporting Goods Shop	Sushi Restaurant	Burger Joint	Furniture / Home Store	Grocery Store	Gym	Beer Store	Supermarket
21	Scarborough	7	Construction & Landscaping	Bar	Yoga Studio	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
22	East Toronto	7	Yoga Studio	Fast Food Restaurant	Restaurant	Butcher	Skate Park	Auto Workshop	Burrito Place	Light Rail Station	Garden
24	Mississauga	7	Coffee Shop	Hotel	American Restaurant	Gym / Fitness Center	Fried Chicken Joint	Mediterranean Restaurant	Sandwich Place	Burrito Place	Dumpling Restaurant
25	Central Toronto	7	Pizza Place	Dessert Shop	Sandwich Place	Sushi Restaurant	Restaurant	Café	Thai Restaurant	Coffee Shop	Italian Restaurant
26	Downtown Toronto	7	Café	Vegetarian / Vegan Restaurant	Mexican Restaurant	Bar	Bakery	Coffee Shop	Dumpling Restaurant	Chinese Restaurant	Vietnamese Restaurant

Figure -2 Cluster 7

	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	East York	9	Fast Food Restaurant	Pizza Place	Pet Store	Gastropub	Café	Bank	Athletics & Sports	Intersection	Breakfast Spot	Gym / Fitness Center
30	Etobicoke	9	Pizza Place	Intersection	Sandwich Place	Coffee Shop	Chinese Restaurant	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant
37	Scarborough	9	Pizza Place	Italian Restaurant	Fast Food Restaurant	Pharmacy	Fried Chicken Joint	Bank	Noodle House	Thai Restaurant	Chinese Restaurant	Doner Restaurant
41	Etobicoke	9	Pharmacy	Pizza Place	Beer Store	Liquor Store	Convenience Store	Café	Dumpling Restaurant	Discount Store	Dog Run	Doner Restaurant
44	North York	9	Pharmacy	Pizza Place	Empanada Restaurant	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
46	Etobicoke	9	Pizza Place	Park	Mobile Phone Shop	Bus Line	Yoga Studio	Drugstore	Discount Store	Dog Run	Doner Restaurant	Donut Shop
53	North York	9	Park	Pizza Place	Japanese Restaurant	Pub	Yoga Studio	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant
59	Etobicoke	9	Grocery Store	Fast Food Restaurant	Pizza Place	Sandwich Place	Beer Store	Coffee Shop	Liquor Store	Fried Chicken Joint	Video Store	Pharmacy
62	Scarborough	9	Fast Food Restaurant	Chinese Restaurant	Breakfast Spot	Grocery Store	Pharmacy	Pizza Place	Coffee Shop	American Restaurant	Thrift / Vintage Store	Sandwich Place
71	North York	9	Pharmacy	Pizza Place	Discount Store	Grocery Store	Coffee Shop	Dumpling Restaurant	Diner	Dog Run	Doner Restaurant	Donut Shop
72	York	9	Pizza Place	Grocery Store	Bus Line	Convenience Store	Yoga Studio	Dumpling Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop

Figure -3 Cluster 9

Among these suburbs, Downtown Toronto and North York are good place to run the restaurant, because the density of existing restaurants are not so big but enough to attract

people.

## Discussion:

For some suburbs, there are not enough Foursquare venue data to provide a detailed description of this place, which means there might be some error with this kind of data. Though the suburbs cannot be clustered are cleaned during the processing, there still exist some data with big variance. Introducing new source of data may help to solve this issue.

## Conclusion:

In this project, we combined the Foursquare data with the geospatial data of Toronto to find out the best place for starting a new restaurant. We applied the K-means clustering to cluster the suburbs of Toronto according to the frequency of venues inside them. Finally, 10 clusters are made and 2 of them are good for our purpose. Inside these two clusters, Downtown Toronto and North York are the most suitable places.