

Introduction/Business Problem

For many business owners who want to succeed in a competitive world, the most asked question at the beginning of the stage is always 'Where should I open my business', because the right place for business will always come with benefits such as more customer or more sale. So business owner often relies on consultant firm or their human instinct to decide where should they open their business. Will it be great if instead of random choice and instinct? We can rely on customer generate data to decide the best place for a business like if somewhere most common place is restaurant type, We can assume that neighborhood is restaurant cluster and customers will likely to go there to find something to eat.

Boston city is the largest city and the capital city of Massachusetts. They have many people live there with multinational and preference. So in this report, we will try to use customer generate location data from Foursquare API to try to find the best neighborhood to open a restaurant in Boston city.

Data

In this report, we will data from Boston city's neighborhood to solve our problem. our data for analysis come from 4 sources :

First from [Analyze Boston](#) website we use Boston Neighborhoods Datasets to obtain a list of Boston neighborhood.

Second from [Statistical Atlas](#) website for our neighborhood population density data.

Third, we use geocoder package to obtain Latitude and Longitude Data from the name of each neighborhood.

and the last is from Foursquare API which we use to obtain data such as venue in each neighborhood and category of it.

Data from [Analyze Boston](#), [Statistical Atlas](#) and latitude longitude from geocode package

	Neighbourhood	Latitude	Longitude	Population Density
0	Roslindale	42.291209	-71.124497	9877.31
1	Jamaica Plain	42.309820	-71.120330	10363.86
2	Mission Hill	42.332926	-71.103214	22637.14
3	Longwood	42.336168	-71.099527	17561.48
4	Bay Village	42.350011	-71.066948	49840.22
5	Chinatown	42.352217	-71.062607	52322.61
6	North End	42.365097	-71.054495	42239.76
7	Roxbury	42.324843	-71.095016	14810.95
8	South End	42.341310	-71.077230	32320.73
9	Back Bay	42.350707	-71.079730	26282.87

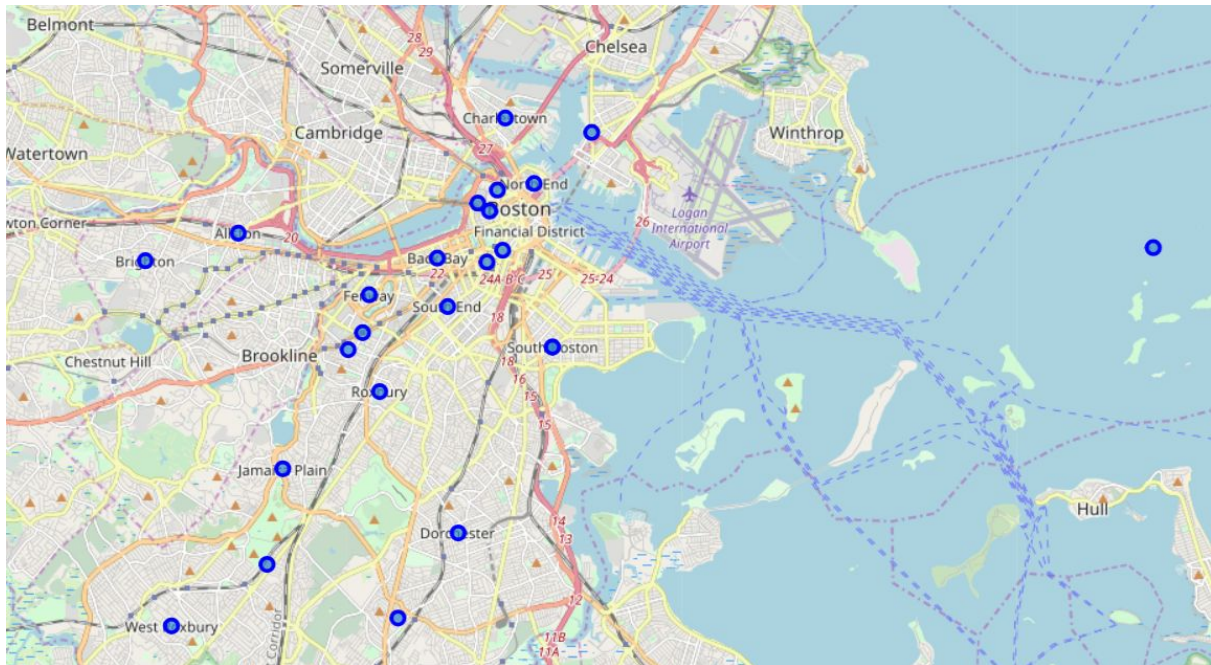
Data from Foursquare API

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roslindale	42.291209	-71.124497	Peters Hill	42.293617	-71.128063	Scenic Lookout
1	Roslindale	42.291209	-71.124497	Roslindale House Of Pizza	42.287989	-71.126549	Pizza Place
2	Roslindale	42.291209	-71.124497	Arboretum Grill	42.291787	-71.122497	American Restaurant
3	Roslindale	42.291209	-71.124497	Target	42.288204	-71.126659	Big Box Store
4	Roslindale	42.291209	-71.124497	Enterprise Rent-A-Car	42.292580	-71.121930	Rental Car Location
5	Roslindale	42.291209	-71.124497	BCYF- Flaherty Pool	42.288133	-71.122913	Pool
6	Roslindale	42.291209	-71.124497	Dunkin' Donuts	42.287912	-71.126991	Donut Shop
7	Roslindale	42.291209	-71.124497	James P. Healey Sq.	42.288553	-71.125753	Plaza
8	Roslindale	42.291209	-71.124497	Washington Street Plumbing	42.287850	-71.126497	Home Service
9	Roslindale	42.291209	-71.124497	Nicholas Sub and Pizza	42.289346	-71.119121	Pizza Place
10	Jamaica Plain	42.309820	-71.120330	Arnold Arboretum: Hunnewell Building	42.307112	-71.120159	Trail
11	Jamaica Plain	42.309820	-71.120330	Vee Vee	42.310210	-71.115143	American Restaurant
12	Jamaica Plain	42.309820	-71.120330	Salmagundi	42.310184	-71.115145	Accessories Store
13	Jamaica Plain	42.309820	-71.120330	Footlight Club	42.310394	-71.115957	Theater
14	Jamaica Plain	42.309820	-71.120330	Fiore Italian Bakery	42.308015	-71.115572	Bakery

Exploratory Data Analysis

Plot map of boston neighbourhood to get a look

First we plot a map boston neighbourhood by each dot is represent one neighbourhood to have an idea what the map look like and we will plot it again after we cluster data later on



Get summary statistics

From data from foursquare api we can see that we have obtained 1096 venue which come from 200 unique venue category the top venue is Dunkin' Donuts (15 frequency) and top venue category is italian restaurant (74 frequency).

	Venue	Venue Category
count	1096	1096
unique	894	200
top	Dunkin' Donuts	Italian Restaurant
freq	15	74

Take a look at unique venue category

We have seen that the Italian restaurant are the top frequency of venue category but are venue category contains just restaurant type? we looked at a sample of unique venue category and found that there is a wide range of type. So we can be sure we do not just obtain venue data that contain just restaurant category.

```
array(['Scenic Lookout', 'Pizza Place', 'American Restaurant',  
      'Big Box Store', 'Rental Car Location', 'Pool', 'Donut Shop',  
      'Plaza', 'Home Service', 'Trail', 'Accessories Store', 'Theater',  
      'Bakery', 'Pet Store', 'Coffee Shop', 'Liquor Store',  
      'Deli / Bodega', 'Seafood Restaurant', 'Art Gallery', 'Library',  
      'Child Care Service', 'Park', 'Bar', 'Italian Restaurant',  
      'New American Restaurant', 'Gastropub', 'Ice Cream Shop',  
      'Sandwich Place', 'Caribbean Restaurant', 'Pub'], dtype=object)
```

Population density statistics

In population density data we can see that it range from 5,640 per square mile to 52,322 per square mile which we think this data can help we define a better place to open the business with more denser in the neighborhood will translate to more customer.

Population Density	
count	24.00000
mean	21205.46500
std	14575.25071
min	5640.36000
25%	11734.99500
50%	15730.40000
75%	27792.33500
max	52322.61000

Preparing data process

One hot encoding

After we have obtained data from Foursquare API. We can't just put the data right away to model because our data of category that we will use for the model come as category data. So first we need to use one hot encoding method to transform our data from category to number value that can use with our model.

	Neighbourhood	Accessories Store	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	...	Tourist Information Center	Trail	Train Station	Vegetarian / Vegan Restaurant
0	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	Roslindale	0	1	0	0	0	0	0	0	0	...	0	0	0	0
3	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
5	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
6	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
7	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
8	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0
9	Roslindale	0	0	0	0	0	0	0	0	0	...	0	0	0	0

Groupby and mean function to calculate frequency

We will use groupby with mean function on one hot encoded data to obtain the frequency of each venue category that appear in each neighborhood. Through this method, we can see which category are most common in some neighborhood and we can use this data to cluster neighborhood together later on.

	Neighbourhood	Accessories Store	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	...	Tourist Information Center	Trail	Train Station
0	Allston	0.000000	0.000000	0.000000	0.012346	0.000000	0.000000	0.024691	0.000000	0.000000	...	0.00	0.000000	0.000000
1	Back Bay	0.020000	0.050000	0.000000	0.010000	0.000000	0.000000	0.000000	0.010000	0.000000	...	0.00	0.000000	0.000000
2	Bay Village	0.000000	0.031250	0.000000	0.000000	0.000000	0.010417	0.020833	0.000000	0.000000	...	0.00	0.000000	0.000000
3	Beacon Hill	0.000000	0.032258	0.000000	0.000000	0.000000	0.000000	0.000000	0.016129	0.000000	...	0.00	0.000000	0.000000
4	Brighton	0.000000	0.020833	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
5	Charlestown	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.029412	0.000000	...	0.00	0.000000	0.000000
6	Chinatown	0.000000	0.020000	0.000000	0.000000	0.000000	0.000000	0.070000	0.000000	0.000000	...	0.01	0.000000	0.000000
7	Dorchester	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
8	Downtown	0.000000	0.021739	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021739	...	0.00	0.000000	0.000000
9	East Boston	0.000000	0.029412	0.000000	0.029412	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
10	Fenway	0.000000	0.030000	0.000000	0.010000	0.000000	0.010000	0.000000	0.010000	0.010000	...	0.00	0.000000	0.000000
11	Hyde Park	0.000000	0.133333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.066667
12	Jamaica Plain	0.055556	0.055556	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.055556	0.000000
13	Longwood	0.000000	0.000000	0.000000	0.000000	0.043478	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
14	Mattapan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.083333
15	Mission Hill	0.000000	0.045455	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
16	North End	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
17	Roslindale	0.000000	0.100000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
18	Roxbury	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
19	South Boston	0.000000	0.024390	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
20	South Boston Waterfront	0.000000	0.024390	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
21	South End	0.022222	0.022222	0.022222	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000	0.000000
22	West End	0.000000	0.022989	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.022989	...	0.00	0.000000	0.011494

Show example of top most common venue in each neighborhood.

----Allston----		
	venue	freq
0	Korean Restaurant	0.08
1	Pizza Place	0.04
2	Chinese Restaurant	0.04
----Back Bay----		
	venue	freq
0	Coffee Shop	0.05
1	American Restaurant	0.05
2	Italian Restaurant	0.05
----Bay Village----		
	venue	freq
0	Sandwich Place	0.05
1	Hotel	0.05
2	Theater	0.04

Modeling

With the problem to find the best neighborhood to open a restaurant business. I choose the cluster model to find a solution to this problem. Because in real-world common business often group together or open near each other and form the cluster that attracts customers to their location. So I think if we can clustering our neighborhood together base on venue category and population density. We can find the cluster of the neighborhood that appropriate to open a restaurant.

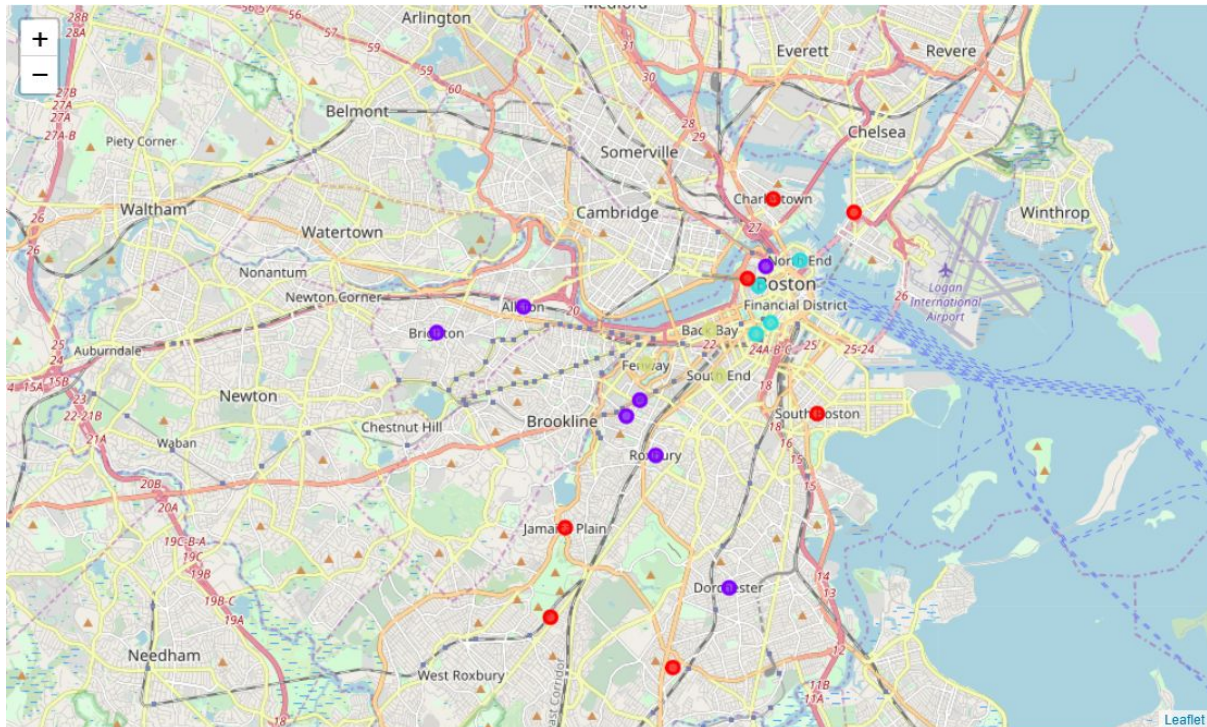
K-Mean Clustering

From cluster approach, we choose K Mean Cluster for clustering the data. The problem of K Mean approach is we have to specify the number of cluster that we want to have so we run this model multiple time with different number of cluster and come up with number 4. We think it the best number of cluster that has the most dissimilar between each cluster and has the most similarity between intercluster.

```
# set number of clusters
kclusters = 4
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(boston_grouped_clustering)
```

Results

After cluster the similar neighborhood together we can plot it on the map as shown below.



Characteristics of each cluster

If we look at each cluster you will see that each cluster have cluster mainly base on population density and we can see the top most common venues for some cluster that dissimilar with each other.

By the results, we can characterize each neighborhood cluster from its most common venue category and population density (for the potential to open business) as follows.

Cluster 1 : Average density , Fast food restaurant , Public venues , and Coffee

	Neighbourhood	Population Density	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	Mission Hill	22637.14	1	Sushi Restaurant	Sandwich Place	Pizza Place	Convenience Store	Greek Restaurant	Caribbean Restaurant	Gastropub	Café	Coffee Shop	Bar
3	Longwood	17561.48	1	Donut Shop	Sandwich Place	Italian Restaurant	Gastropub	Coffee Shop	Metro Station	Falafel Restaurant	Café	Liquor Store	Bookstore
7	Roxbury	14810.95	1	Art Gallery	Plaza	Furniture / Home Store	Gym	Metro Station	Yoga Studio	Dog Run	Fast Food Restaurant	Farmers Market	Falafel Restaurant
12	West End	17420.70	1	Sandwich Place	Pizza Place	Donut Shop	Coffee Shop	Hotel	Bar	Italian Restaurant	Café	Mexican Restaurant	Sports Bar
16	Brighton	16433.47	1	Chinese Restaurant	Pizza Place	Pub	Coffee Shop	Bakery	Dry Cleaner	Bus Station	Grocery Store	Deli / Bodega	Bank
19	Dorchester	15027.33	1	Yoga Studio	Southern / Soul Food Restaurant	Gym	Fried Chicken Joint	Market	Mobile Phone Shop	Donut Shop	Diner	Nail Salon	Pizza Place
21	Allston	19029.08	1	Korean Restaurant	Pizza Place	Chinese Restaurant	Bakery	Thai Restaurant	Dive Bar	Donut Shop	Bubble Tea Shop	Italian Restaurant	Pharmacy

Cluster 2 : High density , Restaurant and Public venues

	Neighbourhood	Population Density	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	Bay Village	49840.22	2	Sandwich Place	Hotel	Steakhouse	Performing Arts Venue	Theater	Italian Restaurant	Seafood Restaurant	Spa	Gym / Fitness Center	Coffee Shop
5	Chinatown	52322.61	2	Chinese Restaurant	Asian Restaurant	Bakery	Coffee Shop	Theater	Sushi Restaurant	Performing Arts Venue	Pizza Place	Sandwich Place	Mexican Restaurant
6	North End	42239.76	2	Italian Restaurant	Pizza Place	Park	Seafood Restaurant	Bakery	Wine Shop	Café	Coffee Shop	Market	Sandwich Place
13	Beacon Hill	47770.65	2	Italian Restaurant	Hotel Bar	Pizza Place	French Restaurant	Museum	Hotel	Gourmet Shop	Gift Shop	Sushi Restaurant	Plaza

Cluster 3 : Above average density , Mostly restaurant and bar

	Neighbourhood	Population Density	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	South End	32320.73	3	Italian Restaurant	Coffee Shop	Gift Shop	Wine Shop	Wine Bar	Park	Bakery	Bar	Latin American Restaurant	Japanese Restaurant
9	Back Bay	26282.87	3	American Restaurant	Coffee Shop	Italian Restaurant	Seafood Restaurant	Hotel	Sporting Goods Shop	Clothing Store	Cosmetics Shop	Spa	Salon / Barbershop
15	Fenway	37978.28	3	Sports Bar	Pizza Place	Lounge	Coffee Shop	American Restaurant	Restaurant	Baseball Field	Thai Restaurant	Brewery	Mexican Restaurant

Cluster 4 : Below average density , Mixed with restaurant and Public venues

	Neighbourhood	Population Density	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	Roslindale	9877.31	4	Yoga Studio	Pizza Place	American Restaurant	Scenic Lookout	Donut Shop	Pool	Big Box Store	Rental Car Location	Dive Bar	Falafel Restaurant
1	Jamaica Plain	10363.86	4	Art Gallery	Bakery	Coffee Shop	Accessories Store	Deli / Bodega	Liquor Store	Library	Donut Shop	Pizza Place	Theater
10	East Boston	8594.21	4	Latin American Restaurant	Sandwich Place	Convenience Store	Pizza Place	Donut Shop	Colombian Restaurant	Pharmacy	Mexican Restaurant	Restaurant	Fast Food Restaurant
11	Charlestown	12192.04	4	Pizza Place	Coffee Shop	Convenience Store	Pub	Liquor Store	Gastropub	Shopping Mall	Donut Shop	Discount Store	Café
14	Downtown	5975.80	4	Italian Restaurant	Hotel Bar	Gift Shop	Museum	Gourmet Shop	Baseball Field	Bistro	Playground	Pizza Place	Other Repair Shop
17	Hyde Park	7474.90	4	American Restaurant	Pizza Place	Bar	Gym	Donut Shop	Discount Store	Platform	Theater	Italian Restaurant	Fast Food Restaurant
18	Mattapan	12753.33	4	Home Service	Bakery	Park	Scenic Lookout	Liquor Store	Furniture / Home Store	Caribbean Restaurant	Rental Car Location	Ice Cream Shop	Cosmetics Shop
20	South Boston	12192.04	4	Chinese Restaurant	Convenience Store	Pizza Place	Italian Restaurant	Cosmetics Shop	Bar	Liquor Store	Sports Bar	Coffee Shop	Ice Cream Shop

Conclusion

Now we will try to answer our business problem that is 'Where should i open a restaurant in Boston city?'. From our problem and results, we are considering two things first is potential for open business which we will evaluate from population density and second is more restaurant as top commonplace in that cluster which mean it will attract customer if the cluster of the restaurant is formed.

The results we think cluster 3 is the most appropriate cluster for open restaurant because it has more than average density and mostly comprises with the restaurant as top commonplace. Follow by cluster 2 that we think the owner of the business should consider too because it has the most density in all cluster but has more public space and non-restaurant venue as top common so we place this cluster as the second. the remain two cluster we think are not worth considering because average to low population density and top commonplace often mixed with the non-restaurant type

Future directions

We use only 2 feature of data for this model we think if we can collect more data about the neighborhood like average income, eating out spending or some other data we can build the more concrete model for answer the problem.

In addition, if you try to run the model with the different data like foursquare data that only contain restaurant type. We can answer a more specific question like what type of restaurant should we open.