Picking a location in MA for Food Supply Chain Company

Introduction

The goal of the project is to provide investors the data support for their decision on the location choice to start their new business. Their purpose is to start a food supply chain company that target on the restaurants' food supply in Massachusetts. Investors would like to know the concentration of different types of the restaurants in different cities and after that they can start researching on their food supply preference and deciding the location for their new company.

Data

The cities data is from Wikipedia page (here) List of Municipalities in Massachusetts
After read the data from the html link, I cleaned it to make it only contains the Cities since the investors would mainly concern to start from a place with higher population-density.

Methodology

Step1. Import the geographical coordinates from the geoby library for all the Cities in our list.

	City	Latitude	Longitude
4	Agawam	42.069539	-72.614812
6	Amesbury	42.857954	-70.930092
16	Attleboro	41.944544	-71.285608
20	Barnstable	41.700113	-70.299467
30	Beverly	42.558428	-70.880049

Step2. Explore the Cities in MA

We used the foursquare API to get the information of all Venues in each city and import them to a data frame like below

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agawam	42.069539	-72.614812	Giovanni's	42.069010	-72.615021	Bakery
1	Agawam	42.069539	-72.614812	Phelp's School Field	42.070885	-72.613872	Baseball Field
2	Agawam	42.069539	-72.614812	Dealmaster Pro	42.069589	-72.615128	Business Service
3	Agawam	42.069539	-72.614812	Utopia	42.069134	-72.614888	Food & Drink Shop
4	Agawam	42.069539	-72.614812	Shell	42.072247	-72.615099	Gas Station

Step3. Analyze the venues for each city

We have a look at all the venues/categories that we got from foursquare. Be aware that the only venues the investors interested are the those of restaurants.

	City	ATM	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bag Sho
0	Agawam	0	0	0	0	0	0	0	0	0	0	0	
1	Agawam	0	0	0	0	0	0	0	0	0	0	0	
2	Agawam	0	0	0	0	0	0	0	0	0	0	0	
3	Agawam	0	0	0	0	0	0	0	0	0	0	0	
4	Agawam	0	0	0	0	0	0	0	0	0	0	0	
5	Agawam	0	0	0	0	0	0	0	0	0	0	0	
6	Amesbury	0	0	0	0	0	0	0	0	0	0	0	
7	Amesbury	0	0	0	0	0	0	0	0	0	0	0	
8	Amesbury	0	0	1	0	0	0	0	0	0	0	0	
9	Amesbury	0	0	0	0	0	0	0	0	0	0	0	
10	Amesbury	0	0	1	0	0	0	0	0	0	0	0	
11	Amesbury	0	0	1	0	0	0	0	0	0	0	0	
12	Amesbury	0	0	0	0	0	0	0	0	0	0	0	
13	Amesbury	0	0	0	0	0	0	0	0	0	0	0	
14	Amesbury	0	0	1	0	0	0	0	0	0	0	0	
15	Amesbury	0	0	0	0	0	0	0	0	0	0	0	
16	Amesbury	0	0	0	0	0	0	0	0	0	0	0	

Step4. Slice the dataframe to leave only the venues of restaurants.

City	African Restaurant	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Caribbean Restaurant	Chinese Restaurant	C Res
scroll output; double click to hide										
Agawam	0	0	0	0	0	0	0	0	0	
Agawam	0	0	0	0	0	0	0	0	0	
Agawam	0	0	0	0	0	0	0	0	0	
Agawam	0	0	0	0	0	0	0	0	0	
Agawam	0	0	0	0	0	0	0	0	0	
Agawam	0	0	0	0	0	0	0	0	0	
Amesbury	0	0	0	0	0	0	0	0	0	
Amesbury	0	0	0	0	0	0	0	0	0	
Amesbury	0	1	0	0	0	0	0	0	0	
Amesbury	0	0	0	0	0	0	0	0	0	

Then we can group rows by cities and by taking the mean of the frequency of occurrence of each category.

click to scroll output; double click to hide

	City	African Restaurant	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Caribbean Restaurant	Chinese Restaurant
0	Agawam	0.0	0.000000	0.000000	0.00	0.0	0.0	0.0	0.0	0.000000
1	Amesbury	0.0	0.148148	0.000000	0.00	0.0	0.0	0.0	0.0	0.074074
2	Attleboro	0.0	0.000000	0.062500	0.00	0.0	0.0	0.0	0.0	0.000000
3	Barnstable	0.0	0.125000	0.000000	0.00	0.0	0.0	0.0	0.0	0.000000
4	Beverly	0.0	0.000000	0.000000	0.00	0.0	0.0	0.0	0.0	0.000000
5	Boston	0.0	0.020000	0.000000	0.01	0.0	0.0	0.0	0.0	0.000000
6	Braintree	0.0	0.000000	0.000000	0.00	0.0	0.0	0.0	0.0	0.000000
7	Bridgewater	0.0	0.100000	0.000000	0.00	0.0	0.0	0.0	0.0	0.050000
8	Brockton	0.0	0.000000	0.166667	0.00	0.0	0.0	0.0	0.0	0.000000
9	Cambridge	0.0	0.000000	0.000000	0.00	0.0	0.0	0.0	0.0	0.000000

After that, we print out each cities along with the top 10 most common venues.

click to scroll output; double click to hide

```
----Agawam----
                        venue  freq
0          African Restaurant   0.0
1       Portuguese Restaurant   0.0
2   Latin American Restaurant   0.0
3    Mediterranean Restaurant   0.0
4          Mexican Restaurant   0.0
5             Meze Restaurant   0.0
6   Middle Eastern Restaurant   0.0
7         Moroccan Restaurant   0.0
8     New American Restaurant   0.0
9          Persian Restaurant   0.0


----Amesbury----
                          venue  freq
0           American Restaurant  0.15
1            Chinese Restaurant  0.07
2  Vegetarian / Vegan Restaurant  0.04
3            Italian Restaurant  0.04
4       Comfort Food Restaurant  0.04
5            African Restaurant  0.00
6         Portuguese Restaurant  0.00
7            Mexican Restaurant  0.00
8               Meze Restaurant  0.00
9     Middle Eastern Restaurant  0.00


----Attleboro----
```

Finally we have the data frame with the above info.

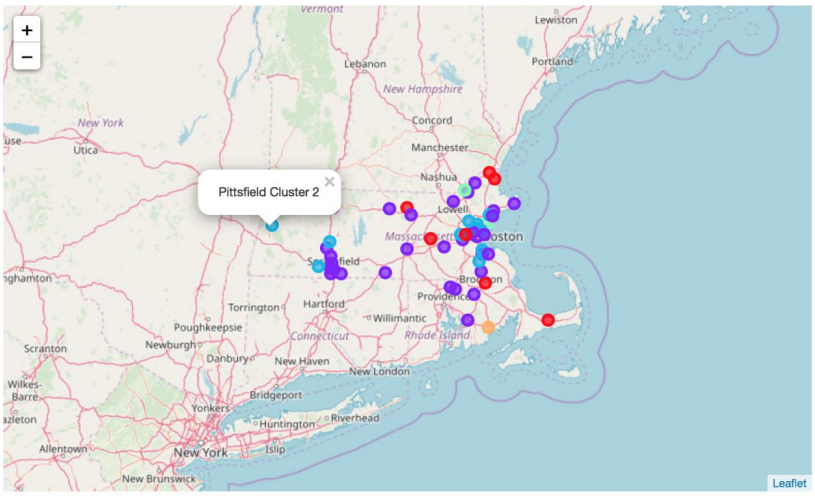
	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mo Comm Venu
0	Agawam	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopi Restaura
1	Amesbury	American Restaurant	Chinese Restaurant	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Caribbean Restaurant	Belgi Restaura
2	Attleboro	Asian Restaurant	Spanish Restaurant	Italian Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Foo Restaura
3	Barnstable	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Food Restaurant	Fala Restaura
4	Beverly	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopi Restaura

Step5: Cluster different types of restaurants, using k-means

Set the k = 5, we can create 5 clusters of the restaurants

	City	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
4	Agawam	42.069539	-72.614812	1	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
6	Amesbury	42.857954	-70.930092	0	American Restaurant	Chinese Restaurant	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Brazilian Restaurant
16	Attleboro	41.944544	-71.285608	1	Asian Restaurant	Spanish Restaurant	Italian Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant
20	Barnstable	41.700113	-70.299467	0	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant
30	Beverly	42.558428	-70.880049	1	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
35	Boston	42.360253	-71.058291	1	Seafood Restaurant	Italian Restaurant	American Restaurant	Restaurant	Falafel Restaurant	Greek Restaurant
40	Braintree	42.222322	-70.999492	1	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
42	Bridgewater	41.990379	-70.975043	0	American Restaurant	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Cuban Restaurant	Halal Restaurant
44	Brockton	42.083433	-71.018379	1	Asian Restaurant	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant
						New	Middle			

Step6. Visualize the Clusters on the Map



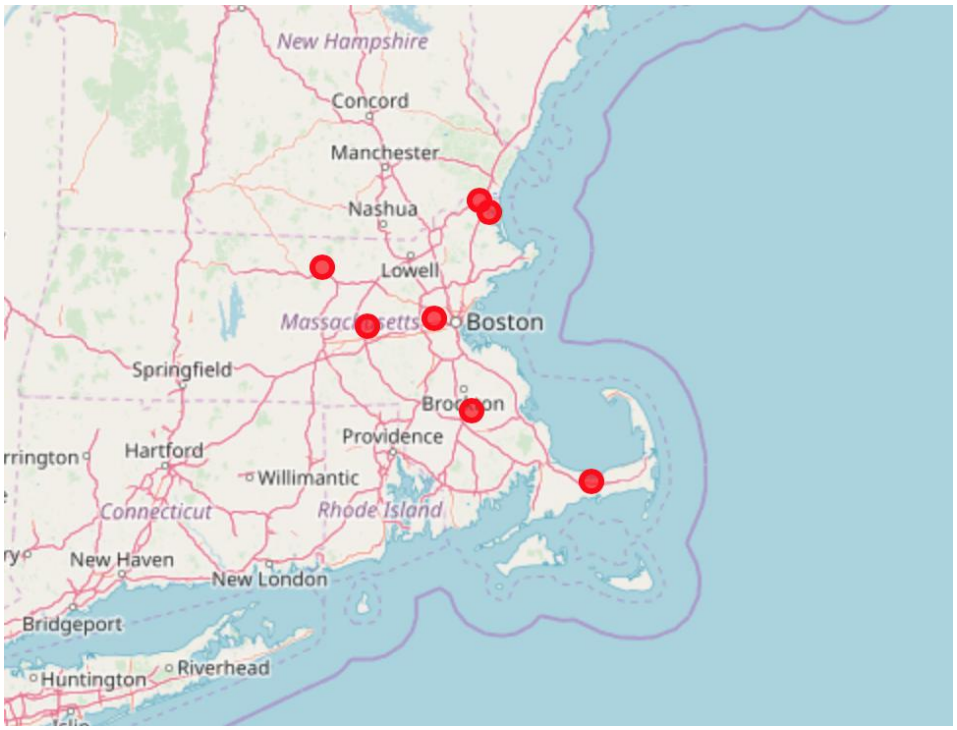
Results

First by check each cluster, we can simply drop the clusters with label 3 and 4 since they had very limited samples thus it's hard to tell the patterns of the restaurants' types.

Cluster Label = 0

This cluster shows the cities with restaurants of types mostly of American and Chinese food restaurants. And the Cities are mainly surrounding the Boston area.

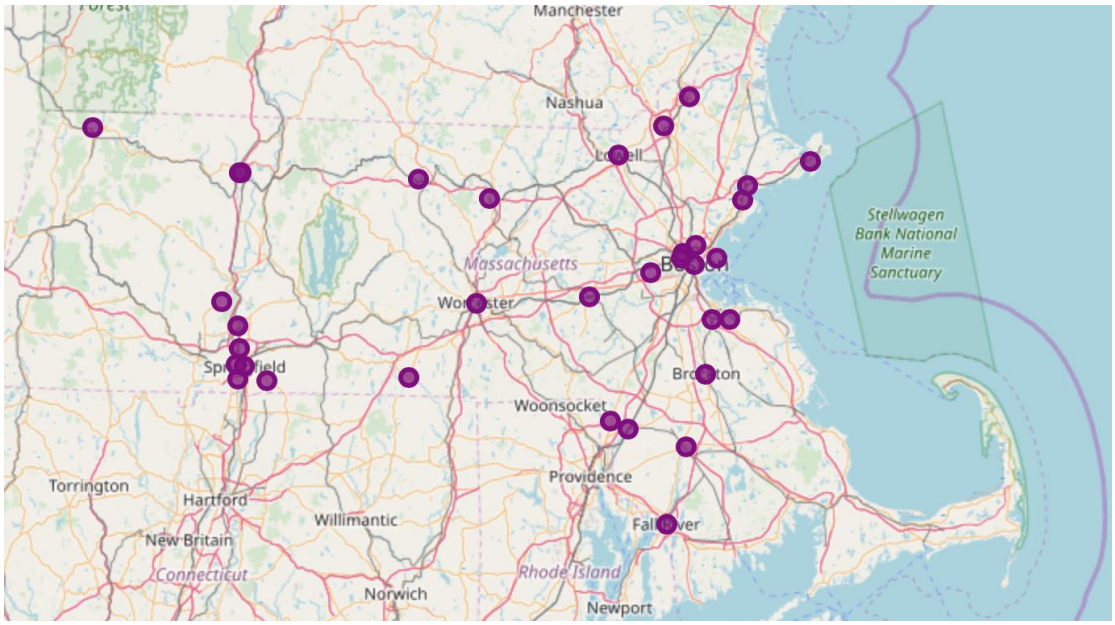
	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Con
6	Amesbury	0	American Restaurant	Chinese Restaurant	Comfort Food Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Carib Resta
20	Barnstable	0	American Restaurant	Seafood Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Resta
42	Bridgewater	0	American Restaurant	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	F Resta
97	Fitchburg	0	American Restaurant	Chinese Restaurant	Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Resta
169	Marlborough	0	American Restaurant	Chinese Restaurant	Thai Restaurant	Seafood Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	F Resta
205	Newburyport	0	American Restaurant	Seafood Restaurant	Italian Restaurant	New American Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	Ge Resta
313	Watertown	0	Persian Restaurant	American Restaurant	Middle Eastern Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant	French Restaurant	Fast Resta



Cluster Label = 1

In Cluster Label 1, The most common venue is Asian restaurants. And the cities are spread out all in MA with pretty large samples. And the density is higher in Central Boston area and around Springfield.

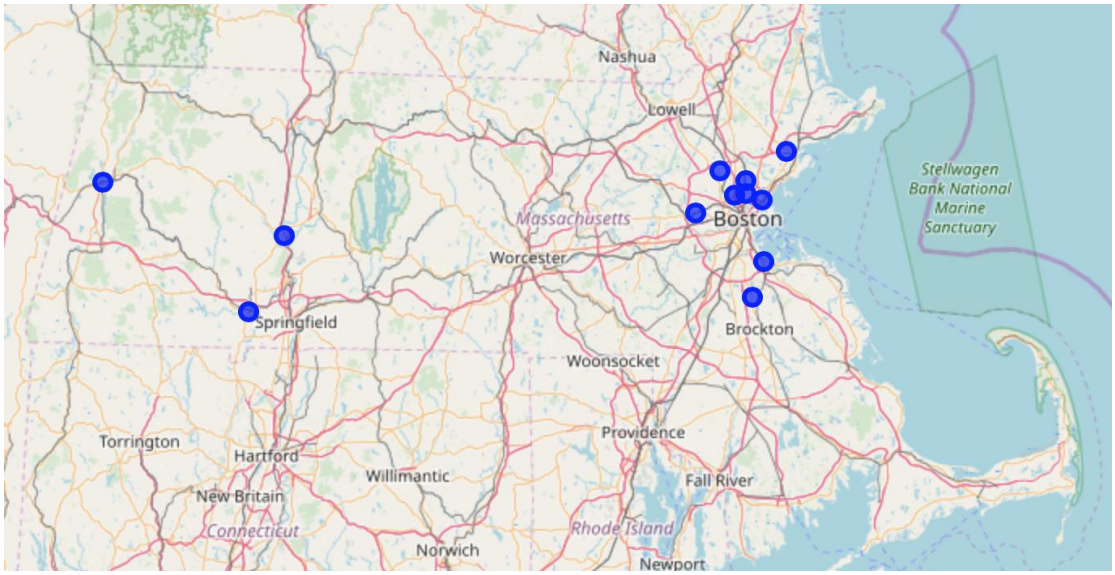
	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
4	Agawam	42.069539	-72.614812	1	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
16	Attleboro	41.944544	-71.285608	1	Asian Restaurant	Spanish Restaurant	Italian Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant
30	Beverly	42.558428	-70.880049	1	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
35	Boston	42.360253	-71.058291	1	Seafood Restaurant	Italian Restaurant	American Restaurant	Restaurant	Falafel Restaurant	Greek Restaurant
40	Braintree	42.222322	-70.999492	1	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
44	Brockton	42.083433	-71.018379	1	Asian Restaurant	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant
49	Cambridge	42.375100	-71.105616	1	Indian Restaurant	New American Restaurant	Middle Eastern Restaurant	Thai Restaurant	Tapas Restaurant	Italian Restaurant
61	Chicopee	42.148704	-72.607867	1	German Restaurant	Vietnamese Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant
85	East Longmeadow	42.064540	-72.512586	1	American Restaurant	Mexican Restaurant	Vietnamese Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant
87	Easthampton	42.266757	-72.668980	1	Indian Restaurant	Mexican Restaurant	Restaurant	Cuban Restaurant	Halal Restaurant	Greek Restaurant



Cluster Label = 2

The most common venues of the cities in cluster labeled 2 are Asian restaurants and Italian restaurants. Cities are mainly concentrated around Malden.

	City	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
164	Malden	42.425096	-71.066163	2	Asian Restaurant	Chinese Restaurant	Vietnamese Restaurant	American Restaurant	Indian Restaurant	Korean Restaurant
175	Medford	42.418430	-71.106164	2	Italian Restaurant	American Restaurant	Chinese Restaurant	Thai Restaurant	Restaurant	Mexican Restaurant
177	Melrose	42.456432	-71.064182	2	Italian Restaurant	American Restaurant	Sushi Restaurant	Seafood Restaurant	Mexican Restaurant	Cuban Restaurant
213	Northampton	42.319023	-72.630634	2	Italian Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	American Restaurant	Thai Restaurant	Asian Restaurant
228	Peabody	42.527873	-70.928661	2	Italian Restaurant	Sushi Restaurant	Mexican Restaurant	Restaurant	Portuguese Restaurant	Comfort Food Restaurant
235	Pittsfield	42.450097	-73.245379	2	Italian Restaurant	Cuban Restaurant	Indian Restaurant	Halal Restaurant	Greek Restaurant	German Restaurant
242	Quincy	42.252877	-71.002270	2	American Restaurant	Italian Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Chinese Restaurant	Korean Restaurant
243	Randolph	42.161800	-71.042484	2	Vietnamese Restaurant	Italian Restaurant	American Restaurant	Ethiopian Restaurant	Indian Restaurant	Halal Restaurant
247	Revere	42.408430	-71.011995	2	Italian Restaurant	American Restaurant	Mexican Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Halal Restaurant



Discussion

From the the above data, we find that the most popular restaurant types are American Restaurants and Asian food restaurants. The investors can start doing research on these types of the restaurant about the their currently adopting food supply chain. If the investors had decided which type of restaurants are their target customers, they can choose the location by checking the visualized clusters of the restaurants. In addition, if the investors do not have a specific preference, I recommend to set the new company in Boston area or the Springfield area since these place has the most restaurants concentration.

Conclusion

In this project, I download all the cities data in Massachusetts from Wikipedia. By using Geopy library and Foursquare API, I export all the common venues in each city. Since the investors will only concern the concentration of different types of restaurants, we drop the unnecessary data to leave only the venues of restaurants. In order to separate the cities by the similarity of the types of restaurants, I use k-means cluster and picked the k as 5. After that, I use the folium maps to visualize each cluster of the cities by which the investors can decide the location on their preference.