

Coursera Capstone
IBM Applied Data Science Capstone

The best neighborhood to open a restaurant in Boston city

by Kessawa Palanund



Introduction:

In the business world, there is the most interesting question when people ask where should they open the business. It is the undeniable fact that the right place for business will always come with benefits such as more customers or more sales. It would be better if we use the customer generate data to decide the best place for a business instead of random choice and instinct. For example, we can find commonplace for restaurant type by using the data generated from Foursquare that defines the cluster by processing the code using python to show where in the neighborhood is restaurant cluster that customers will likely to go to find something to eat.

We will focus on Boston city as it is the largest city and the capital city of Massachusetts. There are also many people with a diversity of multinational.

In conclusion, we will try to use data from the Foursquare API to find the best neighborhood to open a restaurant in Boston city.

Business problem:

Ideally, we want to open a restaurant where there is a high population. This can be done by the use of data science methodologies where in the neighborhood is restaurant cluster that customers will likely to go to find something to eat in Boston.

Data:

In this report, we will use 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.

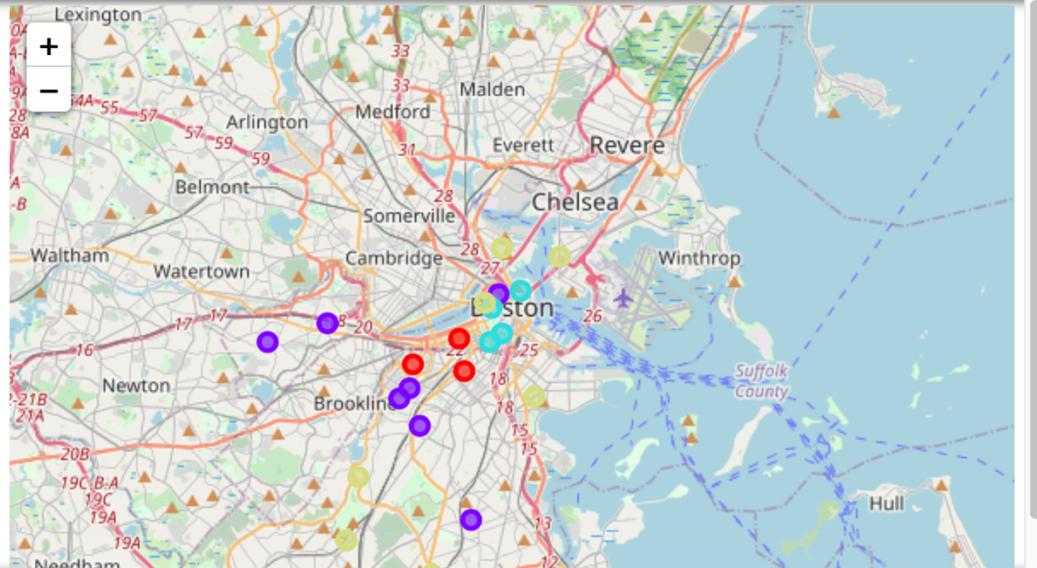
Methodology:

1. First, I scraped data containing the city. After cleaning the data, I appended the latitude and longitude of the cities to the original data frame containing the city. Then, I used a folium module to plot the cities on the map to visualize the data.
2. Using the Foursquare API client ID, I collected the venues near Boston.
3. Using these frequencies, I made 4 clusters to identify which cluster had the highest and lowest restaurant.

Results:

Through clustering, we can find that there are 4 clusters.

Out[51]:



Cluster 1 which most common is Pizza place, café and Korean restaurant.

```
boston_merged.loc[boston_merged['Cluster Labels'] == 1,boston_merged.columns[[1,4,5,6,7,8,9,10,11,12,13,14,15]]]
```

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	Sandwich Place	Pizza Place	Sushi Restaurant	American Restaurant	Greek Restaurant	New American Restaurant	Coffee Shop	Gastropub	Donut Shop	Bar
3	Longwood	17561.48	1	Donut Shop	Sandwich Place	Italian Restaurant	College Cafeteria	Gym	Gastropub	Coffee Shop	New American Restaurant	Bar	Pub
7	Roxbury	14810.95	1	Park	Gym	Plaza	Art Gallery	Furniture / Home Store	Rental Car Location	Yoga Studio	Dive Bar	Farmers Market	Falafel Restaurant
12	West End	17420.70	1	Pizza Place	Donut Shop	Hotel	Sandwich Place	Bar	Italian Restaurant	Café	Coffee Shop	Sports Bar	Mexican Restaurant
16	Brighton	16433.47	1	Pizza Place	Bank	Bakery	Pub	Deli / Bodega	Coffee Shop	Bus Station	Grocery Store	Chinese Restaurant	Dance Studio
20	Dorchester	15027.33	1	Café	Shoe Store	Pizza Place	Southern / Soul Food Restaurant	Vegetarian / Vegan Restaurant	Plaza	Market	Fried Chicken Joint	Gym	Dog Run
22	Allston	19029.08	1	Korean Restaurant	Thrift / Vintage Store	Vegetarian / Vegan Restaurant	Chinese Restaurant	Pizza Place	Bakery	Thai Restaurant	Italian Restaurant	Bubble Tea Shop	Pharmacy

Cluster 2 which most common is Spa, Chinese and Italian restaurant.

```
boston_merged.loc[boston_merged['Cluster Labels'] == 2,boston_merged.columns[[1,4,5,6,7,8,9,10,11,12,13,14,15]]]
```

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	Spa	Seafood Restaurant	Bakery	Sandwich Place	Park	Italian Restaurant	Performing Arts Venue	Pizza Place	Theater	Burrito Place
5	Chinatown	52322.61	2	Chinese Restaurant	Bakery	Asian Restaurant	Sushi Restaurant	Coffee Shop	Theater	Sandwich Place	Seafood Restaurant	Hotel Bar	Pizza Place
6	North End	42239.76	2	Italian Restaurant	Seafood Restaurant	Pizza Place	Park	Bakery	Coffee Shop	Sports Bar	Sandwich Place	Spa	Café
13	Beacon Hill	47770.65	2	Pizza Place	Italian Restaurant	Plaza	French Restaurant	Hotel	Lake	Ice Cream Shop	Irish Pub	Kids Store	Korean Restaurant

Cluster 3 which most common is Pizza place, Bakery and another variety of place such as Yoga, Playground, Pharmacy and home service.

```
boston_merged.loc[boston_merged['Cluster Labels'] == 3,boston_merged.columns[[1,4,5,6,7,8,9,10,11,12,13,14,15]]]
```

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	3	Yoga Studio	Big Box Store	Cuban Restaurant	Pool	Rental Car Location	Scenic Lookout	Donut Shop	Pizza Place	American Restaurant	Harc
1	Jamaica Plain	10363.86	3	Bakery	Donut Shop	Art Gallery	Coffee Shop	Pizza Place	Library	Liquor Store	Noodle House	Deli / Bodega	
10	East Boston	8594.21	3	Pizza Place	Convenience Store	Art Gallery	Pharmacy	Latin American Restaurant	Sandwich Place	Bank	Fried Chicken Joint	Mexican Restaurant	Liquor
11	Charlestown	12192.04	3	Convenience Store	Coffee Shop	Pizza Place	Gastropub	Pub	Donut Shop	Yoga Studio	Plaza	Playground	Hi Mu
14	Downtown	5975.80	3	Playground	Pizza Place	Italian Restaurant	Food Truck	History Museum	Park	Other Repair Shop	Optical Shop	Mexican Restaurant	Mediterranean Resta
17	West Roxbury	5640.36	3	Home Service	Yoga Studio	Doctor's Office	Food Court	Food	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dum Resta
18	Hyde Park	7474.90	3	American Restaurant	Pizza Place	ATM	Bank	Ice Cream Shop	Italian Restaurant	Gym	Grocery Store	Fast Food Restaurant	Donut
19	Mattapan	12753.33	3	Pharmacy	Caribbean Restaurant	Home Service	Liquor Store	Pizza Place	Ice Cream Shop	Train Station	Cosmetics Shop	Bakery	Renta Loc
21	South Boston	12192.04	3	Liquor Store	Pizza Place	Cosmetics Shop	Sports Bar	Bar	Coffee Shop	Italian Restaurant	Chinese Restaurant	Historic Site	Sanc

Cluster 4 which most common is Coffee shop, sports bar and Italian restaurant.

```
boston_merged.loc[boston_merged['Cluster Labels'] == 4,boston_merged.columns[[1,4,5,6,7,8,9,10,11,12,13,14,15]]]
```

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	4	Italian Restaurant	Coffee Shop	Park	Bar	Bakery	Wine Shop	French Restaurant	Gift Shop	Wine Bar	Dog Run
9	Back Bay	26282.87	4	Coffee Shop	American Restaurant	Seafood Restaurant	Ice Cream Shop	Italian Restaurant	Juice Bar	Gym / Fitness Center	Dessert Shop	Mediterranean Restaurant	Restaurant
15	Fenway	37978.28	4	Sports Bar	Pizza Place	Coffee Shop	Lounge	Furniture / Home Store	American Restaurant	Thai Restaurant	Baseball Field	Café	Donut Shop

Discussion:

With the above analysis, it is clear that it depend on the type of the restaurant the population also provide for futher consideration. Cluster 1 may suited for Pizza place, café and Korean restaurant while Cluster 2 may suited for Chinese and Italian restaurant. Cluster 3 may suited for Pizza place, Bakery. And cluster 4 for Coffee shop and Italian restaurant.

Limitations:

The data may not be sufficient since the data obtain by online source may be Lack of consistency to maintain the the accuracy.

Conclusion:

In this project, we have gone through the various potential locations for a new restaurant in Boston through population analysis and the use of API to determine the location of existing restaurant. As per our findings, we can conclude that it depend on the type of the restaurant the population also provide for futher consideration. Cluster 1 may suited for Pizza place, café and Korean restaurant while Cluster 2 may suited for Chinese and Italian restaurant. Cluster 3 may suited for Pizza place, Bakery. And cluster 4 for Coffee shop and Italian restaurant.