Restaurant Expansion from Boston to Manhattan and Chicago

IBM Applied Data Science Capstone Project

By Tiffany Brand

August 2019

Introduction

Successful businesses looking to expand may seek similar cities and neighborhoods to open new locations. The type of neighborhood including types and numbers of venues present can contribute to the success of the business' current location, so looking for similar neighborhoods in other cities can help focus the search for business expansion to new areas. Entrepreneur.com in their article "Expanding with a Second Location" suggests that business owners should look at the strengths of their existing location to inform an expansion to a new location. They recommend considering other shops located nearby that may help generate and sustain traffic and performing market research on proposed new locations. The research conducted in this project will focus on looking at the venue makeup of the current neighborhood and finding a similar composition in potential expansion neighborhoods using clustering techniques.

Business Problem

The purpose of this project is to assist a potential client with a successful restaurant business located in Boston to choose possible neighborhoods to open other restaurant locations in Manhattan and Chicago. This project will look at the numbers and types of venues in Boston, Manhattan, and Chicago and use data science clustering methods to find clusters of similar neighborhoods. For the purposes of this project, our stakeholder currently owns a restaurant located in Boston's Back Bay neighborhood and is looking to expand to Manhattan and Chicago. This stakeholder's trendy restaurant thrives in its current location, so they want to find a neighborhood of similar composition to Boston's Back Bay to hopefully replicate the success of their current business type. This will require looking at all types of venues located in neighborhoods of each city. Using data science methods, we will identify several potential expansion neighborhoods in Chicago and Manhattan, looking at areas of similar venue categories and comparable venue density.

Target Audience

A business owner who wishes to expand a successful business in one city to other cities. In this case, our business owner has a restaurant in Boston's Back Bay neighborhood and is looking to expand to Chicago and Manhattan.

Data

Based on the requirements of the business problem, the following types of data will be necessary:

- geographic data defining neighborhoods in Boston, Chicago, and Manhattan
- the number of venues located in neighborhoods in Boston, Chicago, and Manhattan
- category types of venues located in neighborhoods in Boston, Chicago, and Manhattan

Neighborhood boundaries for Boston, Chicago, and Manhattan will be found in:

Boston: The Boston Neighborhoods dataset can be found here: https://data.boston.gov/dataset/boston-neighborhoods - This dataset is found on the Analyze Boston website created by Boston's GIS department. The data is a combination of zoning neighborhood boundaries, zip code boundaries, and 2010 Census tract boundaries. They state that the neighborhood boundaries are not official but can be used in a broad sense for visualization and planning. The GeoJSON version of the data will be downloaded. The following features will be extracted: Neighborhood name and the latitude and longitude data marking the boundaries of the neighborhood. From the boundaries, centroid coordinates of the neighborhoods will be calculated.

Chicago: The Chicago Boundaries – Neighborhoods dataset can be found here: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9 - This dataset is found on the Chicago Data Portal and was developed by the Office of Tourism to describe the approximate neighborhood names of Chicago. The GeoJSON version of the data will be downloaded. The following features will be extracted: Neighborhood name and the latitude and longitude data marking the boundaries of the neighborhood. From the boundaries, centroid coordinates of the neighborhoods will be calculated.

Manhattan: The New York City neighborhoods data can be found here in a JSON file: https://cocl.us/new_york_dataset (this was the dataset used for the week 3 lab in this capstone course). The dataset is found on the NYU Spatial Dataset Repository https://geo.nyu.edu/catalog/nyu_2451_34572. It was created as a guide to New York City's neighborhoods. The following features will be extracted: Neighborhood name, Borough, Latitude and Longitude. Using the borough feature, the neighborhoods from Manhattan will be extracted.

Foursquare Venue Data:

Using the Foursquare API (https://api.foursquare.com), data on nearby venues to each neighborhood will be extracted. Foursquare houses a global database of 105 million places with more than 70 venue attributes and more than 900 categories. Features extracted will be Venue Name, Venue Category, Venue Latitude and Venue Longitude for each Neighborhood Name, Neighborhood Latitude, and Neighborhood Longitude.

Methodology

Data cleaning and preparation

For each city, the geographic information was obtained from the sources listed. For two of the cities, Boston and Chicago, the neighborhoods were defined by shape coordinates detailing the boundaries of each neighborhood. Centroids of each neighborhood were calculated to use as reference points when obtaining venue information from Foursquare. Data from all cities was combined into a single dataframe containing the features: Location, Neighborhood, Latitude, and Longitude. When combining the city data, an additional feature was added which combined the location and neighborhood into one feature since there were several neighborhood names that were the same in different cities (locations). The final dataframe consisted of 164 samples with 5 features.

Next, venues nearby each neighborhood were extracted using the Foursquare API. Nearby venues were defined as the first 100 venues within a radius of 500 meters of the neighborhood centroid. The features extracted were Location/Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, and Venue Category. The data consisted of 7328 samples with 7 features.

Methods for Analysis

In this project seeks to find neighborhoods that are similar to the stakeholder's current restaurant neighborhood of Boston's Back Bay. We looked for similarities based on the numbers and categories of venues in each neighborhood, as well as the density of venues in each neighborhood.

Exploratory Analysis: After data cleaning and preparation, we did some exploratory analysis looking at the numbers of venues returned for each neighborhood. Some neighborhoods, including our stakeholder's current restaurant location of Boston's Back Bay, returned the maximum 100 venues. Other neighborhoods have many fewer venues. We also discovered that there were 364 unique categories of venues that made up our data.

Because density of venues is a factor in the composition of neighborhoods, we narrowed down the list of neighborhoods to consider by looking at only neighborhoods which have a similar density of venues to Boston's Back Bay. Back Bay returned the maximum number of venues we requested, 100. To find neighborhoods with similar density, we extracted just the neighborhoods that had more than 50 venues returned.

Final Analysis: For the final analysis, clusters were created using k-means clustering in order to find related groupings of neighborhoods based on the common categories of venues. One hot encoding was used to describe the venue categories in each neighborhood and then the data was grouped by neighborhood and the mean of the frequency of occurrence of each category. The top ten venues for each neighborhood were added to a dataframe in order of popularity and this data was used for k means clustering. The elbow method was used to determine the optimal number of clusters for analysis, 5. K means clustering was then performed resulting in 5 clusters of neighborhoods which we listed and mapped.

Results

The cluster analysis resulted in 5 different neighborhood groups. The groups consisted of two groups with a large number of neighborhoods, one group with a smaller number of neighborhoods, and two groups containing a single neighborhood each.

First Cluster (0)

	Neighborhood	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
5	Leather District	0	Chinese Restaurant	Bakery	Asian Restaurant	Food Truck	Coffee Shop	Sandwich Place	Café	Sushi Restaurant	Vegetarian / Vegan Restaurant	American Restaurant
6	Chinatown	0	Chinese Restaurant	Asian Restaurant	Bakery	Sushi Restaurant	Theater	Bubble Tea Shop	Coffee Shop	Café	Japanese Restaurant	Restaurant
27	Chinatown	0	Chinese Restaurant	Cocktail Bar	American Restaurant	Salon / Barbershop	Vietnamese Restaurant	Spa	Bubble Tea Shop	Ice Cream Shop	Dim Sum Restaurant	Bakery
45	East Village	0	Bar	Wine Bar	Ice Cream Shop	Mexican Restaurant	Chinese Restaurant	Pizza Place	Ramen Restaurant	Coffee Shop	Cocktail Bar	Vegetarian / Vegan Restaurant
149	Bridgeport	0	Chinese Restaurant	Pharmacy	Bar	Grocery Store	Pizza Place	Mobile Phone Shop	Korean Restaurant	Wings Joint	Gift Shop	Park

Second Cluster (1)

	Neighborhood	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	1	Theater	Hotel	Sandwich Place	Spa	Seafood Restaurant	American Restaurant	Performing Arts Venue	Coffee Shop	Steakhouse	Italian Restaurant
10	Back Bay	1	Coffee Shop	American Restaurant	Italian Restaurant	Seafood Restaurant	Hotel	Ice Cream Shop	Clothing Store	Cosmetics Shop	Gym	Sporting Good Shop
15	Downtown	1	Coffee Shop	Sandwich Place	Italian Restaurant	American Restaurant	Historic Site	New American Restaurant	Hotel	Park	Falafel Restaurant	Salad Place
34	Upper East Side	1	Italian Restaurant	Exhibit	Art Gallery	Coffee Shop	Bakery	Gym / Fitness Center	Juice Bar	Cocktail Bar	French Restaurant	Sp
38	Upper West Side	1	Italian Restaurant	Wine Bar	Bar	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Bakery	Indian Restaurant	Cosmetics Shop	Coffee Shop	Pu
39	Lincoln Square	1	Gym / Fitness Center	Theater	Plaza	Concert Hall	Café	Italian Restaurant	Indie Movie Theater	French Restaurant	Opera House	Par
40	Clinton	1	Theater	Gym / Fitness Center	Italian Restaurant	American Restaurant	Hotel	Coffee Shop	Spa	Wine Shop	Sandwich Place	Loung
41	Midtown	1	Hotel	Clothing Store	Cocktail Bar	Coffee Shop	Theater	Sporting Goods Shop	Spa	Bookstore	Steakhouse	Baker
42	Murray Hill	1	Coffee Shop	Hotel	Sandwich Place	Japanese Restaurant	Italian Restaurant	Bar	French Restaurant	Gym / Fitness Center	Gym	Restaurar
43	Chelsea	1	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Hotel	Theater	American Restaurant	Seafood Restaurant	Art Galler
44	Greenwich Village	1	Italian Restaurant	Clothing Store	Sushi Restaurant	Cosmetics Shop	Chinese Restaurant	Seafood Restaurant	Café	Indian Restaurant	French Restaurant	Gyr
47	Tribeca	1	Italian Restaurant	Café	Park	Boutique	American Restaurant	Spa	Greek Restaurant	Wine Bar	Wine Shop	Coffe Sho
48	Little Italy	1	Bakery	Café	Italian Restaurant	Clothing Store	Bubble Tea Shop	Mediterranean Restaurant	Salon / Barbershop	Sandwich Place	Yoga Studio	Ice Crear Sho
49	Soho	1	Clothing Store	Boutique	Women's Store	Art Gallery	Shoe Store	Italian Restaurant	Furniture / Home Store	Mediterranean Restaurant	Bakery	Sporting Good Sho

50	West Village	1	Italian Restaurant	Cosmetics Shop	New American Restaurant	Wine Bar	Cocktail Bar	Park	American Restaurant	Bakery	Coffee Shop	Speakeas
54	Battery Park City	1	Park	Coffee Shop	Hotel	Memorial Site	Gym	Clothing Store	Italian Restaurant	Wine Shop	Ice Cream Shop	Grocer Stor
55	Financial District	1	Coffee Shop	Steakhouse	Gym	Wine Shop	Hotel	Event Space	Gym / Fitness Center	Cocktail Bar	Pizza Place	America Restaurar
57	Noho	1	Italian Restaurant	French Restaurant	Cocktail Bar	Grocery Store	Art Gallery	Rock Club	Mexican Restaurant	Coffee Shop	Pizza Place	Boutiqu
58	Civic Center	1	Italian Restaurant	Gym / Fitness Center	Sandwich Place	French Restaurant	Hotel	Park	Yoga Studio	Coffee Shop	Bakery	Sp
59	Midtown South	1	Korean Restaurant	Hotel	Cosmetics Shop	Japanese Restaurant	Dessert Shop	Hotel Bar	American Restaurant	Coffee Shop	Boutique	Yog Studi
60	Sutton Place	1	Gym / Fitness Center	Indian Restaurant	Furniture / Home Store	Italian Restaurant	Juice Bar	Dessert Shop	American Restaurant	Pizza Place	Gym	Ice Crear Sho
61	Turtle Bay	1	Italian Restaurant	Steakhouse	Coffee Shop	Sushi Restaurant	Wine Bar	Ramen Restaurant	Park	Café	Indian Restaurant	Japanes Restaurar
64	Flatiron	1	Gym	Gym / Fitness Center	Japanese Restaurant	Yoga Studio	New American Restaurant	American Restaurant	Clothing Store	Café	Salon / Barbershop	Italia Restaurar
65	Hudson Yards	1	American Restaurant	Café	Italian Restaurant	Gym / Fitness Center	Hotel	Spanish Restaurant	Restaurant	Dog Run	Ice Cream Shop	Theate
67	Printers Row	1	Gym / Fitness Center	Pizza Place	Coffee Shop	Clothing Store	Hotel	American Restaurant	Indian Restaurant	Yoga Studio	Furniture / Home Store	Taco Plac
92	Loop	1	Coffee Shop	Italian Restaurant	Sandwich Place	Bakery	Hotel	Vegetarian / Vegan Restaurant	Theater	Middle Eastern Restaurant	Salad Place	Shoe Stor
102	Magnificent Mile	1	Clothing Store	Hotel	American Restaurant	Department Store	New American Restaurant	Pizza Place	Italian Restaurant	Gastropub	Café	Steakhous
111	Rush & Division	1	American Restaurant	Hotel	Steakhouse	Women's Store	Spa	Clothing Store	Coffee Shop	Boutique	New American Restaurant	Seafoo Restaurar
132	Millenium Park	1	Coffee Shop	Hotel	Park	Italian Restaurant	Exhibit	Plaza	Pharmacy	Bakery	Burger Joint	America Restaurar
161	Gold Coast	1	Italian Restaurant	Bar	Café	Gym	Lounge	Hotel	Restaurant	Coffee Shop	Beach	Fast Foo Restaurar

Third Cluster (2)

	Neighborhood	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
7	North End	2	Italian Restaurant	Park	Bakery	Seafood Restaurant	Pizza Place	Wine Shop	Café	Market	Sandwich Place	Coffee Shop

Fourth Cluster (3)

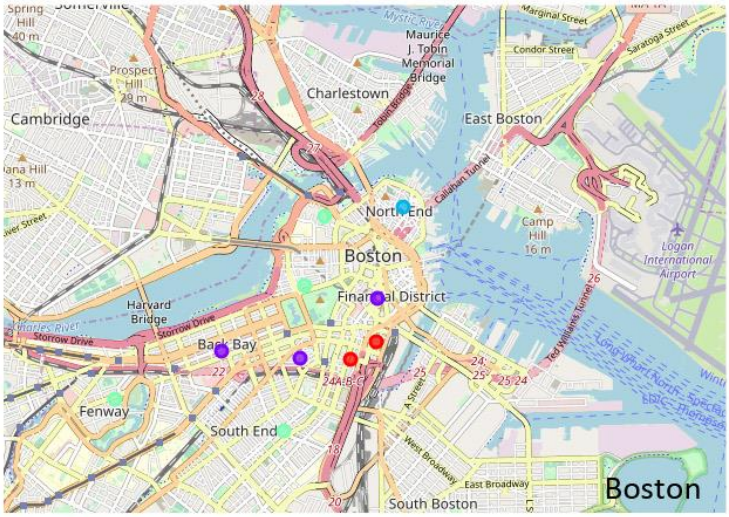
	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	South End	3	Wine Bar	American Restaurant	Italian Restaurant	Mexican Restaurant	Pizza Place	Pet Store	Park	Theater	Café	Gym / Fitness Center
13	West End	3	Sandwich Place	Pizza Place	Hotel	Donut Shop	American Restaurant	Bar	Café	Coffee Shop	Museum	Mexican Restaurant
14	Beacon Hill	3	Pizza Place	Italian Restaurant	Gift Shop	Sandwich Place	Plaza	Playground	Park	French Restaurant	Sushi Restaurant	Gourmet Shop
16	Fenway	3	Sports Bar	American Restaurant	Pizza Place	Coffee Shop	Lounge	Café	Hotel	Baseball Field	Thai Restaurant	Japanese Restaurant
28	Washington Heights	3	Café	Mobile Phone Shop	Bakery	Grocery Store	Coffee Shop	Mexican Restaurant	Park	Latin American Restaurant	Spanish Restaurant	Tapas Restaurant

29	Inwood	3	Café	Mexican Restaurant	Pizza Place	Lounge	Restaurant	Frozen Yogurt Shop	Spanish Restaurant	Chinese Restaurant	Bakery	Park
30	Hamilton Heights	3	Pizza Place	Café	Mexican Restaurant	Coffee Shop	Yoga Studio	Bakery	Sandwich Place	School	Sushi Restaurant	Deli / Bodega
35	Yorkville	3	Italian Restaurant	Gym	Coffee Shop	Bar	Pizza Place	Sushi Restaurant	Wine Shop	Japanese Restaurant	Mexican Restaurant	Deli / Bodega
36	Lenox Hill	3	Coffee Shop	Italian Restaurant	Pizza Place	Sushi Restaurant	Sporting Goods Shop	Cosmetics Shop	Burger Joint	Gym	Gym / Fitness Center	Wine Shop
46	Lower East Side	3	Coffee Shop	Café	Pizza Place	Ramen Restaurant	Cocktail Bar	Art Gallery	Sandwich Place	Bakery	Chinese Restaurant	Japanese Restaurant
51	Manhattan Valley	3	Pizza Place	Coffee Shop	Mexican Restaurant	Indian Restaurant	Szechuan Restaurant	Yoga Studio	Deli / Bodega	Playground	Café	Thai Restaurant
53	Gramercy	3	Bar	Italian Restaurant	Bagel Shop	American Restaurant	Pizza Place	Hotel	Ice Cream Shop	Thrift / Vintage Store	Cocktail Bar	Coffee Shop
56	Carnegie Hill	3	Pizza Place	Coffee Shop	Cosmetics Shop	Café	Bakery	French Restaurant	Spa	Wine Shop	Yoga Studio	Japanese Restaurant
62	Tudor City	3	Park	Café	Mexican Restaurant	Pizza Place	Greek Restaurant	Deli / Bodega	Diner	Restaurant	Sushi Restaurant	Garden
100	Lake View	3	Sandwich Place	Sports Bar	Bar	Sporting Goods Shop	Pizza Place	American Restaurant	Mexican Restaurant	Bakery	Japanese Restaurant	Video Game Store
101	Lincoln Park	3	Bar	Coffee Shop	Sandwich Place	Pizza Place	Mexican Restaurant	Hot Dog Joint	Theater	Indonesian Restaurant	Vietnamese Restaurant	Thai Restaurant
106	West Loop	3	Greek Restaurant	Coffee Shop	Café	Convenience Store	Sandwich Place	Spa	Grocery Store	Sports Bar	Pizza Place	History Museum
108	Andersonville	3	Coffee Shop	Italian Restaurant	Grocery Store	Sandwich Place	Café	Middle Eastern Restaurant	Lounge	Sushi Restaurant	Korean Restaurant	Burger Joint
122	Uptown	3	Coffee Shop	Bus Station	Mexican Restaurant	Pizza Place	Diner	Sandwich Place	Sushi Restaurant	Bar	Mobile Phone Shop	Music Venue
133	Near South Side	3	Park	Pizza Place	Historic Site	Gym	American Restaurant	Bar	Caribbean Restaurant	Rock Club	Coffee Shop	Sandwich Place
143	Wrigleyville	3	General Entertainment	Sports Bar	Pizza Place	American Restaurant	Baseball Stadium	Salon / Barbershop	Hot Dog Joint	New American Restaurant	Bar	Snack Place
147	East Village	3	Mexican Restaurant	Pizza Place	Grocery Store	Café	Bar	Bakery	Coffee Shop	Greek Restaurant	Gym	Korean Restaurant
154	Wicker Park	3	Pizza Place	Clothing Store	Boutique	Bar	Coffee Shop	Bookstore	Shoe Store	Gourmet Shop	Sushi Restaurant	Frozen Yogurt Shop
158	Old Town	3	Comedy Club	Bar	Sandwich Place	Mexican Restaurant	Italian Restaurant	Gym	Boutique	Dessert Shop	American Restaurant	Theater
162	Boystown	3	Gay Bar	Sushi Restaurant	Coffee Shop	Mexican Restaurant	Spa	Dive Bar	Japanese Restaurant	Gym	New American Restaurant	BBQ Joint

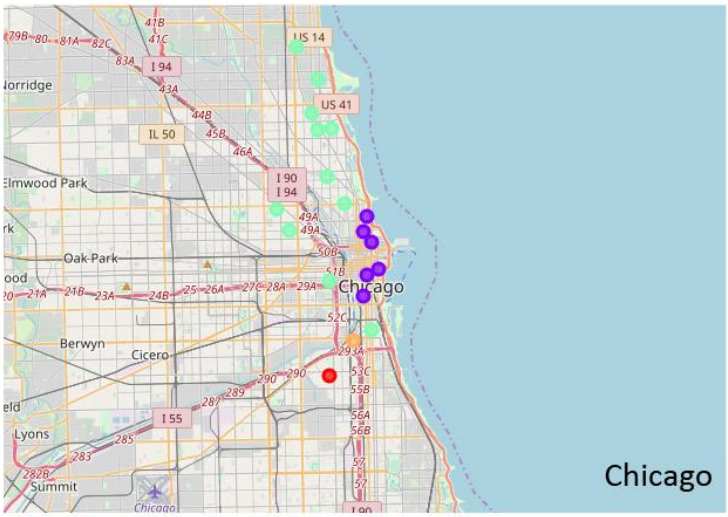
Fifth Cluster (4)

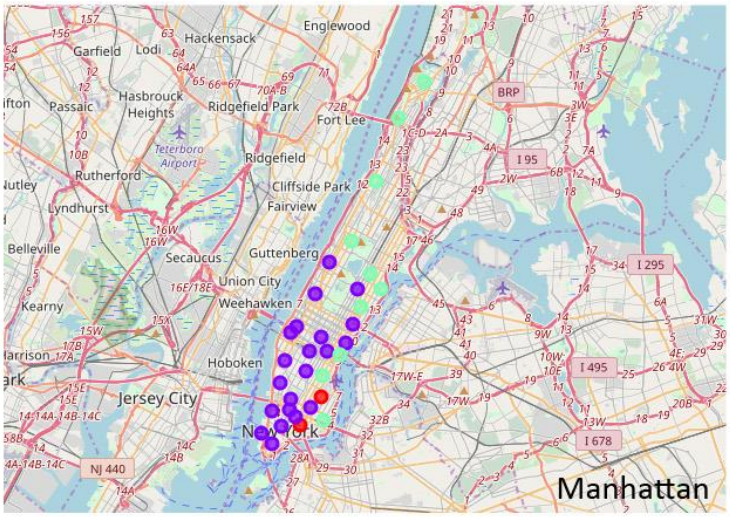
	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
134	Chinatown	4	Chinese Restaurant	Asian Restaurant	Korean Restaurant	Dessert Shop	Bubble Tea Shop	Seafood Restaurant	Bakery	Dim Sum Restaurant	Tea Room	Sculpture Garden

Clusters Mapped:



- First Cluster (0)
- Second Cluster (1)
- Third Cluster (2)
- Fourth Cluster (3)
- Fifth Cluster (4)





When viewing the clusters on a map, one large cluster was concentrated in the centers of the cities, and the other large cluster had neighborhoods in the centers, but also in the more outer areas.

Discussion

In looking at the venue composition of the clusters and their map locations, we observed certain types of venues in the clusters, and also whether the neighborhoods were in the city centers or had more outlying neighborhoods. Two of the clusters only had one neighborhood in them, so we characterized them as being unique. Keeping this in mind, the following labels were chosen for the clusters:

- First Cluster (0) - Asian Restaurants
- Second Cluster (1) - Mixed Variety Restaurants, Hotels, Theater, City Center
- Third Cluster (2) - Unique Italian Neighborhood
- Fourth Cluster (3) - Quick Food - Cafes, Coffee Shops & Pizza
- Fifth Cluster (4) - Unique Chinatown Neighborhood

The model group for our stakeholder is the second cluster (labelled cluster 1). Our stakeholder's current restaurant is located in Boston's Back Bay neighborhood, which is grouped into this cluster. In looking at the venue composition of that cluster, we see a diverse selection of restaurants, hotels, and theaters. When looking at the maps of this cluster, we see the neighborhoods are concentrated in the city centers. Our recommendation would be for our stakeholder to investigate other neighborhoods in this cluster located in Manhattan and Chicago to pursue further inquiry into finding a location for their restaurant expansion.

Recommended Neighborhoods for Restaurant Expansion

Chicago	Printers Row
	Loop
	Magnificent Mile
	Rush & Division
	Millennium Park
	Gold Coast
Manhattan	Upper East Side
	Upper West Side
	Lincoln Square
	Clinton
	Midtown
	Murray Hill
	Chelsea
	Greenwich Village
	Tribeca
	Little Italy
	Soho
	West Village
	Battery Park City

	Financial District
	Noho
	Civic Center
	Midtown South
	Sutton Place
	Turtle Bay
	Flatiron
	Hudson Yards

Conclusion

The purpose of this project was to identify areas in Chicago and Manhattan with venue composition and density comparable to our stakeholder's current restaurant neighborhood in Boston so that they might find potential locations for restaurant expansion. We found data defining neighborhoods in Chicago, Manhattan, and Boston, and used Foursquare data to look at venue composition and density. Using clustering techniques on the data from the three cities, we defined clusters of neighborhoods and identified a target group of neighborhoods for our stakeholder. This target cluster of neighborhoods was characterized by a diverse composition of restaurants, hotels, and theaters located near the city centers. Our stakeholder will be able to use this target cluster of neighborhoods in Chicago and Manhattan to look for potential locations for new restaurants.

References

"Expanding With a Second Location." *Entrepreneur*, 23 Feb. 2003, www.entrepreneur.com/article/47552.