

# Capstone Project: Evaluation of Boston Neighborhoods for New Pizza Store Locations

## Introduction

A pizza company, Bob's Pizza, that is currently well established in the Mid-Atlantic states, is interested in expanding into the Northeast. The company has chosen Boston as a location for their initial expansion. Our company, XYZ Data Analytics, has been hired to identify five locations within the Boston city limits where the company can open its pizza stores.

In order to maximize the probability that these five locations will be the best possible locations, in terms of profitability, for the new stores, Bob's Pizza has provided a set of guidelines that have proven successful in other cities where they have established stores. These guidelines are:

- Target the neighborhoods that have existing fast food stores, including coffee shops, but are underserved by pizza shops.
- Do not target neighborhoods that have a high presence of pizza stores already.
- Target neighborhoods that have a higher concentration of hotels.

## Description of Data

Data on Boston neighborhoods can be downloaded from the following link:

[http://bostonopendata-boston.opendata.arcgis.com/datasets/3525b0ee6e6b427f9aab5d0ald0ala28\\_0.csv](http://bostonopendata-boston.opendata.arcgis.com/datasets/3525b0ee6e6b427f9aab5d0ald0ala28_0.csv)

We are primarily interested in the Boston neighborhood name, which will be used as an input to the geocoder API to determine the coordinates for that neighborhood.

With the coordinates of each neighborhood, the FourSquare API will be queried to determine the concentration of various types of venues within a certain perimeter. This data will be analyzed to identify the best five locations for new pizza stores.

## Methodology

The following steps were taken to prepare the data for analysis:

- The raw Boston neighborhoods data was cleaned to remove unnecessary or unusable data as follows:

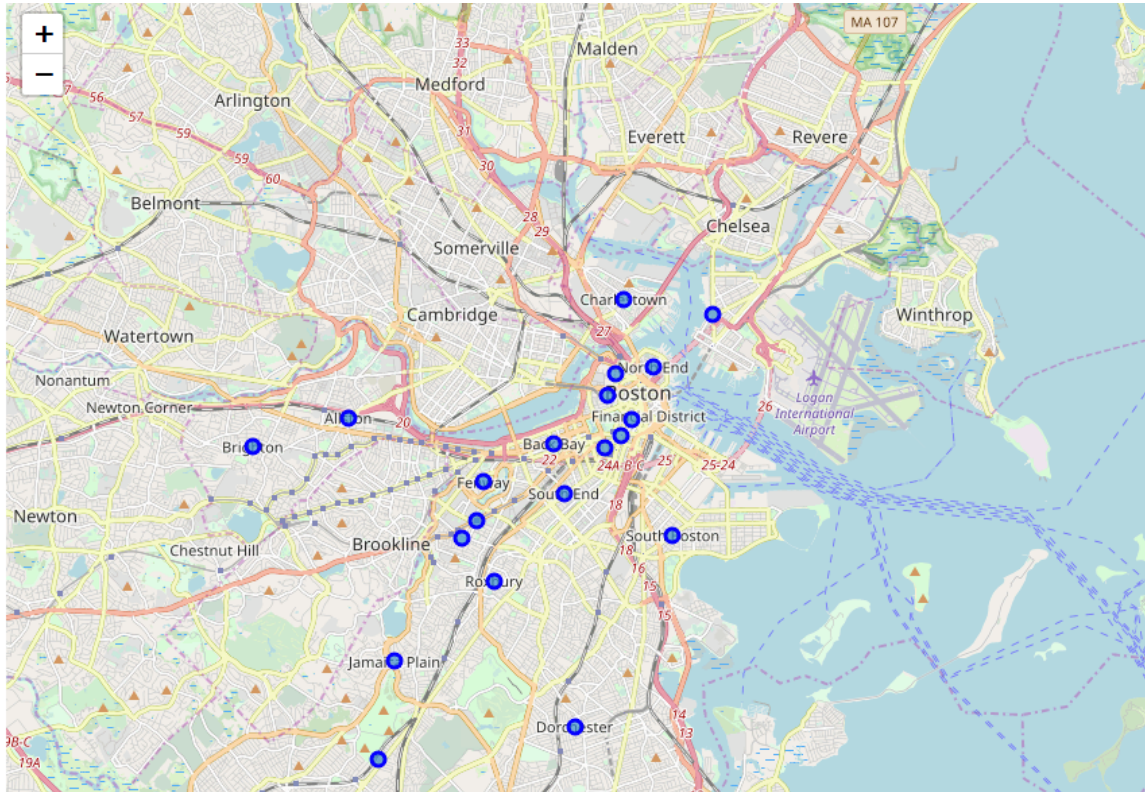
A couple of Boston neighborhoods are not known to the geocoder API and are thus removed from the data set. These neighborhoods are the Leather District and Harbor Islands.

A couple of neighborhoods, South Boston and South Boston Waterfront, produce the same coordinates via the geocoder API and so, South Boston Waterfront is removed from the data set.

Columns other than 'OBJECTID' and 'Name' are removed from the data set. The 'OBJECTID' column is kept so that the 'Name' column does not contain the index value which would interfere with the geocoder queries as the neighborhood name would be something like '1 Roslindale'.

- The Boston neighborhood coordinates were retrieved with queries to the geocoder API using the neighborhood name from the data set's 'Name' column. Two new columns were added to the data set for the latitude and longitude of each neighborhood.

The following map of Boston shows the Boston neighborhoods mapped using their coordinates.



With the Boston neighborhood names and coordinates available, we queried the FourSquare API for the concentration of various types of venues within a 500 meter radius around the neighborhood coordinate.

Our end goal is to identify the five best neighborhoods to locate the new pizza stores. Bob's pizza has recommended the certain criteria for selecting these best locations based on past market research in other cities.

Neighborhoods that already support fast food stores other than pizza stores are considered preferable locations. A large number of fast food stores would suggest that there is sufficient demand for fast food at these locations. A low concentration of pizza stores would suggest that there may be more demand than supply of pizza in that neighborhood.

Neighborhoods that already have a high concentration of pizza stores are not considered preferable locations because of the loyalty that customers typically show to their existing pizza brand.

Finally, the company's market research has shown that pizza locations in the vicinity of hotels usually perform better than those not located near hotels, all other factors being the same. The convenience and economy of ordering pizza for a meal is valued by many hotel guests.

With the above criteria in mind, we are interested in three venue types from the FourSquare data set.

- Pizza Stores
- Fast Food Stores (Coffee Shops, Sandwich Place, Donut Shops)
- Hotels

We initially collect a sample data set of the top 10 popular venues for each Boston neighborhood. This data set is obtained by performing a one-hot encoding on the raw data from the FourSquare API which consists of one line for each venue. Then the venues are grouped together by neighborhood and a mean value is determined to indicate the concentration of the venue type in that neighborhood as shown in the following table.

	Name	Accessories Store	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	...	Tour Provider	Tourist Information Center	Traffic
0	Allston	0.000000	0.000000	0.000000	0.012195	0.00	0.000000	0.024390	0.000000	0.012195	...	0.00	0.00	0.0
1	Back Bay	0.020000	0.050000	0.000000	0.010000	0.00	0.000000	0.000000	0.010000	0.000000	...	0.01	0.00	0.0
2	Bay Village	0.000000	0.032609	0.000000	0.000000	0.00	0.01087	0.021739	0.000000	0.000000	...	0.00	0.00	0.0
3	Beacon Hill	0.000000	0.028986	0.000000	0.000000	0.00	0.000000	0.000000	0.014493	0.000000	...	0.00	0.00	0.0
4	Brighton	0.000000	0.020000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.00	0.0
5	Charlestown	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.027027	0.000000	...	0.00	0.00	0.0
6	Chinatown	0.000000	0.020000	0.000000	0.000000	0.00	0.000000	0.080000	0.000000	0.000000	...	0.00	0.01	0.0
7	Dorchester	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.00	0.0
8	Downtown	0.000000	0.040000	0.000000	0.000000	0.00	0.000000	0.020000	0.010000	0.000000	...	0.00	0.01	0.0
9	East Boston	0.000000	0.028571	0.000000	0.028571	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.00	0.0
10	Fenway	0.000000	0.030000	0.000000	0.010000	0.00	0.01000	0.000000	0.010000	0.000000	...	0.01	0.00	0.0
11	Hyde Park	0.000000	0.142857	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.00	0.0
12	Jamaica Plain	0.050000	0.050000	0.000000	0.100000	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.00	0.0
13	Longwood	0.000000	0.000000	0.000000	0.040000	0.04	0.000000	0.000000	0.000000	0.000000	...	0.00	0.00	0.0

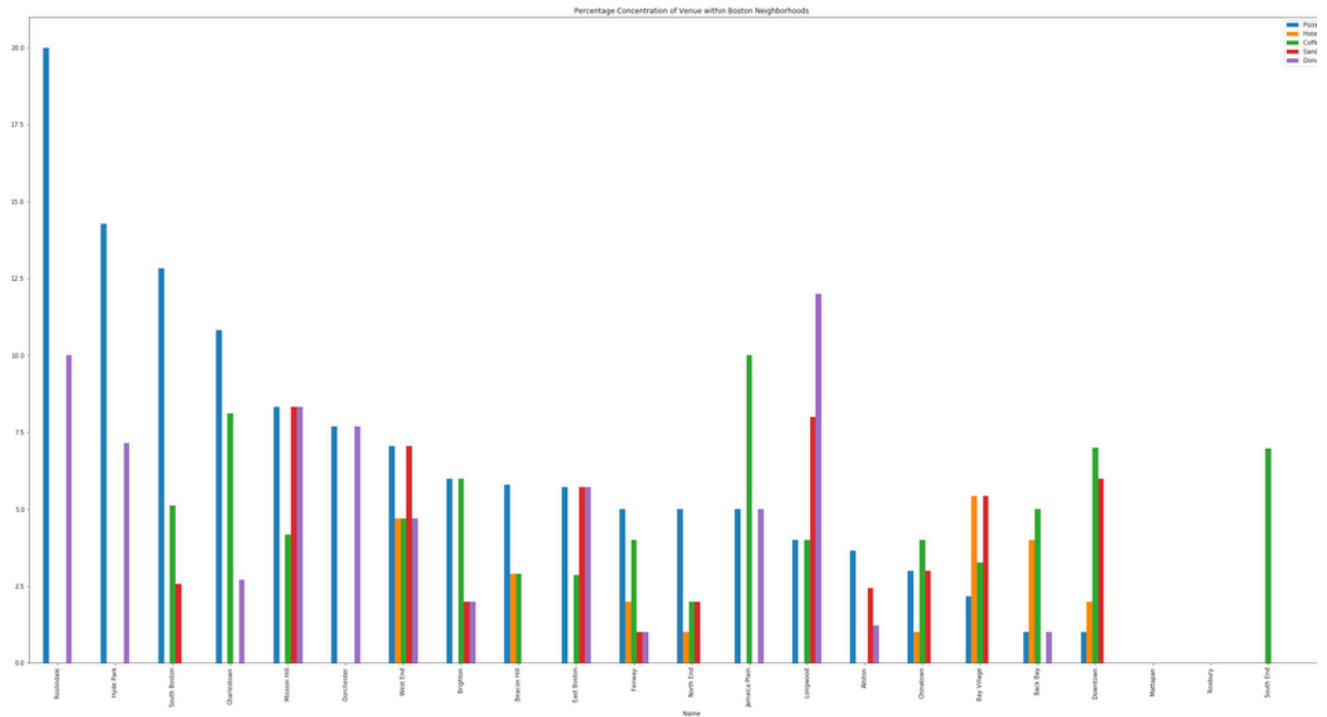
This data set is then converted to a top 10 venues data set with a row for each neighborhood and 10 columns for each of the top 10 places. The final data set looks like the following table.

	Neighbourhood	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	Allston	Korean Restaurant	Bakery	Pizza Place	Vegetarian / Vegan Restaurant	Chinese Restaurant	Thai Restaurant	Sandwich Place	Bubble Tea Shop	Pharmacy	Italian Restaurant
1	Back Bay	American Restaurant	Coffee Shop	Italian Restaurant	Seafood Restaurant	Clothing Store	Hotel	Cosmetics Shop	Spa	Sporting Goods Shop	Ice Cream Shop
2	Bay Village	Sandwich Place	Hotel	Italian Restaurant	Seafood Restaurant	Bakery	Theater	Steakhouse	Chinese Restaurant	Spa	Coffee Shop
3	Beacon Hill	Pizza Place	Italian Restaurant	Hotel Bar	Playground	Museum	Coffee Shop	French Restaurant	Plaza	Hotel	Sushi Restaurant
4	Brighton	Coffee Shop	Bakery	Pub	Chinese Restaurant	Bus Station	Bank	Pizza Place	Grocery Store	Dry Cleaner	Deli / Bodega
5	Charlestown	Pizza Place	Coffee Shop	Liquor Store	Gastropub	Pub	Convenience Store	Italian Restaurant	Mobile Phone Shop	Chinese Restaurant	Plaza
6	Chinatown	Chinese Restaurant	Asian Restaurant	Bakery	Coffee Shop	Sushi Restaurant	Theater	Pizza Place	Sandwich Place	Performing Arts Venue	Seafood Restaurant

7	Dorchester	Market	Donut Shop	Caribbean Restaurant	Diner	Breakfast Spot	Nail Salon	Plaza	Café	Pizza Place	Vegetarian / Vegan Restaurant
8	Downtown	Coffee Shop	Sandwich Place	American Restaurant	Falafel Restaurant	New American Restaurant	Restaurant	Historic Site	Gym / Fitness Center	Chinese Restaurant	Hotel
9	East Boston	Latin American Restaurant	Mexican Restaurant	Convenience Store	Park	Pharmacy	Pizza Place	Sandwich Place	Donut Shop	Seafood Restaurant	Market
10	Fenway	Sports Bar	Pizza Place	Lounge	Coffee Shop	Thai Restaurant	Café	Baseball Field	Restaurant	American Restaurant	Furniture / Home Store
11	Hyde Park	American Restaurant	Pizza Place	Bar	Ice Cream Shop	Discount Store	Italian Restaurant	Platform	Theater	Pharmacy	Donut Shop
12	Jamaica Plain	Coffee Shop	Bakery	Art Gallery	Library	Liquor Store	Donut Shop	Deli / Bodega	Noodle House	Park	Pet Store
13	Longwood	Donut Shop	Sandwich Place	Italian Restaurant	Ice Cream Shop	Coffee Shop	Pub	Rental Car Location	Pizza Place	Sushi Restaurant	Liquor Store
14	Mattapan	Yoga Studio	Bakery	Convenience Store	Pharmacy	Caribbean Restaurant	Mobile Phone Shop	Shoe Store	Southern / Soul Food Restaurant	Food & Drink Shop	Fast Food Restaurant
15	Mission Hill	Pizza Place	Donut Shop	Sushi Restaurant	Café	Sandwich Place	Gastropub	Caribbean Restaurant	Italian Restaurant	Bar	Convenience Store
16	North End	Italian Restaurant	Pizza Place	Park	Bakery	Wine Shop	Seafood Restaurant	Café	Grocery Store	Market	Coffee Shop
17	Roslindale	Pizza Place	Rental Car Location	American Restaurant	Scenic Lookout	Donut Shop	Cycle Studio	Plaza	Pool	Big Box Store	Diner
18	Roxbury	Art Gallery	Gym	Plaza	Rental Car Location	Metro Station	Furniture / Home Store	Yoga Studio	Discount Store	Event Space	Electronics Store
19	South Boston	Pizza Place	Italian Restaurant	Bar	Coffee Shop	Sports Bar	Convenience Store	Park	Nightclub	Restaurant	Chinese Restaurant
20	South End	Italian Restaurant	Coffee Shop	Wine Shop	Wine Bar	Gift Shop	Park	Bakery	Bar	Mediterranean Restaurant	Jazz Club
21	West End	Pizza Place	Sandwich Place	Coffee Shop	Bar	Donut Shop	Hotel	Café	Italian Restaurant	Gym	Mexican Restaurant
22	West Roxbury	Health & Beauty Service	Yoga Studio	Discount Store	Farmers Market	Falafel Restaurant	Event Space	Electronics Store	Dumpling Restaurant	Dry Cleaner	Donut Shop

Based on the market research determined criteria that we outlined earlier, visually examining the above table reveals three neighborhoods that fall in the highly preferable category of a large concentration of fast food stores, a low concentration of pizza stores and a high concentration of hotels. These neighborhoods are Backbay, Bay Village and Downtown.

To validate this finding and to determine the remaining two locations, we visualize the venues concentration data in a bar chart using just five venues, pizza stores, hotels, coffee shops, sandwich places, and donut shops. The bar chart is show below.



The above bar chart is based on all values of the five chosen categories. Hence, it can be considered more accurate than the top 10 venues data set.

In the bar chart, we see that Back Bay, Bay Village and Downtown again contain the best combination of high concentration of fast food stores and hotels and a low concentration of pizza stores.

Other locations that meet the criteria somewhat are West End, Beacon Hill, Fenway and Longwood. The selection of the final two locations will be discussed in the Results section.

## Results

As discussed in the previous section, Back Bay, Bay Village and Downtown met the criteria recommended by the client quite well. So, they will be three of the five locations.

The fourth location chosen is West End which has a higher hotel and fast food concentration relative to the

concentration of pizza stores than other unselected neighborhoods.

For the fifth location, Longwood was chosen. This neighborhood has almost no hotel presence, but the concentration of fast food venues is much higher than the concentration of pizza stores. We expect there will be sufficient demand for a new pizza store over and above the existing demand for pizzas at this location.

The chosen neighborhoods for new Bob's Pizza stores are:

Back Bay

Bay Village

Downtown

West End

Longwood

## Discussion

The ability to use one-hot encoding was critical in processing the data into a form where it was easy to visualize the relative concentrations of venue types for each neighborhood.

Additional categories such as demographics could also be taken into consideration, but we believe existing venues are a better indicator of what works in the neighborhood.

## Conclusion

Boston is a thriving city with many growing neighborhoods. Our analysis was based on data concerning existing businesses alone. Our selection of the five locations for the new stores was guided by the market research determined criteria provided by the client.

We have high confidence in the accuracy of the data and as such, we believe that the proposed five locations are the best options to locate the new Bob's Pizza stores.



