

# Restaurant diversity in Boston neighborhoods

IBM Data Science Professional Certificate  
Capstone Project

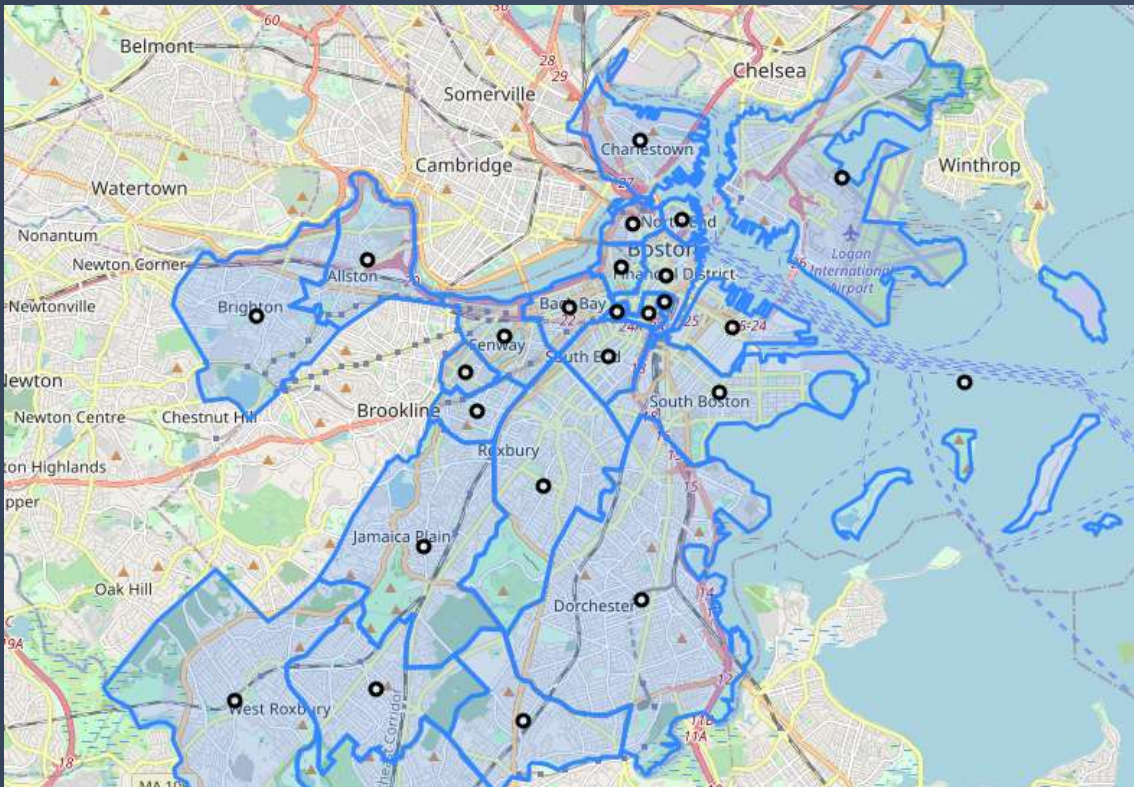
Brian Schaefer  
April 17, 2020

# Objective: Find the neighborhoods in Boston with the most diverse restaurants



Where should someone visit or live if they want access to diverse cuisines, like Taiwanese dumplings and Southern fried chicken and waffles?

# Boston Neighborhoods

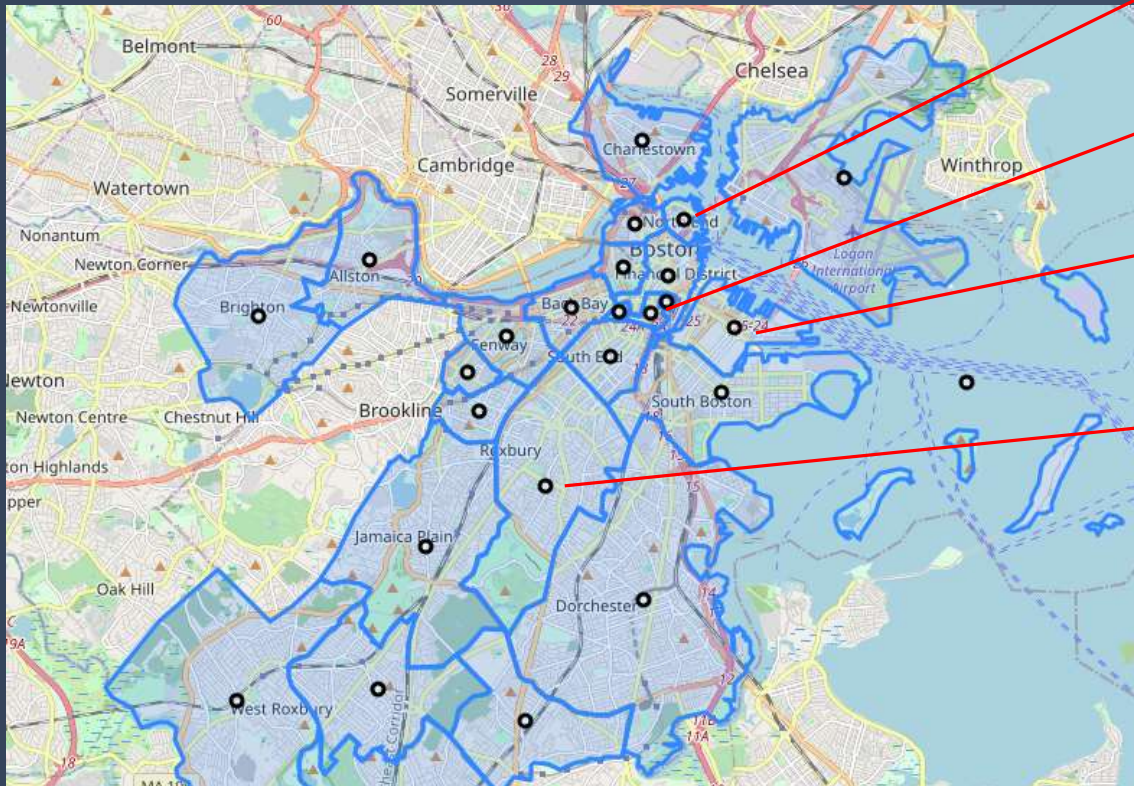


- Geographical data (GeoJSON format):  
<https://data.boston.gov/dataset/boston-neighborhoods>
- Centroids (black) calculated from polygon boundaries



# Boston Neighborhoods

Some neighborhoods are known for a particular cuisine:



North End



Chinatown



Seaport District



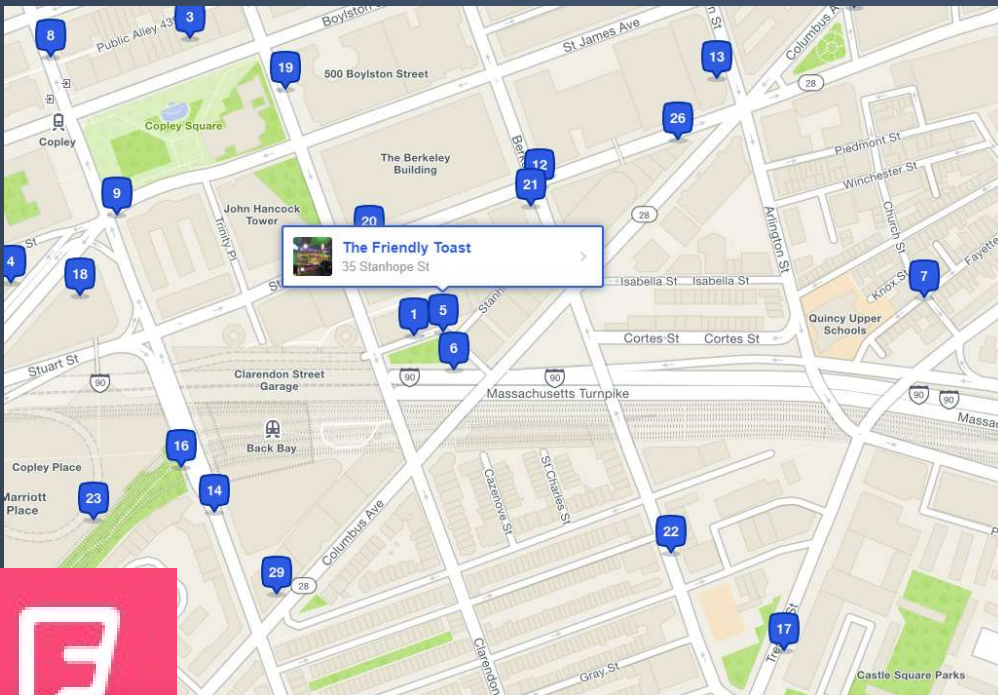
Roxbury

Could this neighborhood have more diverse cuisine?

- Geographical data (GeoJSON format):  
<https://data.boston.gov/dataset/boston-neighborhoods>
- Centroids (black) calculated from polygon boundaries

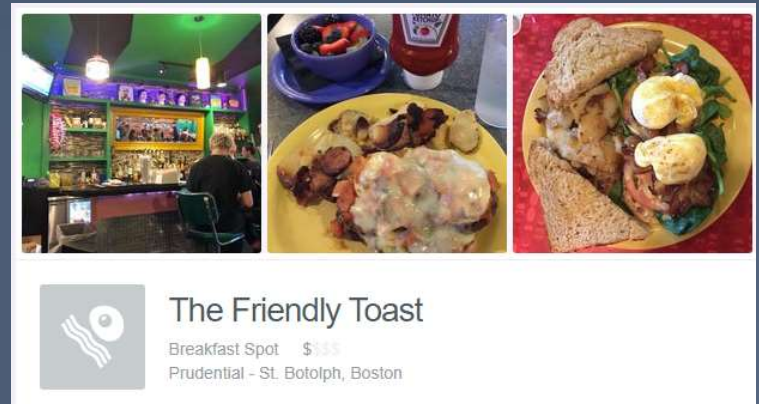
# Exploring restaurants

Use the Foursquare API to search for “Food” venues around the geographical centroid of each neighborhood



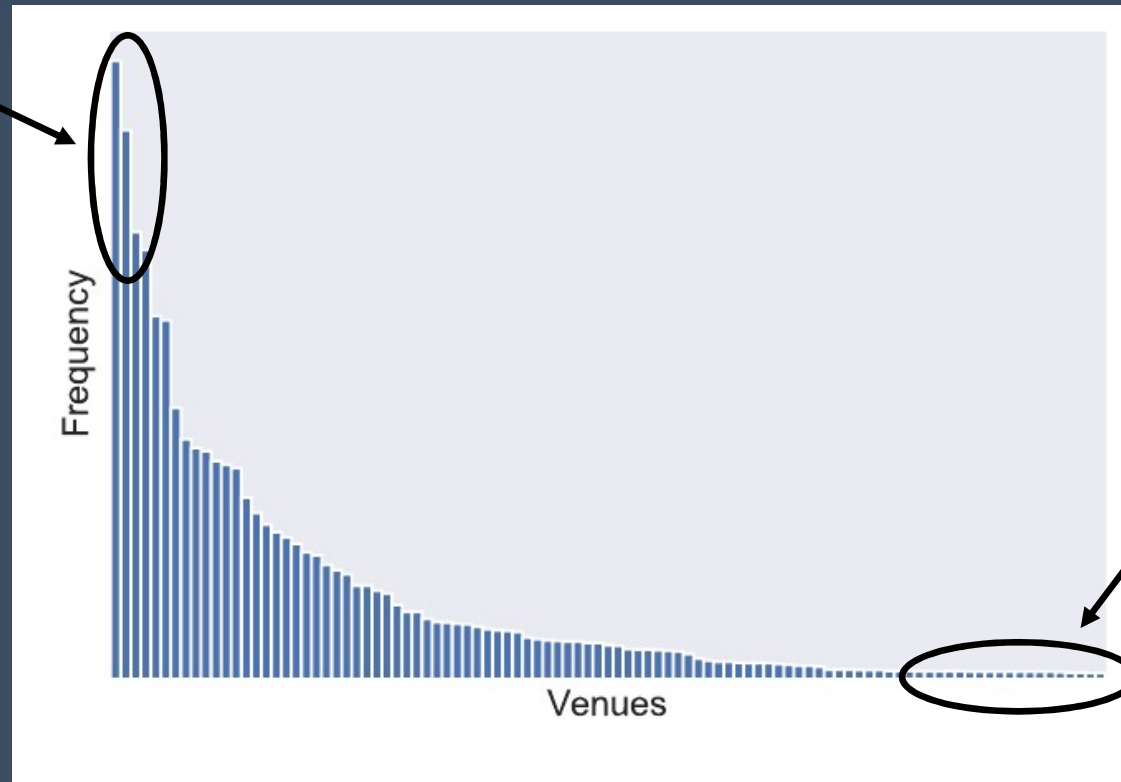
Information for each venue:

- Venue name (“The Friendly Toast”)
- Venue category (“Breakfast Spot”)
- Geographical coordinates
- Address



# Distribution of venue types in Boston

- Pizza place
- Coffee shop
- Café



- Creperie
- Noodle House
- Wings Joint
- ...

# Diversity index

Define a “diversity index”

- Accounts for neighborhoods that might have a large number of restaurants
- Does not count multiple restaurants of the same type
- A useful metric for determining the diversity of restaurants in a neighborhood

$$\text{Diversity} = \frac{\text{Number of unique restaurant types}}{\text{Total number of restaurants}}.$$

# Data classification

## $k$ -means clustering

- Assignment of neighborhoods into clusters based on similarity of their feature vectors
  - Feature vectors consist of information related to the number of restaurants of each type
- Use “elbow” method to determine a useful number of clusters ( $k=3$ )
  - In the end, these clusters could be labeled “high diversity”, “moderate diversity”, and “low diversity”



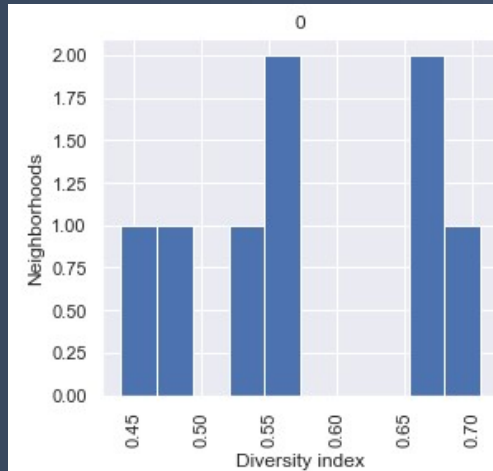


# Clustering with restaurant counts

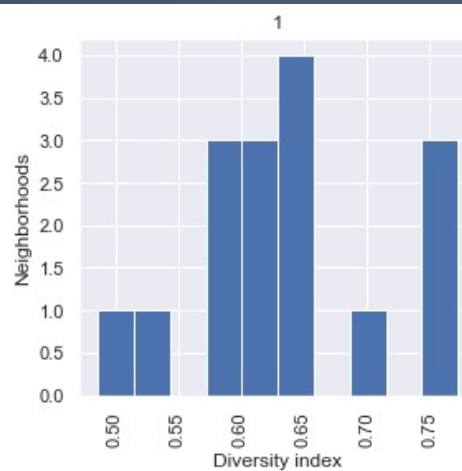
Data: counts of numbers of each type of restaurant in the neighborhood

	neighborhood	African Restaurant	American Restaurant	Asian Restaurant	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	Breakfast Spot
0	Allston	0	0	0	0	0	0	1	0
1	Back Bay	0	3	0	0	0	0	2	1
2	Bay Village	0	0	0	0	0	0	0	0
3	Beacon Hill	0	1	0	0	0	0	1	1
4	Brighton	0	1	0	0	0	0	0	1

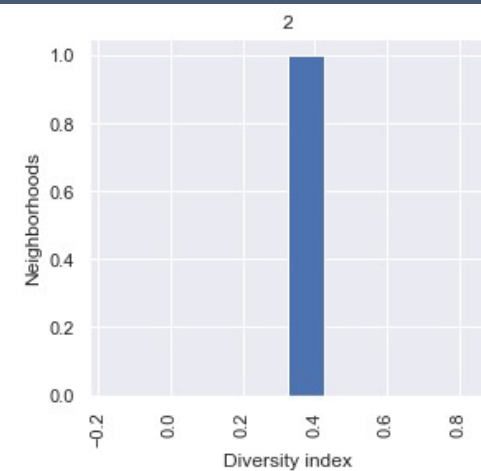
Cluster 0: Pizza places, cafés



Cluster 1: A variety of restaurants



Cluster 2: North End (many, many Italian restaurants) restaurants

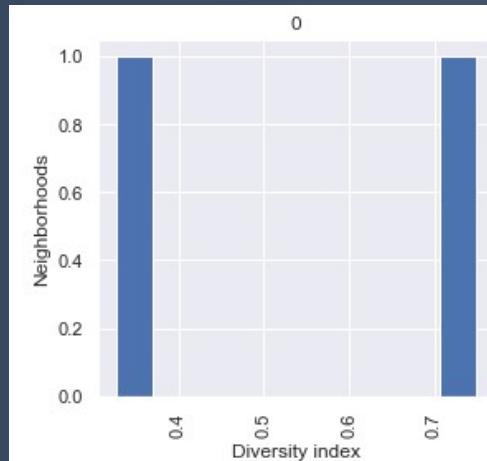


# Clustering with restaurant proportions

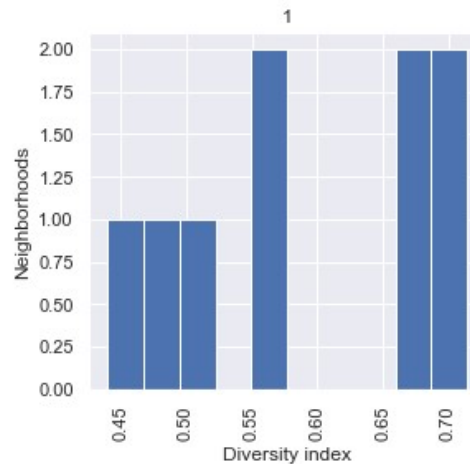
Data: proportion of each type of restaurant out of the total number of restaurants in the neighborhood

	neighborhood	African Restaurant	American Restaurant	Asian Restaurant	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	Breakfast Spot
0	Allston	0.0	0.000000	0.0	0.0	0.0	0.0	0.026316	0.000000
1	Back Bay	0.0	0.081081	0.0	0.0	0.0	0.0	0.054054	0.027027
2	Bay Village	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.000000
3	Beacon Hill	0.0	0.052632	0.0	0.0	0.0	0.0	0.052632	0.052632
4	Brighton	0.0	0.033333	0.0	0.0	0.0	0.0	0.000000	0.033333

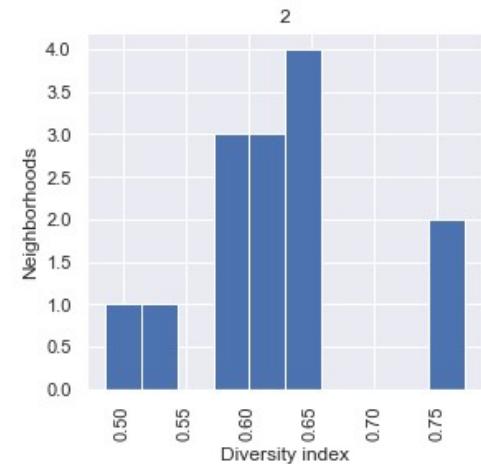
Cluster 0: Italian most popular



Cluster 1: Pizza places, cafés



Cluster 2: A variety of restaurants



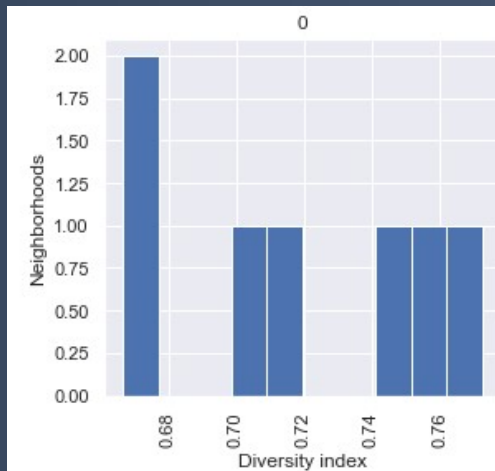
# Clustering with diversity index

Data: diversity index for each neighborhood

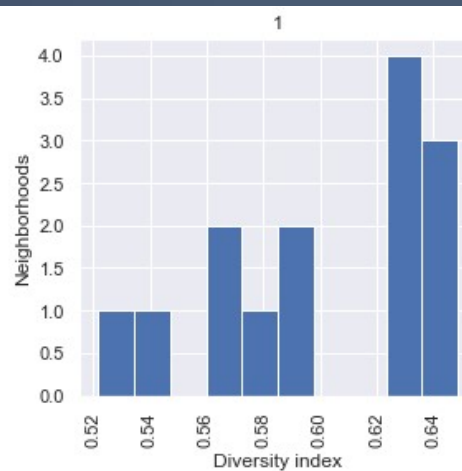
$$\text{Diversity} = \frac{\text{Number of unique restaurant types}}{\text{Total number of restaurants}}$$

	neighborhood	diversity
0	Allston	0.634146
1	Back Bay	0.585366
2	Bay Village	0.750000
3	Beacon Hill	0.714286
4	Brighton	0.666667

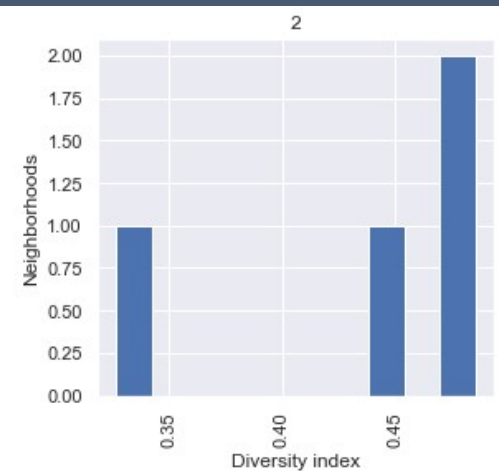
Cluster 0: Most diverse



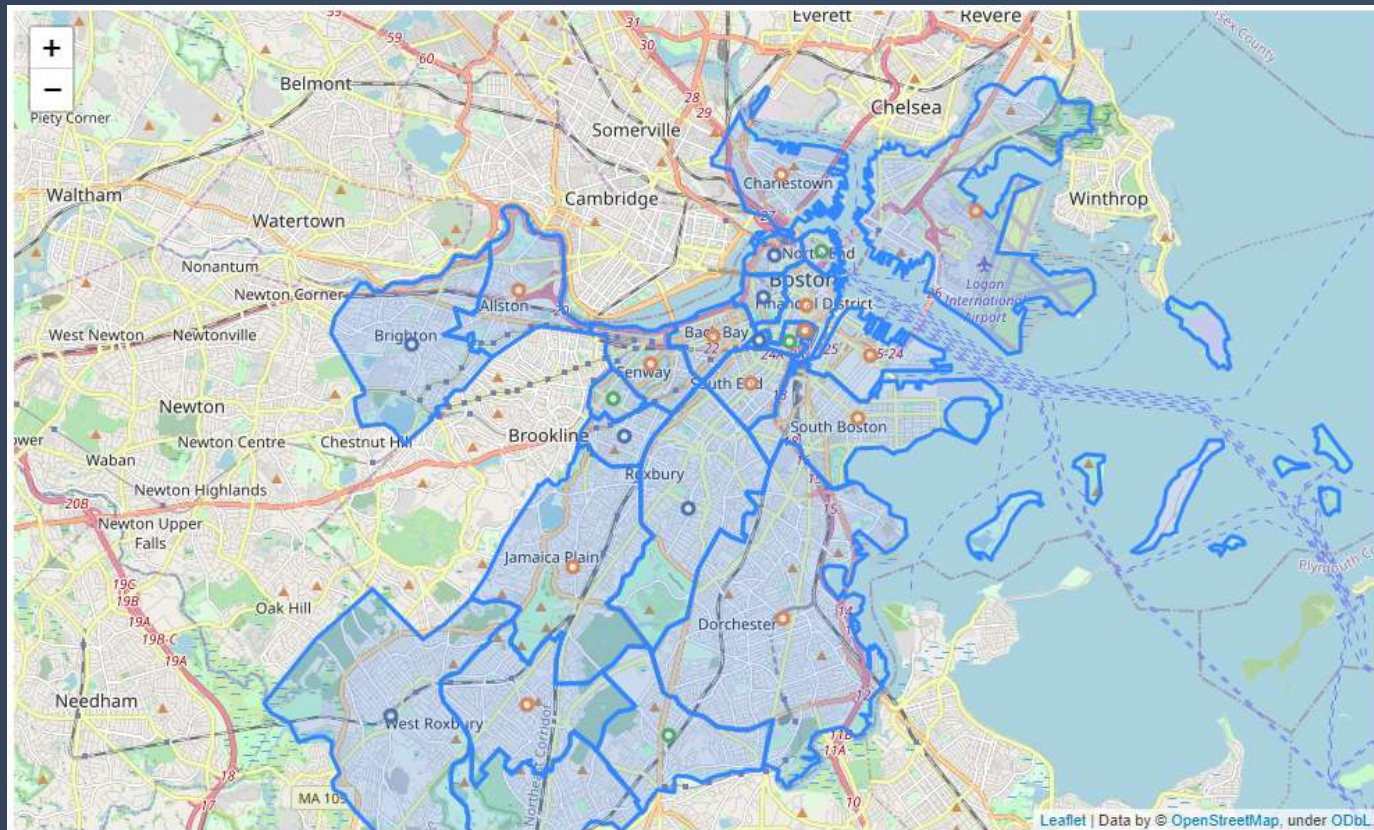
Cluster 1: Moderately diverse



Cluster 2: Least diverse



# Final clustering



## Least diverse

- Chinatown (Chinese)
- North End (Italian)
- Longwood (a medical campus)
- Mattapan (Afro-Caribbean)

## Moderately diverse

## Most diverse

- Bay Village
- Beacon Hill
- Brighton
- Mission Hill
- Roxbury
- West End
- West Roxbury



# Conclusions

Used Foursquare data and “diversity index” to classify Boston neighborhoods on variety of restaurants

Great neighborhoods for the adventurous palate:

- Bay Village
- Beacon Hill
- Brighton
- Mission Hill
- Roxbury
- West End
- West Roxbury

