

Segmenting and Clustering Towns in Massachusetts Based on Venues

Vivian Xia

January 4, 2021

1. Introduction

1.1 Background

Massachusetts is the 15th most populous state in the United States. It is also home to over one hundred colleges and universities including Harvard University, MIT, Tufts University, and my own undergraduate college, Bentley University. Many college students end up residing in Massachusetts after completing their education. Therefore, it would be useful information to understand the many towns of Massachusetts.

1.2 Problem

This project aims to segment and cluster similar towns based on venues in Massachusetts. The types of venues that are frequented the most in each town are used to cluster towns with one another. The project will locate towns that are similar to other towns and their amenities to allow for a more informed decision on where to move to or stay.

2. Data acquisition and cleaning

2.1 Data sources

The list of towns in Massachusetts was scraped from a [Wikipedia page](#) for a total of 351 towns. The list of latitude and longitude points of the state and each town were found through the Google Maps Geocoding API. The venues within a 500-meter radius of each town were found through the Foursquare API.

2.2 Data cleaning

The data scraped from the Wikipedia page was prepared by removing missing values. Those missing values were associated with rows that were not classified as towns or cities, thus insignificant to the goal of the project and removed. The column heading containing towns was changed from “Municipality” to “Neighborhood” to allow for more ease in merging the venue information from Foursquare to the main dataframe. Additional columns in the dataframe are added including the state, to be used in the creation of a map, as well as the latitude and longitude, to be used to get the nearby venues of each town.

3. Methodology

3.1 Exploratory data analysis

The location of each town is denoted by its latitude and longitude. Using its latitude and longitude, a map of Massachusetts, Figure 1, is created and each town is marked.

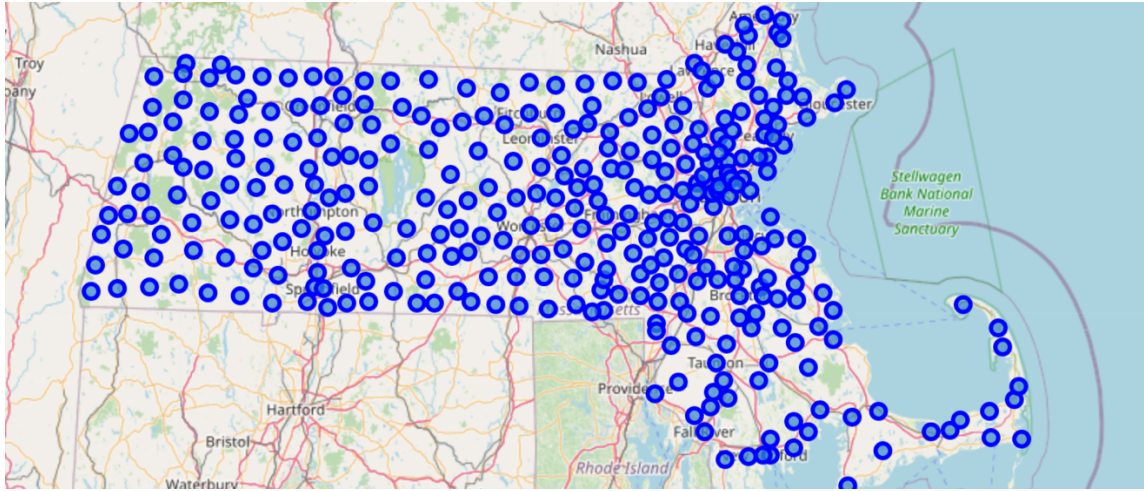


Figure 1. Map of the towns.

3.2 K-means clustering

To figure out the optimal number of clusters to use, the elbow method is used to visualize and evaluate the best value of k . The k is chosen where the graph bends at the “elbow” where the distortion and inertia decrease. As seen from figures 2, and 3, the optimal number of clusters to use is 6.

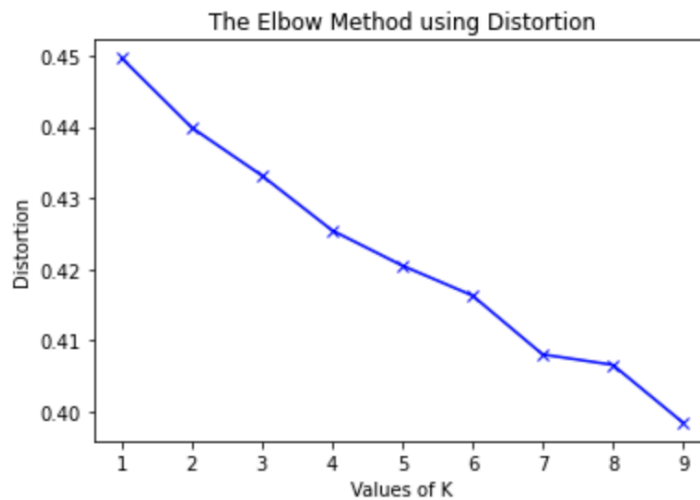


Figure 2. Elbow method using distortion visualization.

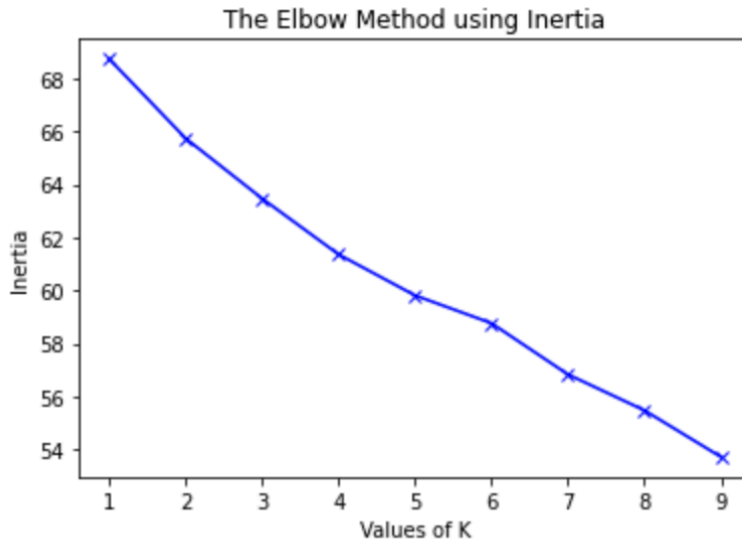


Figure 3. Elbow method using inertia visualization.

The k-means clustering method is used as an unsupervised machine learning to segment towns based on their top 10 most frequented venues. The venues are organized by their venue category and the frequency of occurrence in each category by town is averaged. Using the venues, each town is assigned one of the six clusters.

4. Results

The resulting six clusters are shown in Figures 4-9.

	Neighborhood	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	Weston	Coffee Shop	Diner	Bagel Shop	Market	Peking Duck Restaurant	Pharmacy	Grocery Store	Pizza Place	Fishing Spot	Fish Market
3	Sherborn	Convenience Store	Pizza Place	Sandwich Place	Donut Shop	Pet Store	Women's Store	Fast Food Restaurant	Event Space	Exhibit	Fabric Shop
4	Sudbury	Italian Restaurant	Pizza Place	Donut Shop	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
10	Wayland	Home Service	Thai Restaurant	Athletics & Sports	Chinese Restaurant	Pizza Place	Pharmacy	Event Space	Exhibit	Fabric Shop	Farm
16	Westwood	Cosmetics Shop	Café	Chinese Restaurant	Pizza Place	Home Service	Donut Shop	Women's Store	Fast Food Restaurant	Exhibit	Fabric Shop
...
342	Brockton	Chinese Restaurant	Lounge	Pub	Pizza Place	Automotive Shop	Bus Station	Asian Restaurant	Financial or Legal Service	Fabric Shop	Farm
346	Orange	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
348	Chelsea	Pizza Place	Spanish Restaurant	Mexican Restaurant	Donut Shop	Fast Food Restaurant	Bagel Shop	Thai Restaurant	Discount Store	Bank	Pharmacy
349	Holyoke	Pizza Place	Spanish Restaurant	Art Gallery	Bar	Bakery	Fast Food Restaurant	Storage Facility	Park	Trail	American Restaurant
350	Amherst	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

152 rows x 11 columns

Figure 4. Cluster 1

	Neighborhood	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
1	Dover	Post Office	Deli / Bodega	Donut Shop	Grocery Store	Flower Shop	Food	Flea Market	Fishing Store	Fishing Spot	Fish Market
2	Carlisle	Sandwich Place	Women's Store	Film Studio	Event Space	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Financial or Legal Service
5	Wellesley	Coffee Shop	Pizza Place	Arts & Crafts Store	Mediterranean Restaurant	Italian Restaurant	Thai Restaurant	Yoga Studio	Men's Store	Mattress Store	Tapas Restaurant
6	Winchester	New American Restaurant	Italian Restaurant	Coffee Shop	Cupcake Shop	Donut Shop	Pharmacy	Café	Gastropub	Supermarket	Chinese Restaurant
7	Manchester-by-the-Sea	Café	Grocery Store	Diner	Bank	New American Restaurant	Other Great Outdoors	Train Station	Seafood Restaurant	Bookstore	Pharmacy
...
344	North Adams	Pharmacy	American Restaurant	Pizza Place	Mexican Restaurant	Art Museum	Discount Store	Hot Dog Joint	Multiplex	Bistro	Liquor Store
345	Fall River	Donut Shop	Café	Bus Station	Bar	Portuguese Restaurant	Gym	Pharmacy	Pub	Doctor's Office	Restaurant
347	New Bedford	Breakfast Spot	Fried Chicken Joint	American Restaurant	Laundromat	Fast Food Restaurant	Women's Store	Fish & Chips Shop	Fabric Shop	Farm	Farmers Market
351	Springfield	Hotel	Hockey Arena	Shipping Store	American Restaurant	Café	Donut Shop	Coffee Shop	Sports Bar	Sculpture Garden	Science Museum
352	Lawrence	Caribbean Restaurant	Donut Shop	Shoe Store	Pizza Place	Convenience Store	Pub	Bank	Pharmacy	Café	Mexican Restaurant

172 rows × 11 columns

Figure 5. Cluster 2

	Neighborhood	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
27	Bolton	Park	Business Service	Women's Store	Film Studio	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Financial or Legal Service
34	Richmond	Business Service	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop
201	Ashburnham	Business Service	Bed & Breakfast	Women's Store	Fish & Chips Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service	Fish Market
238	Blackstone	Vineyard	Business Service	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio
270	South Hadley	Furniture / Home Store	Business Service	Bed & Breakfast	Women's Store	Fish & Chips Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service
297	Buckland	Business Service	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop

Figure 6. Cluster 3

	Neighborhood	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
53	Reading	Park	Playground	Women's Store	Film Studio	Event Space	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
90	Newbury	Park	Butcher	Women's Store	Film Studio	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Financial or Legal Service
95	Burlington	Park	Baseball Field	Pizza Place	Gas Station	Women's Store	Financial or Legal Service	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
135	Kingston	Baseball Field	River	Park	Museum	Home Service	Event Space	Exhibit	Fabric Shop	Farm	Farmers Market
153	Deerfield	River	Women's Store	Event Space	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service
159	Hampden	Post Office	Park	Italian Restaurant	Film Studio	Event Space	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
165	Charlton	Park	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop
209	Bourne	Post Office	Pharmacy	Park	Bar	Flea Market	Fishing Store	Fishing Spot	Fish Market	Flower Shop	Fish & Chips Shop

225	Swansea	Park	Dessert Shop	Business Service	Women's Store	Film Studio	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
242	Freetown	Convenience Store	Park	Women's Store	Film Studio	Event Space	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
254	Somerset	River	History Museum	Park	Harbor / Marina	Baseball Field	Fishing Spot	Farm	Event Service	Event Space	Flower Shop
275	Bernardston	Park	Ski Chalet	Women's Store	Film Studio	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Financial or Legal Service
278	Oak Bluffs	Beach	Park	Construction & Landscaping	Art Gallery	Women's Store	Financial or Legal Service	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
290	Leominster	Dance Studio	Construction & Landscaping	Park	Business Service	Women's Store	Financial or Legal Service	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant
327	Palmer	Park	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop

Figure 7. Cluster 4

	Neighborhood	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
96	Hull	Scenic Lookout	American Restaurant	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio
110	New Ashford	American Restaurant	Women's Store	Fish & Chips Shop	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service
286	Oxford	American Restaurant	Women's Store	Fish & Chips Shop	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service

Figure 8. Cluster 5

	Neighborhood	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
56	Wenham	Breakfast Spot	History Museum	Women's Store	Financial or Legal Service	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio
220	Berkley	Breakfast Spot	Grocery Store	Women's Store	Fish & Chips Shop	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service
289	Russell	Breakfast Spot	Women's Store	Fish & Chips Shop	Exhibit	Fabric Shop	Farm	Farmers Market	Fast Food Restaurant	Film Studio	Financial or Legal Service

Figure 9. Cluster 6

A map of the resulting clusters is shown in Figure 10.

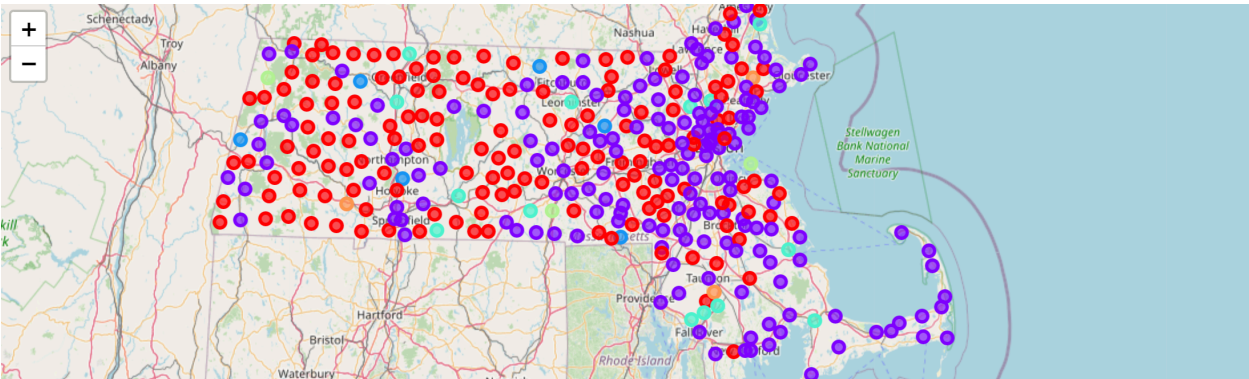


Figure 10. Map of clusters.

5. Discussion

The towns within each cluster have similar frequented venues. Cluster 1 has the most towns in their cluster which is denoted by the purple markers in Figure 10. It can be easily seen from the map that Cluster 1, which includes Boston and Brookline, is the largest cluster and highly concentrated in the east. Cluster 2 is denoted by red markers and are more concentrated in the midwest and west of Massachusetts. Cluster 3 and 4 are more sporadically scattered in the southeast or west and have more specific venues that are not as frequented in Cluster 1 and 2 towns. Cluster 5 and 6 have the most unique frequented venues compared to the other clusters as they only three towns each.

6. Conclusion

For Massachusetts residents who are looking to move, it may be helpful to see which town is similar in venues to their current location, especially if they are looking to go somewhere similar. The results can also be used to find towns that are not as similar to their current location as well. With other factors that may associated with choosing a town to live in, it may also be beneficial for another project to assess cost of living, job prospects, safety, and more for each town.