Comparing Neighborhood Similarities between Toronto, ON and Boston, MA

By Anthony Oppong-Gyebi

https://bestlifeonline.com/sleepless-cities/

Background

- The US sees many people around the globe travel in and out annually for diverse purposes.
- Immigrants would preferably want to situate in states which presents with similar amenities as their original homes, mainly for ease of adjustment to the new environment.
- An instance is when persons travelling from coastal areas will wish to find homes and jobs in places where they can have access to the beaches as well as enjoy sea food, all year.

Problem

- One major challenge immigrants face is choosing neighborhoods which suitably fit their interests and are like places they had migrated from, considering unavailability of comprehensive comparative information for neighborhoods.

- The aim of this project, therefore, is to highlight the similarities in Toronto and Boston neighborhoods for Toronto, ON residents who intend to travel to Boston, MA and vice versa.

- This will be addressed by segmenting and clustering the neighborhoods using machine learning techniques to compare the two cities.

Methodology

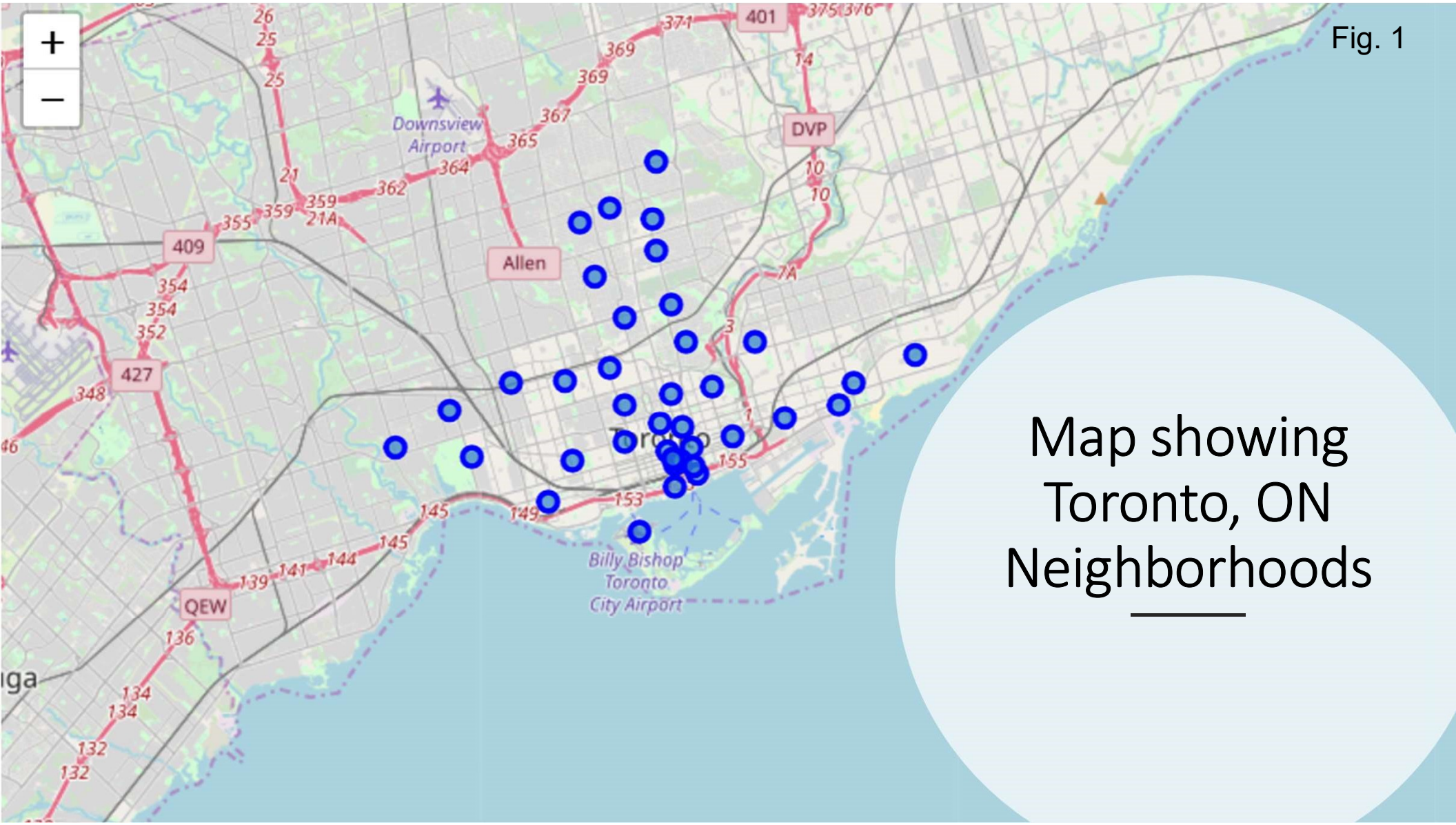
Source

- The json file for Boston neighborhood will be obtained from the link https://github.com/dj/boston/blob/master/data/boston-neighborhoods.json and web scraping will be used for the link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M to glean the neighborhoods for Toronto.

Data Cleaning

- The Canada dataset will be cleaned to extract the boroughs and neighborhoods for Toronto. Boston json file will be used to obtain the boroughs and neighborhoods for Boston.

Fig. 1


Map showing
Toronto, ON
Neighborhoods
Downsview Airport
Allen
DVP
401
409
427
QEW
Billy Bishop Toronto City Airport

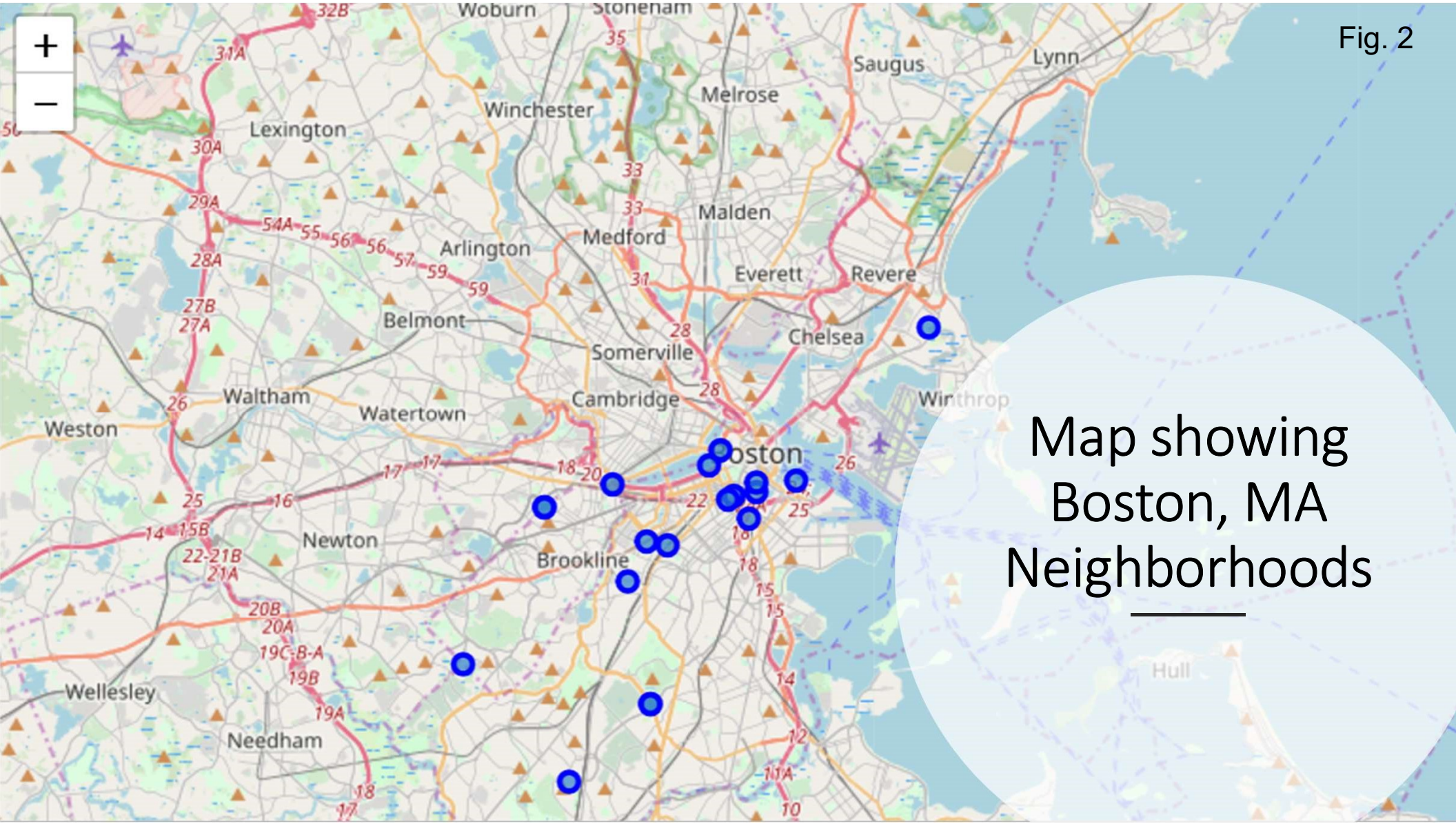


Fig. 2
Map showing
Boston, MA
Neighborhoods
Woburn
Stoneham
Saugus
Lynn
Melrose
Winchester
Lexington
Malden
Medford
Arlington
Everett
Revere
Belmont
Somerville
Chelsea
Waltham
Watertown
Cambridge
Weston
Newton
Brookline
Wellesley
Needham
Hull

Feature selection

- Coordinates of each neighborhood, the latitudes and longitudes were extracted to obtain the venues in the neighborhoods.
- Foursquare API was used to generate the locations/ amenities within the specified vicinities followed by one-hot encoding to obtain the required features from the distinct categories.
- Mean of the total number of venues per neighborhood was determined as representative of the fraction of distinct categories.
- K Means clustering was used to segment neighborhoods to determine the key similarities between the two cities, Toronto and Boston.

Toronto had a higher mean number of venues than that of Boston

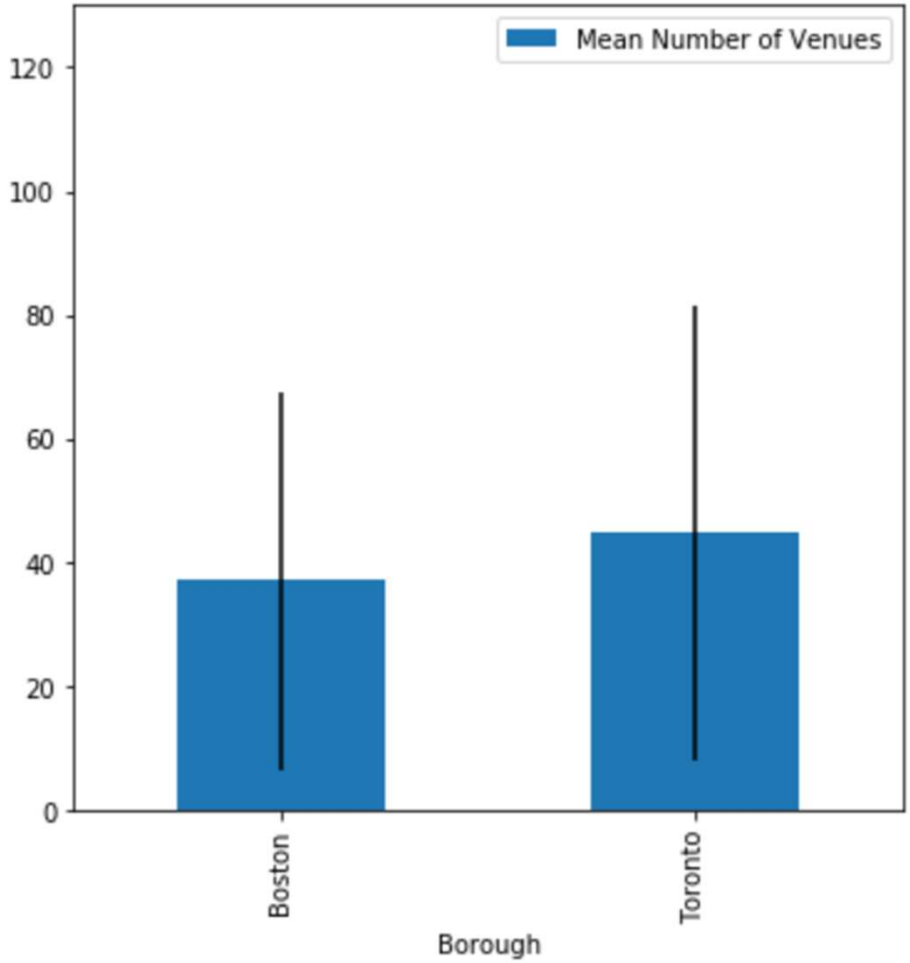


Fig. 3 Mean number of venues between Boston and Toronto

Mean proportions of first 25 venues showed significant divergence between Toronto and Boston

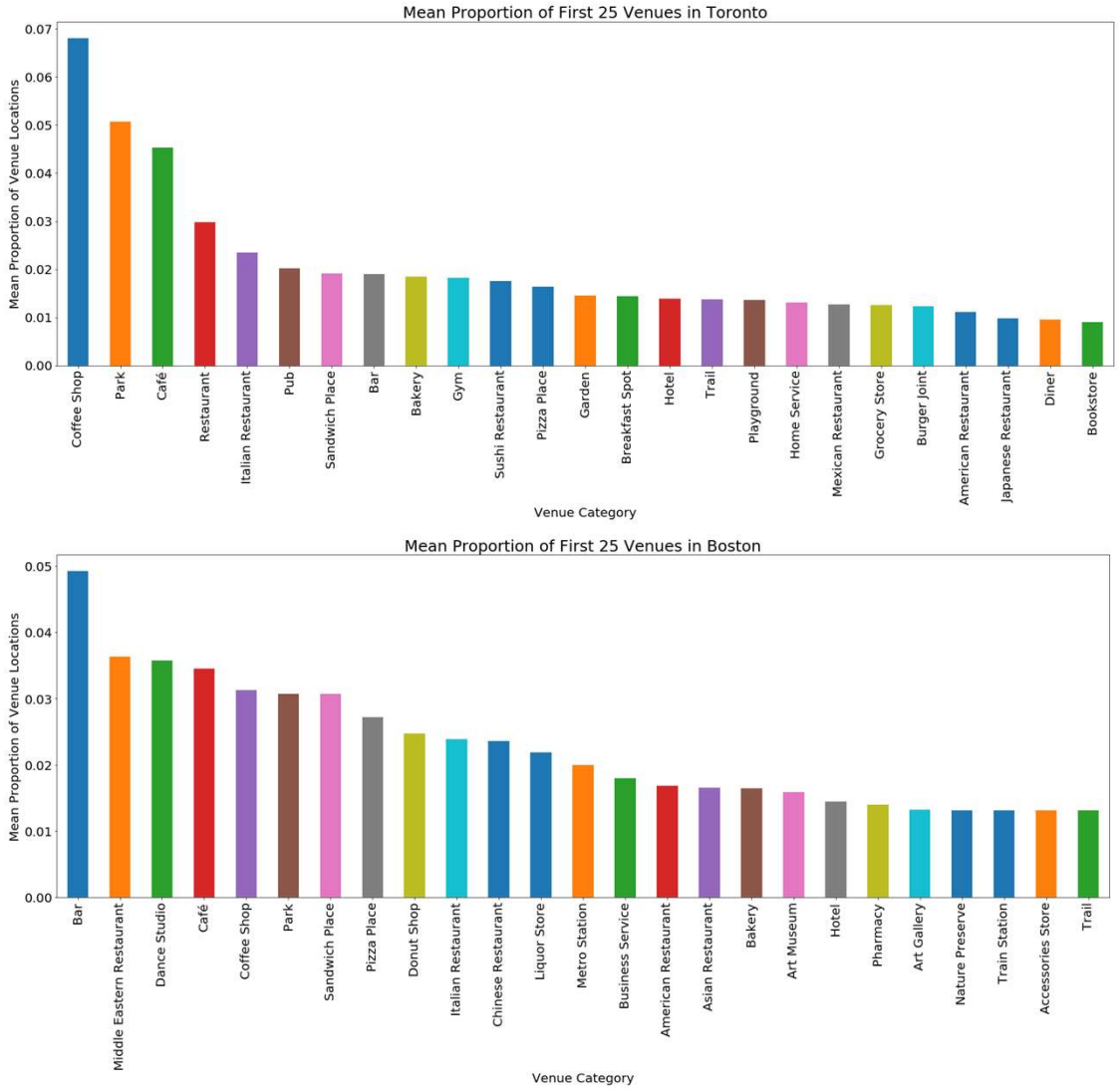


Fig. 4 Mean Proportion of First 25 Venues in Boston and Toronto

Nine-cluster model showed the best model with the smallest inertia

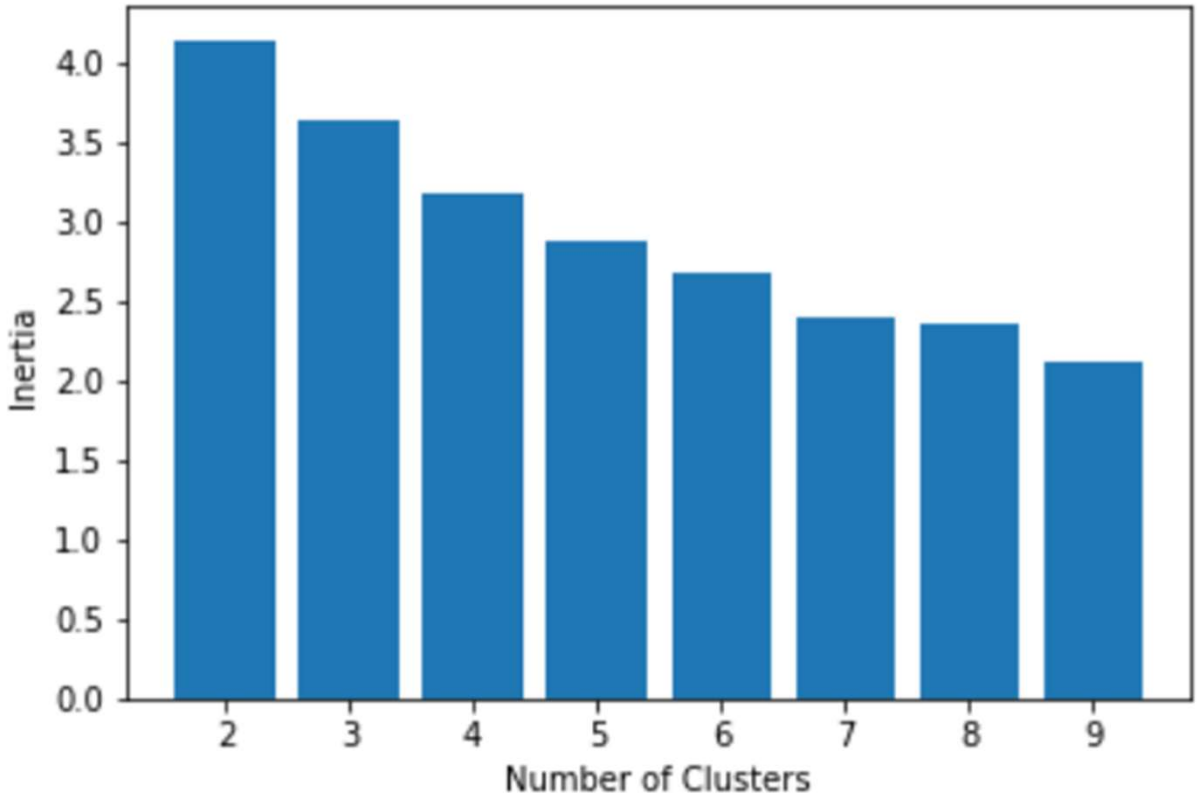


Fig. 5 Choosing Best K Means Model

Toronto and Boston had shared similar neighborhood structures in two clusters

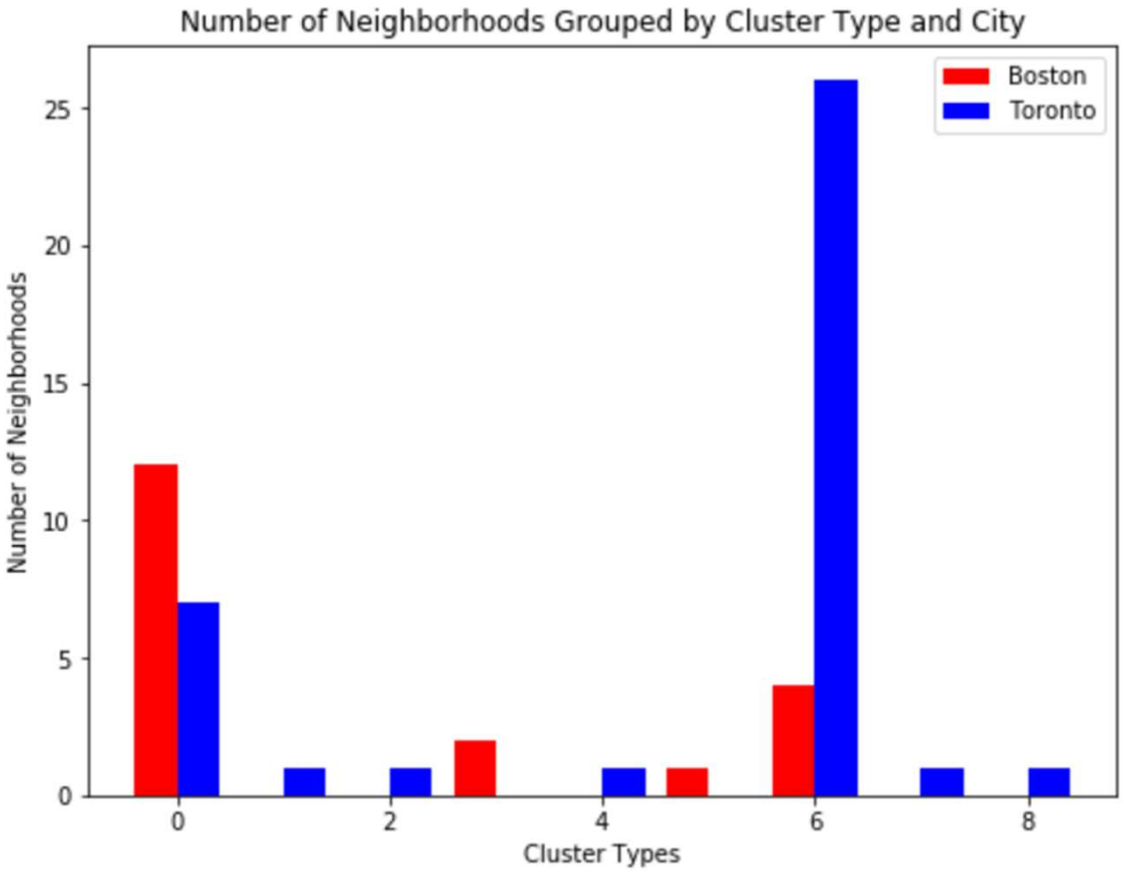
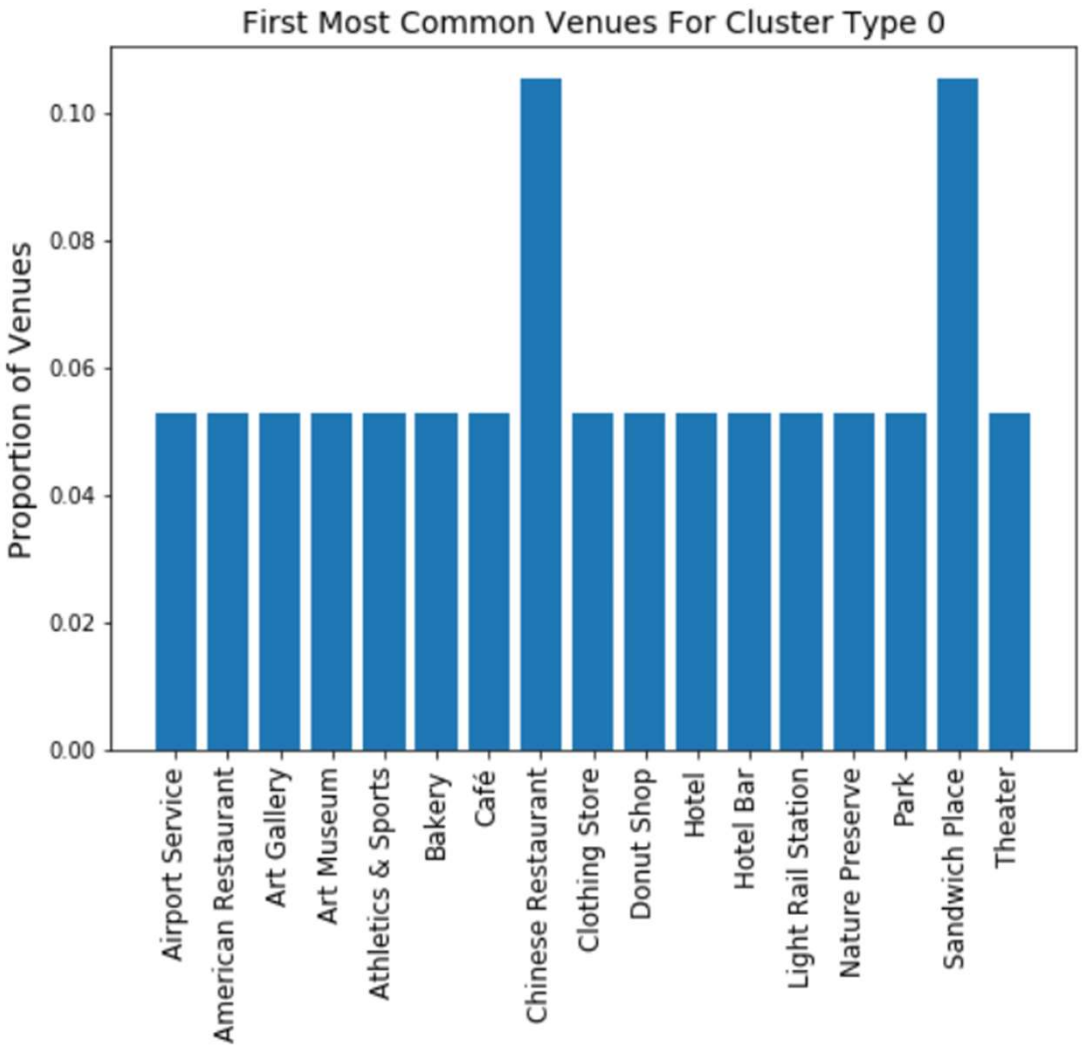
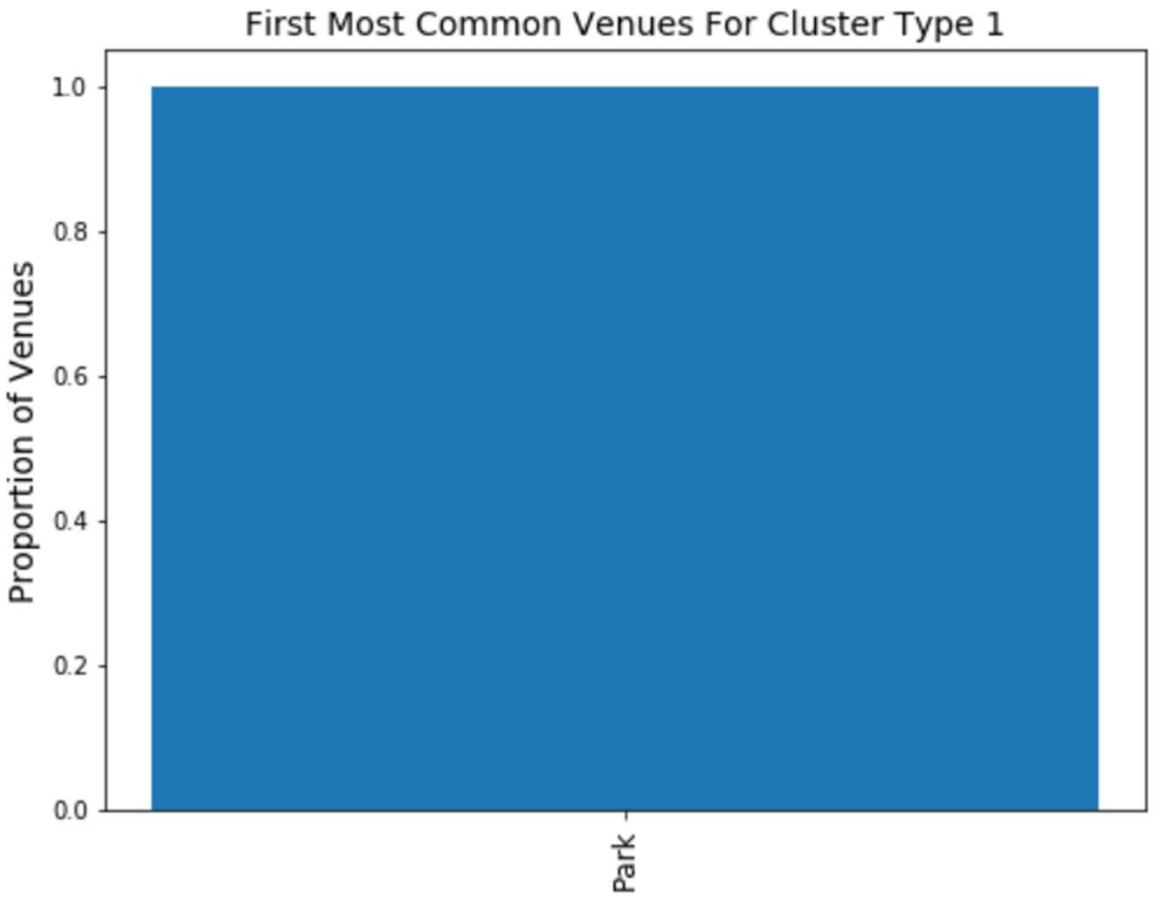


Fig. 6 Number of Neighborhoods by Cluster and City

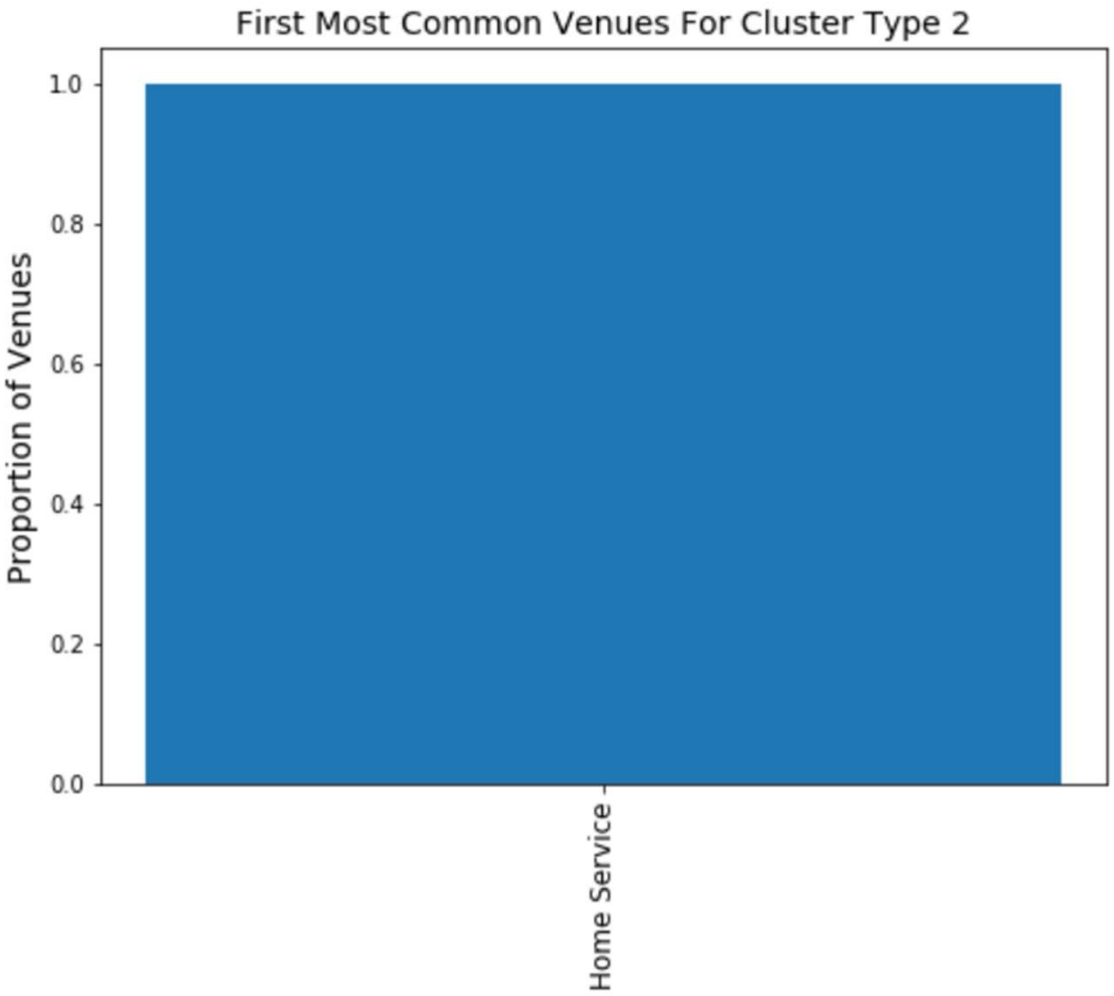
Cluster Type 0



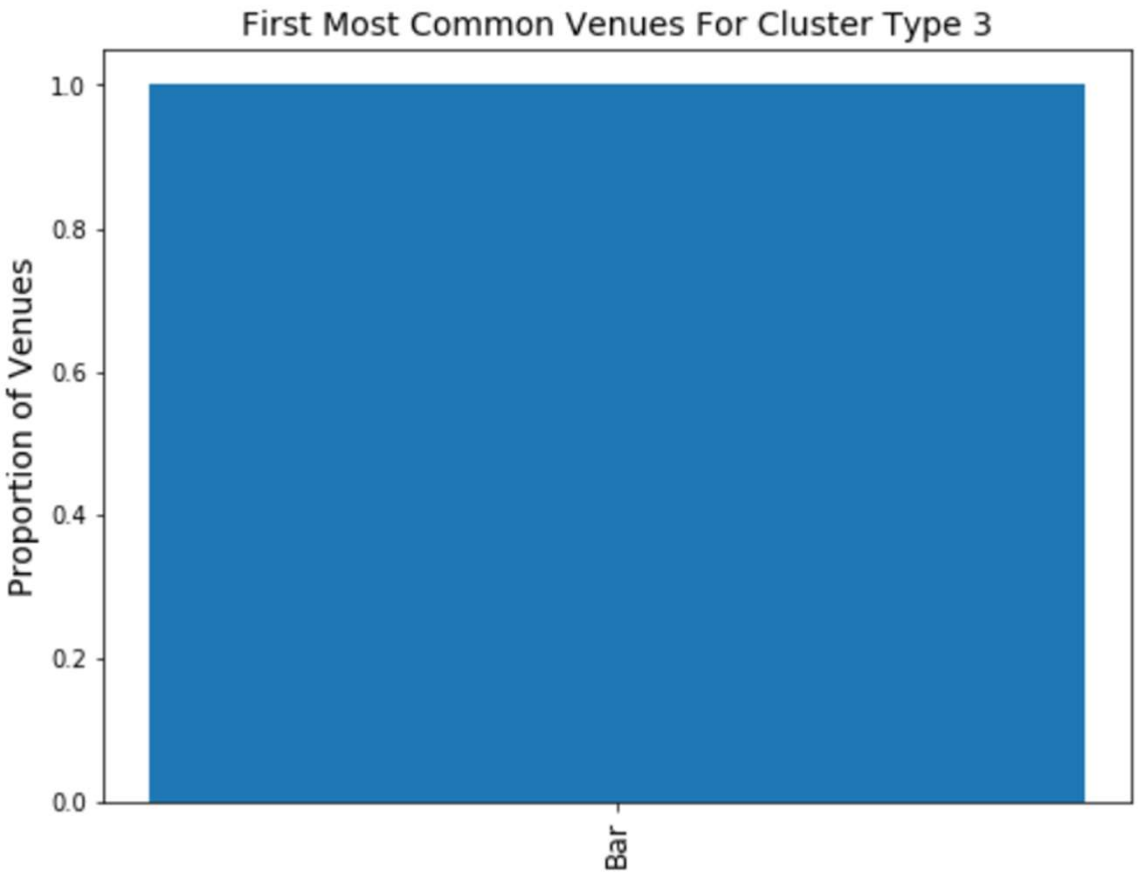
Cluster Type 1



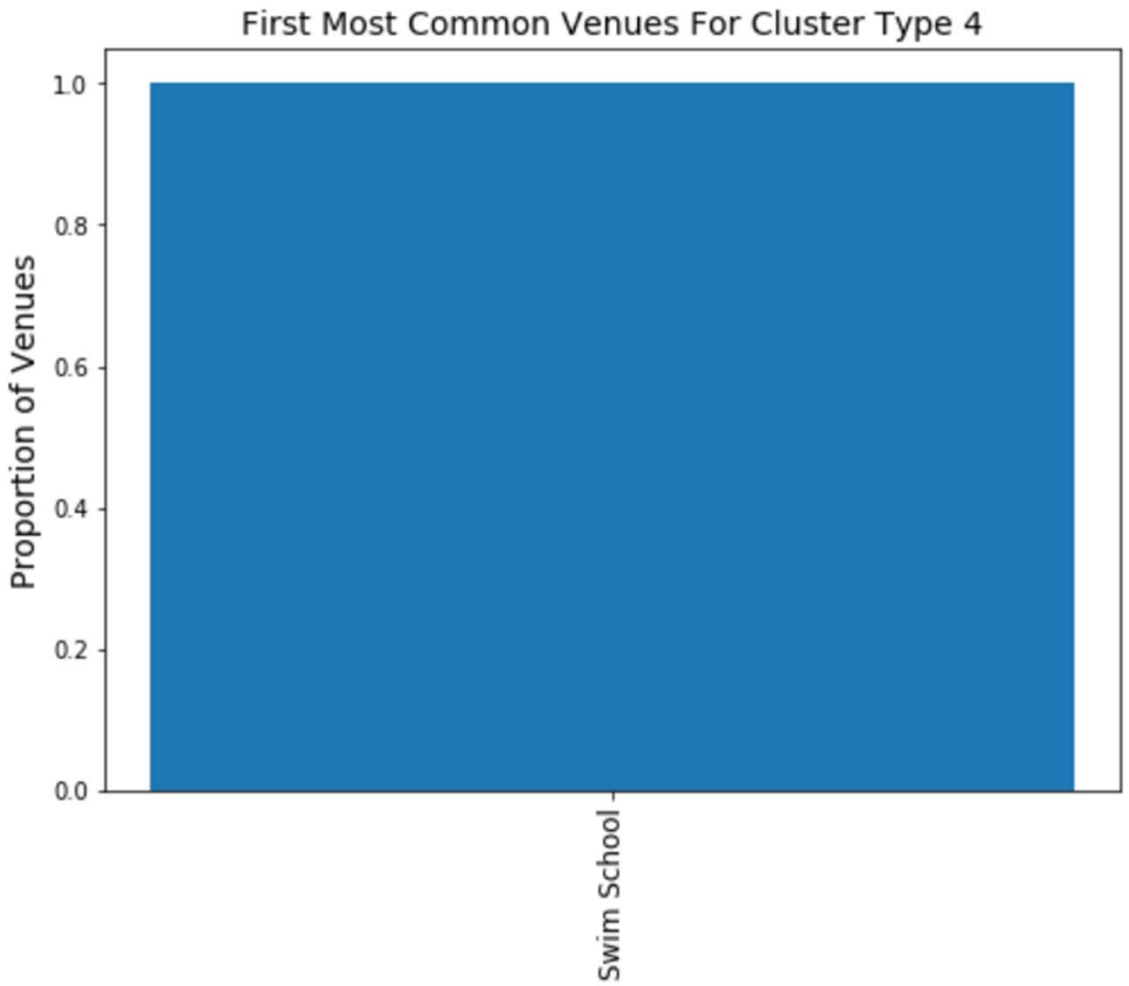
Cluster Type 2



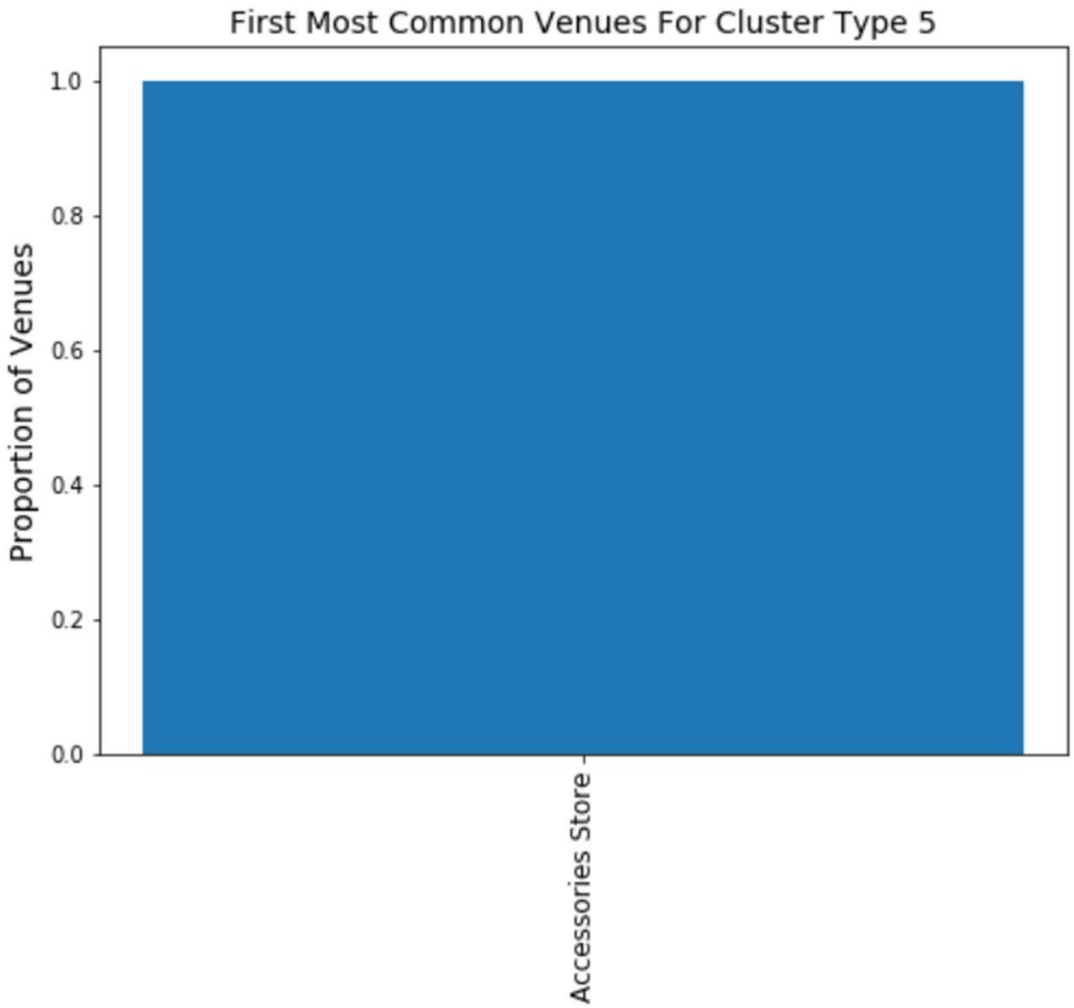
Cluster Type 3



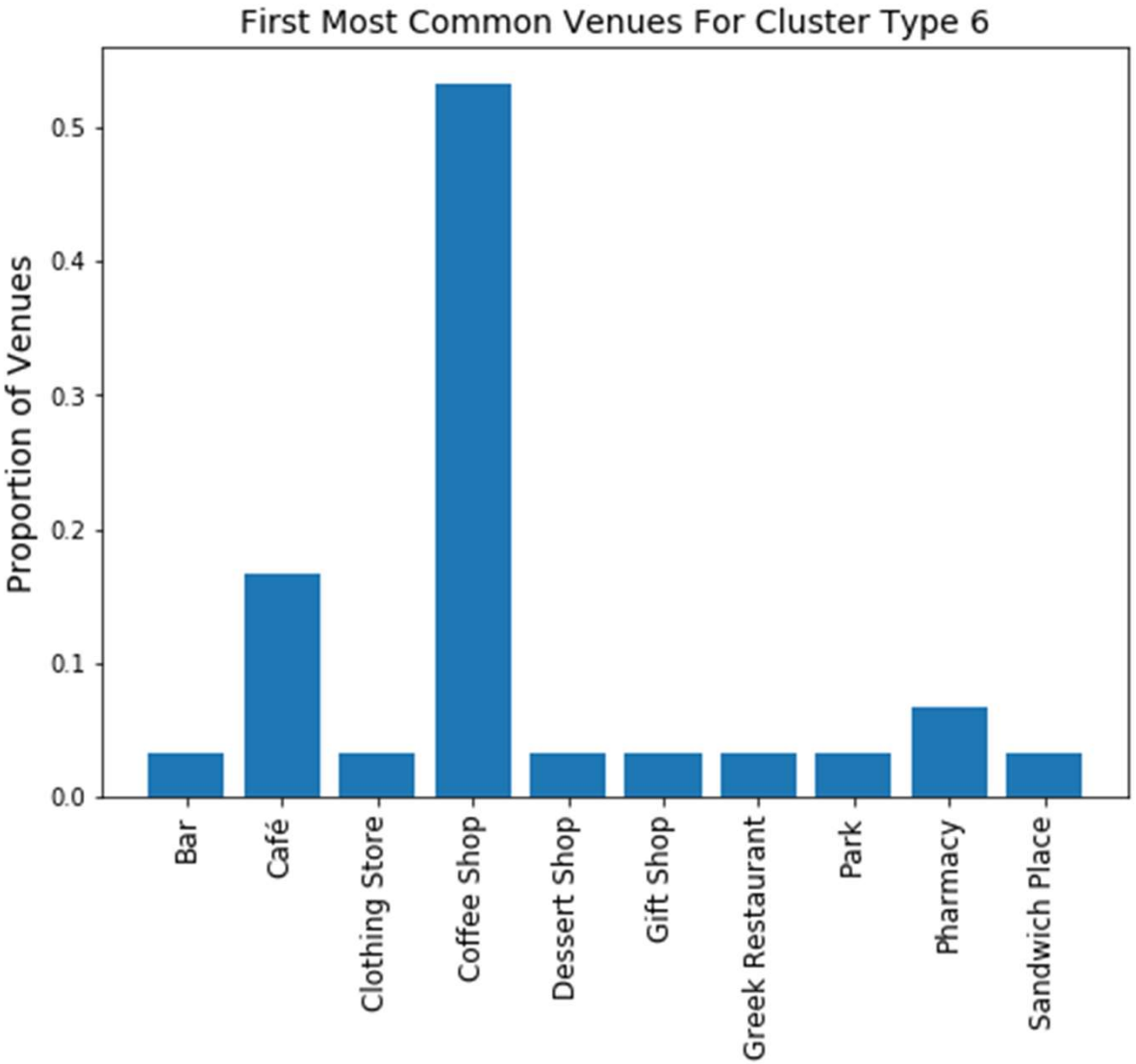
Cluster Type 4



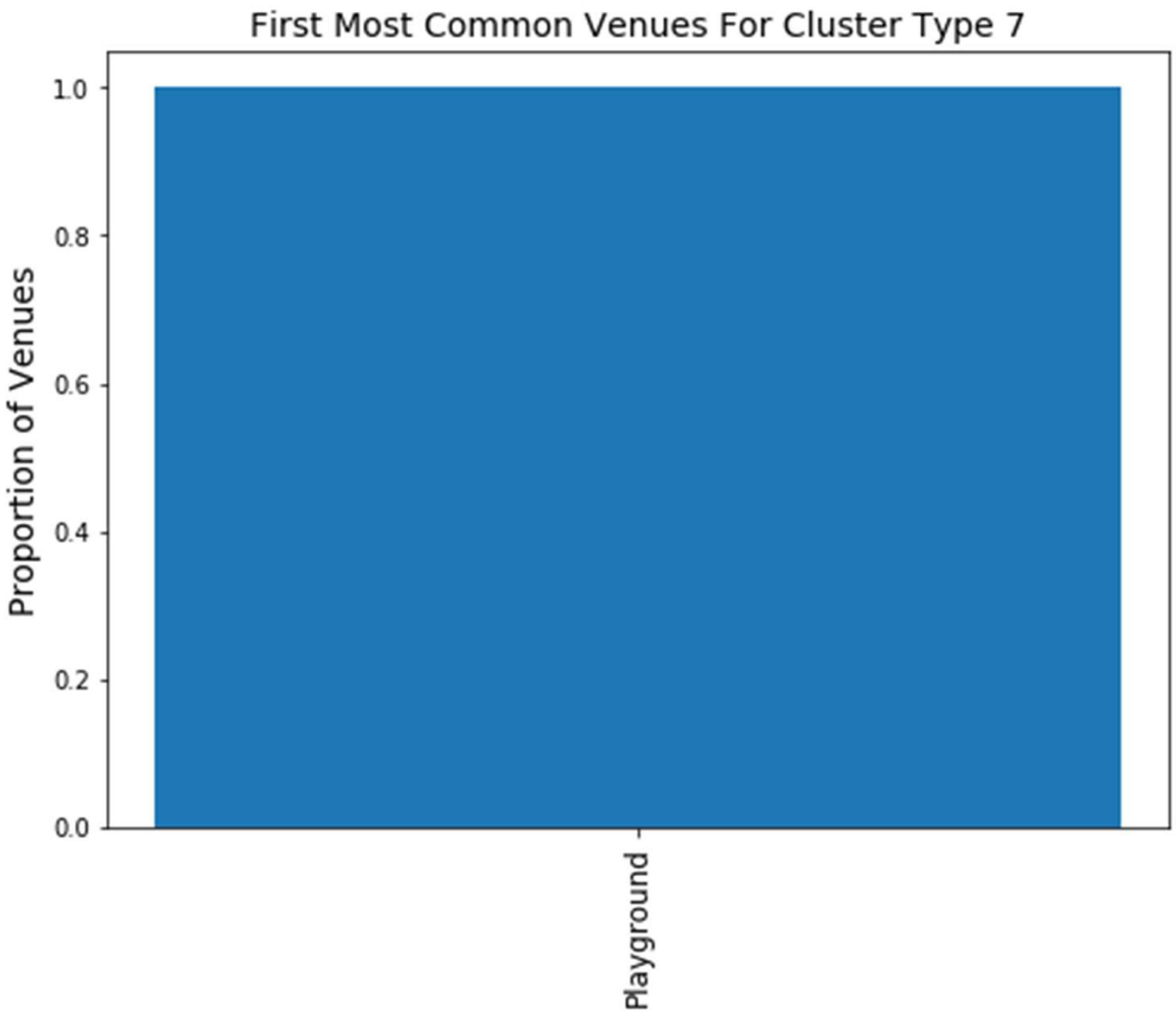
Cluster Type 5



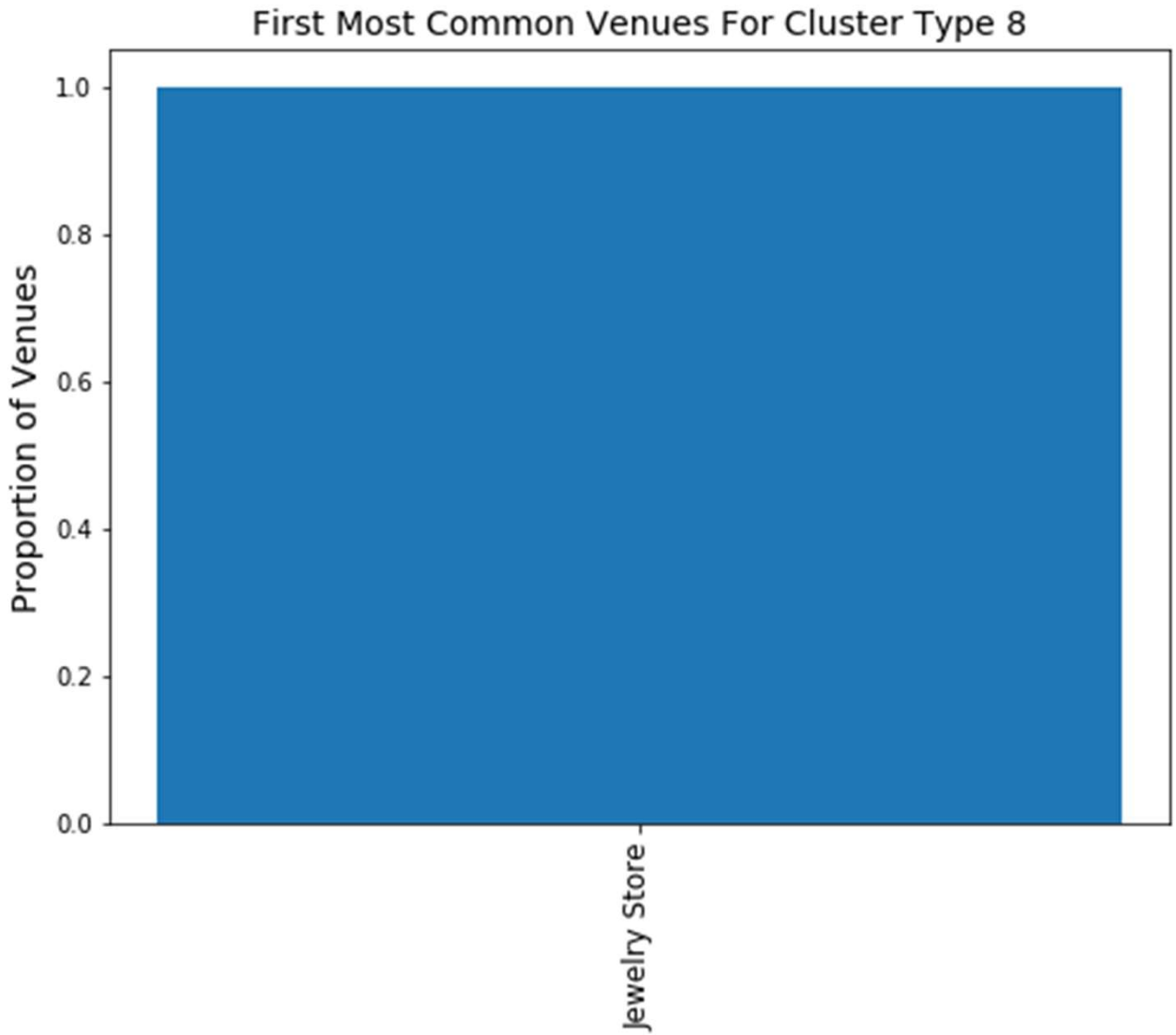
Cluster Type 6



Cluster Type 7



Cluster Type 8



Conclusion

- We have seen from this project that, two cities can be compared using Foursquare API and machine learning to reveal any similarities in the amenities available in the neighborhoods.

- The relevance of this strategy is to inform immigrants about the convergence or divergence of their new/prospective environments with their original neighborhoods.

Future Direction

- Moving forward, there will be a need to further explore the subtle similarities and differences like:
  - mortgage rates,
  - seasonal weather conditions and
  - rate of crime