Battle of the Neighborhoods

By Amy Adamson

February 9, 2020

1. Introduction

1.1 Background

Moving is one of the most stressful events in people's lives. Determining where to move can be costly and time consuming. A friend is considering moving to a new city and wants to know where she should start her search. She does not want to own a car and she currently lives in the East Village of New York City. She enjoys the variety of restaurants, bars and other establishments within her neighborhood. She would like to live in a similar neighborhood. Having a starting point for her search could save her money, time and frustration.

1.2 Problem

What neighborhoods in what cities should my friend start researching for a future move? This project will review venue data in different neighborhoods and predict which neighborhoods will be most like the East Village of New York City. This project will also research which cities have the best public transportation and take in any end user feedback to restrict the list before reviewing the venue data.

1.3 Application

Besides helping my friend, it is possible that this could be used at a larger scale for individuals looking for the same information. This project will allow for research into if this is feasible to create a future app to help people decide where they might want to move.

2. Data acquisition and cleaning

2.1 Data Sources

Since my friend does not want to own a car, she will need to move to an area with excellent public transportation. Using the US News article, "The 10 Best Cities for Public Transportation", and user input, the list of potential cities was narrowed down to 5 cities: San Francisco, Boston, Philadelphia, Chicago and Portland, OR. Wikipedia was used to get a list of neighborhoods for the 5 cities and www.usgs.gov was cross referenced to get the latitude and longitudes for the neighborhoods. Unfortunately, not all neighborhood locations were available on the website, so Google was used to complete the list.

Finally, all venues within 500m around the neighborhood locations were pulled from FourSquare.

2.2 Data Cleansing

For each city, a different subset of venue categories has been created in FourSquare. There is overlap in categories in each city. The East Village contained 57 different venue categories. The venue data was restricted to the 57 venue categories from the East Village.

The restricted list was then investigated for neighborhoods with less than 5 venues. See table 1 for counts. Based on the number of neighborhoods in each city, it was determined that removing the neighborhoods with less than 5 venues would leave ample neighborhoods to investigate.

Table 1 – Investigating Neighborhoods with less than 5 Venues.

City	Number of Neighborhoods with less than 5 Venues	Number of Neighborhoods by City	Number of Neighborhoods with more than 4 Venues
Boston	4	20	16
Chicago	31	69	38
Philadelphia	27	50	23
Portland OR	46	77	31
San Francisco	16	81	65

This final venue list was used to create a record for each neighborhood with an entry for each of the 57 categories. The percent of venues within that category in that neighborhood were placed in the category entry (Table 2). This list was used to create a list of neighborhoods with the top 10 most popular venue types (Table 3).

Table 2 – Snippet of created table showing percent of venue category by neighborhood.

	City	Neighborhood	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Bagel Shop	Bakery	Bar	Beer Bar	Beer Store	Burger Joint
0	Boston	Allston	0.125000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.375000	0.000000	0.000000	0.125000
1	Boston	Back Bay	0.111111	0.00	0.000000	0.000000	0.000000	0.000000	0.027778	0.083333	0.000000	0.000000	0.055556
2	Boston	Bay Village	0.108108	0.00	0.000000	0.000000	0.027027	0.027027	0.054054	0.000000	0.000000	0.000000	0.000000
3	Boston	Beacon Hill	0.130435	0.00	0.000000	0.000000	0.000000	0.000000	0.043478	0.043478	0.000000	0.000000	0.000000
4	Boston	Brighton	0.066667	0.00	0.000000	0.000000	0.000000	0.000000	0.200000	0.066667	0.000000	0.000000	0.000000
5	Boston	Charlestown	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Boston	Chinatown	0.117647	0.00	0.000000	0.000000	0.029412	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000
7	Boston	Downtown	0.058824	0.00	0.000000	0.000000	0.000000	0.000000	0.088235	0.000000	0.000000	0.000000	0.058824
8	Boston	East Boston	0.083333	0.00	0.000000	0.166667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 3 – Snippet of created table showing 10 most popular venues by neighborhood.

	City	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	Boston	Allston	Bar	Pizza Place	American Restaurant	Burger Joint	Coffee Shop	Cocktail Bar	Gourmet Shop	Gift Shop	Garden Center	French Restaurant
1	Boston	Back Bay	American Restaurant	Ice Cream Shop	Italian Restaurant	Bar	Coffee Shop	Vietnamese Restaurant	Burger Joint	Seafood Restaurant	Mexican Restaurant	Bakery
2	Boston	Bay Village	Spa	American Restaurant	Seafood Restaurant	Italian Restaurant	Park	Bakery	Coffee Shop	French Restaurant	Mexican Restaurant	Arts & Crafts Store
3	Boston	Beacon Hill	Italian Restaurant	American Restaurant	Pizza Place	French Restaurant	Gift Shop	Ice Cream Shop	Korean Restaurant	Mexican Restaurant	Dessert Shop	Park
4	Boston	Brighton	Pizza Place	Bakery	Chinese Restaurant	Coffee Shop	Wine Shop	Bar	Dessert Shop	Greek Restaurant	American Restaurant	Spa

3. Methodology

K Means clustering was used with the 10 most popular venues by neighborhood table to cluster the neighborhoods into 10 clusters using 30 iterations. Less iterations were tried but the data continued to move with less iterations. With 10 clusters, the similarities within the clusters were easy to see. The following are snippets of 3 of the 10 clusters:

- Cluster 0 – Most have American Restaurant and Pizza Place as first and second most common Venue.

City	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
Boston	Hyde Park	42.255654	-71.124496	0	American Restaurant	Pizza Place
Chicago	Chicago Lawn	41.775033	-87.696441	0	Pizza Place	American Restaurant
Chicago	Montclare	41.927878	-87.815590	0	American Restaurant	Bakery
Chicago	Rogers Park	42.008643	-87.666724	0	American Restaurant	Bar
Chicago	South Shore	41.761979	-87.577825	0	American Restaurant	Pizza Place
Philadelphia	Chestnut Hill	40.077055	-75.207400	0	American Restaurant	Coffee Shop
Philadelphia	Northeast Philadelphia	40.050674	-75.066800	0	American Restaurant	Pizza Place
Philadelphia	Powelton	39.958446	-75.191012	0	Pizza Place	American Restaurant
San Francisco	Presidio Heights	37.788693	-122.457541	0	American Restaurant	Italian Restaurant

- Cluster 1 – Most have Italian Restaurant as first or second most common venue.

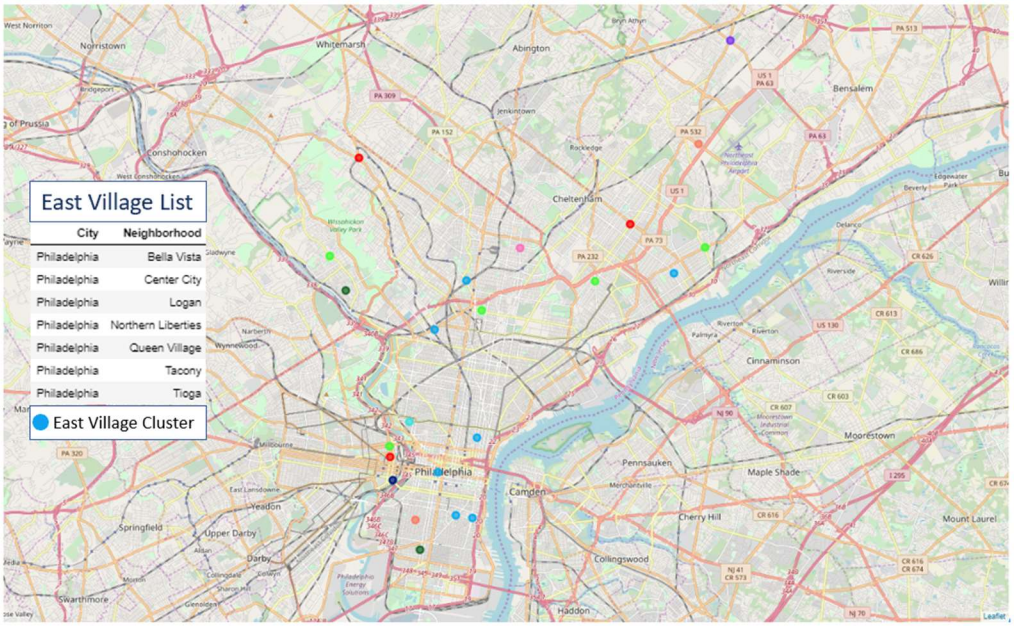
City	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Boston	Beacon Hill	42.358708	-71.067829	1	Italian Restaurant	American Restaurant	Pizza Place
Boston	North End	42.365097	-71.054495	1	Italian Restaurant	Pizza Place	Bakery
Boston	West End	42.364746	-71.070421	1	American Restaurant	Italian Restaurant	French Restaurant
Chicago	Beverly	41.718089	-87.673661	1	Italian Restaurant	Juice Bar	Bakery
Chicago	Forest Glen	41.991701	-87.793395	1	American Restaurant	Bar	Italian Restaurant
Chicago	Portage Park	41.957809	-87.765059	1	Italian Restaurant	Dog Run	Pizza Place
Chicago	Pullman	41.957809	-87.765059	1	Italian Restaurant	Dog Run	Pizza Place
Chicago	Riverdale	41.957809	-87.765059	1	Italian Restaurant	Dog Run	Pizza Place
Philadelphia	Somerton	40.123443	-75.014892	1	Italian Restaurant	French Restaurant	Bagel Shop
San Francisco	Mission Terrace	37.724358	-122.446749	1	American Restaurant	Flower Shop	Dessert Shop
San Francisco	Polk Gulch	37.792353	-122.429786	1	Wine Shop	Beer Store	Ice Cream Shop
San Francisco	Telegraph Hill	37.801667	-122.410576	1	Italian Restaurant	Pizza Place	Coffee Shop

- Cluster 2 – Most have Coffee Shop as first most common venue.

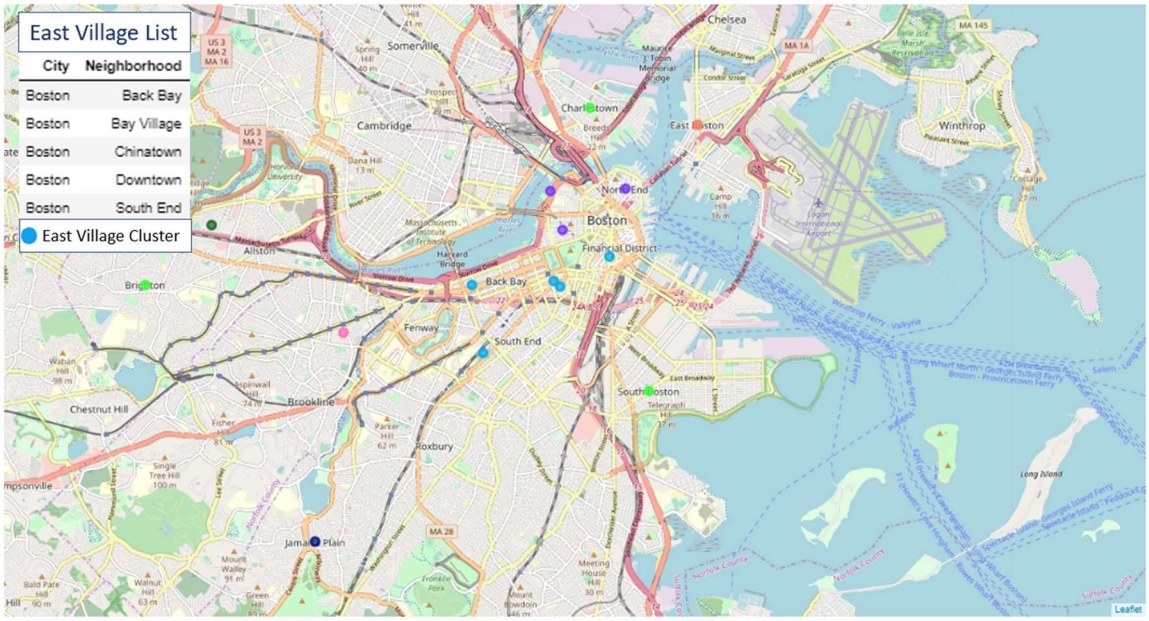
City	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
Boston	Jamaica Plain	42.309820	-71.120330	2	Art Gallery	Bakery
Chicago	Grand Boulevard	41.813923	-87.617272	2	Coffee Shop	Art Gallery
Chicago	Hyde Park	41.794201	-87.593938	2	Coffee Shop	Wine Shop
Chicago	The Loop	41.878971	-87.637882	2	Coffee Shop	Italian Restaurant
Philadelphia	University City	39.949279	-75.189624	2	Coffee Shop	American Restaurant
Portland OR	Brooklyn	45.490397	-122.637317	2	Coffee Shop	Cocktail Bar
Portland OR	Creston Kenilworth	45.494136	-122.637626	2	Coffee Shop	Cocktail Bar
Portland OR	Hosford Abernethy	45.505286	-122.670658	2	Coffee Shop	American Restaurant
Portland OR	Kenton	45.581784	-122.681486	2	Coffee Shop	Pizza Place
Portland OR	Multnomah	45.469563	-122.707318	2	Coffee Shop	American Restaurant
San Francisco	Cole Valley	37.764859	-122.458546	2	Dog Run	Coffee Shop
San Francisco	Little Hollywood	37.711963	-122.402699	2	Coffee Shop	Park
San Francisco	North of Panhandle	37.776317	-122.446729	2	Coffee Shop	Mexican Restaurant
San Francisco	Rincon Hill	37.786871	-122.396164	2	Coffee Shop	American Restaurant

The following are the maps of the 5 cities with the clustering:

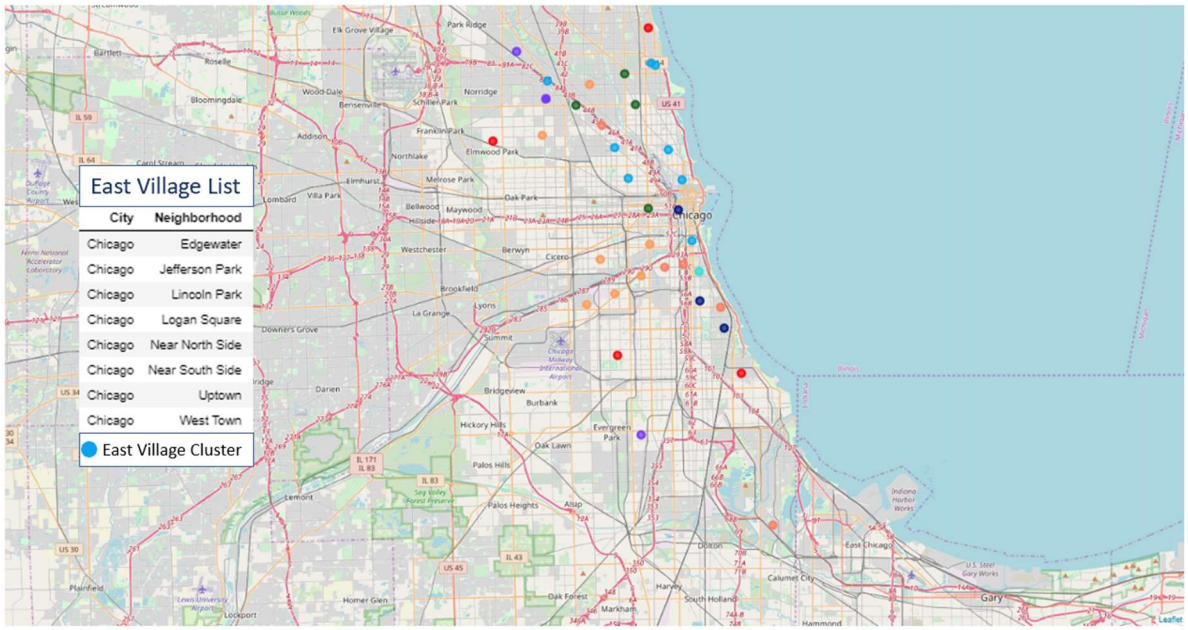
Philadelphia with K Means Clusters



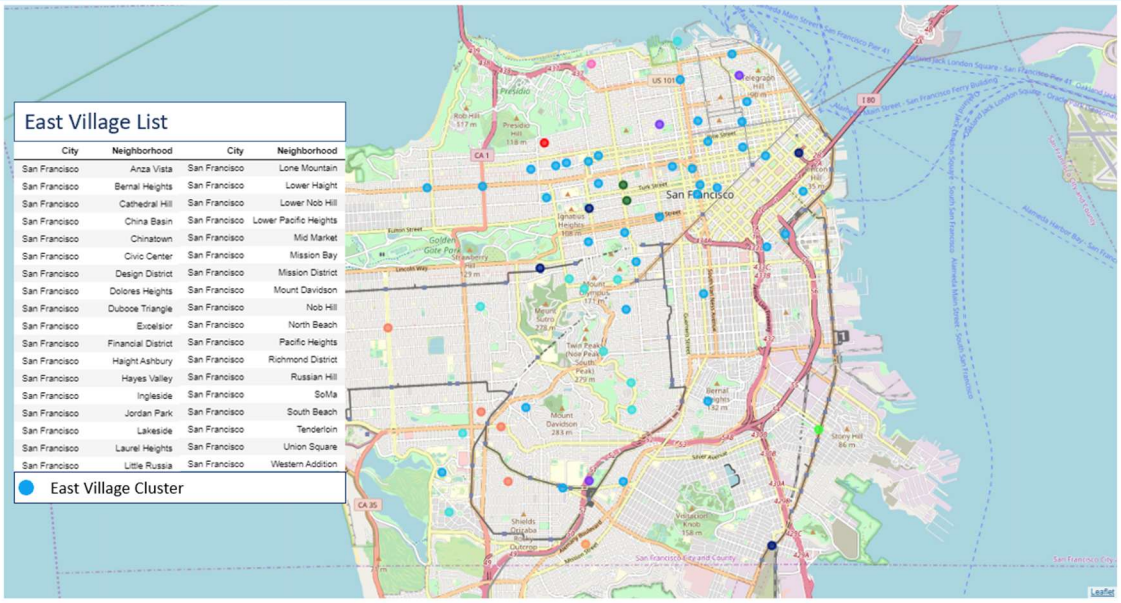
Boston with K Means Clusters



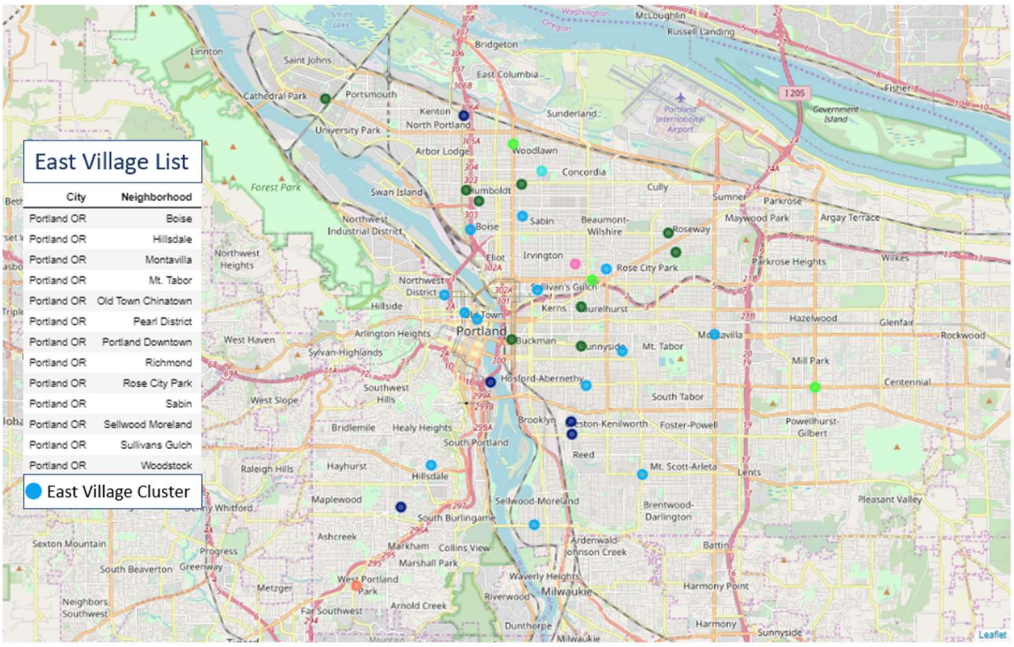
Chicago with K Means Clusters



San Francisco with K Means Clusters



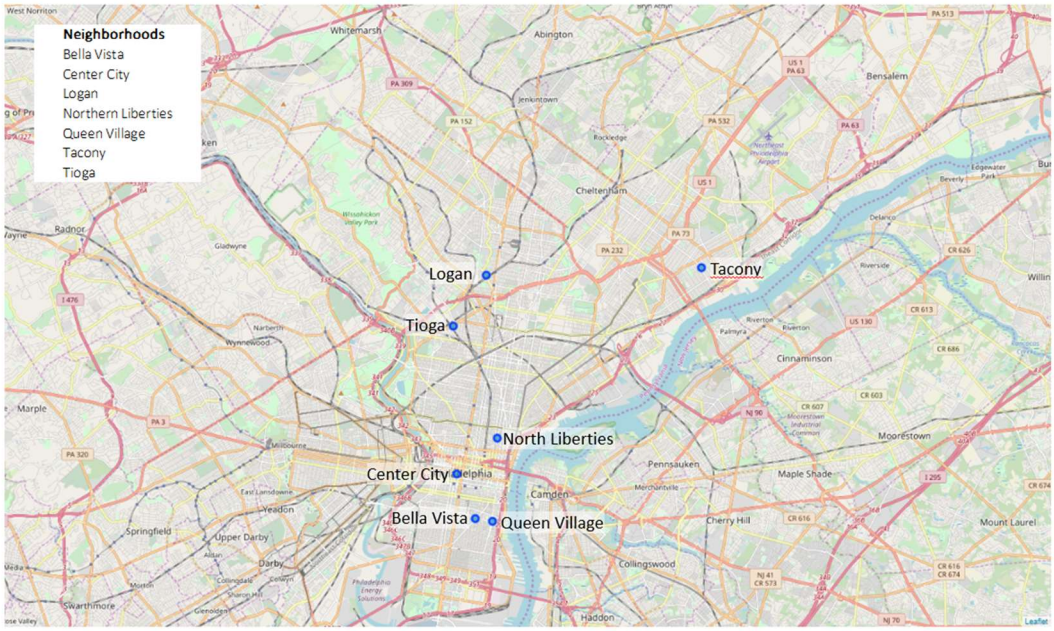
Portland, OR with K Means Clusters



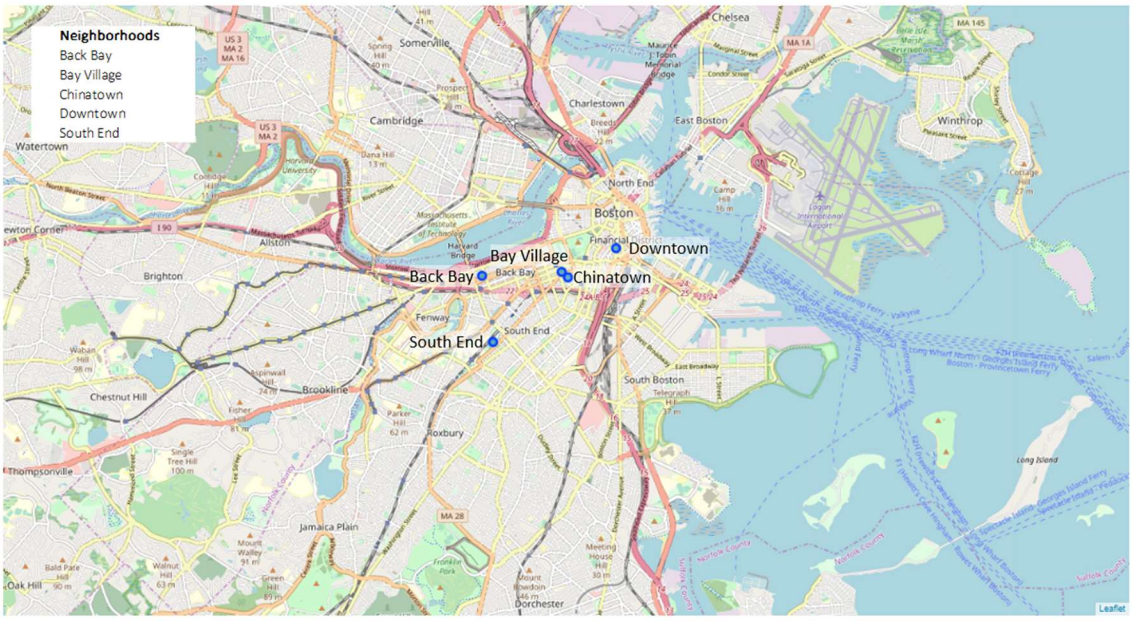
4. **Results**

For the results, 5 city maps were created with the neighborhoods that are like the East Village as blue circles and labelled.

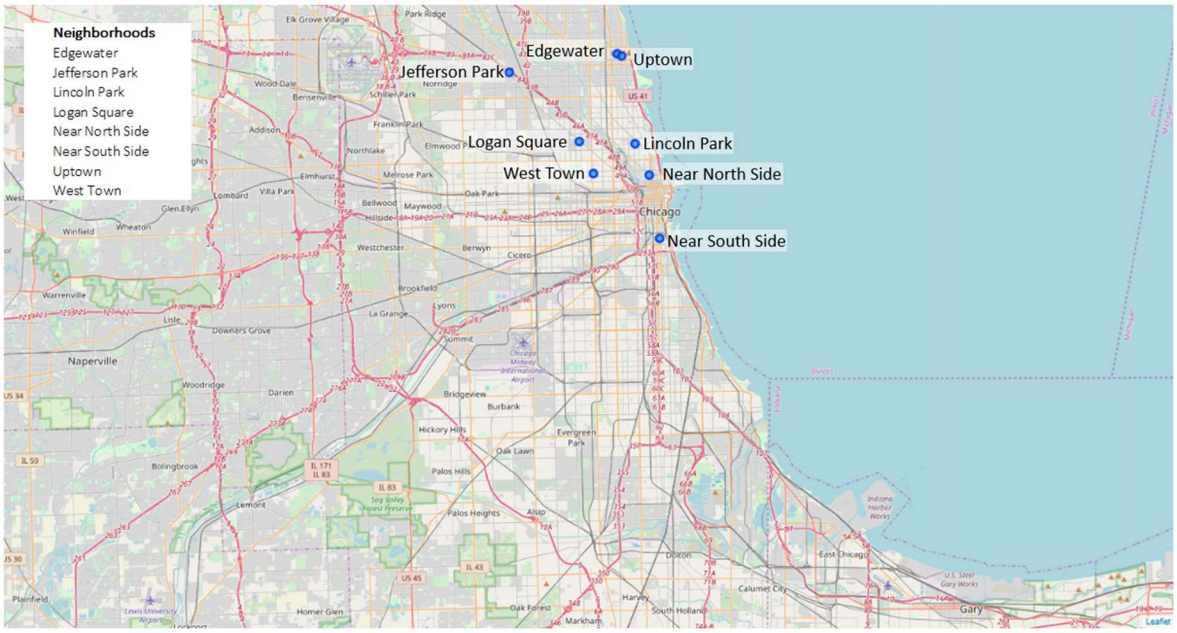
Results – Philadelphia Neighborhoods like EV



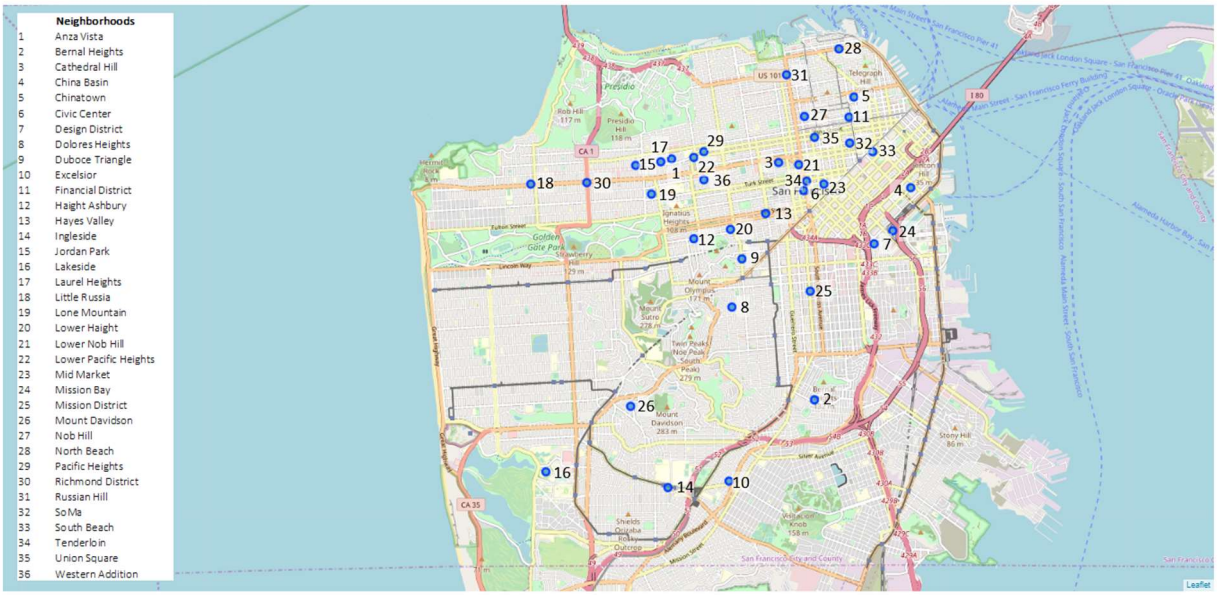
Results – Boston Neighborhoods like EV



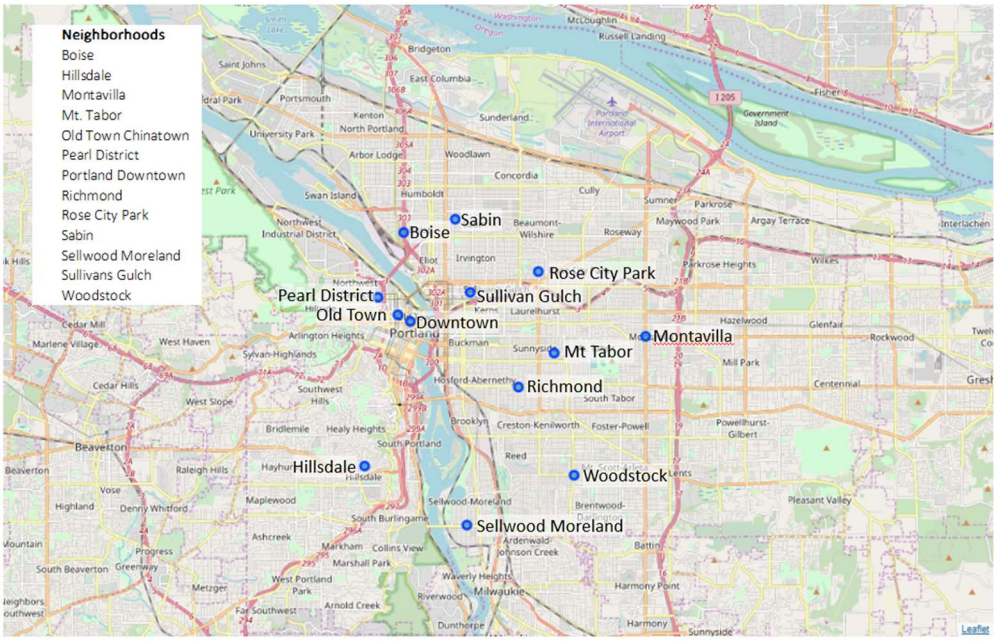
Results – Chicago Neighborhoods like EV



Results – San Francisco Neighborhoods like EV



Results – Portland, OR Neighborhoods like EV



5. Conclusions

Within the 5 cities on the top 10 public transportation list (San Francisco, Boston, Philadelphia, Chicago and Portland, OR), the method was able to analyze 8,583 venues to predict neighborhoods that are like the East Village in New York City.

The process came back with 70 neighborhoods for my friend to consider.

6. Future Considerations

Although 57 venue categories were used, data used from FourSquare might not have translated between the 5 cities since many venue categories were excluded. More data would have been helpful such as age, rent/home costs, and median income. This data was not available, for free, at the granularity that was required. Looking at preference data from FourSquare also could have made a stronger model.