Battle of the Neighborhoods: East meets West

An Analysis of the Neighborhood Venue Composition of Seattle and Boston using Clustering

Intro

An examination of differences or similarities between the neighborhoods of two different cities using Foursquare location data would prove beneficial in various scenarios. To start with, finding neighborhoods in another city that are similar to a particular neighborhood, would be of great use to a business seeking to expand outside the city, whose business depends on the environment of the neighborhood it is located in. Finding a similar neighborhood in a different city would be a step forward in replicating the business's success. Furthermore, travel websites and travelers themselves may be interested in finding out which neighborhoods are similar to ones someone has visited in the past, to serve as recommendations for new potential travel destinations.

Boston and Seattle have been selected as the two cities for comparison. Their most noticeable difference so far is that one is located on the east coast, while the other is on the west. This analysis aims to first examine the unique structure of the clusters of each city's neighborhoods. Afterward, a holistic approach will be used, taking account of the neighborhoods of both cities, to see if similar neighborhoods from both cities can be clustered together, and the implications of such for a target audience.

Data

The primary data source for this project would be venue data provided by Foursquare. This data will be used to determine the frequency of each type of venue for each neighborhood. Then it can be used to cluster the neighborhoods of each city, so analysis can be performed on the makeup of each city's clusters. Next, venue data from both cities will be pooled together and clusters will be calculated again, this time to see which neighborhoods from both cities are most similar to each other according to the clustering algorithm.

The secondary data sources for this project consist of the two geojson files of each city. The Boston geojson file provides the neighborhood outlines and neighborhood names that will be used. The Seattle geojson file originally contained data for 180 neighborhoods in the surrounding area of Seattle. So first any neighborhoods that were not affiliated with Seattle were

removed. Additionally, since Seattle's neighborhoods often are divided into many sub-neighborhoods, some of which location data could not be easily found for. So, for the purpose of simplicity, only non-sub neighborhood names will be used, basically, the larger neighborhoods that have been broken down into sub-neighborhoods.

Methodology

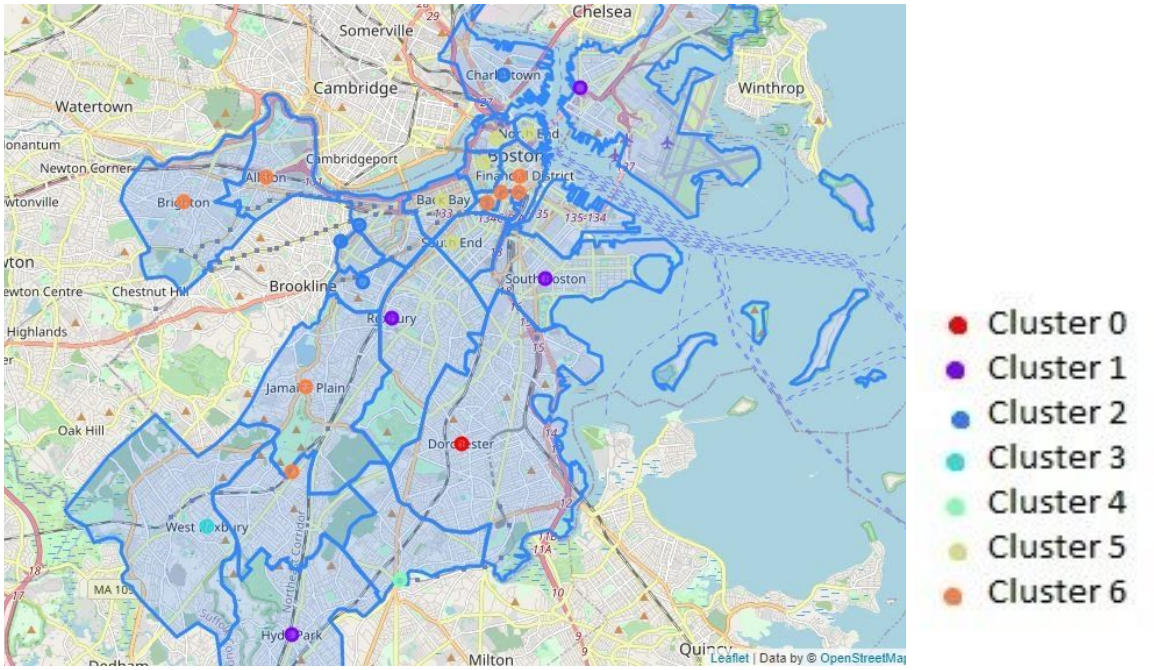
To start with, this project has two objectives, the first of which is to analyze the venue composition of each of the city's neighborhoods and see if they can be grouped together with other similar neighborhoods within the city. To do this, location data for each neighborhood can be gathered using the neighborhood names from the geojson files and the coordinates of each neighborhood can be looked up using geocoding. The resulting coordinates from this, which will now be known as location data, can be used to look up venues from each neighborhood using Foursquare API. Next the venue data from Foursquare will processed to create a neighborhood ranking list of the top 10 most common venue types for each neighborhood. Lastly, the kmeans clustering algorithm with be used on the processed location data to create k=7 clusters, as well as create a map of the neighborhoods in the cluster.

Now to analyze the types of venues that are prevalent in each cluster, let's count the number of times a venue category shows up in the top 10 most common venues of a neighborhood for all the neighborhoods in a cluster, and we'll refer to this as the number of occurrences. For example, if Pizza Place is in the 1st Most Common Venue for a neighborhood, it will count as showing up once, and if Park is in the 10th Most Common Venue it will also count as showing up as once, so we will disregard the positioning of the venue type in the neighborhood ranking list. The end goal, after all, is to find shared venue types among the neighborhoods of a cluster, and by counting the number of occurrences of a venue type among neighborhoods in a cluster, similarities can be established among neighborhoods of a cluster. For single neighborhood clusters, just looking at the ranking list of the top 10 most common venues in a neighborhood is sufficient.

The second objective of this project is to perform analysis on the pooled neighborhood data of both cities. Basically repeating the process used to analyze the neighborhoods of each individual city, but this time using data from both cities pooled together. Kmeans clustering will be used as well, with k=7 clusters, and this time a map containing the resulting neighborhood clusters of both cities will be produced. Analysis of the venue types in each cluster, by counting the number of occurrences of a venue type in a neighborhood's ranking list, will also be performed.

Results and Discussion

Boston Neighborhood Cluster Composition Analysis



Boston Clusters with more than one neighborhood

Out of the 7 clusters, 4 clusters consisted of more than one neighborhood. Let's take a look at the types of venues that made up each of these clusters. Taking only into account the top 10 most mentioned types of venues that showed in a cluster's neighborhood venue ranking.

Cluster 1 has 4 neighborhoods

	Venue Type	Number of Occurrences	Percentage
24	Sandwich Place	3.0	0.75
22	Pizza Place	3.0	0.75
20	Park	3.0	0.75
15	Italian Restaurant	3.0	0.75
14	Gym	2.0	0.50
7	Convenience Store	2.0	0.50
1	Art Gallery	2.0	0.50
16	Latin American Restaurant	1.0	0.25
27	Supermarket	1.0	0.25
26	Storage Facility	1.0	0.25

Cluster 2 has 4 neighborhoods

	Venue Type	Number of Occurrences	Percentage
3	Café	4.0	1.00
5	Coffee Shop	4.0	1.00
7	Donut Shop	4.0	1.00
19	Park	4.0	1.00
21	Pizza Place	2.0	0.50
0	American Restaurant	1.0	0.25
15	Lounge	1.0	0.25
25	Sushi Restaurant	1.0	0.25
24	Sports Bar	1.0	0.25
23	Sandwich Place	1.0	0.25

Cluster 5 has 5 neighborhoods

	Venue Type	Number of Occurrences	Percentage
5	Coffee Shop	5.0	1.0
14	Italian Restaurant	5.0	1.0
0	American Restaurant	4.0	0.8
12	Hotel	4.0	0.8
17	Park	4.0	0.8
23	Seafood Restaurant	3.0	0.6
21	Sandwich Place	2.0	0.4
9	French Restaurant	2.0	0.4
19	Pizza Place	2.0	0.4
1	Bakery	2.0	0.4

Cluster 6 has 8 neighborhoods

	Venue Type	Number of Occurrences	Percentage
4	Bakery	8.0	1.000
11	Coffee Shop	8.0	1.000
10	Chinese Restaurant	6.0	0.750
29	Sandwich Place	5.0	0.625
24	Pizza Place	4.0	0.500
3	Asian Restaurant	4.0	0.500
17	Hotel	4.0	0.500
32	Sushi Restaurant	4.0	0.500
22	Park	3.0	0.375
23	Performing Arts Venue	3.0	0.375

Boston Cluster 1

The 4 neighborhoods in cluster 1 seem to have a fair degree of commonality, with 75% of the neighborhoods having a sandwich place, pizza place, park, and Italian restaurant in their top 10 most common venues ranking. Gym, convenience store, and art gallery are honorable mentions, showing up in only half of these neighborhood's rankings

Boston Cluster 2

Cluster 2 also has 4 neighborhoods, the same as cluster 1, but cluster 2's neighborhoods have a more defining commonality that they share with each other in terms of venue rankings. All of the neighborhoods in the cluster have a cafe, coffee shop, donut shop, and park somewhere in each of their venue rankings, which is quite significant and perhaps a defining characteristic of this cluster.

Boston Cluster 5

Cluster 5 has 5 neighborhoods and they share a fairly high degree of commonality with a coffee shops and Italian restaurants showing up in all of the neighborhood’s top 10 venue rankings. It is also notable to point out that 80% of the neighborhoods also have American restaurants, hotels, and park in their rankings. Additionally, all of cluster 5’s neighborhoods are grouped together near the downtown Boston area, unlike the other clusters whose neighborhoods are spread out over the city, after all it is logical that adjacent neighborhoods would be similar.

Boston Cluster 6

Cluster 6 has 8 neighborhoods, the most out of all these clusters, and any similarities between neighborhoods in this cluster is more significant due to its size. Bakery and coffee shop show up in all 8 of these neighborhoods' venue rankings, which is quite significant, and perhaps having this pair of venues in every neighborhood is what sets this cluster apart from clusters 2 and 5 which have coffee shop paired with cafe and Italian restaurants respectively. It is also worthy to point out that Chinese restaurants, which have not shown up in the top 5 occurrences ranking of any cluster, are present in 75% of these 8 neighborhoods.

Boston Clusters with only one neighborhood

We will be using the top 10 most common venues of a neighborhood for comparison against the top 10 number of occurrences in a cluster.

Single Neighborhood Clusters

==========Cluster 0 ==========

	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
21	Dorchester	Platform	Pharmacy	Bank	Gym	Golf Course	Fried Chicken Joint	Market	Dry Cleaner	Pizza Place	Plaza

==========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
18	West Roxbury	Park	Pizza Place	Pub	Veterinarian	Liquor Store	Gym	Donut Shop	Women's Store	Dive Bar	Doctor's Office

==========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
20	Mattapan	Metro Station	Bakery	Pharmacy	Playground	Plaza	Caribbean Restaurant	Donut Shop	Mobile Phone Shop	Fast Food Restaurant	Southern / Soul Food Restaurant

Boston Cluster 0: Dorchester

Cluster 0 stands out with having platform being its most common venue. Bank, golf course, and fried chicken joint are also venues that seem to be a distinguishing characteristic of this cluster, since none of these venues showed in the the top 10 number of occurrences list of the multi-neighborhood clusters.

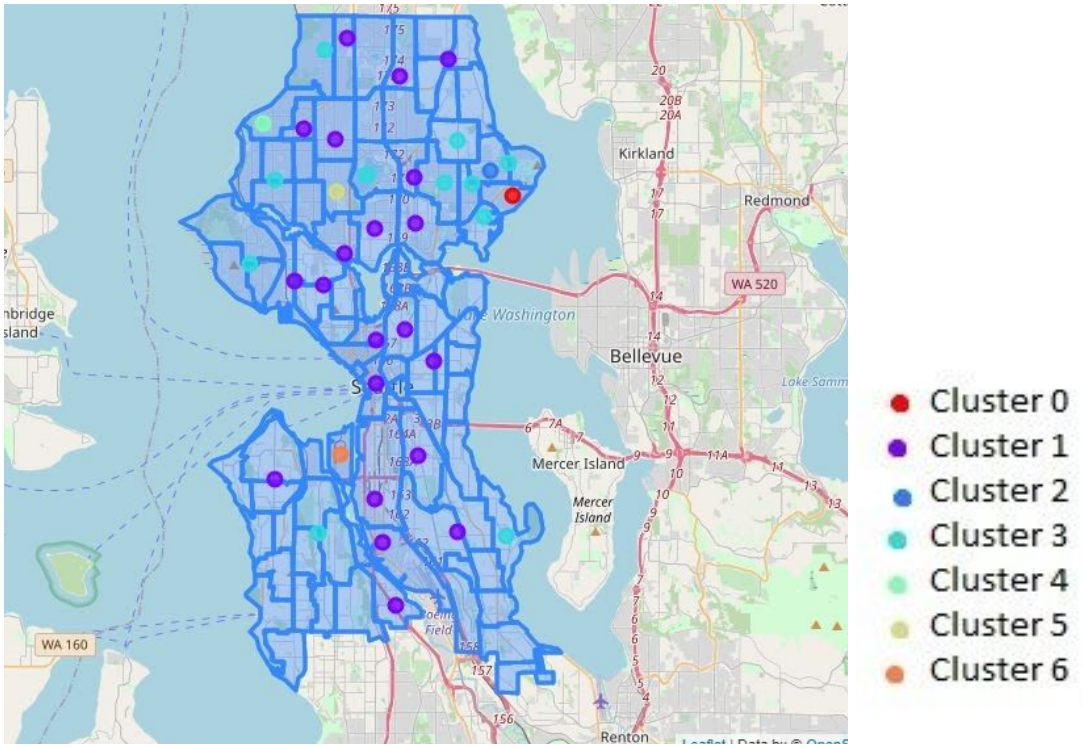
Boston Cluster 3: West Roxbury

While the venues of park, pizza place, and gym did show up in other clusters, the appearance of pub, veterinarian, and liquor store set this neighborhood apart from the rest.

Boston Cluster 4: Mattapan

Similar to cluster 0, having a metro station as its most common venue makes it immediately stand out from the rest of the neighborhoods. However, this does make some sense since Mattapan one of the neighborhoods located furthest away from the downtown Boston area. Hence, it would not be a stretch to conclude that many city commuters might reside here, leading to frequent presence of metro stations.

Seattle Neighborhood Cluster Composition Analysis



Seattle Clusters with more than one neighborhood

Out of the 7 clusters, only 2 clusters consisted of more than one neighborhood, compared to Boston where 4 of the 7 clusters had more than one neighborhood. However, these clusters are much larger than Boston's clusters, which will make for an interesting comparison.

```
Cluster 1 has 21 neighborhoods

         Venue Type  Number of Occurrences  Percentage
21       Coffee Shop                   17.0    0.809524
55       Pizza Place                   11.0    0.523810
8                Bar                   11.0    0.523810
12           Brewery                    8.0    0.380952
17              Café                    8.0    0.380952
40     Grocery Store                    7.0    0.333333
75   Thai Restaurant                    7.0    0.333333
52              Park                    6.0    0.285714
57               Pub                    6.0    0.285714
6             Bakery                    6.0    0.285714
============================================================
Cluster 3 has 11 neighborhoods

         Venue Type  Number of Occurrences  Percentage
46              Park                   11.0    1.000000
61             Trail                    5.0    0.454545
17       Coffee Shop                    4.0    0.363636
62       Video Store                    3.0    0.272727
4               Bank                    3.0    0.272727
33     Grocery Store                    3.0    0.272727
59   Thai Restaurant                    3.0    0.272727
28        Food Truck                    3.0    0.272727
14              Café                    3.0    0.272727
48          Pharmacy                    2.0    0.181818
============================================================
```

Seattle Cluster 1

Cluster 1 most noticably contains neighborhoods with a high frequency of coffee shops, with about 81% of the 21 neighborhoods in this cluster containing coffee shops in their venue ranking. However, occurrences of pizza place and bars in all these neighborhood rankings can not be overlooked, each making an appearance in 52% of the neighborhood venue rankings. Additionally, looking at the placement of the purple map markers denoting cluster 1, it seems that cluster 1 neighborhoods tend to be toward the center area of Seattle and spread out from there toward the north and south, following major Seattle roads,

Seattle Cluster 3

Cluster 3 neighborhoods tend to fall outside the central area of Seattle. A majority of the teal map markers are congregated in the north, mixed in with some of the single neighborhood clusters. While only two are in the south. As for venue composition, every neighborhood in cluster 3 has an appearance of park somewhere in their venue ranking list, which is pretty significant since there are 11 neighborhoods in this cluster. Aside from that, these neighborhoods in this cluster don't share a lot of common venues in their venue ranking list. Trail is an interesting venue in that it is a fairly unique venue type and that it shows up in about 45% of neighborhoods' venue ranking lists, which is worth an honorable mention.

Seattle Clusters with only one neighborhood

Single Neighborhood Clusters

==========Cluster 0 ==========

	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
24	Windermere	Beach	Sculpture Garden	Harbor / Marina	Pizza Place	Snack Place	Greek Restaurant	Bus Station	Zoo Exhibit	Fast Food Restaurant	Fair

==========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
20	View Ridge	Park	Harbor / Marina	Zoo Exhibit	Filipino Restaurant	Fair	Farmers Market	Fast Food Restaurant	Field	Fish Market	Event Space

==========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
34	North Beach - Blue Ridge	Home Service	Pilates Studio	Beach	Park	Food Stand	Food Service	Food Court	Flower Shop	Fish Market	Fabric Shop

==========Cluster 5 ==========

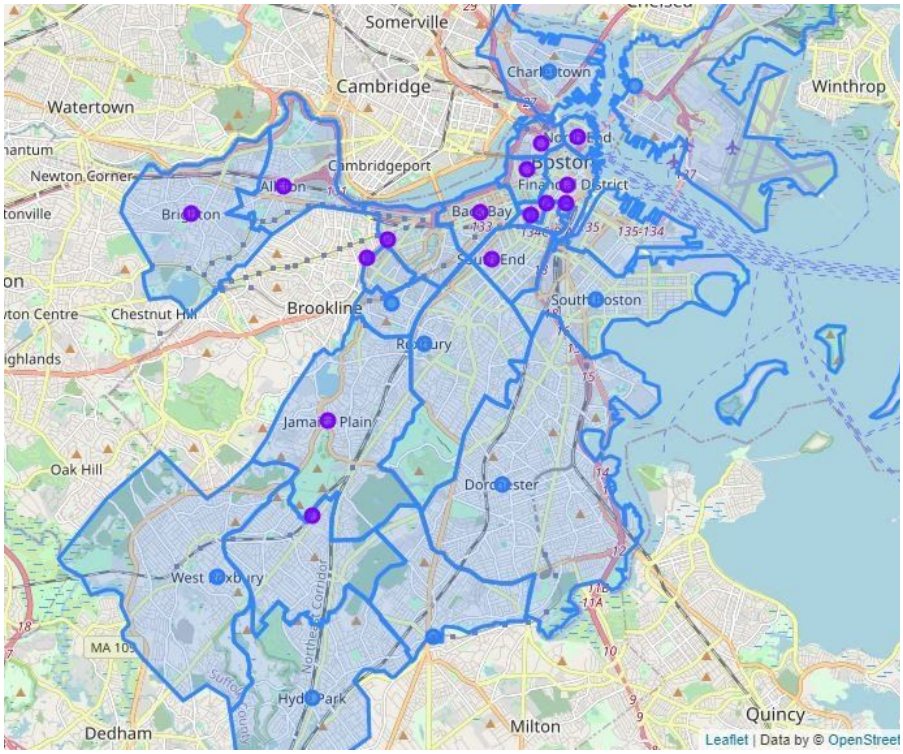
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
16	Phinney Ridge	Zoo Exhibit	Pizza Place	Café	Caribbean Restaurant	Food Truck	Mexican Restaurant	Bowling Green	Soup Place	Burger Joint	Bus Station

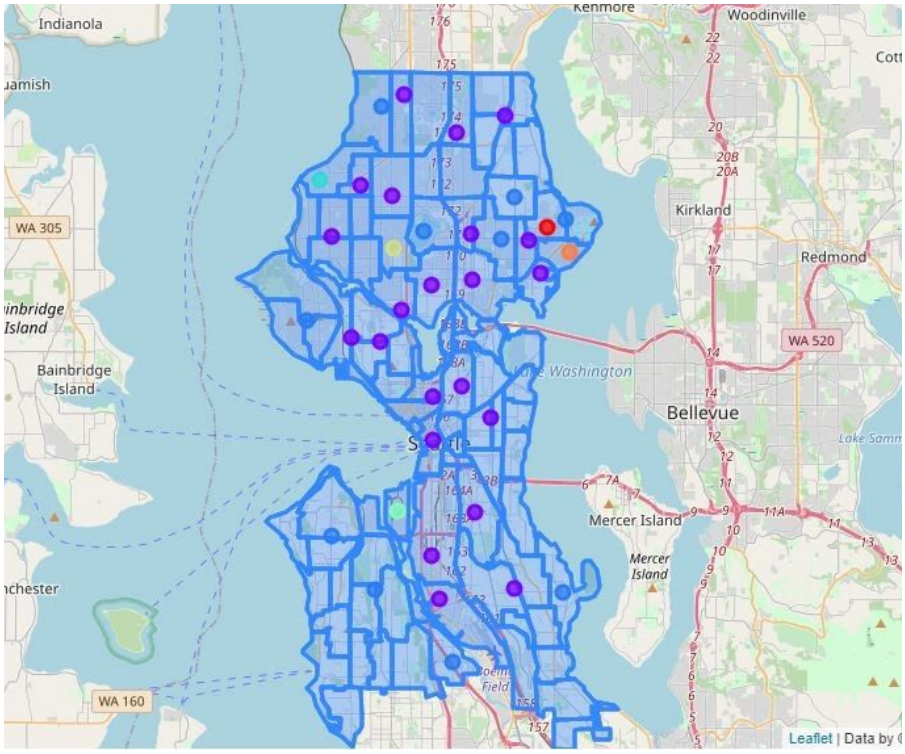
==========Cluster 6 ==========

	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
29	Harbor Island	Boat or Ferry	Pier	Zoo Exhibit	Flower Shop	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish Market	Food Court

Each of the neighborhoods here contains a diverse range of venues. Cluster 6, Harbor Island, seems to be the cluster that stands out the most in terms of interesting venues, with boat/ferry and pier being the top two venues, which makes sense given its unique geographical location in Seattle, followed by a wide range of other types of venues. Cluster 0, Windermere, is also worth mentioning due to its top two venues being beach and sculpture garden. It is also worth pointing out some of the similarities between these neighborhoods, with 4 of the neighborhoods having zoo exhibits in their venue rankings, and 3 of them having fish market.

Combined Neighborhood Cluster Composition Analysis





Combined Clusters with more than one neighborhood

```
Cluster 1 has 37 neighborhoods
Percentage of Boston Neighborhoods in this cluster:    40.54%
Percentage of Seattle Neighborhoods in this cluster:    59.46%
             Venue Type  Number of Occurrences  Percentage
24          Coffee Shop                   33.0    0.891892
7                Bakery                   19.0    0.513514
84          Pizza Place                   16.0    0.432432
80                 Park                   16.0    0.432432
9                   Bar                   13.0    0.351351
58                Hotel                   13.0    0.351351
105     Thai Restaurant                   11.0    0.297297
93       Sandwich Place                   10.0    0.270270
21   Chinese Restaurant                    9.0    0.243243
19                 Café                    9.0    0.243243
============================================================
Cluster 2 has 19 neighborhoods
Percentage of Boston Neighborhoods in this cluster:    47.37%
Percentage of Seattle Neighborhoods in this cluster:    52.63%
             Venue Type  Number of Occurrences  Percentage
61                 Park                  15.0    0.789474
65          Pizza Place                   9.0    0.473684
22           Donut Shop                   7.0    0.368421
41                  Gym                   7.0    0.368421
19          Coffee Shop                   7.0    0.368421
72       Sandwich Place                   6.0    0.315789
45  Italian Restaurant                    6.0    0.315789
40        Grocery Store                   5.0    0.263158
17                 Café                   5.0    0.263158
88                Trail                   4.0    0.210526
============================================================
```

Combined Cluster 1

Cluster 1 is the largest cluster, covering 37 neighborhoods of these two cities. Within this cluster, about 40% of the neighborhoods come from Boston, with the remaining 60% coming from Seattle. The venue type that shows up the most in the venue rankings for these neighborhoods is coffee shop, with almost 90% of the neighborhoods having this venue in their ranking lists. Bakery, pizza place, and park get an honorable mention for showing up in a noticeable percentage of these neighborhoods. And by looking at the distribution of venue types across neighborhoods of this cluster, one can see indicators that there are commonly occurring venue types that are shared among various subsets of neighborhoods in this cluster. Additionally, it seems that cluster 1's neighborhoods in the Seattle neighborhood analysis also fall into this cluster in the combined analysis, accounting for the purple map markers scattered in the middle of Seattle. While it looks that the neighborhoods from cluster 5 and 6 in the Boston analysis also made it into this combined cluster, resulting in the purple map markers concentrated in the northern, downtown area of Boston.

Combined Cluster 2

While not as large as cluster 1, cluster 2 has a decent number of neighborhoods, accounting for 19 of the neighborhoods between the two cities. Seattle still occupies about 53% of the neighborhoods in the cluster with Boston accounting for 47%, but the gap between them is less conspicuous compared to cluster 1. Park is the most commonly seen venue type that shows up in 79% of the neighborhoods' venue ranking lists. While it is interesting to see that pizza place, park, and coffee shop also show up in the breakdown of cluster 1, perhaps it is the presence of other venue types present in the cluster breakdown results that cause certain neighborhoods to

be categorized as such. The Boston neighborhoods here tend to scattered around the outskirts of the downtown area, similar to how the Seattle neighborhods here also are scattered outside of the central area of Seattle.

Single Neighborhood Combined Clusters

	Neighborhood	City	Cluster Labels	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
20	View Ridge	Seattle	0	Park	Harbor / Marina	Food Stand	Food Service	Electronics Store	Ethiopian Restaurant	Event Space	Fabric Shop	Fair	Falafel Restaurant

==========Cluster 3 ==========

	Neighborhood	City	Cluster Labels	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
34	North Beach - Blue Ridge	Seattle	3	Park	Beach	Home Service	Pilates Studio	Field	Ethiopian Restaurant	Event Space	Fabric Shop	Fair	Falafel Restaurant

==========Cluster 4 ==========

	Neighborhood	City	Cluster Labels	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
29	Harbor Island	Seattle	4	Boat or Ferry	Pier	Fish Market	Event Space	Fabric Shop	Fair	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field

==========Cluster 5 ==========

	Neighborhood	City	Cluster Labels	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
16	Phinney Ridge	Seattle	5	Zoo Exhibit	Pizza Place	Café	Food Truck	Sandwich Place	Salon / Barbershop	Pub	Playground	Pie Shop	Park

==========Cluster 6 ==========

	Neighborhood	City	Cluster Labels	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
24	Windermere	Seattle	6	Greek Restaurant	Harbor / Marina	Beach	Pizza Place	Snack Place	Sculpture Garden	Bus Station	Zoo	Farmers Market	Event Space

All of the single neighborhood clusters here are from Seattle. And if that weren't interesting enough, all of these neighborhoods also appeared as single cluster neighborhoods in the Seattle cluster analysis. It can be speculated the uniqueness of the venue types that show up in these neighborhoods caused the algorithm to single them out and leave them on their own because they don't really fit in with the rest, which does make some sense, since coffee shop didn't make an appearance at all in these neighborhood ranking lists, and while pizza place and park occasionally did, they were overshadowed by the appearance of more unique venue types. Furthermore, if the number of clusters was reduced, perhaps some of these neighborhoods could be clustered together since there are some shared venue types among them.

Conclusion

In conclusion, both Seattle and Boston have neighborhoods with a diverse range of venues. Boston neighborhoods have a large emphasis on food with lots of food related venues such as coffee shops, pizza places, sandwich places, and restaurants. While Seattle's neighborhoods also have a high frequency of coffee shops, no surprise since Starbucks was founded and is headquartered there, they also contain a variety of other venues to get a drink from, such as pubs, bar, and cafes. Additionally, Seattle neighborhoods placed a greater emphasis on the great outdoors, with all 11 neighborhoods in a cluster having parks as a venue in their top 10 most common venue ranking list. When it came to the combined analysis, all of Boston's neighborhoods fit into 2 clusters with the majority of Seattle's neighborhoods. The most populated combined neighborhood cluster, placed great emphasis on food venues, while the other had a fair mix of venues, making it pretty well rounded in terms of venues.

The takeaway from all this analysis and information? A coffee shop owner can narrow down location choices for a new location by taking a look a cluster which doesn't have a high prevalence of coffee shops in its neighborhoods. For instance, looking at neighborhoods in combined cluster 2 instead of combined cluster 1, whose neighborhoods already have a high prevalence of coffee shops. A travel website algorithm could recommend users who've visited a neighborhood in a certain cluster, other neighborhoods in that cluster to visit. And this works with neighborhoods with a city and neighborhoods among cities. So neighborhoods with Boston cluster 2 could be recommended as well as neighborhoods in combined cluster 2.