Recommending Restaurant Locations with Clustering

IAN CARVER

Purpose

Large cities like New York have an abundance of restaurants, which can make it difficult for prospective restaurant owners to choose a location.

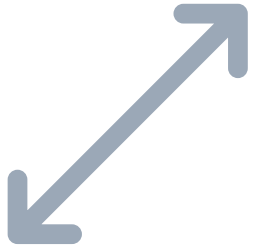
If they choose an area that already has many restaurants, they may be in direct competition with surrounding restaurants that serve similar food and will be unable to find a niche.

If they choose an area that has very few restaurants, there may not be a demand for food in that area.

Solution

Using a K-Nearest-Neighbor algorithm, it is possible to group neighborhoods by their restaurant density and the types of cuisine they have already.

By selecting the cluster with high average density, a prospective restaurant owner should be able to find out what type of restaurant isn't common in the area already, which means that it wouldn't be in direct competition with nearby restaurants.



In the future, once the first location is successful, the owner can select other neighborhoods in the same cluster for additional locations to expand into.

Data

Neighborhood data, including latitude and longitude.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Restaurant data, including the venue category.

Venue	Venue Latitude	Venue Longitude	Venue Category
Roselle Desserts	43.653447	-79.362017	Bakery
Tandem Coffee	43.653559	-79.361809	Coffee Shop
Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center

Derived Features

Scaled frequency of each venue category in each neighborhood.

	Neighborhood	American Restaurant	Bagel Shop	Bakery	Burger Joint	Café
0	Allerton	0.047619	0.000000	0.142857	0.0	0.0
1	Annadale	0.166667	0.000000	0.083333	0.0	0.0
2	Arden Heights	0.000000	0.000000	0.000000	0.0	0.0
3	Arlington	0.250000	0.000000	0.000000	0.0	0.0

Restaurant density, based off the scaled number of restaurants, in each neighborhood.

	Density
Neighborhood	
Allerton	0.277778
Annadale	0.152778
Arden Heights	0.013889
Arlington	0.041667
Arrochar	0.138889

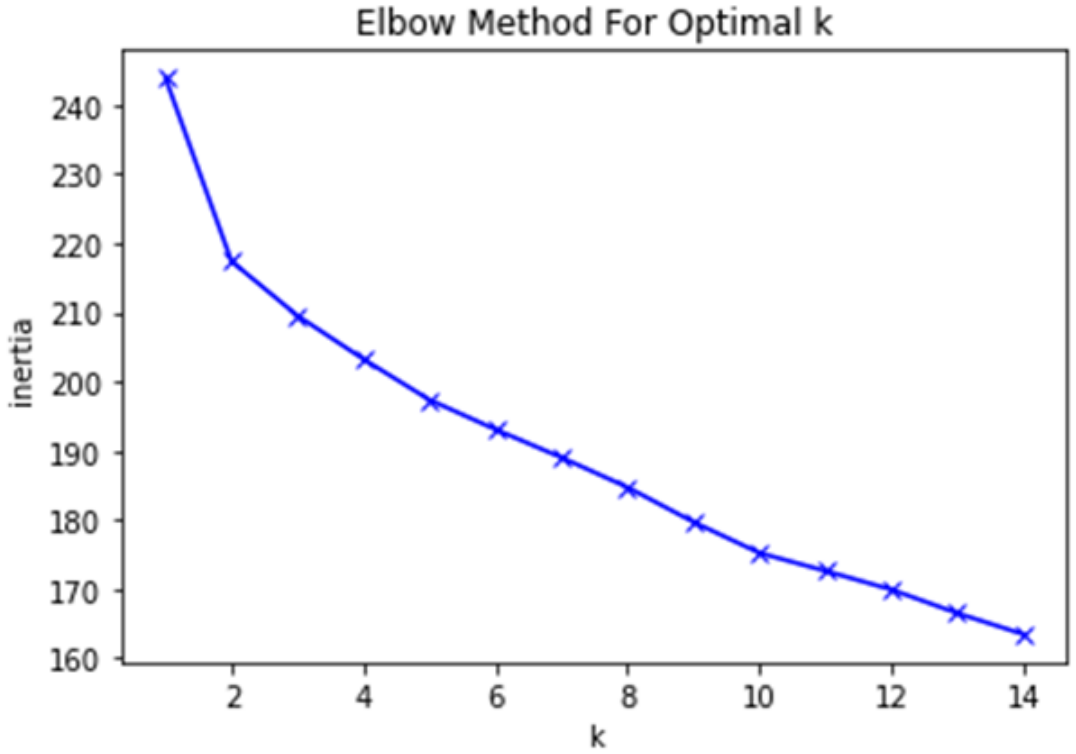
	Neighborhood	Density	American Restaurant	Bagel Shop	Bakery	Burger Joint	Café	Chinese Restaurant	Deli / Bodega	Diner	...	Food	Food Truck	Italian Restaurant
0	Allerton	0.277778	0.047619	0.000000	0.142857	0.0	0.0	0.142857	0.190476	0.000000	...	0.095238	0.000000	0.000000
1	Annadale	0.152778	0.166667	0.000000	0.083333	0.0	0.0	0.000000	0.083333	0.083333	...	0.083333	0.000000	0.000000
2	Arden Heights	0.013889	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.500000	0.000000	...	0.000000	0.000000	0.000000
3	Arlington	0.041667	0.250000	0.000000	0.000000	0.0	0.0	0.000000	0.250000	0.000000	...	0.000000	0.000000	0.000000
4	Arrochar	0.138889	0.000000	0.181818	0.000000	0.0	0.0	0.000000	0.181818	0.000000	...	0.000000	0.090909	0.181818

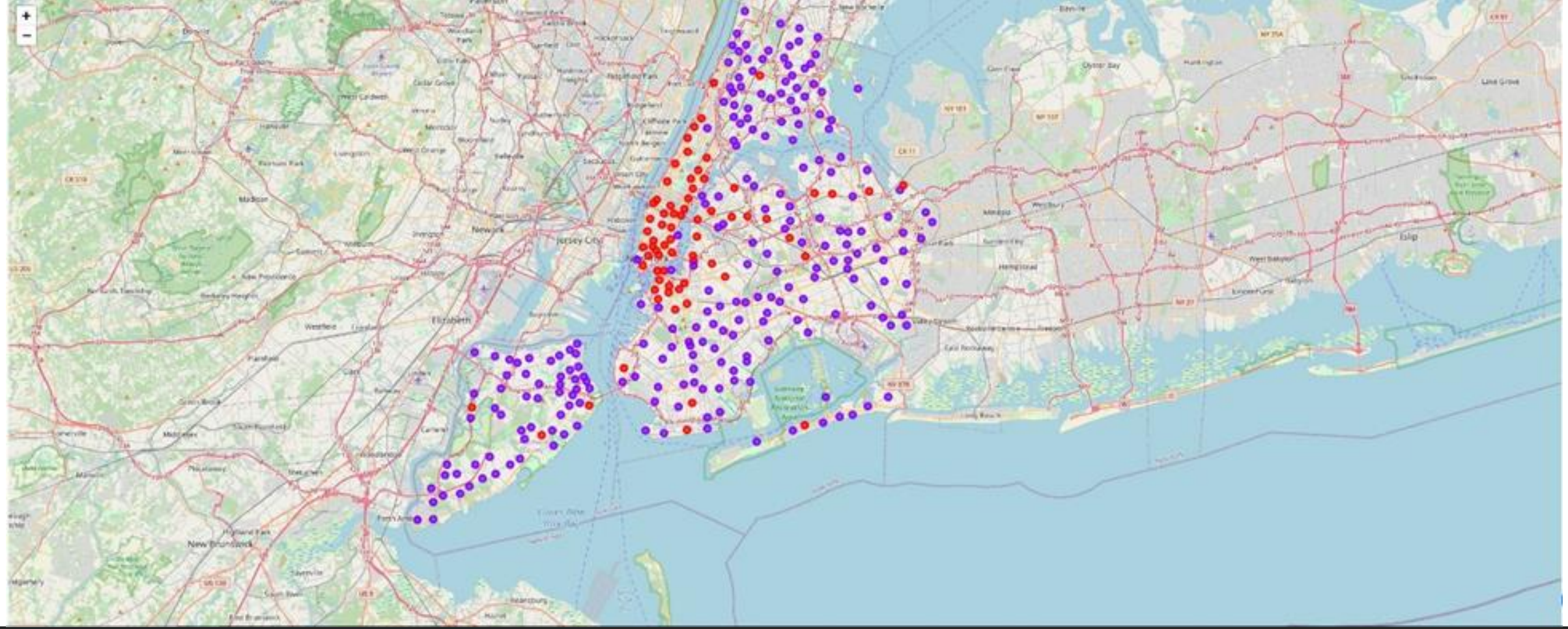
Final Dataframe

All features were Min-Max scaled for performance.

Optimal Hyperparameter

Using the elbow method for finding the optimal value of K for KNN, the optimal value was K=2.





Location Clusters

The clusters seem to be divided into one group that is primarily Manhattan neighborhoods, and then all non-Manhattan neighborhoods.

Optimal Restaurant Selection

Cluster 1 has a higher average density.

By counting the number of times each category is in the top 5 least common restaurants, we can determine the optimal cuisine.

Then the optimal location can be selected using the highest density that lacks the optimal cuisine.

	1st Least Common Venue	2nd Least Common Venue	3rd Least Common Venue	4th Least Common Venue	5th Least Common Venue	Totals
Vietnamese Restaurant	26.0	0.0	0.0	0.0	0.0	26.0
Vegetarian / Vegan Restaurant	0.0	7.0	9.0	5.0	3.0	24.0
Korean Restaurant	0.0	0.0	6.0	7.0	7.0	20.0
Latin American Restaurant	0.0	4.0	4.0	7.0	4.0	19.0
Greek Restaurant	1.0	6.0	4.0	3.0	3.0	17.0

	Neighborhood	Density	Borough	Latitude	Longitude	Cluster Labels	1st Least Common Venue	2nd Least Common Venue	3rd Least Common Venue	4th Least Common Venue	5th Least Common Venue
174	Midtown South	1.000000	Manhattan	40.748510	-73.988713	0.0	Vietnamese Restaurant	Seafood Restaurant	Noodle House	Middle Eastern Restaurant	Mexican Restaurant
150	Lenox Hill	1.000000	Manhattan	40.768113	-73.958860	0.0	Vietnamese Restaurant	Snack Place	Food	Fast Food Restaurant	Falafel Restaurant

Conclusion

THE OPTIMAL RESTAURANT TO OPEN IN NYC BASED UPON THIS TECHNIQUE WOULD BE A VIETNAMESE RESTAURANT IN MIDTOWN SOUTH OR LENOX HILL IN MANHATTAN.

Future Improvement

Using restaurant data from a different source may result in better category separation, i.e. Vietnamese versus Vietnamese fusion.

Changing the neighborhoods to a different division to control for geographic area or population density may produce a different result.