

Clustering of Neighborhoods in Newport County, Rhode Island

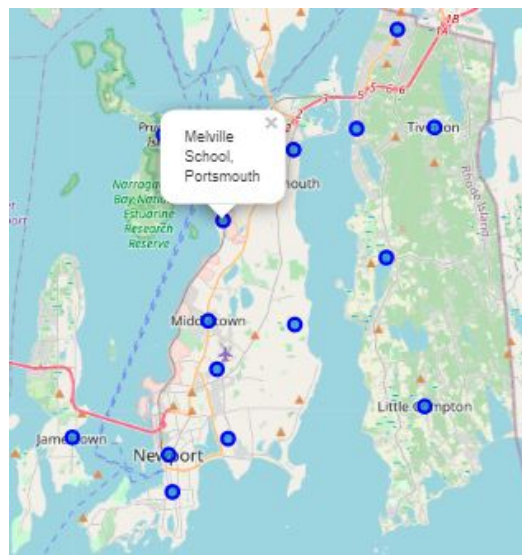
Gregory Decoteaux

November 6, 2020

1. Introduction

1a. The Business problem

Joe is in the U.S. Navy and has been stationed with his family in Newport County, Rhode Island. Having no idea about the area, Joe found a place in the Melville area of Portsmouth, RI on the same Island as the base. After visiting the area before moving in, Joe found that he prefers the Newport area due to its old historic neighborhood and scenic beaches, but his rental that he picked out is very close to the elementary school and he would like to live near an elementary school.



1. Joe wants to find a similar neighborhood to the one with Melville Elementary School.
2. Joe wants to find a neighborhood close to an elementary school.
3. Joe wants to find a neighborhood much closer to the First and Second town beaches on southern Aquidneck island.

1b. Background info on the Area

Newport County is one of the five counties in Rhode Island, in addition to Bristol, Kent, Providence and Washington Counties. According to wikipedia, It was “one of the two original counties of the Colony of Rhode Island and Providence Plantations.” (“Newport County, Rhode Island”) It currently is made of one city, Newport, and five towns. Most of the county is on Rhode Island, which is usually referred to as its original name (Aquidneck Island) to avoid confusion with the State name. Newport is a tourist attraction due to its beaches and old colonial buildings and sailing heritage. The island has both a Naval Base and the Naval War college, so many US Sailors are stationed there.

2. Data acquisition and cleaning

I first needed data on the location of elementary schools in Newport county. I found the list I decided to take the list of Elementary schools in Rhode Island from Wikipedia and only add the ones from Newport County to my data frame. Using Geopy, and reading about the schools on their webpages, I found some of the info on the wiki to be incorrect, such as school no longer being open. To deal with this I changed the name of the school to the city name and dropped some redundant rows where the same name appeared. I was left with 15 neighborhoods.

	School	City	State
0	Aquidneck School	Middletown	Rhode Island
1	Carey School	Newport	Rhode Island
2	Elmhurst School	Portsmouth	Rhode Island
3	Forest Avenue School	Middletown	Rhode Island
4	Fort Barton School	Tiverton	Rhode Island
5	Howard Hathaway School	Portsmouth	Rhode Island
6	Jamestown	Jamestown	Rhode Island
7	Middletown	Middletown	Rhode Island
8	Melville School	Portsmouth	Rhode Island
9	Nonquit School	Tiverton	Rhode Island
10	Pocasset School	Tiverton	Rhode Island
11	Prudence Island School	Prudence Island	Rhode Island
12	Tiverton	Tiverton	Rhode Island
13	Newport	Newport	Rhode Island
14	Little Compton	Little Compton	Rhode Island

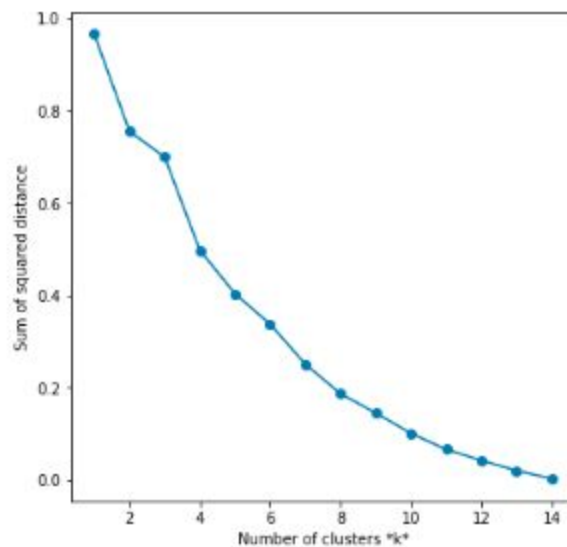
S.

I found the locations of the three beaches using google maps and created a data frame with the names and addresses.

	Beach	Address
0	Eastons Beach	175 Memorial Blvd
1	Second Beach	315 Sachuest Point Rd
2	Third Beach	Hanging Rock rd

3. Methodology

In order to get a sense of Joe's neighborhood I used foursquare to collect information on the venues near each of the schools. I used machine learning to find the k means and to cluster the neighborhoods. Once done I would be able to use the cluster containing Melville to find appropriate neighborhoods to fill Joe's first criteria and second criteria. In order to determine the correct number of clusters, I employed the elbow method and calculated the results with clusters. The graph showed that the Sum of squared distance from the data points to the clusters decreases nearly linearly until reaching k=6. At that point, there is a non linear change in the slope. Seeing this change in data, I used k=6 as the "elbow" and ran the clustering algorithm with that value for k.



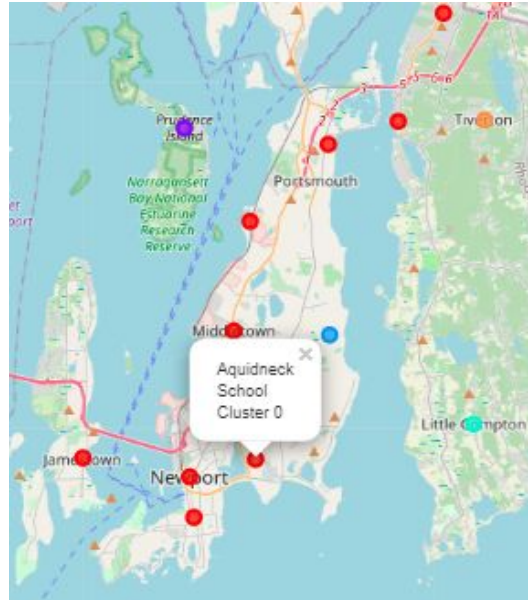
The third criteria depended on first finding neighborhoods in the same cluster and then comparing distance. I used geopy's distance method to calculate the distance between each neighborhood and each southern beach. I then was able to find the average distance between each neighborhood and each beach. I used the elbow method and rested

4. Results

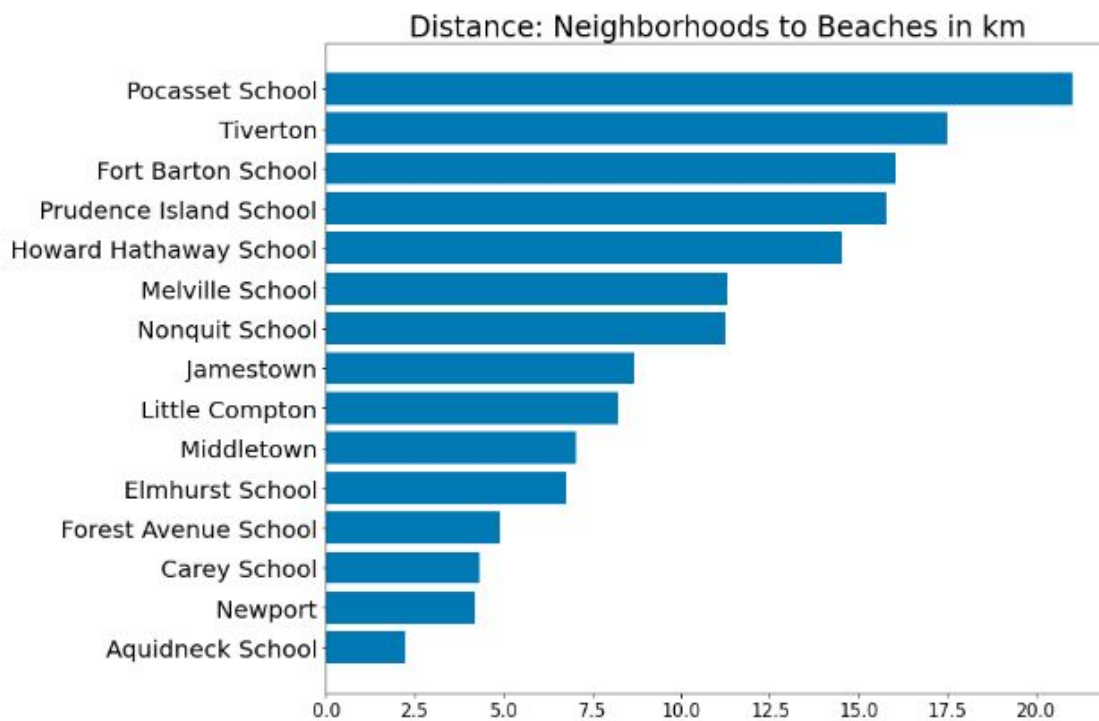
As can be seen from the figure below most of the neighborhoods most common attractions are american and seafood restaurants. Other forms of restaurants, hotels and historic sites show up frequently.

	School	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Aquidneck School	American Restaurant	Seafood Restaurant	Beach	Pizza Place	Scenic Lookout
1	Carey School	American Restaurant	Seafood Restaurant	Hotel	Historic Site	Pub
2	Elmhurst School	Trail	Park	Vineyard	Beach	Athletics & Sports
3	Forest Avenue School	Hotel	Airport	Bakery	Donut Shop	Gym / Fitness Center
4	Fort Barton School	American Restaurant	Harbor / Marina	Café	Coffee Shop	Liquor Store
5	Howard Hathaway School	American Restaurant	Hostel	Beach	Grocery Store	Liquor Store
6	Jamestown	Harbor / Marina	Hotel	Ice Cream Shop	Beach	City
7	Little Compton	Pizza Place	Gift Shop	Breakfast Spot	Post Office	Thrift / Vintage Store
8	Melville School	Harbor / Marina	American Restaurant	Pizza Place	Museum	Café
9	Middletown	Airport	Sporting Goods Shop	Gas Station	Cosmetics Shop	Coffee Shop
10	Newport	Seafood Restaurant	American Restaurant	Bar	Hotel	Pub
11	Nonquit School	Deli / Bodega	Cheese Shop	American Restaurant	Art Gallery	Arts & Crafts Store
12	Pocasset School	American Restaurant	Pharmacy	Pizza Place	Convenience Store	Food & Drink Shop
13	Prudence Island School	Beach	Convenience Store	Boat or Ferry	Other Great Outdoors	Candy Store
14	Tiverton	Sandwich Place	Ice Cream Shop	Donut Shop	Playground	Spa

After creating clusters, it was shown that most of the places along the western and southern shores of the island all fell into the same cluster. Using this map, It was easy to see that there were lots of options for Joe's Selection.



The distance graph shows the Carey School and Aquidneck Schools with the lowest average distance to the beaches. Using the data analysis, a recommendation can be given to Joe that the Aquidneck School is the best choice.



5. Conclusions

The Aquidneck school is similar to the Melville School according to the machine learning algorithms used as both schools fall into the same cluster, based on the venues located within 500 meters of the schools. The aquidneck school is on average 2.1 km closer to the beaches than the Carey school, while the Melville school is 11.3 kilometers away from the beaches.

Given all of this info I suggest Joe seek housing in the Aquidneck School's neighborhood to fulfill Joe's first requirement. Second, the aquidneck school is clustered with the Melville school, indicating that its neighborhood is similar. Third, of all the elementary schools in the county, the Aquidneck School is closest. The Carey School falls at a close second option if the housing market near aquidneck does not provide an ample selection. The Carey School neighborhood falls second, only due to its slightly farther distance from the beaches.

Bibliography

"Newport County, Rhode Island." *Wikipedia*, 2020,

https://en.wikipedia.org/wiki/Newport_County,_Rhode_Island. Accessed 11 11 2020.