Establish a repository and distribution center for a book publisher in New York State

By Haamed Momeni

GitHub: https://github.com/haamedmomeni

Introduction

The goal of this project is to find the best place to establish a repository and distribution center for a book publisher in New York State. Now online shopping is a common way to buy merchandise from home. Then, finding the best point is a problem for suppliers to minimize the cost of shipping. In this project, all counties of New York state evaluated as a candidate. The factors to find the best place based on the overall displacement of the Target Community. For this purpose, the population, number of schools, and existing book stores are collected for each county. Then, considering the displacement between counties to all other counties calculated. Using all these factors, the best candidate is chosen between all counties.
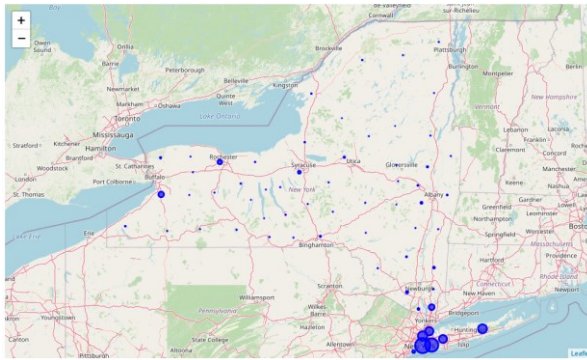
Dataset

In this section, two available datasets are used which have a detailed data of all counties in New York State. One of them is a general dataset from Wikipedia[1] and another one from State of New York official website[2]. The geolocation data of these counties are extracted from FourSquare[3]. The final extracted dataset is presented in the page.
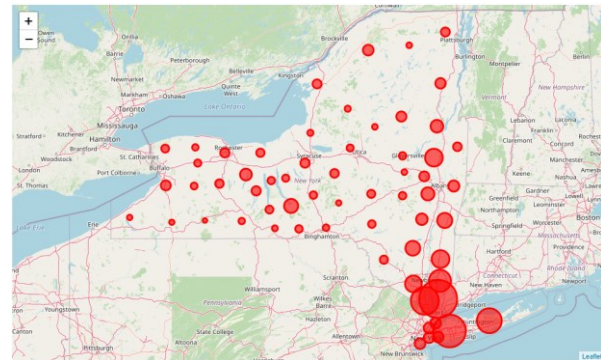
	County	Population	Density (Pop./mi2)	Area (mi2)	Latitude	Longitude	LandPrice
0	Albany	304204	570.74	533.00	42.598690	-73.984400	217500.0
1	Allegany	48946	47.34	1033.92	42.244606	-78.041928	75500.0
2	Bronx	1385108	24118.20	57.43	40.850485	-73.840404	NaN
3	Broome	200600	280.56	715.00	42.145562	-75.840411	110000.0
4	Cattaraugus	80317	61.31	1310.01	42.223482	-78.647710	87250.0

For the missed values, the average of land price of NY State is used.

The distribution of population and also the price of land are shown in the next figures.

(a) distribution of population in NYS (b) Land price in NYS

---

[1] https://wikipedia.org/wiki/List_of_counties_in_New_York
[2] https://www.tax.ny.gov/research/property/assess/sales/resmedian.htm
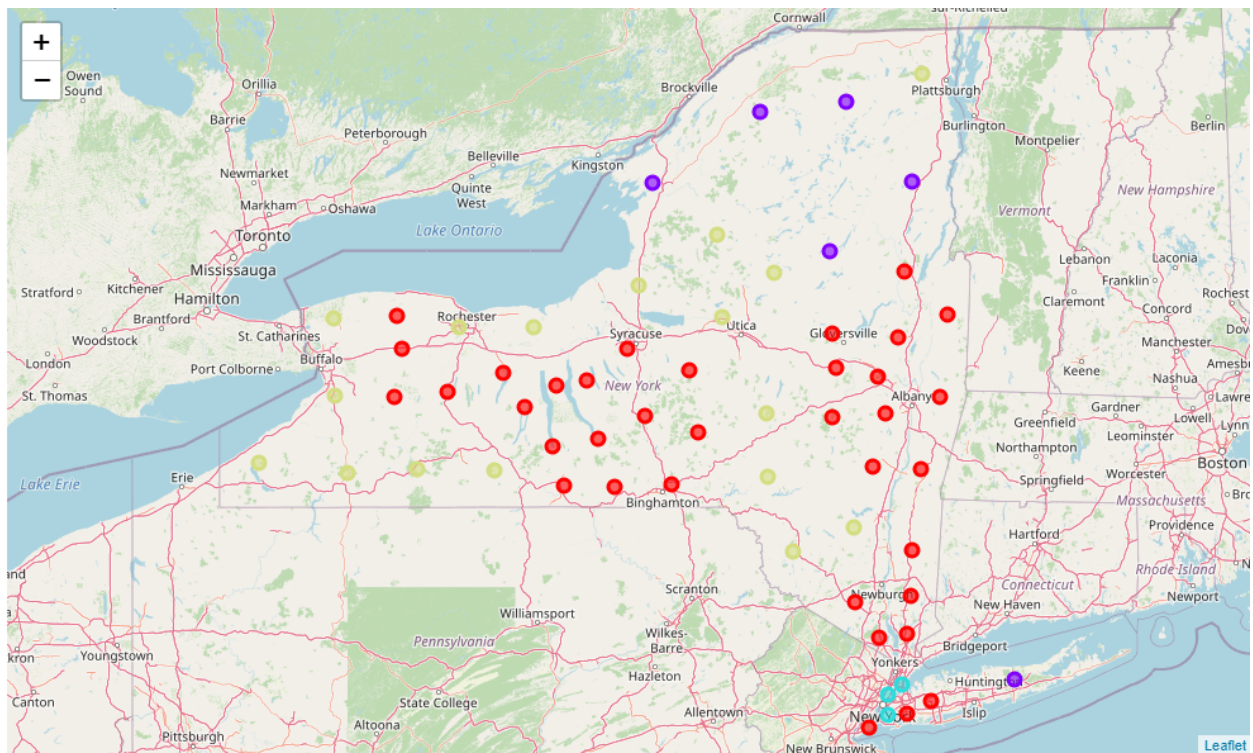[3] https://foursquare.com/

Clustering

Using this dataset, a clustering with 4 number of clusters is done to have a better comprehension about the counties.

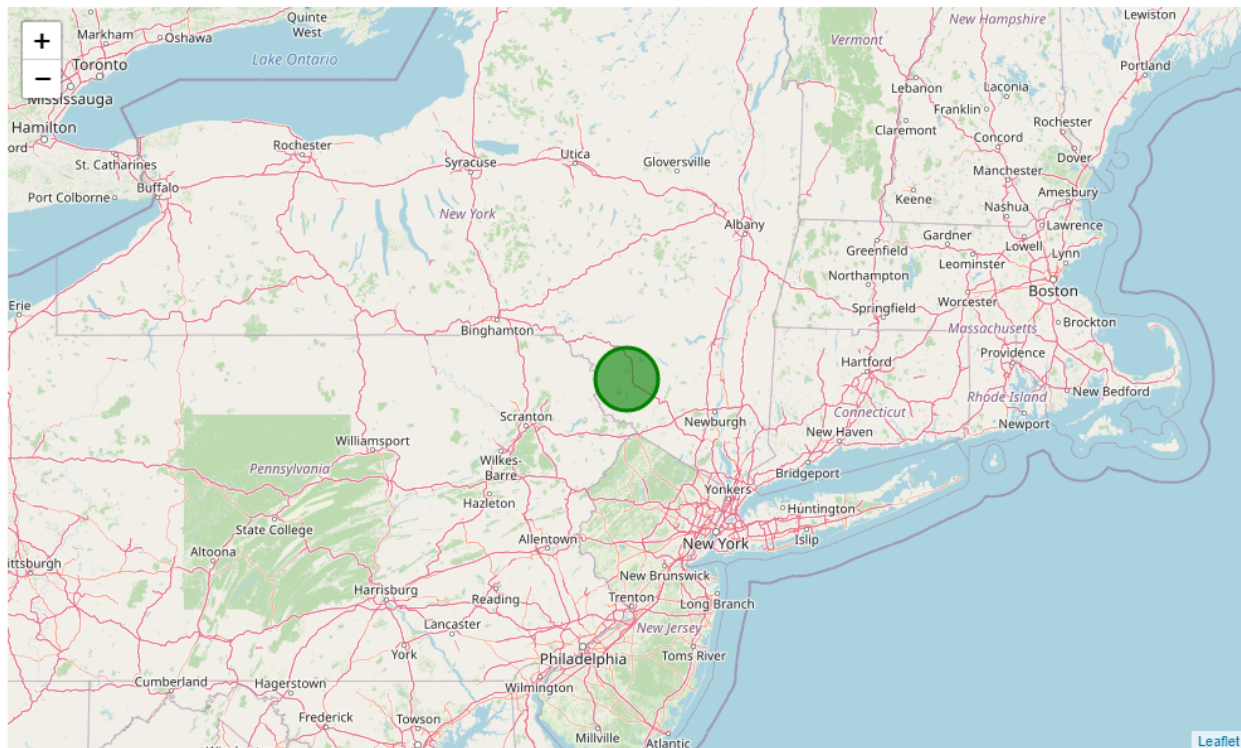|	County	Population	Density (Pop./mi2)	Area (mi2)	Latitude	Longitude	Labels	WeightedDistance
0	Albany	304204	570.74	533.00	42.598690	-73.984400	0	4.047503e+07
1	Allegany	48946	47.34	1033.92	42.244606	-78.041928	3	7.303513e+07
2	Bronx	1385108	24118.20	57.43	40.850485	-73.840404	2	2.675696e+07
3	Broome	200600	280.56	715.00	42.145562	-75.840411	0	4.474017e+07
4	Cattaraugus	80317	61.31	1310.01	42.223482	-78.647710	3	8.265323e+07

Finding best point for book distributer hub center

In this section an indictor which shows the overall cost of establishing a book distributer hub calculated in each county. For this purpose, a weighted average of all other counties to the related county is calculated. But, in each step, the population of destination county is multiplied by the population of that counties. This indicator shows us the overall cost of shipment of books for each candidate point. The other factor is the price of buying a land in each county. Considering these two factors an overall cost indicator is calculated and all the counties sorted by lesser values.

Results

The results are showed in the next table. As you can see, Sullivan county is the best place to establish a book distributer hub center

	County	Population	Density (Pop./mi2)	Area (mi2)	Latitude	Longitude	Labels	WeightedDistance	Cost
52	Sullivan	77547	77.78	997.00	41.715631	-74.780432	3	3.400705e+07	3.813437e+09
30	New York	1585873	46961.00	33.77	40.781018	-73.959300	2	2.671660e+07	3.879229e+09
2	Bronx	1385108	24118.20	57.43	40.850485	-73.840404	2	2.675696e+07	3.881732e+09
28	Montgomery	50219	122.49	409.98	42.894127	-74.409975	0	4.362945e+07	3.883026e+09
8	Chenango	50477	56.16	898.81	42.478457	-75.613028	0	4.469691e+07	3.887209e+09

Discussion and Conclusion

In this project all counties in New York State are evaluated as candidate for the best place to establish a repository and distribution center for a book publisher used. An indicator is established to find the best point. The results show us Sullivan county is the best choice which has a good position to ship the books to the customers and also it has a suitable land price. In this project, two main indicators are used which are the distance of population to the candidate points and land price. To have a more reliable results other parameters such as labor cost tax fees and etc. should be involved.