

Segmenting neighborhoods of Toronto and Manhattan using Data from FourthSquare

1.Introduction

The aim of my study is to help an entrepreneur Mr. X to grow its business. X opened a restaurant in Manhattan last year and had seen a huge success. Hence, Mr. X wishes to expand fastly and to open a second restaurant which serves the same customer base but this time in another city of another country, let's suppose it is Toronto.

2.Data & Methodology

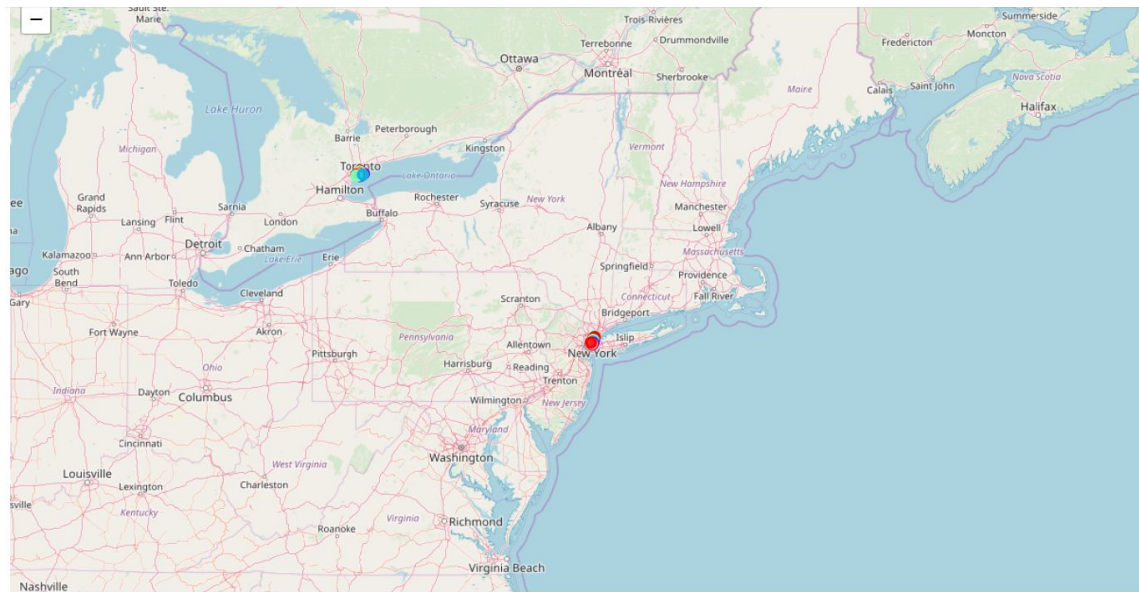
As explained above, it is required to explore, segment, and cluster the neighborhoods in Manhattan and Toronto. We download two datasets from the week 3 of this course. In Dataset 1 & Dataset 2, a list of all the neighborhoods in Manhattan and Toronto respectively with their geospatial information is available.

We combine the 2 Datasets which results in a single dataset containing information about both cities' neighborhoods. Having the geospatial information, we link the data to FourthSquare, hence for each neighborhood we find the venues (restaurants/ coffee shops...) which are nearby. So far, for each neighborhood we have the venues which are nearby, which implies if we have 100 venues close by, we will have 100 rows corresponding to these venues in the dataset. We groupby the data, so instead of having 100 rows for a neighborhood we get a single line with the neighborhood and the average number of each of the venues categories nearby. We apply the k-means clustering algorithm based on the previous information. As a result, for each neighborhood we get a label and the neighborhoods are now clustered. To check where Mr. X must open its restaurant, we look up where its old restaurant was, check to which cluster does it belong, and then find neighborhoods in Toronto which belong to the

same cluster.

36	Manhattan	Tudor City	40.746917	-73.971219
37	Manhattan	Stuyvesant Town	40.731000	-73.974052
38	Manhattan	Flatiron	40.739673	-73.990947
39	Manhattan	Hudson Yards	40.756658	-74.000111
40	East Toronto	The Beaches	43.676357	-79.293031
41	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188
42	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572
43	East Toronto	Studio District	43.659526	-79.340923
44	Central Toronto	Lawrence Park	43.728020	-79.388790
45	Central Toronto	Davisville North	43.712751	-79.390197
46	Central Toronto	North Toronto West	43.715383	-79.405678
47	Central Toronto	Davisville	43.704324	-79.388790
48	Central Toronto	Moore Park,Summerhill East	43.689574	-79.383160
49	Central Toronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	43.686412	-79.400049
50	Downtown Toronto	Rosedale	43.679563	-79.377529
51	Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675
52	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
53	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636
54	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
55	Downtown Toronto	St. James Town	43.651494	-79.375418

3. Results & Discussion



As a result, the neighborhoods in Toronto and Manhattan are clustered into 5 groups. So, it would be easy to use the map and find similar neighborhoods to open a new restaurant.

4. Conclusion

This study is a live example of how data science can be used to tackle daily life problems.