

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.

3. Results & Discussion

4. Conclusion