Study on Similarity or Dissimilarity of New York City and Toronto Neighborhoods

1. Introduction

1.1. Background

New York City and Toronto are both prior cities in North America, and there are some common culture features between the two cities. Both cities accommodate millions of tourists every year. The hypothetical scenario here is that a successful tourist company based in New York City wants to expand their business to Toronto. Shareholders of the company want to know if they need similar or different expertise, management pattern or marketing strategy in future business in Toronto.

1.2. Problem

Data about mixture of most popular ventures in New York City and Toronto may be analyzed for the purpose of the hypothetical company. This project aims to investigate how similar or dissimilar the two cities are for the tourist.

1.3. Interest

The hypothetical company would be very interested in the report, and potential tourists of New York City and Toronto may be also interested in the report.

2. Data

2.1. Data Source

Data about ventures of New York City and Toronto can be found in Foursquare Location Database. Version “20180605” is used in Foursquare Location API.

2.2. Data Cleaning and Feature Selection

Data downloaded from Foursquare includes Venture Names, Categories and Popularities of the ventures.

The top 10 popular ventures and their categories of each neighborhood are processed to the training data. The percentage of each category of 10 ventures in all categories in each neighborhood are calculated and analyzed as training data.

There are 140 observations and 399 features in training data set.

3. Clustering Model

K-means method is applied on the training data set and the number of clusters is set to 5. The result is presented in following tables.

Table 1. Clusters of Toronto Neighborhoods

Clusters	Neighborhood Numbers	Percentages
0	2	2%
1	90	87%
2	9	9%
3	1	1%
4	1	1%

Total	103	100%

Table 2. Clusters of New York City Neighborhoods

Clusters	Neighborhood Numbers	Percentages
0	0	0%
1	40	100%
2	0	0%
3	0	0%
4	0	0%
Total	40	100%

Result of the clustering model is also visualized in maps of two cities as following.

Figure 1. Clustering of Toronto Neighborhoods

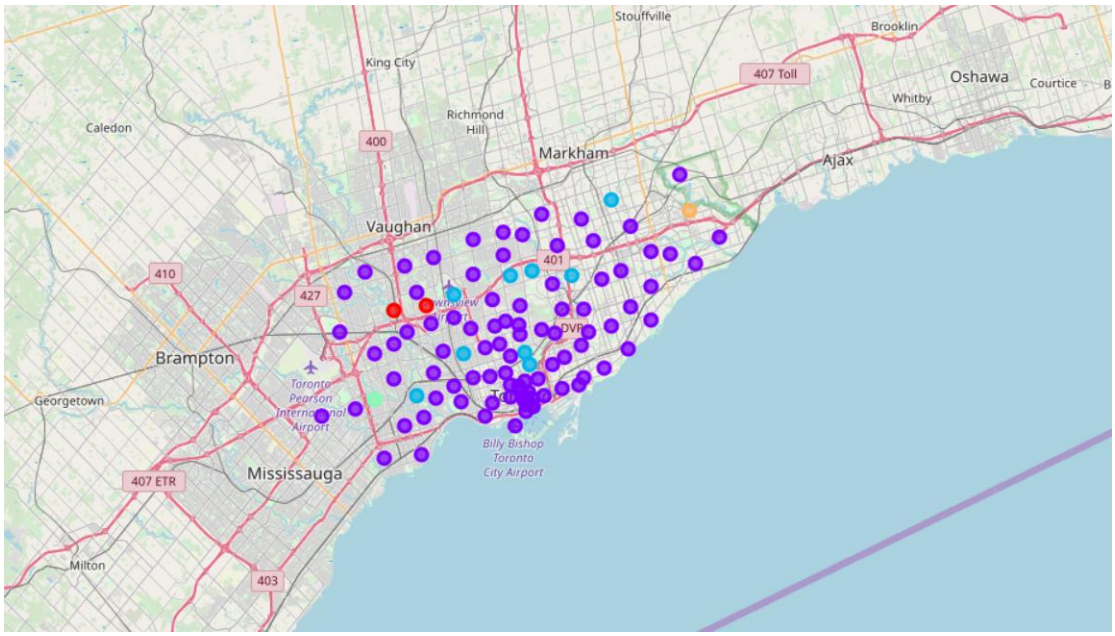
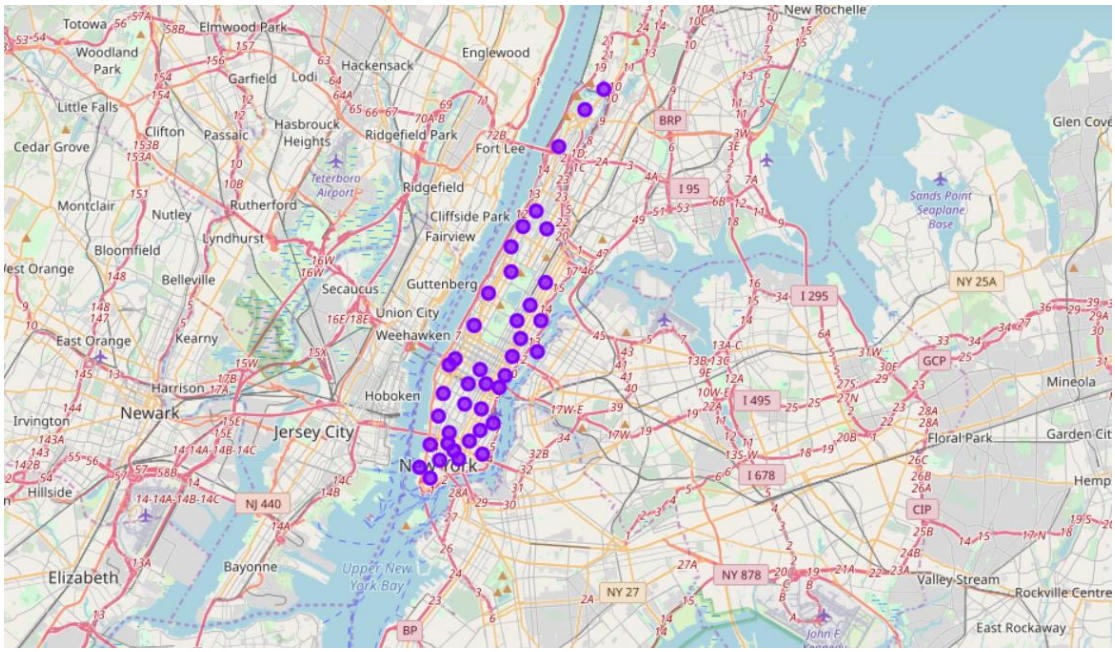


Figure 2. Clustering of New York City Neighborhoods



As it appears in above tables and figures, Toronto neighborhood characteristics for tourists tends to be very different from that of New York City.
Another interesting implication is that Toronto neighborhoods are more diversified from each other than New York City.

4. Conclusion

In this study, k-means method is applied on a dataset of ventures in New York City and Toronto. I used percentage mixture of top 10 popular venture categories in neighborhoods as training data. The clustering result implies that characteristics of Toronto neighborhoods are different from those of New York City for tourists, and Toronto neighborhoods are more diversified.

5. Discussion and Future Direction

For the hypothetical scenario, the company might want to apply different expertise, management pattern and marketing strategy in the new business of Toronto, and it might be more difficult for the new business because of the diversity in Toronto neighborhoods.