Comparison of two cities: Toronto and New York

Introduction

New York City is the most populous city in the Unites States with a population of over 8 million people. It is also one of the most popular cities in the United States and a centre stage for many cultural events. In the same way Toronto is the most populous city in Canada wit a population of over 2 million people. Toronto is deemed as an international centre of business and culture and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Problem and Interest

Two major cities of two major and different countries. Information on how similar and dissimilar they are is not currently available.

An analysis comparing these two major cities would be of interest to people looking to settle in one of these cities. For people looking to have a vibrant and culturally stimulating life.

Data

The data obtained comes from different sources. First, For the analysis, information on the list of neighbourhoods and the coordinates is needed to obtain information on the various venues present in each neibourhood and each city.

Initially, for Toronto, a table with neighbourhoods and postcodes was obtained from wikipedia. The table was cleaned to eliminate rows with 'not assigned' borough and for rows with 'not assigned' neighbourhoods, the name of the borough was used instead. There was the instance of more than one neighbourhood per postcode. In this case the rows where merged to create one row with one postcode and the name of both neibourhoods separated by a comma.

We obtain the coordinates from http://cocl.us/Geospatial_data . The file contained the postcodes and their respective coordinates. We then merged both tables.

Methodology

Once we have obtained the neighbourhoods and venues data, we can move onto analyzing the information. Because the sheer number of the information, we will calculate the frequency of types of venues and and then segmentate the neighbourhoods using Kmeans to cluster the data. We will then compare the clusters between the cities; looking for insights into the cities.

Analysis

In order to use KMeans, we expand the Venue Category column. Using get_dummies, we create new columns with the categories on 'Venues Categories. These columns are filled with 0s and 1s so that it can be divided into clusters by KMeans. We then group the table by Neighbourhood to get the frequency of shops per neighbourhood. We then create a table of the top ten most frequent types of venues per neighbourhood.

Results and discussion

Segmenting both cities using KMeans yielded 5 clusters per city.

The clusters for Toronto:

- Cluster 1

There are no elements in this cluster

- Cluster 2

It is the largest cluster with 81 neighbourhoods out of 103. It shows a variety of restaurants from a variety of ethnicity and cafes.

- Cluster 3

There are not enough neighbourhood to form an opinion. Although similarly to cluster 1, there are a lot event related venues. Cafes seem to be very popular in the neighbourhoods of this region.

- Cluster 4

The elements of this cluster have a major frequency of Park areas and Discount stores.

- Cluster 5

The neighbourhoods of this cluster have all, at first or second place, Pizza place. They also have in similar women's store and distribution centre.

The clusters for New York City :

- Cluster 1

Every neighbourhood boots in its top ten: parks, event services, event space, and exhibits.

- Cluster 2

They all seem to be situated along beaches.

- Cluster 3

There are not enough neighbourhood to form an opinion. Although similarly to cluster 1, there are a lot event related venues.

- Cluster 4

This cluster has a preponderance of Deli/Bodegas. It also shows a large variety of restaurants from many different ethnicity. A statement of the multicultural property of New York City.

- Cluster 5

It is the largest cluster with 249 neighbourhoods out of 306. It shows a variety of restaurants and leisure venues.

Conclusion

Although both cities are the most populous cities of their country, New York City has 4 times more inhabitants that its counterpart, Toronto. This is reflected also in the number of neighbourhoods: 103 and 306 for Toronto and New York respectively. But it is in the venues where the difference is more seen; the 2154 of Toronto venues against the 10097

of New York City. The number of venues in New York City is 5 times that of the venues in Toronto.

The segmentation was very looped with one cluster having about 80% of the neibouroods and one or two clusters having 0-2 neighbourhoods. But some characteristics are still observable. Both cities show their multicultural cuisine and the importance of leisure activities. The importance of the culture of the Deli/Bodegas in New York City is seen in cluster 4. And while there is a variety of cuisines, Pizza places and cafes seem to be the most important in Toronto as seen in cluster 2 and 5.

We have outlined some similarities and differences of both this cities based on the types and frequency of venues. For people deciding to transfer, there are still others deciding factors as housing and rental prices, social end economic interactions, etc.