

Capstone Project - How similar is New York to Toronto

Introduction

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area, of which the majority is within the Greater Toronto Area, held a population of 5,928,040, making it Canada's most populous CMA. Toronto is the fastest growing city in North America, and is the anchor of an urban agglomeration, known as the Golden Horseshoe in Southern Ontario, located on the northwestern shore of Lake Ontario. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. [1]

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important centre for international diplomacy. [1]

From the description above, both of these two megacities are centres of business, finance, arts, culture, etc. Except for the fact that the population of New York is almost quadruple of that of Toronto, one may expect these two financial centres of North America to bear much similarities between them. In this project, we will use the techniques of Data Science to find out if these two cities are similar and how similar they are. In this process we may find unexpected answers that are only possible to obtain from data.

Business investor/people interested in moving to these two cities might be interested in this question. The differences in these two cities will provide insights of make financial decisions and relocation.

Methodology

A) Data required:

The venue exploration data from the Foursquare API, as well as Geometric data for these two cities would be utilized for this project. From the distribution of different venues, analysis will be carried out to classify the neighbourhoods and measure the similarities between two cities.

1) **FourSquare API data** would be used to obtain venue information from Foursquare database for Toronto and New York.

2) Python package **geopy geocoder** provides the geometric data for these two cities which will then be referred to in calling the Foursquare database.

B) Data analysis

K-means clustering algorithm will be used to classify all combined neighbourhoods from both cities into 10 categories. These categories will provide insights to our question.

Scikit-Learn package in python provides k-means clustering function, which is implemented in `sklearn.cluster.KMeans`.

References:

- **[1]** Toronto, New York - Wikipedia