Comparing Neighborhoods of New York and Toronto

By : Qamar Zia

1 Introduction

1.1 Background

New York city is the most densely populated city in United States with an estimated population of 8,336,817 people living in the city. New York is also known as the cultural, financial, and media capital of world. It is known for the diversity in culture in its all Five boroughs including Brooklyn, Queens, Manhattan the Bronx and Staten Island. The city is so much diverse in culture that as much as 800 different languages are spoken in it. On the other hand Toronto is the provincial capital of Ontario and the most densely populated city in Canada with an estimated population of 5,928,040 . It is the 4th most populated city in North America. Toronto like New York is also the business hub and is very diverse in culture. People have travelled from all around the world and make Toronto their home . Toronto encompasses of several municipalities including East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston and York. Comparing the neighborhoods of both cities will give us great deal of insight like similarity and differences between two cities. And on the bases of these insights we can help people in making a choice of moving in a different city yet in similar surroundings.

1.2 Problem Statement

In todays world people are often moving from one place to other for work or for change of life and they also want to be in a place that is much like home. For instance a person comes too you saying that he is moving from New York to Toronto and he is so much confused that where should he find a home in Toronto. By Clustering the two cities on the bases of features we can help the person in his confusion and help him make a decision.

2 Data Gathering and Data Wrangling

2.1 Data Acquisition

For this project I will need two data sets one containing the data of boroughs of New York and the other one containing Data for Toronto. Luckily the data for New York was available on internet with its longitude and latitude values at Here in a geo json file. For

the data of Toronto it was not directly available and the names and postal codes were available on Wikipedia Here.

2.2 Data Wrangling

For the data of New York which was already available on Internet in a Geo Json file consisted of different features which I did not need for this project after exploratory analysis of the file using I came to know that the important values are under “features” category so I extracted the data and made a Data frame using Python Pandas library the final Data frame Consisted of 4 columns “Borough, Neighborhood, Latitude and Longitude.

As for the data of Toronto which was not directly available as a file I had to work for data accusation, Searching for thee internet I came across the Wikipedia page containing a table of Postal codes of Toronto as well as borough and neighborhoods. I started web scrapping for data and done it by using Python Beautiful Soup Library , After extraction of table and converting it into data frame I saw a problem of missing values as many of the borough were not assigned with any value so I dropped the rows which consisted of “Not Assigned” values . Further analysis revealed that a single neighborhood consisted of multiple postal codes with unassigned values so I combined those values in a single row. The new data frame consisted of three columns Postal Code , Borough and Neighborhood. I still needed two more columns like Longitude and Latitude for completing the data. For those two columns I assigned Each column with respective longitude and latitude.