**Capstone Project**

**Toronto vs New York – Battle of Neighbourhoods**

**By- Pranish Naoghare**

May 15, 2020

**1. Introduction**

**1.1. Background**

Today Tourism is one of the pillars of the economy and the people most often visits those countries who are rich in heritage and developed enough from a foreign perspective, like friendly environment.

Toronto and New York are the famous places in the world. They are diverse in many ways. Both are multicultural as well as the financial hubs of their respective countries. We want to explore how much they are similar or dissimilar in aspects from a tourist point of view regarding food, accommodation, beautiful places, and many more.

**1.2. Problem**

Every city is unique in its own way and give something new. And now the information is so common regarding location of every place around the world on your fingertips which make it easier to explore. Therefore, tourists always eager to travel to different places on the basis of available information, and the comparison (the part of the information) between the two cities always assist to choose the specific places or according to their choice. This project analyses the two boroughs from each of the cities- Downtown Toronto, Toronto and Manhattan, New York for their appeal to attract tourist.

**1.3. Interest**

This project will interest the tourists who are planning to visit these two cities but are confused as to which tour will be economical for them and what are the gains and pains that they will be facing during a tour.

**2. Data Acquisition and cleaning**

**2.1. Data Sources**

For this problem, we will get the services of Foursquare API to explore the data of two cities, in terms of their neighbourhoods. The data also include the information about the places around each neighbourhood like restaurants, hotels, coffee shops, parks, theatres, art galleries, museums and many more. We selected one Borough from each city to analyse their neighbourhoods. Manhattan from New York and Downtown Toronto from Toronto.

For Downtown Toronto the neighbourhoods data is not readily available on the internet. But there exists a Wikipedia Page exists that has all the information that we need to explore and cluster the neighbourhoods in Toronto. This data will have to be scraped from the page, wrangle and clean it and then read into the pandas dataframe.

For Manhattan, a saved data file was explored through Foursquare API. In this all boroughs along with their coordinates are readily available.

**2.2 Data Cleaning**

Data for Toronto scraped from the Wikipedia page required wrangling and cleaning it and then read into the pandas dataframe. Cleaning involved dropping the rows to which Borough was "Not Assigned". Also, later on coordinates of each borough were obtained from the internet and then merged into the previous dataframe.

**3. Methodology**

We will use machine learning technique, “Clustering” to segment the neighbourhoods with similar objects on the basis of each neighbourhood’s data. These objects will be given priority on the basis of foot traffic (activity) in their respective neighbourhoods. This will help to locate the tourist’s areas and hubs, and then we can judge the similarity or dissimilarity between two cities on that basis.

Both the neighbourhoods were then explored separately as venues and venue categories and then they were subjected to machine learning model of "Clustering" using "k-means" technique.

Using the k-means clustering, each neighbourhood will be categorized into 5 number of clusters based on their similarities.

**Exploratory Data Analysis**

Here, under EDA, the number of venues for each of the neighbourhood in the city is calculated.

The Venues per Neighbourhood is restricted to 20 so as to limit the dataset.

1. Venues in Downtown Toronto-

|  |  |
| --- | --- |
| **Neighborhood** | **Venue** |
|  |  |
| Berczy Park | 20 |
| CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport | 16 |
| Central Bay Street | 20 |
| Christie | 16 |
| Church and Wellesley | 20 |
| Commerce Court, Victoria Hotel | 20 |
| First Canadian Place, Underground city | 20 |
| Garden District, Ryerson | 20 |
| Harbourfront East, Union Station, Toronto Islands | 20 |
| Kensington Market, Chinatown, Grange Park | 20 |
| Queen's Park, Ontario Provincial Government | 20 |
| Regent Park, Harbourfront | 20 |
| Richmond, Adelaide, King | 20 |
| Rosedale | 4 |
| St. James Town | 20 |
| St. James Town, Cabbagetown | 20 |
| Stn A PO Boxes | 20 |
| Toronto Dominion Centre, Design Exchange | 20 |
| University of Toronto, Harbord | 20 |

2. Venues in Manhattan, New York-

|  |  |
| --- | --- |
| **Neighborhood** | **Venue** |
| Battery Park City | 20 |
| Carnegie Hill | 20 |
| Central Harlem | 20 |
| Chelsea | 20 |
| Chinatown | 20 |
| Civic Center | 20 |
| Clinton | 20 |
| East Harlem | 20 |
| East Village | 20 |
| Financial District | 20 |
| Flatiron | 20 |
| Gramercy | 20 |
| Greenwich Village | 20 |
| Hamilton Heights | 20 |
| Hudson Yards | 20 |
| Inwood | 20 |
| Lenox Hill | 20 |
| Lincoln Square | 20 |
| Little Italy | 20 |
| Lower East Side | 20 |
| Manhattan Valley | 20 |
| Manhattanville | 20 |
| Marble Hill | 20 |
| Midtown | 20 |
| Midtown South | 20 |
| Morningside Heights | 20 |
| Murray Hill | 20 |
| Noho | 20 |
| Roosevelt Island | 20 |
| Soho | 20 |
| Stuyvesant Town | 17 |
| Sutton Place | 20 |
| Tribeca | 20 |
| Tudor City | 20 |
| Turtle Bay | 20 |
| Upper East Side | 20 |
| Upper West Side | 20 |
| Washington Heights | 20 |
| West Village | 20 |
| Yorkville | 20 |

**4. Result**

After clustering the data of the respective neighborhoods, both cities (Boroughs) have venues which can be explored and attract the Tourists. The neighborhoods are much similar in features like Theaters, opera houses, food places, clubs, museums, parks etc. As far as concern to dissimilarity, it differs in terms of some unique places like veterinary hospitals, historical places and monuments

**5. Discussion and Recommendation**

When I compared the tourist places, I observed that the historical place is only situated in Downtown Toronto and the Monument or landmark venue is in Manhattan neighbourhoods. Similarly, Airport facility, Harbor, Sculpture garden and Boat or ferry services are also available in Downtown Toronto while venues like Nightlife, climbing gym and Museums are present in Manhattan.

As far as concern to recommendations, I recommend Downtown Toronto Neighbourhoods will be considered first to visit. The tourists have an easily travelling access due to Airport facility, which not only saves time but also helps to save money. This saved money can be utilized to explore more, the attracting venues.

**6. Conclusion**

The Downtown Toronto and Manhattan neighbourhoods have more like similar venues. As we know that every place is unique in its own way, so that’s argument is present in both neighbourhoods. The dissimilarity exists in terms of some different venues and facilities but not on a larger extent