Capstone Project - The Battle of Neighborhoods

Rattanakorn Rerkdee

Coursera: IBM Applied Data Science Capstone

April 4, 2020

Coursera course: IBM Applied Data Science Capstone (IBM Data Science)

Email: paee45@gmail.com

PROBLEM & 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. Every city is unique in its own way and give something new. And now the information is so common regarding the location of every place around the world on your fingertips which makes 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. As we have selected two cities Borough to explore their neighborhoods. The data exploration, analysis, and visualization for both boroughs are done in the same way but separately. When we compare the tourist places, we observe that the historical place is only situated in Downtown Toronto and the Monument or landmark venue is in Bangkok neighborhoods. 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 Bangkok.

DATA DESCRIPTION

For this problem, we will get the services of Foursquare API to explore the data of two cities, in terms of their neighborhoods. The data also includes information about the places around each neighborhood like restaurants, hotels, coffee shops, parks, theaters, art galleries, museums and many more. We selected one Borough from each city to analyze their neighborhoods. Bangkok from Thailand and Downtown Toronto from Toronto, Canada. We will use the machine learning technique, "Clustering" to segment the neighborhoods with similar objects on the basis of each neighborhood data. These objects will be given priority on the basis of foot traffic (activity) in their respective neighborhoods. 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.

METHODOLOGY

As we have selected two cities Borough to explore their neighborhoods. The data exploration, analysis, and visualization for both boroughs are done in the same way but separately using segmentation, clustering, then analyze both boroughs neighborhoods through one hot encoding (giving '1' if a venue category is there, and '0' in case of venue category is not there). On the basis of one hot encoding, we calculate the mean of the frequency of occurrence of each category and picked the top ten venues on that basis for each neighborhood. It means the top venues are showing the foot traffic or the more visited places.

EXPLORATION

For the Downtown Toronto case, we have extracted the table of Toronto's Borough from Wikipedia page. Then we arrange the data according to our requirements. In the arrangement phase, which applied multiple steps including but not limited to, eliminating "Not assigned" values, combine neighborhoods that have the same geographical coordinates at each borough and sorted against the concerned borough. For data verification and further exploration, we use Foursquare API to get the coordinates of Downtown Toronto and explore its neighborhoods. The neighborhoods are further characterized as venues and venue categories.

For Bangkok, we used a saved data file which is already explored through foursquare API in which we have extracted all the boroughs of New York and then sorted against the concerned borough. Then we explored the Bangkok neighborhoods as venues and venue categories