The Battle of the Neighborhoods (Coursera-capstone-project)

Introduction/Business Problem

Every city is built on different circumstances, environment and different constraints. I am going to compare the neighborhoods of of 2 major columns and how can they be compared with respect to places to eat, better connectivity to several useful regions and how are they distributed around both cities. The places I will be considering are airports, metro, coffee shop, restaurants, schools, college, general stores, hospitals etc. The audience will be tourists who consider travelling through financial capital of the 2 countries analyzed and search for better neighborhoods suited for their needs.

Data Section

The data used to calculate is available for New york and the other half i.e. for toronto has been been scraped for wikipedia. Both of the dataset consist of the boroughs, neighborhoods and the locations of them. The foursqaure api will be used to analyze the places nearby these neighborhoods and see the proximity of important places from the corresponding neighborhoods.

Methodology Section

For analyzing the neighborhoods for toronto as well as manhattan first we find the neighborhoods popular places using the explore section of thr foursquare api and then for each neighborhood for both of these cities we get the top 10 most common places and try to visualize then to get the similarities among them. Then, from the scikit learn library we use the k-means algorithms to cluster the places and see similarities for different neighborhoods for each city and find similarities among them. The visulization for all both cities is done through the folium library. To find similarities a unsupervised learning is used which helps in clustering places and is quite effective.

Result Section

The results which were found out after visualization the neighborhoods of both cities and applying unsupervised learning algorithm i.e. K-means clustering. The results obtained after applying the K-means algorithm on both the dataset was there was a similarity for both cities first few common places in the 1st cluster. The common places for both the cities had coffee shops, cafe's and restaurants for certain clusters and others for different clusters. The analysis for different clusters are shown in the end of the Battle of Neighborhoods notebook. Both cites common places are compared among different clusters and some similarities in distributions is apparent in both the cities.

Discussion Section

Certain similarities I noted which is the top ten common place similarity is calculated through K-means and according to those 10 most common places does the clusters form. For example,

Cluster 0 for New York and Cluster 3 for toronto has first two places pretty much same as coffee shops and cafe's yet they are not classified in the same cluster as the remaining places differ very extremely.

Conclusion Section

Thus, after analyzing both the cities most common places by applying machine learning algorithm (K-means clustering), I realized that many common places are present in both the cities but are distributed differently for both them and it's the nature and circumstances as well as arrangement which leads to these different distribution in both toronto as well as New York.