Clustering Neighborhoods in New York and Toronto

Nikunj Rathod

March 9,2020

1. **Introduction**

   Every city is built on different circumstances, environments, and different constraints. This project is about to compare the neighborhoods off 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 which will be considered 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.

2. **Data Section**

   The data used to calculate is available for New York and the other half i.e. for Toronto has been scraped for Wikipedia. Both datasets consist of the boroughs, neighborhoods and the locations of them. The foursquare API will be used to analyze the places near these neighborhoods and see the proximity of important places from the corresponding neighborhoods.

3. **Methodology Section**

   For analyzing the neighborhoods for Toronto as well as Manhattan first we find the neighborhoods popular places using the explore section of the 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 visualization for all both cities is done through the folium library. To find similarities an unsupervised learning is used which helps in clustering places and is quite effective.

4. **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, cafes and restaurants for certain clusters and others for different clusters. The analysis for different clusters is 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.

5. **Discussion Section**

   Certain similarities is 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.

6. **Conclusion**

   Thus, after analyzing both the city's most common places by applying machine learning algorithm (K-means clustering), it is noted 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.