The Battle of Neighborhoods of two cities

Coursera Capstone

IBM Data Science Professional Certificate

By: Rajeev Kavety

Introduction

For this project, I have chosen to compare the precedence of food courts among the most popular venue categories in both Toronto, CA and Manhattan, NY. As people like to visit most places that have multiple shopping choices; with niche visitors preferring to visit the location that has food courts as well, and also for the start-up entrepreneurs who are looking to start a business where there is a lucrative opportunity in places that have high traffic of shoppers.

Problem

The problem with the top 10 most common venues is that, it misses the niche market, such as, food courts. This information can be useful for the entrepreneurs who want to setup a business in the top most venues, where possibility of running a business can have good return-of-investment (ROI), or is profitable. Hence, the challenges of identifying these locations, and overlaying the top most common venues having food courts on the maps of Toronto and Manhattan, can address both preferences of the public.

Data

For the Toronto neighborhoods, Wikipedia link will be used.
https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=935851093

Using web scrapping techniques, the Boroughs and Neighborhood information is extracted, into DataFrame. For this beautiful soup and python results package will be used.

For Toronto map coordinates, the information in a CVS file will be downloaded from the link http://cocl.us/Geospatial_data , and will be merged with the neighborhoods DataFrame.

The Battle of Neighborhoods of two cities

For the Manhattan neighborhoods dataset, the link is in the json format.

https://geo.nyu.edu/catalog/nyu_2451_34572

Foursquare API

For the purpose of getting the venues information, Foursquare API can be used. Foursquare API is a location data provider that uses RESTful API calls to retrieve data about most common venue categories in each neighborhood. The results explored using Foursquare venue category is in hierarchy. The venues retrieved for all the neighborhoods are categorized into top 10 most common venues having various businesses.

Finally, using the k-means clustering algorithm, and with the help of Folium library, the maps will be generated showing the clusters for the top 10 most common venues, along with focus on food courts for visualization for the neighborhoods in both Toronto and Manhattan cities.