CAPSTONE PROJECT – BATTLE OF THE NEIGHBOURHOODS

1. Introduction

This capstone project will aim to compare the neighbourhoods in both Toronto, Canada and New York City, USA. This is done with the perspective of a new F&B business client in mind. The client wants to set-up his business near a vibrant and happening neighbourhoods where they are the go-to places for the city residents. I will play a role of a consultant where I will analyse the similarities and dissimilarities between the neighbourhoods in both cities, and thus, determine which neighbourhoods are the best fit for the client.

2. Sources of Data

The data of both cities used for this project will be acquired from Wikipedia. The derived datasets consist of the postal codes, neighbourhood names, latitude, and longitude information for each neighbourhood. Foursquare API search feature will be used to collect neighbourhood venue information. Details about local venues and locality will be provide insight into the qualities of a neighbourhood. In addition to Foursquare, various python packages will be used to create maps and machine learning models (e.g. k-means clustering) to further provide insights into our neighbourhood battle project. The datasets can be found on below links.

- Toronto Neighbourhoods: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. Toronto Latitude and Longitude: http://cocl.us/Geospatial_data
- New York City Neighbourhoods: https://geo.nyu.edu/catalog/nyu_2451_34572
- New York City Latitude and Longitude: Python Geolibrary

3. Workflow

HTTP requests would be made to this Foursquare API server using zip codes of the Seattle city neighbourhoods to pull the location information (Latitude and Longitude). Foursquare API search feature would be enabled to collect the nearby places of the neighbourhoods. Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 700. Folium or Python visualization library would be used to visualize the neighbourhoods cluster distribution of Seattle city over an interactive leaflet map. Extensive comparative analysis of two randomly picked neighbourhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn. Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighbourhoods. These clusters from each of those two chosen neighbourhoods would be analysed individually collectively and comparatively to derive the conclusions.

4. Results and Discussion

K-means was used to group the neighbourhoods in Scarborough into 3 clusters. Cluster_0 has 15 neighbourhoods and the most common venues are skating rinks, international cuisine

restaurants and breakfast spots. Cluster 1 has 1 neighbourhood 1 neighbourhood, and the most common venues are pizza place and noodle house. Cluster 2 has 1 neighbourhood, and the most common venues are Chinese restaurants and discount stores.

K-means was again used to group the Queens borough into 5 clusters. Cluster_0 has 81 neighbourhoods and consist of many international cuisine restaurants and grocery stores. The most common venues are pizza places, deli, and Chinese restaurants. Cluster_1 has 1 neighbourhood and the most common venue is a dance studio. Cluster_2 has 5 neighbourhoods and the most common venue are donut shops and international cuisine restaurants. Cluster_3 has 2 neighbourhoods and the most common venues are the beach and a bakery. Cluster_4 has 2 neighbourhoods and the most common venues are gyms and donut shops.

Toronto has 11 boroughs and 103 neighbourhoods. The geographical coordinate of Toronto, Canada are 43.7170226, -79.4197830350134. In Scarborough borough, found 85 venues in 17 neighbourhoods In Scarborough borough, the neighbourhoods with the most venues are Lamoreaux West and Steeles West. There are 79 distinct venues in 50 categories. New York City has 5 boroughs and 306 neighbourhoods. The geographical coordinate of New York City is 40.7308619, -73.9871558. Foursquare found 2108 venues in 81 neighbourhoods in Queens borough. Many of the neighbourhoods are homogenous and are very similar to each other. Both Scarborough and Queens borough consist of neighbourhood cluster that contain majority of the neighbourhoods, and the remaining cluster had 1-5 neighbourhoods. Queens borough had a significant a greater number of neighbourhoods and venues than Scarborough.

5. Conclusion

In conclusion, based on the quantity of venues and variety of venues, I would choose Queens over Scarborough as a choice to set-up the client's business. This is because Queens offer way more in choices for restaurants, gyms, grocery stores, and extracurricular activities that will result to a lot more customer pools to be reached by client's F&B Venture.