Capstone Project:
The Battle of Neighbourhoods: Toronto Renting Solutions

1. Introduction

1.1 Background

Toronto, Canada is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. In 2010, it was estimated over 100,000 immigrants arrive in the Greater Toronto Area each year. However even within the diversity and the hustle and bustle familiarity and oneness is what one seeks. We, Good Host are an upcoming start-up providing housing solutions that go beyond just the rent.

1.2 Business Problem

Given the massive immigrant/ expat population in Toronto, we want to leverage our data analysis expertise and tap into this market and offer a more personalised approach to those seeking housing options. While searching for accommodation avoiding culture shocks to seek familiarity is most often observed. We wish to build an app providing such solutions providing highly customised criteria for selection of preferred locations to live. We are still in the development stage of the app and have released a pilot model using a sample study approach. The sample chosen is from India and is looking for accommodation having nearby access to Indian food, Indian supermarkets, and Indian neighbours.

Given the criteria we are going to consider one parameter initially. The presence of Indian restaurants, which would indicate that it is within a locality wherein its main customers, mostly Indians, live/work nearby. Procurement of ingredients would also be from within the vicinity, indicating India markets nearby.

2. Data Collection and Extraction:

The input data required to perform the analysis and clustering are as follows:

- List of neighbourhoods in Toronto, Canada
- Latitude and the longitude coordinates of each neighbourhood.
- Venue data for Indian restaurants.

To explore and cluster the neighbourhoods the following will have to be performed:

- Scrapping, wrangling and cleaning of data on Toronto neighbourhoods from the link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Including the latitude and longitude coordinates of the neighbourhoods using Geocoder Package
- Venue data of neighbourhoods using Foursquare API

3. Methodology for Analysis:

The first raw data input was obtained by extracting the list of neighbourhoods from wikipedia page

(" https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M "). The data required web scrapping which was done utilizing pandas html table scraping method.

This data provided only a list of neighbourhood names and postal codes. For venue data utilizing the Foursquare API the latitude and longitude coordinates are required to pull the list of venues near these neighbourhoods. This is usually obtained using the Geocoder package however the calls were not going through. A csv file with the required data provided by IBM team to match the coordinates of Toronto neighbourhoods, was used instead.

Using the Foursquare API, a list of top 100 venues within 500 meters radius as obtained for the coordinates. This provided the names, categories, latitude, and longitude of the venues. Further this allowed determination of unique categories

An analysis of each neighbourhood by grouping the rows by neighbourhood and

taking the mean on the frequency of occurrence of each venue category would further the clustering to be performed. The unique category also threw up specifically "Indian restaurants", which is what the prime parameter to match rental properties and the custom criteria.

Lastly, a clustering of the data was performed using k-means clustering. K-means clustering

algorithm identifies k number of centroids, and then allocates every data point to the nearest

cluster, while keeping the centroids as small as possible. Dividing the neighbourhood into 3 clusters to obtain the frequency of occurrence of "Indian Restaurants"

4. Results:
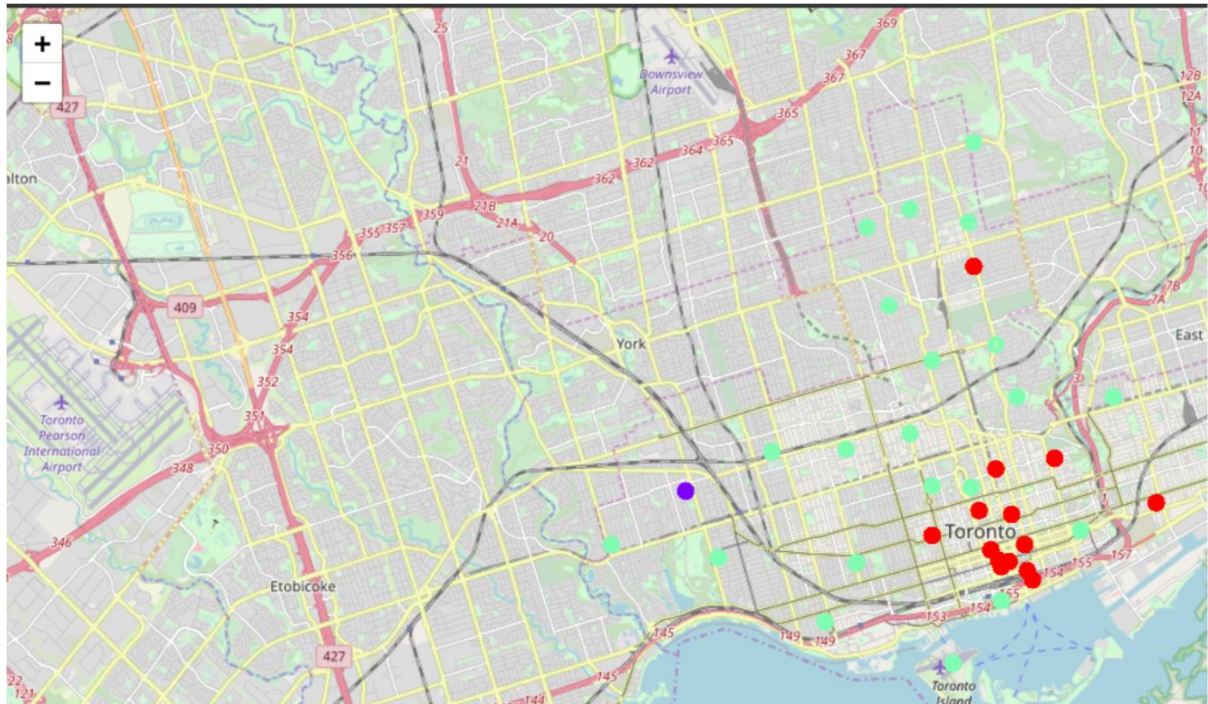
The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Indian restaurants are in each neighborhood:

● Cluster 0: Neighborhoods with no Indian restaurants

● Cluster 1: Neighborhoods with high number of India restaurants

● Cluster 2: Neighborhoods with little or no of Indian restaurants

The results are visualized in the above map with Cluster 0 in red purple, Cluster 1 in green colour and Cluster 2 in red color.

4. Observations:

As observed from the Clustering, cluster 1 shows more of a density of Indian restuarants especially in the neighbourhoods of Central Bay Street, The Annex, North Midtown and Yorkville. Hence suggestions of properties from these neighbourhoods should be priorities

for the sample client. The areas are relatively dispersed and hence should provide a sufficient coverage to obtain rental properties.

5. Limitations:

The sample taken considers only one parameter to fulfil the conditions. Further testing should be conducted to consider more diverse preferences. Which will have to consider various parameters.

6. Conclusion:

The first trial of producing results for rental solutions using customised criterion and preferences enables us to obtain concrete results to work with. With more input data on other factors would help our model grow before we can launch it in the market.

7. References:

List of neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
Foursquare Developer Documentation: https://developer.foursquare.com/docs