Business Expansion in Toronto

Business Problem:

A Chinese restaurant brand based in Manhattan, New York is looking to expand their business in Toronto, Canada. The client wants to understand the neighbourhoods in Toronto so it can conclude where their next restaurant in Toronto could be. The client is also interested whether the neighbourhoods in New York and Toronto are similar or dissimilar so it can target the neighbourhood accordingly ultimately leading to profit.

Target Audience:

The stakeholders of the company will get a clear understanding at the end of project as to which neighbourhood(s) to target for their next restaurant in Toronto.

Objectives:

1) Compare New York and Toronto cities to determine how similar or dissimilar they are.
2) Explore category venues and trending venues in a particular neighbourhood in Toronto and targeting accordingly.

Exploring Data:

- The data includes feature such as Borough, neighbourhood.
- Use geopy library to get the latitude and longitude values of the neighbourhoods in New York City and Toronto.

1)New York:

- The data is in json format.

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork DS0701EN-SkillsNetwork/labs/newyork_data.json

2)Toronto:

- For the Toronto neighbourhood data, the below Wikipedia page exists that has all the information we need to explore and cluster the

neighbourhoods in Toronto. We will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format like the New York dataset. The BeautifulSoup package will help in web scraping.

https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&direction=prev& oldid=926287641 Foursquare API

- In this project we will construct a Foursquare URL to send a request to the API to search for a specific type of venues, to explore a particular venue, to explore a geographical location, and to get trending venues around a location. Also, you will learn how to use the visualization library, Folium, to visualize the results.

Analyse

Lastly based on the venue categories we will cluster the data and provide insightful results.

Methodology:

Foursquare API provides a range of tools for developers to incorporate the up-to-date location data to enhance their projects. Foursquare, which is the location data provider we will be using in this course, and its API. Essentially, we create a Foursquare developer account, and use credentials to search for nearby venues of a specific type, explore a particular venue, and search for trending venues around a location.

ETL - After the Extraction, Transformation process on the data we analyse the neighbourhoods in Toronto and Manhattan. We use K-machine learning algorithm to segment neighbourhoods.

K-Means is a type of partitioning clustering, that is, it divides the data into K non-overlapping subsets or clusters without any cluster internal structure or labels. This means, it's an unsupervised algorithm.
Objects within a cluster are very similar, and objects across different clusters are very different or dissimilar.

Observations:

Neighbourhoods in Cluster 0 and 2 does not have any Chinese restaurants. We could target these neighbourhoods. These neighbourhoods are mostly central and away from coastal area of Toronto. Opening a restaurant in these neighbourhoods could prove an advantage and avoid competition from existing restaurants. The Chinese restaurants in Manhattan city are over 43 whereas in Toronto there overall only 7 mostly in the heart of the city. This opens an opportunity for Toronto as a new upcoming market for opening up Chinese restaurants.

Cluster 0 and 1: Zero Chinese Restaurants
Cluster 1: Most Chinese Restaurants
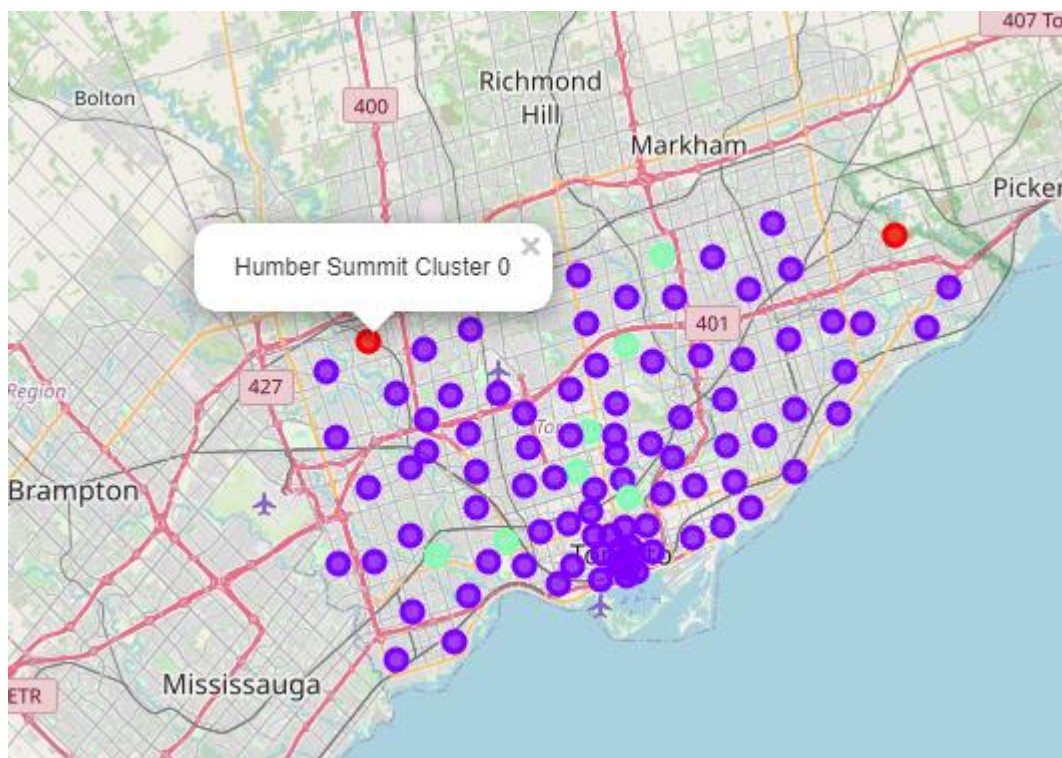
Future Scope

Data about the people such as population and income of residents and other factors of locality could fetch more accurate results in deciding the location of the restaurants. The project used Foursquare Sandbox Tier account that came with the limitations as to the number of API calls and results returned. Better results could be obtained by going premium.