Introduction

This is the part of the Capstone Project for the IBM Data Science Professional Certificate.

In this document, I have to setup hypothetical business scenario using Neighbourhoods of the large city to analyse and search suitable location for the set criteria.

I have to apply Machine Learning/Data science methods to meet my set objectives described in the next section.

Project Description

Business Background:

XYZ. Ltd is a business in Hampshire, UK, distributing Asian food ingredients to the Local Small Shops and Supermarkets in UK. It owns its own warehouse and packing facility, sourcing ingredients mainly from South East Asia and South America.
They have food processing facility in Gujarat, India and Distribution facility in Dubai for Middle East and African markets.

XYZ is now looking for Organic Growth on the North American Continent and have identified Toronto, Canada for expansion, targeting both wholesale and retail sectors.
Toronto was selected, on bases of

  a) Presence of a large Indian, Pakistani and Chinese communities
  b) Stable socio-economic and political systems
  c) Close Trade and Political relations with UK.
  d) Close proximity and amicable trade relations with USA for future expansion into USA.

Business Objectives:

To select a site to set up a Supermarket where,
  • There are other small and large businesses, to obtain a good foot fall.
  • Ideally based around Central Toronto for easy access from all neighbourhoods of Toronto.
  • With good public transport links and good Car Parking facilities.

I will use internet search to source the relevant data. I will then analyse the data and convey my interpretations and suggestions to XYZ.

Procedure for implementing the setting of Objectives:

 To set objectives in Project Description, I will employ following resources and procedures to set up the project.

Data Sourcing:
- Source Toronto Neighbourhood data with its location in Toronto.
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
- Geocoding data for linking Longitude and Latitude to the Neighbourhoods.
  http://cocl.us/Geospatial_data and Geocoder
- Sourcing Venue data using Foursquare.
- Internet search for Asian population in Toronto, socio/economic data of Neighbourhoods.
  Demographics of **Toronto** - Wikipedia

  https://en.wikipedia.org/wiki/Demographics_of_Toronto

Resources Use:
- Install Python-3 on Jupyter Notebook using Mac/Windows operating system.
- Install all standard libraries for Python.
- Install BeautifulSoup, Geocoder, Request, Folium, Jason and other software as require.
- Test, Jupyter Notebook, Foursquare and Github are functioning properly.

Implicate Data Science and Machine Learning:
- Set up New Jupyter Notebook.
- Import and setup all libraries and software.
- Scrape the Wikipedia for Toronto data using BeautifulSoup and transform the data to pandas dataframe
- Download Longitude and Latitude for the Neighbourhoods and merged it.
- Clean and explore the dataset for processing.
- Capture Venues and Categories data using Foursquare.
- Segment and cluster the data for evaluation.
- Draw conclusion and recommend to XYZ.

Prepared by Dipak Jani