Exploring Toronto Neighborhoods – to open a Chinese Restaurant

As a part of the IBM Data Science program capstone project, I worked on the real datasets to get an experience of what a data scientist goes through in the real life, Main object of this project were to define business problem, using the web data through web scraping, and using Foursquare location data to compare different neighborhoods of Toronto to figure out which neighborhood is profitable for starting a new restaurant. In this project, I will go through the whole process step by step manner from problem targeting, data preparation to final analysis result. Last but not the least, I will provide my suggestion which can be leveraged by the business stakeholders to make their decisions.

1. Description of the Business Problem

In this project, I will go through all steps to make a decision whether it is a good idea to open a Chinese restaurant.
I analyze the neighborhoods in Toronto to identify the most profitable area to place it.
Target Audience:
Business personnel who wants to invest or open a Chinese restaurant in Toronto.
Freelancers who loves to have their own restaurant as a side business.
Chinese crowd who wants to find neighborhoods with lots of option for Chinese restraurants.

2. Data acquisition & cleaning:

   2.1 Data Sources

   I used (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) wiki page to get all the information about the neighborhoods present in Toronto. This page has the postal code, borough & the name of all the neighborhoods present in Toronto.
   Then I used ("https://cocl.us/Geospatial_data")csv file to get all the geographical coordinates of the neighborhoods.
   To get information about the distribution of population by their ethnicity I'm using "Demographics of Toronto" (https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity) wiki page. Using this page, I'm going to identify the neighborhoods which are densely populated with Indians as it might be helpful in identifying the suitable neighborhood to open a new Chinese restaurant.
   To get location and other information about various venues in Toronto I'm using Foursquare's explore API.(https://developer.foursquare.com/docs)

   2.2 Data Cleaning

   Firstly, I scraped Toronto neighborhoods table from Wikipedia.
   Then I added geographical coordinates to the neighborhoods by extracting the data present in the Geospatial Data csv file.
   Thirdly, I scraped the distribution of population from Wikipedia, examining those neighborhood's population to identify the densely populated neighborhoods with Chinese population.

Fourthly, I got location data from Foursquare, by choosing 100 popular spots for each neighborhood within a radius of 1km.

3. Exploratory Data Analysis

   3.1 Visualization

   The next step, I used Folium Library and Leaflet Map to draw an interactive map using coordinate data.

   3.2 Relationship between neighborhood and Chinese restaurants

   I used violin plots to identify the boroughs with densely populated Chinese restaurants.

   3.3 Relationship between neighborhood and Chinese population

   I analyzed the neighborhoods and identified the neighborhoods with the highest number of Chinese population.

   3.4 Relationships between Chinese population and Chinese restaurants

4. Predictive Modeling

   4.1 Clustering neighborhoods of Toronto

   I used elbow methods on the Toronto dataset with Chinese restaurant percentage, aiming to do k-means cluster.

   4.2 Examining the clusters