Coursera Capstone Project

IBM Applied Data Science

# Accommodation for new immigration in Toronto, ON, Canada

**by: Shuo Liu**

**July,2020**

## Introduction

Leo is an owner of the Hotpot restaurant chain in China. Recently, his daughter came to Toronto to pursue her master's degree, so Leo decides to expand his business to Canada and move to Toronto with her daughter to take good care of her. Since Leo is not very good at English, he'd like to live in a neighbourhood where many Chinese people gather. However, since he also needs to make a living, he didn't want to open his hotpot restaurant in the places where too many Chinese restaurants existed, because too many competitors will make the business to be hard-hitting.So we need to help him find the place where is good to live and the other place for his business where is not very far away from home but without too many Chinese restaurants.

## Data

Data List:

- List of postal code, neighbourhoods and borough in Toronto
- Latitude and longitude coordinates of those neighbourhoods.
- Foursquare location data which contains venues in each of the neighbourhoods. We will use this data to perform clustering and estimate whether it is a suitable neighbourhood for Leo to live.

## Methodology

To solve Leo's problem, we will use the Foursquare location data and postal-code list data from Wikipedia and Geospatial Coordinates data in combination. We firstly scraped the postal-code list from
URL: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The list from Wikipedia contains Canadian Postal code with the beginning of M, and the corresponding Borough and Neighborhood. However, this is just a list of names, we need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we combine the postal codes list table with Geospatial Coordinates which includes each Neighborhoods’longitude and latitude. After these processes have been done, we begin to explore the neighbourhoods in Toronto by using Foursquare

location data. For instance, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then we will analyse each neighbourhood by grouping the rows names “Downtown Toronto” and taking the mean of the frequency of occurrence of each venue category. Since we are analysing the “Chinese Restaurant” data, we will filter the “Chinese Restaurant” as venue category for the neighbourhoods. Then according to the data frame, we can figure out the appropriate neighbourhood for Leo to live Where the Chinese like to gather and develop a hotpot business where isn't many competitors that exist.

## Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence of “Chinese Restaurant”:

- Cluster 0: Neighborhoods with very small number of Chinese Restaurant
- Cluster 1: Neighborhoods with relatively small number of Chinese Restaurant
- Cluster 2: Neighborhoods with relatively higher concentration of Chinese Restaurant

The results of the clustering are visualized in the map below with cluster 0 in red cluster 1 in purple and cluster 2 in green.

## Discussion

According to our observation from resulting map above, most of the Chinese restaurants are concentrated in the west and northeast area of Toronto with the higher number in cluster 2 and moderate number in cluster 1, also, cluster 0 has very low concentration to Chinese restaurant in the neighborhoods.  This represents a great opportunity for Leo to open a new Hotpot restaurant in the cluster 0 or cluster 1 area with relatively small competition, and Leo is advised to avoid neighborhoods in cluster 2 which already have high concentration of Chinese restaurant and suffering from intense competition.

## Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarity and lastly providing recommendations to the relevant stakeholders. By looking at the location data and foursquare data we explored, we could find out that the neighborhood where Chinese people always hand out is St. James Town and Church and Wellesley since Chinese restaurant is the No.3 most popular venues in these areas, so both of them are very suitable places for Leo to live. However, since Leo needs to develop his own hotpot business, Church and Wellesley seems like a better choice for him because comparing to St. James Town, starting business in Church and Wellesley will meet fewer competitors with lower concentration. In conclusion, we will suggest Leo move to Church and Wellesley to begin his new life in Toronto.