

**Project Title:** Exploring Toronto Neighborhoods and Identifying Suitable Location for Opening Chinese Restaurant

## 1. Introduction and Background

This coursera IBM capstone project aims at exploring a business problem, business understanding and providing scientific solution by adopting the data science methodologies on real dataset. In this project, we follow the data science methodologies such as problem understanding, data preparation, data modelling and evaluation.

### 1.1 Problem Statement

*“Exploring Toronto Neighborhoods and Identifying Suitable Location for Opening Chinese Restaurant”*

### 1.2 Problem Description

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016. The demographics of Toronto, Ontario, Canada make Toronto one of the most multicultural and multiracial cities in the world. In 2016, 51.5% of the residents of the city proper belonged to a visible minority group. Toronto also has established ethnic neighbourhoods such as the multiple Chinatowns, Corso Italia, Little Italy, Little India, Greektown, Koreatown, Little Jamaica, Little Portugal and Roncesvalles, which celebrate the city's multiculturalism. According to the ethnic breakdown of 2016 census Chinese ethnic group is on top (12.5% of total population) among the most prevalent ethnic origins in the City of Toronto.

Keep in mind the population of Chinese ethnic group, we aim at identifying suitable location for opening a new Chinese restaurant. For this purpose, we will collect spatial and non-spatial data of Toronto neighborhoods and analyse the dataset. The data analysis includes exploring the population distributions in top visible minority groups in neighborhoods, exploring relationship between venues in the neighborhoods, identifying the relationship between neighborhoods and existing Chinese restaurants, identifying relationship between Chinese population and Chinese restaurants, and clustering analysis for identifying suitable location for opening new Chinese restaurant.

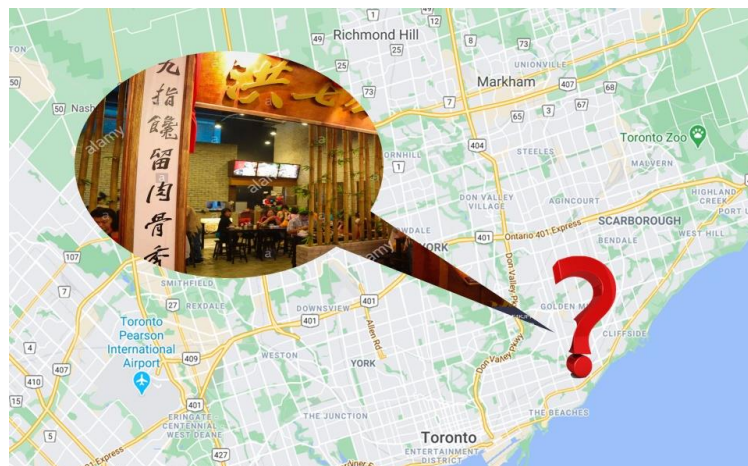


Abbildung 1: Finding suitable Chinese restaurant location in the neighborhoods of Toronto, Canada.

The findings will give an overview for investors interesting in opening new Chinese restaurant by taking into account Chinese population and number of existing restaurants in the vicinity. The analysis will also give information to customers to find neighborhoods with lots of option for Chinese restaurants.

## 2. Data Acquisition and Pre-processing

### 2.1 Data sources.

- we will use Wikipedia data “List of Postal code of Canada: M” ([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)) to get all the information about the neighborhoods including their postal code, borough & the name of all the neighborhoods present in Toronto.
- The population information will be extracted from Wikipedia page [https://en.wikipedia.org/wiki/Demographics\\_of\\_Toronto](https://en.wikipedia.org/wiki/Demographics_of_Toronto). We will use demographic data to get population information about Chinese ethnic group in Toronto neighborhoods. The population information together with number of existing restaurants will allow us to explore the relationship between population and existing restaurants.
- Spatial information such as latitude and longitude of Toronto neighborhoods will be extracted from “[https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)”
- Information about various venues in Toronto and their spatial information will be extracted using Foursquare’s API. Detail on API can be found on <https://developer.foursquare.com/docs>. We are interested in existing Chinese restaurants. Therefore, we will extract the number of existing restaurants in each neighborhood. This will allow us to explore the relationship between neighborhoods and Chinese Restaurant.

### 2.2 Pre-processing and Data Cleaning

The pre-processing and data cleaning involve scraping Toronto neighborhood data, extracting latitude and longitude of neighborhoods, scraping venue items, extracting Chinese population in neighborhoods and number of existing Chinese restaurants. All the extracted data will be converted into data frames for further analysis. Each pre-processing and data cleaning steps will be discussed, and documented as markdowns in the final jupyter notebook.