

# **Coursera Capstone**

**IBM Applied Data Science Capstone**

## ***Opening a New Chinese Restaurant in Singapore***

By: Junli Wu  
September 2019



# 1 Introduction

## 1.1 Background

According to Wiki Demographics of Singapore, Singapore is a multiracial and multicultural country with ethnic Chinese (76.2% of the citizen population), Malays (15.0%), and ethnic Indians (7.4%) making up the majority of the population. The 76.2% Chinese population decides that there is a great potential demanding for Chinese Restaurant. For investors, the location of the new restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

## 1.2 Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Singapore to open a new Chinese Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Singapore, if a property developer is looking to open a new Chinese Restaurant, where would you recommend that they open it?

# 2 Data Acquisition

In this project, we will need the following data:

- List of neighbourhoods in Singapore. This defines the scope of this project which is confined to the city of Singapore
- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to Chinese Restaurant. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page ([https://en.wikipedia.org/wiki/Planning\\_Areas\\_of\\_Singapore](https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore)) contains a list of neighbourhoods in Singapore, with a total of 55 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and BeautifulSoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare API will provide many categories of the venue data, we are particularly interested in the Chinese Restaurant category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.