

# IBM APPLIED DATA SCIENCE CAPSTONE

## Opening a New Chinese Restaurants in Los Angeles, California.

by:  
**Stephen Kofi Acheampong**  
June 12, 2020



## Introduction

Los Angeles, with one of the largest diasporic Chinese populations outside Asia, is widely regarded as an epicenter for Chinese cuisine. Although the best chinese restaurants are mostly concentrated in the ethnic enclaves of the San Gabriel Valley, there are increasingly more in neighborhoods throughout L.A., from historic Chinatown to the Westside.

Combining exotic flavors with great services and availability, Chinese food slowly climbed the rungs of the ladder to emerge as the top ethnic cuisine, and the most popular worldwide.

For chinese restaurants owners are taking advantage of this trend to build more restaurants to cater for the rising demand.

As a result, there are many chinese restaurants in the city of Los Angeles. But as with any business decision, opening a new chinese restaurants requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurants is one of the most important decisions that will determine whether the chinese restaurants will be a success or a failure.

## Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Los Angeles, California to open a new chinese restaurants. 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 Los Angeles, California, if a chinese restaurants owner or franchisee is looking to open a new chinese restaurants, where would you recommend that they open it?

## Target Audience of this project

This project is particularly useful to chinese restaurants owners or franchisee looking to open or invest in new chinese restaurants in Los Angeles, California. This project is timely as a lot of people in the city are craving for chinese restaurants and more of such can benefit both investors and customers immensely.

## Description of Data

Data of different venues in different cities in Los Angeles from Wikipedia will be used. In order to gain that information we will use "Foursquare" locational information to get a list of all venue and filtered to get a list of chinese restarants in such locations.

We will need the following data:

- List of neighbourhoods in Los Angeles. This defines the scope of this project which is confined to the city of Los Angeles, California.
- 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 restarants. We will use this data to perform clustering on the neighbourhoods.

### **Sources of data and methods to extract them.**

This Wikipedia page contains a list of 88 neighbourhoods in Los Angeles. We will use web scraping techniques to extract the data from the Wikipedia page. 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 has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the chinese restarants 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).