IBM Data Science Specialization:
Capatone Project

Opening a new Chinese Restaurant in Mumbai, India

By: Siddhartha Patra

July 2019

Introduction

Mumbai is fast becoming one of the places where people prefer to have Chinese cuisines and good service. Chinese food is always favoured by people of all ages-Soups, momos, noodles and dim sums to name a few. People in Mumbai are no different. Nowadays most teenagers in Mumbai savour eating Chinese noodles which could range from Hakka, chilli chicken, tom yum soups etc. The following business case aims to study the density and population of such restaurants all over the city of Mumbai in an attempt to suggest the best possible location for a business owner to start & set up a new Chinese restaurant in Mumbai.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Mumbai, India 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 Mumbai, India, if a business owner is looking to open a new Chinese restaurant, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to business owners and investors looking to open or invest in new Chinese restaurants in the largest city of India i.e. Mumbai. This project is timely as the city is currently suffering from oversupply of Chinese restaurants. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing restaurant space, and the agency predicted that total occupancy may dip below 86 per cent. The local newspaper The Mumbai Mirror also reported in March last year that the true occupancy rates in restaurants may be as low as 40 per cent in some areas, quoting a Financial Times (FT) article cataloguing the country's continued obsession with building more commercial space despite chronic oversupply.

Data

To solve the problem, we will need the following data:

- List of neighbourhoods in Mumbai. This defines the scope of this project which is confined to the city of Mumbai, the largest city of the country of India in Asia.
- 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 restaurants. 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/Category:Neighbourhoods_in_Mumbai) contains a list of neighbourhoods in Mumbai, with a total of 70 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 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 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.