Introduction

New York City is a large and ethnically diverse city. It is the largest city in the United States with most dense population. New York City was home to nearly 8.5 million people in 2018, accounting for over 40% of the population of New York State and a slightly lower percentage of the New York metropolitan area, home to approximately 23.6 million.

It has been said as the cultural, financial, and media capital of the world, significantly influencing many areas including entertainment, research, technology, education, politics, fashion, and even sports.

Over the last decade, the city has been growing faster which being the most linguistically diverse city, the New York region continues to be the leading metropolitan gateway for immigrants admitting into the United States.

This project explores the best locations for restaurants to be opened in the city of New York. Opening restaurant mostly consider factors such as rent, revenue etc, however, location/ neighbourhood is also one of the important factor, which it will determine the success or a failure of the business.

So this project will attempt to answer the questions including
1. Where should the investor open a Japanese Restaurant?
2. What are neighbourhood beside?

Business Problem

The objective of this Capstone project is to analyze location data and select the best locations in the city of New York to open a restaurant. We will use Data Science methodology and instruments such as Data Analysis, Mining and Visualization. We will provide solutions to answer where in the city of New York should the investor open a restaurant.

Target Audience

This project is particularly useful immigrants or investors looking to open or invest in a restaurant in New York. As New York is a ethnicity and cultural diversity place, food are also with this characteristics. There are many restaurants in New York which many belonging to different categories like Chinese, Indian, French, etc. Therefore it can help them to make decision or provide support evidence for them to consider.

Data using

1. New York City data containing the neighbourhoods and boroughs
2. Latitude and longitude coordinates of those neighbourhoods
3. Venue data, particularly data related to restaurants.

   We are going to use these data to perform analysis of the neighbourhoods. We will obtain from an open data source: https://cocl.us/new_york_dataset.
   After that, we will get the geographical coordinates of the neighbourhoods (latitude and longitude) using Geocoder package.
   Finally, Foursquare API will be used to retrieve venue data for the neighbourhoods extracted at the previous step. Many categories of the venue data can be utilised to solve the business problem defined above.