IBM Applied Data Science Capstone Project
-- Battle of Neighborhoods(Week 1)

INTRODUCTION

New York, as the largest city in the United State, is composed of people with different races and varied religious and literary backgrounds. According to the 2018 ACS, the racial composition of New York City was:

- White: 42.67%
- Black or African American: 24.27%
- Other race: 15.12%
- Asian: 13.95%
- Two or more races: 3.51%
- Native American: 0.43%
- Native Hawaiian or Pacific Islander: 0.05%

With the very diverse population, all types of cuisines can be found in NYC. American, British, Caribbean, Chinese, French, Greek, Indian, Italian, □□. You name it, you find it! Among all the cuisines, Chinese food is one of the most popular foods. Color, smell and taste are the three traditional aspects used to describe Chinese food, as well as the appearance and nutrition of the food. Not to mention, the big serving amount. You can easily feed yourself to full with a reasonable price.

Therefore it could be a great opportunity for an entrepreneur who is based on NYC. He/She might think of opening a Chinese restaurant. With the purpose in mind, finding a good location to open such a restaurant is one of the most important decisions to make. Meanwhile, with 5 boroughs and 306 neighborhoods in NYC, it is also not an easy decision to make. I am designing this project to not only satisfy my IBM Data Science Professional Certificate program capstone course requirement, but also help the entrepreneur find the most suitable location.

BUSINESS PROBLEM

The objective of this capstone project is to find the best location for an entrepreneur to open an authentic Chinese restaurant in NYC. By using data science methods and tools with machine learning algorithms such as clustering, the project is aiming to provide solutions to answer the business question: In NYC, if an entrepreneur targets on

opening a new Chinese restaurant, where is the most suitable place should he/she consider?

TARGET AUDIENCE

Someone who wants to find the location to open a new Chinese restaurant in NYC.

DATA

To solve the problem, we need data described as below:

- List of boroughs and neighborhoods in NYC
- Latitude and Longitude of the neighborhoods
- Chinese restaurant venue data

METHODOLOGY

- Loading and exploring NYC boroughs and neighborhoods data from a certain dataset
- Getting Latitude and Longitude data of the boroughs via Geocoder package
- Using Foursquare API to get Chinese venue data related to the neighborhoods
- Running *k*-means to cluster the neighborhoods