

1. Introduction:

New York City, as one of the most populous cities in the World, is known for the flamboyant Times Square, stunning skyline and a massive array of shops, museums and more. While what attracts me most is the over 20,000 restaurants in the city. Especially, New York City's food culture is greatly influenced by the city's immigrant history. Italian immigrants brought New York-style pizza and Italian cuisine into the city, while Jewish immigrants and Irish immigrants brought pastrami and corned beef, respectively. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city.

In this project, we are going to explore the restaurants in the New York City. We will cluster its 306 neighborhoods into 5 groups based on the similarity of their restaurants' category. We will also conduct a detailed analysis for the Chinese restaurants. This project will be specifically helpful for:

1. Business personnel who would like to invest or open a Chinese restaurant in New York City. This project will help them to find out the best location and name for the restaurant, as well as obtain a general idea of the overall New York City Chinese restaurants' status, etc.;
2. Chinese food lovers who recently moved to New York City. This project will be helpful for them to easily find out the best Chinese restaurants;
3. Data scientists who want to implement some of the most used data analysis techniques (e.g. *k-means* clustering), data visualization techniques (e.g. Folium Map), and get familiar with some very useful databases (e.g. Foursquare).

2. Data:

For the neighborhood clustering problem, we obtained the following data set:

1. New York City Neighborhood dataset, which contains the name, latitude, longitude, and borough information for all of the 306 New York City neighborhoods. It was downloaded as a json file from https://geo.nyu.edu/catalog/nyu_2451_34572.
2. New York City restaurants dataset, which was retrieved from the Foursquare location database. For each New York City neighborhood coordinate, we retrieved the top 100 restaurants within a radius of 800 meters to it. The restaurant information includes its name, latitude, longitude and category.

For the Chinese restaurant analysis, we obtained the following data set:

1. New York City Chinese restaurant dataset, which was retrieved from the Yelp database. This dataset includes 1000 restaurants information consisting of restaurant name, price level, rating, review counts, latitude and longitude.