

# **Applied Data Science Capstone Project**

**Metis XU**

## **Introduction**

New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. New York City was home to nearly 8.5 million people in 2014, 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. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.

New York City has also been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and/or have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, and French etc.

For many Chinese residing in the US or particularly in the New York City, it's hectic to find Chinese cuisine restaurant. As a result, a lot of Chinese students and workers look for their home cuisine in the foreign countries, and this brings to the development of a lot of Chinese restaurants in the US. For retailers, the central location and the large crowd, low neighboring restaurants provides a great business. As a result, there are many Chinese cuisine restaurants in the New York city, USA and many more are being built. Opening restaurants in perfect place allows restaurant managers to earn a lot. Particularly, the location of the restaurant is one of the most important factors that determines whether the restaurant will be success or a failure.

# Problems

Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the main business question: What is best location in New York City for a Chinese Cuisine?

Along with some sub-questions like:

1. Which areas have potential Chinese Restaurants?
2. Which areas lack Chinese Restaurants?
3. Which is the best place to stay if you prefer Chinese Cuisine?