# IBM Applied Data Science Capstone

# Capstone Project - The Battle of Neighborhoods

# **Opening a Japanese restaurant in New York**

## Introduction/Business Problem

New York is the most populous city in the United States and one of the largest metropolises in the world. Its food culture is influenced by the history of urban immigration. For example: Central and Eastern European immigrants brought bagels, cheesecakes and hot dogs to the city. Italian immigrants brought pizza and pasta into the city, and Chinese and other Asian restaurants are all over the city.

As shown in the picture, New York City has attracted many people to start businesses in the food industry. Among them, the sushi buffet is very popular because of its exquisite health and convenience.

This report explores which neighborhoods and districts in New York City have the most and best Japanese restaurant. In addition, I will try to answer the question "Where should I open a sushi buffet?" And "If I want great sushi buffet, where should I go?"

## Data

In order to answer the above questions, data on New York City neighborhoods, boroughs to include boundaries, latitude, longitude, restaurants, and restaurant ratings and tips are required.

New York City data containing the neighborhoods and boroughs, latitudes, and longitudes will be obtained from the data source: https://cocl.us/new_york_dataset

New York City data containing neighborhood boundaries will be obtained from the data source: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

All data related to locations and quality of Italian restaurants will be obtained via the FourSquare API utilized via the Request library in Python.

## Methodology

1. Data will be collected from https://cocl.us/new_york_dataset and cleaned and processed into a dataframe.
2. FourSquare be used to locate all venues and then filtered by Japanese restaurants. Ratings, tips, and likes by users will be counted and added to the dataframe.
3. Data will be sorted based on rankings
4. Finally, the data be will be visually assessed using graphing from various Python libraries.