**Introduction**

New York City is multicultural, the largest city and the financial hub of the USA. An international restaurant chain AAA wants to know the best location to open its first restaurant in New York. AAA sells spicy dumplings with 3 different levels of hotness.

New York has more than 1.2 million Asian population: Chinese, Filipino, Korean and Japanese (total 949 K) including 315K South Asians. Since most of the Asian people are familiar with the dumplings, the dumplings from the company AAA tastes different than what is served in other Asian countries. AAA believes once people get the taste of its dumplings, they will love them and convert to its regular customers. In addition, certain percentage of people from other culture who like international food will also become its customer as well.

**Objective**

In this project, I will study in detail using segmentation and clustering to answer following business question:

Best location to open the restaurant:

* Highest Asian population – they are familiar with the dumpling without any spices, boiled in water. They are expected to like the spicy dumplings once they eat it.
* Highest South Asian population since they are familiar with the taste of the spices used in the dumplings.

**Data**

I will be using Wikipedia data for New York City.

1. New York population
2. New York demographics
3. List of New York locations by per capita income
4. Foursquare API

<https://en.wikipedia.org/wiki/List_of_New_York_locations_by_per_capita_income>

h[ttps://en.wikipedia.org/wiki/Cuisine\_of\_New\_York\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City)

h[ttps://en.wikipedia.org/wiki/List\_of\_Michelin\_starred\_restaurants\_in\_New\_York\_City](https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City)

<https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City>

* Population and demographics data will be used to find the locations with highest number of Asian immigrants.
* Per capita income will be used to further narrow down the location with high income so that they can dine out frequently, and expectation is they can make multiple dining in the restaurant per week.
* Foursquare data will be used to find the location with higher number of Asian cousins, which in conjunction with high per capita income will help to determine the location to open the restaurant.

Methodology

This will include the following components of the analysis…

1. Get Data
2. EDA to understand the data
3. Data wrangling/preprocessing and preparation
4. Modeling/segmentation and clustering
5. Review/modify the model as needed.