Using Location Data and K-Means Clustering to determine the best locations for a new Restaurant Chain and its Distribution Centres in New York City

Introductory Section / Business Problem

There are already a multitude of restaurants, bars and lifestyle locations in New York City. Fierce competition can therefore be expected when opening new restaurants**.** To ensure business success it is crucial to determine the right location for the new venue taking into account the characteristics of the neighborhood, population and restaurant variety. The search for white spots where the expected demand is high, but the restaurant density that corresponds to the demand is still rather low increases the chances of a successful start.

This report examines how census and location data can be used to determine the locations of a new Chinese restaurant chain that wants to gain a foothold in the New York restaurant scene. These are the main questions that will be answered:

- **Which districts are the most promising to open up new Chinese restaurants?** This question will use data about density of existing Chinese food locations and Chinese population data.
- Having determined the new restaurant locations, the next step is to deal with logistic questions. A well-organized supply chain will not only ensure smooth supply of the restaurants but also reduce costs. Therefore, the question will be answered: **Which are the best places to set-up the distribution centres for supplying the restaurant chain?**

Methodology and Data Selection

This section describes the methodology and data that will be used for answering the business problem. To determine the best locations for opening up new Chines restaurant, the following data is collected:

- **Foursquare location data** to retrieve the existing Chinese restaurants per New York District. This data will be accessed using the public Foursquare API. The data will include address, coordinates, name and category of the restaurant.
- **Census Data** to examine the Chinese population density per New York District. The Official Website of the City of New York offers access to the decennial census data, which is collected once a decade. This data includes the Asian population and selected subgroups per community district. For the analysis the data from 2000 is used which can be downloaded here: https://www1.nyc.gov/site/planning/data-maps/nyc-population/demo-tables-2000.page.

  In total there are 55 New York community districts in the data set. The following table shows an example of the census for selected Brooklyn districts:

Table SF1 P-110: Total Asian Population by Selected Subgroups

New York City Community Districts, 2000

Community District	Total Asian Only				
	Total Asian	Chinese	....	Thai	Vietnamese
Brooklyn 1	5.850	3.566		19	134
Brooklyn 2	4.604	1.901		59	91
Brooklyn 3	1.471	453		6	26
Brooklyn 4	3.336	1.648		12	89
Brooklyn 5	6.013	1.778		8	65
Brooklyn 6	4.739	1.748		47	86
Brooklyn 7	21.252	17.362		24	388
Brooklyn 8	1.567	358		6	55

To map the census data on the location data, a list of neighborhoods by community district is used. This data can be accessed via the following Wikipedia website: https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City

To estimate the likelihood of success for a new restaurant the following ratio is applied: **Number of Chinese restaurants per 10,000 Chinese inhabitants.**

Having determined the most promising districts to open the restaurants, the next step is to find the best location to set-up the distribution centres for supplying the restaurant chain. This is done by applying K-means clustering. The coordinates of the selected restaurant location are used to cluster the restaurants and define the best locations for their distribution centres.

https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm