Data

As our problem statement is about opening a Thai restaurant in New York City by knowing the location having lesser number of Thai food restaurants already present and having the population which is interested in Thai food. For this we should know the boroughs and neighborhood in New York and the people's preference for the cuisines. The preference of cuisines can be obtained by fetching the data from Foursquare API in which the top 10 preference of venues are mentioned for each neighborhood and we need to select the neighborhoods having Thai food restaurants as their 1st preference.

As I am going to work on New York City neighborhood we need a dataset that contains the boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. It will help us to see the locations and visualize the distribution of the Thai food restaurants over the particular location and help us in selecting the location for our new Thai Restaurant. The dataset for New York City having names and location coordinates of Neighborhood exists for free on the web.

Link to the dataset is: https://cocl.us/new_york_dataset

New York City geographical coordinates can be obtained by geocoder. It will be used by Folium library to make a map. The data of locations of neighborhood from the New York City dataset will be utilized as input for the Foursquare API that will be leveraged to provision venues information for each neighborhood. I am going to find out the best location for starting a Thai restaurant so I will retrieve the venues for Thai category. Thai category Id 4bf58dd8d48988d149941735 is used for retrieving data from Foursquare API.

Of the neighborhoods with Chinese, Ramen, Vietnamese, Korean, Japanese, Sushi or Thai Restaurants in the top 10 venues, I remove those neighborhoods that already have Thai restaurant in the top 10 venues. I then use Foursquare to find Thai restaurant in the 0.5 km vicinity of these areas that are not listed in the top 10 venues. This enables me to choose only those locations with no Thai restaurant in the specified vicinity. I exclude locations with Thai restaurant.

Now the neighborhoods of these venues are explored by the data obtained from foursquare API. The neighborhoods are grouped by neighborhood keeping a count of venue by its category so that we can know about which venue is most popular in that particular neighborhood. So, after getting this preference of venue data we can cluster the locations based on the venue preferences of the people to find out the location with least number of pre established Thai restaurant and having 1st preference of people of that location. In this way we can find out the desired location for our new Thai restaurant.