

## **Coursera Capstone Project**

**By Sana Syed**

### **The Battle of the Neighborhoods**

#### **Part of the IBM Data Science Professional Certificate**

#### **Introduction / Business Problem**

Chai San wants to open his first ever Thai Restaurant in England but, having never visited the country before, has no idea as to where would be the best location to begin looking in. We are tasked with identifying the best location for his new business venture.

The goal put-forward to us is to identify a location which houses zero other Thai restaurants in close proximity. The location should also not already be overcrowded with restaurants of any type and should be in an area with a high Thai population.

In order to identify which location would suit Chai San's restaurant the best, we will gather and analyze publicly available data of residents of England and facilities for each district. This will ensure that the location selected is optimal for the target demographic with a high density of potential customers and meets all of the goals set-forth.

#### **Data**

Based on the given business problem, we can determine that we will require the following data to perform our analysis:

- Population data of ethnic groups across England
- List of existing restaurants in the neighbourhood
- List of existing Thai restaurants in the neighbourhood
- List of nearby venues and their types in the neighbourhood
- List of the boroughs in the chosen county

The following sources will be used to gather the data required:

- Foursquare API for restaurant/venue information

- Office of National Statistics for Ethnic Populating Data
- <http://martinjc.github.io> For GeoJSON Data of England
- Wikipedia for a list of London boroughs and lat/lon values