# IBM DATA SCIENCE CAPSTONE PROJECT

**INTRODUCTION**

This capstone project is for IBM Data Science course. In this project I am assuming myself as an entrepreneur to start a chain for ‘Thai Restaurant’ as people in Toronto like Thai food. Which place is suitable for the start-up restaurant in Toronto?

**BUSSINESS PROBLEM**

The aim of this problem is to find the most suitable place to open the Thai restaurant in Toronto where I should setup my first restaurant. By using Data Science tools I will be solving this problem by clustering the restaurants according to the neighbourhood, the areas having less number of restaurants or opening at a new place with no Thai restaurants. Which place is profitable ?

**DATA SECTION:**

For this project we need the following data:

1. Toronto city data that contains neighbourhoods and boroughs along with latitude and longitude.

* Data source for neighbourhoods and borough <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* Data source for latitude and longitude 'http://cocl.us/Geospatial\_data'

1. Thai Restaurant in each neighbourhood of Toronto

* Data source: Foursquare API
* This will provide all the venues in each neighbourhood. We can use only Thai restaurant venues of our choice.

**Observations**

A glance at the generated map reveals that the majority of the restaurants in HCMC are situated up North and East of the city. Cluster 1 has almost no recognized restaurants while cluster 2 has the highest number of recognized restaurants. Last not least, cluster 0 has a moderate number of recognized restaurants. As a result, Opening a restaurant in districts in cluster 1 and 0 are more challenging due to the higher competitiveness. More importantly, it is significant that although most of districts in Cluster 1 are closers to the city center, they have very few recognized restaurants. Therefore, it is better to open a restaurant in Cluster 1.