IBM DATA SCIENCE CAPSTONE PROJECT
**Find the Best Location for Living and Opening a
Vietnamese Restaurant in City of Toronto**

Hai Nguyen

nhhaidee@gmail.com

1. Introduction

1. 1 Problem Description

For this capstone project, assuming that immigrant family is planning to migrate to Canada and choose Toronto as our destination. One of the biggest concerns for immigrant families is to choose area which has very crime rate but large population to open a business which is a Vietnamese Restaurant.

1. 2 Project Requirement

Propose the best area for the immigrant family with following requirements:

- Area must have low crime rate.
- Area must not have the common venues are Asian Restaurants so low number of competitors.
- Population must be greater than 25000.

2. Data

For analysis, we use the following data sources :

- Toronto Crime by Neighbourhood: this source of data is used to find neighborhoods which have low crime rate.

  https://opendata.arcgis.com/datasets/af500b5abb7240399853b35a2362d0c0_0.geojson

- Toronto Neighbourhoods Geo: this source of data is used to plot Choropleth diagram to crime rate distribution among Toronto Neighborhoods. The data contains geographical coordinate (latitude, longitude) of 140 neighborhoods of Toronto.

  https://open.toronto.ca/dataset/neighbourhoods/

- Foursquare Developers Access to venue data: this source of data is used to find venues for each Toronto Neighborhood which is then used for clustering method.

  https://foursquare.com/

3. Data Analysis Tools

Python libraries such as: Pandas, Scikitlearn, Numpy, matplotlib, Folium, Foursquare API will be used for data analysis