# Capstone Project - The Battle of Neighborhoods

1. **Introduction**

Problem background:

Pho, a World Famous Vietnamese Cuisine Phenomenon! In 2011, along with Banh My, Pho was added to Oxford Dictionary as a representative of Vietnamese cuisine. For people who didn't know about Pho, it is a Vietnamese soup that is normally made with a bone-beef broth, Banh pho noodles, and thinly sliced beef, that's often served with bean sprouts and other fresh herbs on the side. Not to be confused with Japanese ramen, which is usually made with wheat noodles, Pho is made with rice noodles.

Problem description:

Pho Ong Hung is one of the most famous brand in Vietnam. Now, the brand has a plan to expand its business in Toronto. The first job is finding the most suitable place for its.

1. **Data**

Based on definition of our problem, factors that will influence our decision are:

* number of existing restaurants in the neighborhood (any type of restaurant)
* number of and distance to Vietnamese restaurants in the neighborhood, if any
* distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered on city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

* centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
* number of restaurants and their type and location in every neighborhood will be obtained using foursquare API
* coordinate of Toronto center will be obtained using Google Maps API geocoding of well-known Toronto location

1. **Methodology**

In this project we will direct our efforts on detecting areas of Toronto that have low restaurant density, particularly those with low number of Vietnamese restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required data: location and type (category) of every restaurant within 6km from Toronto center (Royal Museum). We have also identified Vietnamese restaurants (according to foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Toronto - we will use heat maps to identify a few promising areas close to center with low number of restaurants in general (and no Vietnamese restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Vietnamese restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

1. **Results**

Our analysis shows that although there is a great number of restaurants in Toronto (~1300 in our initial area of interest which was 12x12km around Royal Museum), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected south from Royal Museum, so we focused our attention to areas north, west and east, corresponding to boroughs Etobicoke, Scarborough. Another borough was identified as potentially interesting, but our attention was focused on Etobicoke and Scarborough which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Royal Museum) we first created a dense grid of location candidates; those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Vietnamese restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Vietnamese restaurants particularly.