

# Capstone Project - The Battle of Neighborhoods

---

## Part I (Week 1)

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

This submission will eventually become your **Introduction / Business Problem** section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it.

## Part 2 [Week 1]

---

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

This submission will eventually become your **Data** section in your final report. So I recommend that you push the report (having your **Data** section) to your Github repository and submit a link to it.

# Section 1: Introduction

---

Utilizing the Foursquare location data, crime statistics and restaurants availability, schools and other convenient shopping stores in Vancouver for homebuyer's decisions and house price predictions.

## Background

---

[FourSquare](#) is one of the best sites that will tell you all about places to go, things to see, restaurants to eat at, bars to drink in, nightclubs to part the night away in and then where to go in the morning to get breakfast and a strong coffee. However, Foursquare alone can't give you hints of which areas are perfect to settle or live. It also can't predict the real price of the home in an area. If we combine data from several sources we might be able to build a model that will help homebuyers to take such a life decision with accurate if not close to accurate about the real price. That is no need even to consult a real state agent or spend a lot of time searching if that is the real value of the house you decide to buy or not. Imagine the following scenario:

1. You like to plan ahead and always review your options and make your choices about where you will settle and live.
2. You are in a city and you plan to move to a different one (example Vancouver) to settle after you got a job offer.
3. You start exploring the chances to buy a home in an area in Vancouver.
4. But you know no one in Vancouver to show you around to all the best areas and even if so they don't have much of information about the house prices or they don't know what you really looking for.
5. Also you want to avoid directly asking a real state agent and even if so you need a full and accurate estimate of the prices based on your own criteria.

**What do you do ... ?**

## Project Idea

---

My idea for the Capstone Project is to show that when driven by venue and location data from FourSquare, backed up with open source crime data in Vancouver, that it is possible to predict the price of a home based on your own criteria.

A high level approach is as follows:

1. The homebuyers decides on a city location [in this case Vancouver]
2. The ForeSquare website is scrapped for the top venues in the city
3. From this list of top venues the list is augmented with additional geographical data
4. Using this additional geographical data the top nearby restaurants, schools, shopping mall are selected and scored.
5. The historical crime within a predetermined distance of all venues are obtained
6. A map is presented to the to the homebuyers showing the selected venues and crime statistics of the area.
7. The base estimate of a home in a neighborhood is predicted based on some criteria.

**Note: The prediction model can be interactive and the homebuyers can add futures to see how the price can be affected (example number of rooms, bathrooms, size,...,etc). But because of time limitation of this project I will only achieve till point seven where a basic estimation of a single family home will be predicted.**

### *Who is this solution targeted at*

This solution is targeted regular homebuyers and any real state investors. They want to visualize the home prices on the map based on the prediction model giving all previously mentioned criteria

There are many data science aspect of this project including:

1. Data Acquisition
2. Data Cleansing
3. Data Analysis
4. Machine Learning
5. Prediction

## Section 2: Data

### Data Description

---

In this section, I will describe the data used to solve the problem as described previously.

As noted below in the Further Development Section, it is possible to attempt quite complex and sophisticated scenarios when approaching this problem. However, given the size of the project and for simplicity only the following scenario will be addressed:

1. Query the FourSqaure website for the top sites in Vancouver
2. Use the FourSquare API to get supplemental geographical data about the top sites
3. Use the FourSquare API to get top restaurant recommendations closest to each of the top site
4. Use open source Vancouver Crime data to provide the user with additional crime data.
5. Use the average price in the city of Vancouver as a basic price.
6. Then use the criteria to estimate the predicted price based on all previously mentioned factors.
7. Visualize the home prices on the map.