

**Capstone Project**  
**The Battle of the Neighborhoods**  
**(Leveraging location data for problem solving)**

**Hazem Alaa Shams**

**6 March 2020**

## 1. Introduction

### 1.1. Background

Relocating for work can be intimidating – even more so when you have to relocate to another country which you have never been to before and now you have to take your partner or family into account as well. It's an overwhelming life event that makes you start thinking - what should I do, where do I begin and how do I approach this? The idea for this Capstone Project is to show how leveraging location data from FourSquare and other data sources can assist you with making decisions based on your individual needs when having to relocate. The approach will be to follow and apply the systematic data science methodology to our scenario where I will:

1. Understand the problem and identify our approach
2. Identify the required data
3. Collect and understand the data
4. Prepare the data
5. Analyze the data
6. Model the data
7. Evaluate the model

### 1.2. Problem

In this scenario the individual needs to relocate to Chicago, United States where he will be working. He needs to take his partner (who is a chef) into consideration as she will be relocating with him and will also have to look for a new job. He also needs to take his child into consideration for a school.

He starts off by listing his and his family's basic wants and needs to identify how to approach the problem and to identify what data is required.

<b>Security</b>	<b>To live in a safe environment</b>
<b>Close to working locations</b>	Must have loads of restaurants in the area, Close to Museum
<b>Public Elementary school</b>	Neighborhood close to school

## 2. Data

### 2.1. Data acquisition

From the basic wants and needs list in the scenario description the following data is required

1. Crime data
2. School data
3. FourSquare data on restaurants and work location

Open source data from Analyze Chicago is obtained which is the City of Chicago's open data hub. From here the crime data is obtained as well as school data. This will assist in identifying safe neighborhoods to live in as well as where the options are for Elementary schools.

The FourSquare API to query is used for geographical data on restaurants. It will be ideal to stay in the neighborhood or close to the neighborhood where there are many restaurants as a potential work option for the wife. FourSquare is also used to identify the location of the museum so that it can be taken into consideration when identifying possible neighborhoods to minimize extensive travel where possible. For each dataset used in this study a similar approach is followed to first explain where the data comes from, what is contained in the data and how the data is prepared in order to start working with a clean dataset.