**The Battle of Neighborhoods**

**Part 1 [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.

**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.

**Part 3 [Week 2]**

In this week, you will continue working on your capstone project. Please remember by the end of this week, you will need to submit the following:

1. A full report consisting of all of the following components (**15 marks**):
   * Introduction where you discuss the business problem and who would be interested in this project.
   * Data where you describe the data that will be used to solve the problem and the source of the data.
   * Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.
   * Results section where you discuss the results.
   * Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
   * Conclusion section where you conclude the report.
2. A link to your Notebook on your Github repository pushed showing your code. (**15 marks**).
3. Your choice of a presentation or blogpost. (**10 marks**)

# Section 1: Introduction

In this section I will clearly define the idea of my choosing, where I leverage the Foursquare location data to solve the imagined business opportunity.

## Background

There are 100's, maybe even 1000's, of travel sites on the Internet, including [FourSquare](https://github.com/davidcolton/Coursera_Capstone/blob/master/www.foursquare.com), 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. The problems with these sites is that they are one dimensional. If you want to find out all this information about a city you plan to visit next month, **you** have to do the hard work. Also, just because a venue is the hottest place to go for a night out does not always mean that the unwitting tourist should just ramble in unprepared. The areas surrounding this new venue might be riddled with crime including muggings, car theft and assault, for example. Approach the venue from any direction other than from the north and you could be putting your life in danger. This is when my idea comes in.

Imagine the following scenario:

1. You like to plan ahead and always review your options and make your choices about where you will visit and eat up front before you travel.
2. You are flying to San Francisco for a Data Science Conference.
3. You arrive in san Francisco the day the conference starts but you've managed to convince your boss to delay your return by a few days giving you time to explore.
4. But you know no one in san Francisco to show you around to all the top sites and to bring you to the best restaurants.
5. Also the last time you went to a conference you were mugged and had you passport. money and credit cards stolen so you're now nervous of going somewhere without first researching the venue and the surrounding area.
6. The conference is next week and you don't have time to do all the research you'd like.

**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, that it is possible to present the cautious and nervous traveller with a list of attractions to visit supplementd with a graphics showing the occurance of crime in the region of the venue.

A high level approach is as follows:

1. The travelers decides on a city location [in this case san Francisco ]
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 are selects
5. The historical crime within a predetermined distance of all venues are obtained
6. A map is presented to the to the traveler showing the selected venues and crime statistics of the area.
7. The future probability of a crime happening near or around the selected top sites is also presented to the user

#### Who is this solution targeted at

This solution is targeted at the cautious traveler. The want to see all the main sites of a city that they have never visited before but at the same time, for whatever reasons unknown, they want to be able to do all that they can to make sure that they stay clear of trouble i.e. is it safe to visit this venue and this restaurant at 4:00 pm in the afternoon.

Some examples of envisioned users include:

* A single white female traveler
* An elderly traveler that has had previous back experiences when travelling

There are many data science aspect of this project including:

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

Now that the conference is over the Data Scientist can explore san Francisco and feel much safer.

# 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 san Francisco
2. Use the FourSquare API to get supplemental geographical data about the top sites
3. Use the FourSquare API to get top restaurent recommendations closest to each of the top site
4. Use open source san Francisco Crime data to provide the user with additional crime data

# Conclusions and Discussions

Although all of the goals of this project were met there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be developed into a fully pledged application that could support the cautious traveler in an unknown location.

Of the contributing data the san Francisco Crime data is the one where more data would be good to have. Also not every city in the world makes this data freely available so that is a drawback.

FourSquare proved to be a good source of data but frustrating at times. Despite having a Developer account I regularly exceeded my hourly limit locking me out for the day. This is why Pickle was used to store the captured data.