

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](#), 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 Chicago for a Data Science Conference.
3. You arrive in Chicago 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 Chicago 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 your 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 traveler with a list of attractions to visit supplemented with a graphics showing the occurrence 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 Chicago]
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 grographical data
4. Using this additional geographical data the top nearby restaurents 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 reaons 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 Chicago and feel much safer.