

# The Evaluation of Criminal Threat for Venues in San Francisco

Tai-Wei Chao

4/20/2021

## 1. Introduction

### a. Background

There were many public venues unexpectedly suffered a loss caused by criminal incidents every day. Are stakeholders of every single venue aware of the threat level in their vicinities? Do they evaluate the public security is based on their feeling or data driven?

### b. Problem

We already know there are some statistics available, and which is divided by district. The district is a broad scope with hundreds or thousands of venues, a venue where is sited in a good public security district would be less threat level than others in a poor public security district? Is this a fair judgement? In addition, the type of criminal incidents affect business differently, thus what type of criminal incidents occurred nearby is an important indicator too. This project aims to provide an evaluation of the criminal threats to each venue in the radius of 50 meters.

### c. Interest

Most insurance companies would be immensely interested to know the accurate threat level for each of venues; they can offer customized rates based on criminal types and frequency within the 50-meter radius. It means insurance companies can generate a strategic marketing plan and offer aggressive rates to their existing or potential customers with less risks as they have massive information to support the decision. Besides, real estate agencies and merchants will show huge interest in this project too.

## 2. Data acquisition and Solution

### a. Data sources

First of all, we need the incident report from Police Department of San Francisco which shall include incident date, type, coordinates, and so on. It can be found from DataSF (<http://data.sfgov.org>), however I chose to retrieve dataset from IBM Developer Skills Network where provides data narrowed down to what I need mostly. ([Data links](#))

I would use 2016 dataset to instantiate the project. In addition to the incident report, we also need venues information including name, location, and coordinates which we can retrieve through Foursquare API.

b. Solution

Once we get the incident report and venue location, we can search the incidents occurred within the 50-meter radius of each of venues then we can display the threats to every single venue. We will further cluster venues into different group by using K-means, it would be able to tell us the threat level for each of venues. We will also score the threat level, so clients can easier set pricing or plan based on the cluster and the score of each of venues.