Predicting Crime Levels in San Francisco Based on Foursquare Data

Colin Howarth
June, 2019

Problem

Summary

There are a variety of factors that may contribute to the crime level in different neighborhoods of a city. This project aims to identify whether Foursquare data can be used to predict crime levels in a neighborhood, and whether particular types of Foursquare location are particularly predictive of crime.

Audience

The output of this project may be of interest to city employees who are involved in planning for future zoning and development. It may also be of interest to police officers for planning staffing levels or for choosing locations for future police stations.

Data

Data Sources

Two primary data sources will be used:
1. The Foursquare Search API, which allows retrieval of a categorize list of venues for a given location.
2. San Francisco Police Department (SFPD) incident reports from 2018 onwards, available from the city's open data site.

Data Retrieval and Cleansing

SFPD Incident Reports

The SFPD data was retrieved and cleaned by:
- Removing un-needed columns (such as the incident ID and date-time of each incident)
- Removing incident reports that don't have latitude and longitude included.
- Removing incident report types that don't relate to a crime (such as "Lost Property" and "Non-Criminal")

The city of San Francisco was then divided into a 20x20 grid, and each incident mapped to one of the grid squares. The total number of crime reports of each type were then calculated for each grid square.

Foursquare Venues Lists

For each grid square, the Foursquare API was called to retrieve the list of venues in that grid square across each of Foursquare's ten top level categories. The count of the venues of each type was thus calculated for each grid square.

Combined Dataset

The two datasets were then combined to produce a single table listing the number of incident reports (by type) and the number of Foursquare venues (by type) for each grid square.

Methodology

Results

Discussion

Conclusion