Project Proposal Capstone 1

February 23, 2018

Description

Houston Texas with its 2.4 million residents is the fourth most populous city in the United States, just behind New York, Los Angeles, and Chicago. As with any large city, Houston has a rich sporting culture with five professional major league teams and four Division I college athletic programs. With so many sporting events through the year, *does crime increase around a given sporting event?* This project will show any connections between given sporting event and crime rate around it for the city.

The crime data is provided by the Uniform Crime Reporting program or UCR. It compiles official data collected by law enforcement agencies across the United States. UCR criminal offenses are divided into two major groups, Part I and Part II.

Part I offenses are considered to be serious and are broken into two categories: violent and property crimes, those include murder, rape, robbery, aggravated assault, burglary, theft, and auto theft.

Part II offenses are all crimes classifications other than those defined as Part I, some of those include: other assaults, forgery, fraud, vandalism, prostitution, disorderly conduct.

Datasets from Part I offenses will be used to find the type of crime around a sporting event given its stadium location.

The client would be Houston police department; the outcome would result in the rate change of patrols on a given sporting event.

Data Collection

Datasets will include crime statistics of the last eight years, and schedule of sporting events for the five professional Teams and four Universities within that time frame.

[Police Department Crime Statistics](#)

- date: *date of offense, includes month/date/year*
- hour: *approximate time when it occurs, value from 0-24*
- offense type: *Part I offense type*
- beat: *territory that police officer patrols.*
- premise: *type of location, road, apartment, parking lot.*
- block range: *value range of street.*
- street name: *name of street where the offense occurs*
- type: *street type, rd., Blvd.*
- suffix: *N, S, E, W*
- offences: *times offense happen within the time frame.*

Sports dataset **not yet determined** will include for each of a five major league sports and College athletic programs:

- *date of event*
- *type of sport*
- *stadium location (GPS location)*
- *playoffs?*
- *winner?*

Solution Techniques

The first approach will be to attain the GPS location of all the stadiums in question along with given sporting event date, then match that with crime around the given stadium.

The second will be to create a predictive model yet to be determined to answer the question that would satisfy the client within the appropriate time frame.

Last will be to deliver the project with original datasets, techniques used (logistic regression, clustering...) along with slide deck with given results.