Houston crimes: predicting crime risk probabilities

Problem

How safe is the city we live in regarding crimes? How has the crime rate changed during the past years? Which types of crimes have increased and which has decreased and why? How is crime rate correlated with economic status (unemployment rate) and demographics (population change)? Other interesting questions include: what's the correlation between crimes and weather and season; what are the common features among victims? The police department gathered a lot of information over time about crimes in the city. We propose to use the HPD crime data enriched by data from various sources regarding economic status, demographics and weather conditions to predict crime risk probabilities in a city by time, date and location based on historical crime data in the city.

Clients

Who wants to become victims of criminals? Noboday? So let's fight against crimes! Police board, government, and general public would be beneficiaries. The police officers can use this model to be better deployed. The government can take precausions more efficiently. Residents can use this model to better protect their lives and properties.

Data

The crime data will be acquired from police beat crime statsics. The data is broken down by police districts and beats and shows reports of the following types of crimes: murder, rape, robbery, aggravated assault, burglary, theft, auto theft on a monthly basis. It contains crime informaiton on date, hour, offense type, beat, premise, blockrange, streetname, type and number of offenses. In this project, we will investigate the data collected during 2010-2017. So we need to scrape the data and stack them as a whole. It's also useful to find correlations between crimes and weather conditions, economic status (unemployment rate) and demographics (population change). We will aquire data for the selected city only.

Method

This is a supervised machine learning problem. We will acquire the data and do data cleaning followed by some exploratary data analysis and build a predictive machine learning model. We will evaluate the model by future data. We can visualize the occurences of crimes on a map where the location is geocoded by combining the blockrange and streetname features.

Deliverables

The source codes along with a map and a presentation report will be shard through Github.