

**Jacquelyn Andrade**

**Yao Zhang**

**CS591 Project Proposal**

1. For this project we will be using the Crime Incident Reports (July 2012 - August 2015) dataset from the City of Boston open data portal. We will be retrieving the dataset from its end-point using the Socrata Open Data API. This dataset contains field values for the crime incident type, time, day, month, and year of incident, weapon used, true false values for whether shooting was involved and if the incident was domestic, along with the coordinates, street name, reported area, and district. Pre-processing will be necessary as we need to eliminate row values that are null as well as incident types that are not relevant to our study.
  
2. We will be analyzing the types of crime that occur in given locations around Boston on specific times of day and categorizing them using nearest neighbor classification techniques. We are only concerned with the crimes that threaten public safety. For this technique we plan to take a portion of the data as training data, and then use the remaining data to illustrate the accuracy of our prediction based on time and location. We plan on also using anomaly detection to account for types of crimes that may be considered uncommon in particular areas in Boston. We want to generalize our analysis so outliers will not skew the data.
  
3. Our categorical and predictive analysis will hope to benefit the Boston Police department in granting more insight into the nature of crimes prevalent in certain areas and improving police response. Given a specific time and location, we hope to be able to provide an appropriate prediction for police to account for the type of crime they should expect.

4. We expect to be able to predict when and the type of crime most likely to occur in a certain area with a good amount of accuracy. Our findings should closely correlate to “typical” crimes associated with certain times of the day. For example, we do not expect to find a substantial amount of crimes such as DUI in the morning, whereas robberies are more likely to occur at any time of day.