**Capstone Proposal**

**Problem:**

According to City-Data.com, for the period of 2002-2016, Chicago has had, for the most part, crime rates above the U.S. average (<http://www.city-data.com/crime/crime-Chicago-Illinois.html>). Judging by how much above average Chicago is, it is likely that Chicago has had above average crime rates for a much longer period than 2002-2016. In an effort to mitigate crime rates in Chicago, I would like to figure out any spatial/temporal factors that would be useful in predicting the type of crime committed.

The Chicago police department would by my main client. If they were aware of connections between spatial/temporal factors and the type of crime committed, perhaps officers may be better able to anticipate the type of crime that may occur at a given area/time and either prevent the crime from happening or more easily catch the perpetrator. Depending on what type of crime would be most prevalent, the police department could act accordingly to better prepare themselves; for example, deciding how many officers are required to patrol in a certain area and what specific skills they would need to deal with the prevalent crime in that area. In being prepared, this could help cut down on police injuries/fatalities.

In addition, the Department of Transportation could use data found in this study. For example, if there was an area that had a significant amount of theft, perhaps the installation of more streetlights could help the problem.

**Data:**

Crime reports from 2001-present, minus the most recent 7 days, (<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>) from the Chicago Police Department’s CLEAR (Citizen Law Enforcement Analysis and Reporting) system will be primarily used. This dataset contains over 6 million crime reports with information such as date/time of the report, primary type of crime, block, community, and latitude/longitude.

If time permits and if no sufficient relationships between spatial/temporal factors and the type of crime are found, hourly weather data (<https://mrcc.illinois.edu/CLIMATE/welcome.jsp>) may be used. This data would be used to see if certain weather conditions a few hours prior to the crime could be an indicator to the type of crime committed. Another dataset that may be used would be Chicago community demographic, social, and economic data (<http://www.robparal.com/ChicagoCommunityAreaData.html>).

**Method:**

1. Download the latest dataset of crime reports from the aforementioned link.
2. Prepare data for analysis by:
   1. Looking for reports that may be missing primary type of crime and discarding them.
   2. Trying to figure out a way to estimate any other features that may be missing from the reports.
   3. Removing any duplicate reports.
   4. Creating new temporal columns:
      1. Time of day (morning/afternoon/evening)
      2. Day
      3. Weekday/weekend
      4. Holiday
      5. Months
      6. Seasons
3. Perform exploratory data analysis to find any significant correlations between spatial/temporal data and the type of crime.
4. Divide crime reports into training data and testing data. Based on the above findings, create a model to predict the type of crime using training data.
5. Test the accuracy of my model using testing data.

**Deliverables:**

1. Python script(s) used for the project.
2. Paper detailing my methodology and findings.
3. Slide deck presenting my methodology and findings.