**Analyzing the crime in Mexico**

**Abstract**

Crime in Mexican cities can be attributed to two separate processes: (1) collecting past reported crimes and the context that surrounded those crimes, sorted by time and location, and separated into severity and types of crimes; (2) training collected data in order to accurately represent how dangerous certain city areas can be at any month. The utility of this crime forecasting model ranges widely: through the extrapolation to other major cities to collect and train data based on similar metrics, incorporation into law enforcement patrol patterns to deter crime occurrences, or even incorporation into a mobile application that civilians can use at own discretion to gauge safety through the different places.

**Data:**

We plan to collect our data through scraping several websites that contain criminal, economic, geographical and demographic information at a municipality and state level for Mexico. That information is available at a monthly and yearly time basis. Also, the information is in CSV format so the challenge is to clean the data, standardize it and merge it in order to obtain a unique table composed of several variables that explain crime rate. The predicted variable in this case is going to be the crime rate and we will analyze the variables (demographic, environmental, economic etc.) that are correlated with the predicted variable. The number of independent variables we plan to obtain will be between 30 and 50.

**Analysis:**

Once we have scrapped, merged, and sorted the data, we will perform analyses on that, which can better inform law enforcement agencies on various issues. For example, one of the questions that we aim to answer is whether there are certain municipalities in Mexico that are more crime prone than others. If yes, then what are some of the most underlying factors that could be attributed to crime rates in those areas. We can perform several techniques to find that out such as correlation coefficients, multivariate regression analysis etc.

Once we have the analysis, we will create a dashboard, which will show heat maps of different municipalities of Mexico, and the user will have the opportunity to see the heatmaps according to different dates and crime activity types.

We can build upon the exiting model to create a crime forecasting model. This model, based on our training data, will tell a user the probability of a crime occuring in the near future in a particular area. For this, we could use the technique of LASSO, which was utilized by the researcher Chris Blattman in several research papers about cover crime forecasting in Liberia, Colombia and Indonesia.

**Visualization**

We will visualize the outcomes of the predictions, as well as descriptive statistics of the independent and dependent variables through a dashboard coded on Jango or Shiny (R based package to create visualizations).

*The Dashboard will include:*

Linear graphs

Heat maps

Tables (with a friendly user filters to query manually for specifically values)

Histograms

Scatter Plots

Gap minders style graphics (up to 6 variables represented at the same time)

Text to describe the methodology

Text to describe the variables used in the forecasts

Variables to manipulate: data, time, area, crime activity type

**Timeline**

Our timeline will closely follow the proposition below:

4th Week: Gather data

5th Week: Clean collected data

6th Week: Standardize and merge the data.

7th Week: Examine patterns and crime forecasting metrics

8th Week: Develop the dashboard

9th Week: Attempt to extrapolate to another city, hopefully Chicago, and final results.