David Hatem

Foundations of Data Science Project Proposal

P4P Nigerian Delta Data Analysis: Creating a Nigerian Peace Index

1. The problem I would like to solve is that of inefficient distribution of law enforcement and security resources throughout Nigeria. Resources are wasted at a national level when security forces and infrastructure are not distributed where needed most. This project is not meant to try and predict patterns of violence presented through organized efforts, such as attacks by Boko Haram or other organized crime syndicates.
2. My client would be the NGO Partners for Peace in the Nigerian Delta. With my project, I would hope to give them an analytic tool that would hopefully help predict which regions of Nigeria will be most likely to see spikes in interpersonal violent crime at certain times of year and times of day. This information could help Nigerian security forces distribute resources more efficiently.
3. The data I will use for this project is the Partners for Peace interactive crime distribution map. Several NGO’s regularly contribute crime reports to this map, and have made their data easily exportable to Microsoft Excel. I will use Excel tables containing the information from their reports for my project.
4. My approach to solving this problem will first be to compare crime figures from 2014 and 2015 by day of year and time of day. In the data frames detailing these figures, Region, Type of Crime, Time of Day, Day of the Week, and Month will be column names. I will then apply the difference in crime figures to assess changes between those years based on those criteria. I can then look at these differences to see how closely the change in crime rates between those years applies to 2015 and 2016. The difference in projected rates (the prior year’s rates) and actual rates will be applied to the projection for the following year. The algorithms that I will write will apply the adjusted projections based on the time/regional factors listed as columns. The algorithms will automatically calculate updates projections based on newly uploaded data. As time goes on, the accuracy of my projections can be assessed and scrutinized.
5. My deliverables will be the code written to both create and organize my data frames, as well as to get the difference in annual crime figures and apply them to the coming year. Additionally, I will provide a slide presentation showing precedent that validates looking at individual crime through a seasonal and temporal lens. Also, my project presentation will set the expectations for what this project is meant to accomplish, as well as define the limitations of the project.