David Hatem

Capstone Project Milestone Report

**Nigerian Seasonal Peace Index**

Introduction

At a cursory level, the problem of interpersonal violence in the Nigerian Delta has life-or-death ramifications for millions of people. Beyond immediate physical harm, widespread violence in Nigeria deters economic investment, infrastructure development, and adequate aid distribution.

Recognizing the severity of the problem, several international NGO’s have created and fostered the growth of the “Partners for Peace- Nigerian Peace Map,” which tracks the prevalence and location of human rights violations, demographic pressures, personal insecurity measures, economic pressures, collective violence, political violence and upheaval, public services dysfunction, and refugee displacement.

While the Partners for Peace organization has done a remarkable job of chronicling instances of reported violence throughout the country, there is a dearth of analysis on the raw data collected, providing little actionable direction to prevent future violence.

As organizations continue to discover the value in tracking and analyzing their data, the role of the data scientist continues to develop. I posit that the role of a data scientist has boundless potential to helping solve the most fundamental problems facing humanity. As such, my capstone project will be to create a set of algorithms that will aim to predict general patterns of different types of interpersonal violence throughout the country. I will do this by utilizing the ARIMA time-series analysis package within R. I plan on inputting crime statistics from 2014 and 2015 to see if they are statistically significant indicators of 2016’s crime statistics in terms of predicting prevalence of specific types of crime by month and region. My goal is to identify and track any predictive patterns apparent within those variables.

A Closer Look

The available data has all of the necessary information needed to track crime metrics by week, month, and region. The exact date, exact longitude and latitude, which of the 36 states/capital municipality, and the local government area in which the crime occurred in are all detailed by the NGO’s contributing to the data set. At present, the data sets appear to contain enough entries to comprise a representative sample size that will allow results to be statistically significant. All information necessary to conduct a time-series analysis on the data is present in the data sets.

In terms of the limitations of the data, a few extraneous variables need to be acknowledged. For one, premeditated crime on behalf of organizations such as Boko Haram would be problematic to create a predictive model for. The sample size of recorded attacks by this group may not be large enough to draw any significant inferences on when they will conduct attacks. Also, the model I plan to make does not account for political motivations. I cannot take into account sudden political turmoil arbitrary to the factors I’m examining, and the crime that said turmoil may spur.

Another factor to consider in modeling crime within Nigeria is the country’s explosive population growth. As of 2016, the country’s projected annual population growth rate was 2.44%. Being that Nigeria is the world’s fourth most populous country with an estimated population of 186,053,386, the country is set to experience a monumental boost to their total population in the coming years[[1]](#endnote-1). Any model I create will have to account for differences in population on a year to year basis. There will likely be more reported crime annually in the coming years, even if that does not translate to more crime per capita.

In terms of data wrangling and cleanup, I am fortunate in that the data provided by Partners for Peace is relatively orderly already. That’s not to say the data will not need some restructuring, however. There is a feature on the Peace Map website allowing users to filter through duplicate incidents reported by the same NGO, but it does not catch all duplicate occurrences. Therefore, I will need to personally eliminate any remaining duplicates. I plan on doing this by selecting all repeating instances of exact longitudinal and latitudinal coordinates as well as the date the crime took place, and then examining the qualitative accounts of the incident provided by the NGO’s in order to confirm that events are, in fact, duplicates. Also, I plan on removing the qualitative accounts of specific crime incidents from the data I will be working with, as I will not be applying code to them.

Preliminary Explanation

Upon briefly inspecting the data, a few things become clear. It appears that crime tends to be concentrated in the densely populated coastal cities and states, and in the less stable northern state of Borno. In the coastal states where crime appears to be more prevalent, this is not at all surprising, as coastal areas such as Lagos, Rivers State, and Akwa Ibom State are amongst the most populous and most densely populated in the country. Additionally, the prevalence of paramilitary group Boko Haram likely accounts for much of the interpersonal crime in the less densely populated Borno State. Also, there is a larger concentration of oil-related crime occurring in the southern states. This usually amounts to damaging pipelines, stealing oil, or piracy of off-shore oil tankards. With the recent discovery of between 0.5-1 billion barrels of oil off the coast of Nigeria, it would not surprise me to see a stark increase in oil-related crime[[2]](#endnote-2). I will be sure to monitor this in the course of the project.

Moving Forward

My plan is still to conduct a time series analysis that applies retroactive crime figures from 2014 to those of 2015 in order to see how they reflect 2016’s figures. Using R’s ARIMA modeling package, I will use the existing crime data to assess whether or not there is a statistically significant correlation between reported crime frequency and season for each Nigerian state. I also plan on integrating any applicable information covered in the upcoming “Data Analysis In Depth” section of the course. In moving forward, I will certainly need to adjust for actual population estimate differences occurring year to year, as I did not know how significant Nigeria’s population growth rate was before initial background exploration into the subject.

1. https://www.cia.gov/library/publications/the-world-factbook/geos/ni.html [↑](#endnote-ref-1)
2. http://www.ibtimes.com/nigeria-oil-discovery-2016-exxon-mobil-finds-billion-barrel-reserve-africa-2438189 [↑](#endnote-ref-2)