Baltimore Crime Data

Ion Barbus

With 58 murders per 100,000, Baltimore crime levels rival and exceed some of most troubled regions of the world regarding crime (Sean, 2020). The skyrocketing levels of crime have put the city on edge and has depleted the resources of the weary Baltimore police department that has been rocked by scandals and corruption. As such, it is important to be able to identify the key predictors of crime, characteristics of crime by neighborhood, and to be able to forecast the situation into the future to better allocate resources. More specifically the project will focus on using the data provided by the Baltimore police department to forecast crime into the future, using time series forecasting to predict trends.

The main goals of the forecasts will be to predict what might at first seem unusual spikes in crime. These spikes might be due to things such as scandals within the Baltimore police department, events in the city, or other events and to identify more traditional patterns as a result of weather patterns, holidays or annual pattern occurrences. I also want to identify neighborhoods that are at risk and understand the key characteristics of what makes them unique. This might lead to different policing and management practices based on the neighborhood. Finally, I want to be able to identify unique predictors of crime and readjust the model based on those.

The tools used for this project will primarily include MongoDb to store the 293k rows of police department data, and to update the data monthly. I will also use the PyMongo to connect the mongo database to the anaconda distribution and more specifically jupyter notebooks where the analysis and machine learning models will be created. The primary data will be the crime data from the police department, however additional datasets will be identified and used for the models. Datasets such as weather patterns, demographic datasets, city activity events and news articles will be used to supplement the main data. I will be primarily using time series forecasting to forecast the crime incidents into the future but will also utilize the coordinates in the dataset to cluster neighborhoods into different entities and better understand each. I will also be classifying which neighborhoods are most at risk by the values from the original dataset along with the supplementary data.

Citation: Kennedy, Sean. “'The Wire' Is Finished, but Baltimore Still Bleeds.” *The Wall Street Journal*, Dow Jones & Company, 7 Feb. 2020, www.wsj.com/articles/the-wire-is-finished-but-baltimore-still-bleeds-11581119104.