

Presented by:
Anindya Shrivastava
Vincent Amestoy
Utkarsh
Jada Ford

Executive Summary : Crime Time in Baltimore

The city of Baltimore has been reputed as a community rich in culture and history. The beauty of Baltimore is a marvel, challenged only by its increasing crime. In recent years, Baltimore has been plagued by a reputation of violence and a penchant for unscrupulous legislators. The environment has led many local citizens to question what and when can something be done. Our study looks to understand some of the basis of the crimes and the locations of their occurrences. We attempt to use random forest and K-means clustering to analyze the data, from which conclusions can be used to develop specific task force and development of new police protocols.

We selected to view the crime statistics of Baltimore to elaborate on the coincidence and recurrences in time and location of crime. The city of Baltimore has provided the public with their data collected through longitudinal studies completed by the local police departments. We estimate that there is a significant correlation between crime and specific measures of time. This include time of year and the day of the week.

The first step in our analysis has been to sanitize the data. The original data set came with 276,282 variables, among 16 different categories. Years without adequate crime information, and crime without specific details were removed from the analysis. Remaining data was plotted using K-Means clustering, which was then inferred to formulate our conclusion. K-means clustering and decision trees were both attempted unsuccessfully.

Further analysis should be done to gain a better accuracy rate. If accuracy is improved, solid conclusions can be made. If more data was provided for September through December, results may improve. Given the categorical nature of the data and the novice understanding of the team, regressions were a challenge to implement. With improved skill and use, the accuracy of the analysis can be improved.

Summarily, we can conclude that time and location could have an impact on crime, though further analysis is needed.. As we know, correlation doesn't equal causation so no explanation to explain why these times are significant can be deduced from this data alone. Some of the other conclusions we can make preemptively include: larceny and common assault are the two most common crimes; and the Northeast and Southeast are the most common areas for crime. This information is important to understanding the occurrences of crime so that they can be adequately prevented.