Jakub Zurakowski, Michael Sibbald and Alan Cahill

C00205431, C00206817, C00174533

4th Year Data Science Design Document

Investigating felony rates among youth in relation to avg. household income and population density in Baltimore

Institute of Technology, Carlow

Table of Contents

[Hypothesis 1](#_Toc529963102)

[Introduction 1](#_Toc529963103)

[Overview 1](#_Toc529963104)

[Proposed Solution 1](#_Toc529963105)

[Household Income 1](#_Toc529963106)

[Population Density 2](#_Toc529963107)

[Overall Probability of a Felony Crime 2](#_Toc529963108)

[Probability of a Felony Crime based on Age 2](#_Toc529963109)

[Compare Results and Predict 2](#_Toc529963110)

[Project Purpose 2](#_Toc529963111)

[Process 3](#_Toc529963112)

[Conclusions 4](#_Toc529963113)

[References & research documents 5](#_Toc529963114)

# Hypothesis

People in lower income areas of Baltimore are most likely to begin committing felony crimes at age 20-21.

# Introduction

Research is undertaken to find at what age younger population of Baltimore is most likely to commit felony crimes. The research involves looking at population density, income areas and study of previous committed crimes. (J.Z) The following document will detail the current situation in overview, the solution to the hypothesis in proposed solution and the process to be used to solve the issues in process.

# Overview

At this moment in time the Baltimore Police Department[3] provide several data sets for public use. Two of these data sets will be utilised in this project. They are the Incidents and Arrests data sets. There are also two additional data sets available that will not be used in this project. These are Calls for Service and Executive Summary. In addition to these data sets on the Baltimore P.D website there is a service available for developers to upload applications they have developed using the data sets made available to the public. Maryland State Commission[4] provides a pdf detailing offences in law by means of an offence table. Finally City-Data.com [2] provides information on its website about income and population density in Baltimore with each neighbourhood getting its own detailed breakdown of statistics. (A.C)

# Proposed Solution

The project proposes to solve a number of issues. These issues all relate to the level of crime in Baltimore. The following will be solved:

* Investigate crime levels in relation to household income in a neighbourhood.
* Investigate crime levels in relation to population density in a neighbourhood.
* Investigate the probability of a crime committed in Baltimore will be a felony crime.
* Investigate the probability of a crime being a felony based off of the age of the offender.
* Compare the results to predict the age people begin committing felony crimes as opposed to non felony crimes.

## Household Income

It is found that low household income can lead to increase in crime levels. City-data.com [2] outlines the average household income per neighbourhood, which can be used in conjunction with the data on incidents and arrests to pinpoint the troublesome areas. (J.Z)

## Population Density

Population density refers to the average number of people living in a square mile. High population density usually leads to lower standards of housing / living, which in turn leads to an increase in criminal activity. City-data.com [2] outlines the average population density per neighbourhood which can be used in conjunction with the incidents and crimes data to pinpoint the troublesome areas. (J.Z)

## Overall Probability of a Felony Crime

The project will investigate the probability that any crime committed in Baltimore will be a felony crime. All of the crime data will be passed through a Naive Bayes and will give a result detailing the probability that that a crime will be a felony. In order to do this the arrests data set must be appended with a label of “yes” or “no” based on whether it is a felony or not. This will be done by referencing the data set of the scraped data from the offence table pdf. (A.C)

## Probability of a Felony Crime based on Age

The probability of a crime committed being a felony crime will be investigated by splitting the data set into smaller datasets based upon the age of the offender. Each of these secondary datasets will be run through the Naive Bayes algorithm and a probability of felony vs misdemeanour will be recorded. (A.C)

## Compare Results and Predict

The results of the overall probability and the probability based upon age will be compared to each other. The comparison will show a pattern that will allow a prediction on the age people begin committing felonies to be made. This prediction will be made taking into account several factors that could impact the results. (A.C)

## Project Purpose

The main purpose of this project is to help reduce the rate at which felony crimes are committed by young people. It was decided that this would be done by distinguishing the age that these young people move on from misdemeanour to felony crimes. It has been a proven study that if there is an outreach towards young people by police and educational departments, crime will be reduced overall.[5] By using these studies, if the most common age that felonies are committed by people from twenty to twenty-one years of age, these age groups can be targeted for support from these departments, and in turn reduce the overall number of felony crimes. (M.S)

# Process

The datasets must be downloaded from the Baltimore Police Department website. This is straightforward and is completed without much problems. Secondly the data from the Offense Table pdf and the City Data website must be scraped. The offense table data will be manipulated in PDFElement Pro 6 [1] which will transfer the data into a Microsoft Excel table. This data will then be loaded into Python and cleaned. The data from the City Data website will be scraped using a python crawler. It will be cleaned and saved into a csv file.

To determine how average household income per neighbourhood affects crime levels, the household income in a given neighbourhood will be directly compared to the sum of incidents and arrests committed in a given neighbourhood. The data for incidents and arrests is provided by the Baltimore P.D [3] and the data for the neighbourhood income is scraped from City Data website [2] using python. (J.Z)

How population density affects crime levels in a given neighbourhood can be calculated by directly comparing the population density of a given neighbourhood with the sum of incidents and crimes for that neighbourhood, similarly as the impact of average household income per neighbourhood affects crime is calculated. The data for incidents and arrests is provided by the Baltimore P.D [3] and the data for the neighbourhood population density is scraped from City Data website [2] using python. (J.Z)

The process for predicting the age that crime committed will begin to be felony crimes is based off probability. To find the probability a Naive Bayes algorithm will be used on Rapid Miner [1]. The probability of felony crimes being committed will be derived using a Naive Bayes algorithm. The algorithm will give probability results of crime being a felony crime based off the overall data set and the sun datasets of age groups.  (A.C)

The idea for predicting the age that a person in Baltimore will begin to commit felony crimes is that the result of the overall probability of a felony crime will closely match up with the probability of a particular age group committing a felony crime. This theory will be subject to change however based on results. A big jump in probability could be considered an indicator that people are beginning to commit felony crimes even if it does not match the overall probability. (A.C)

There could also be issues regarding the dilution of a probability at a certain age or the dilution of the overall probability. This is to say that the amount of misdemeanour crime could impact the results in a positive and negative way and as such the methods used for prediction will be adjusted accordingly. (A.C)

# Conclusions

In conclusion, the primary goal of this project is to reduce the overall crime rate of the city of Baltimore, Maryland. It aims to do this by looking at possible common causes of crime in relation to social settings such as, household income, population density and rate at which felony crimes are committed. The data sets used detail the arrests and incidents of the city over the past two years. As well as these sets, there is a table provided by Maryland State Commission [4] that distinguishes whether a crime is a misdemeanour or a felony. As well as these sources, the website City-Data.com [2] provides information on the average income and population density of the city based off the city’s neighbourhoods. (M.S)

Recent studies have shown that if young people are exposed to the crime rate statistics and receive outreach from police and education departments it can significantly reduce the likelihood of these young people committing crimes [5]. In these studies, it was also shown that if previously convicted but rehabilitated criminals are involved in these outreach activities the results are even more prominent. (M.S)

In order to visualise the data, it will be loaded onto Rapidminer as a csv file, where it can generate graphs to represent the data. Rapidminer will also use the Naive Bayes Algorithm to predict the likelihood of a crime being a felony committed by a young person. These predictions will be based off of the population density, age and household income of the neighbourhoods. (M.S)

The results of this research can be forwarded to Baltimore P.D [3] with recommendations on how to more effectively prevent crime - especially among younger people. (J.Z)

# References & research documents

1. Best PDF Editor for both Windows & macOS. 2018. *Best PDF Editor for both Windows & macOS.* [ONLINE] Available at: <https://www.wondershare.net/ad/pdf-editor/?gclid=Cj0KCQiAw5_fBRCSARIsAGodhk_5c_W4HhxLMxSyBcKN_pydYYhTdNkw5gMgbmlY9YmEYN0zE1f8y58aApFaEALw_wcB>. [Accessed 11 November 2018]
2. Baltimore, MD Neighborhood Map - Income, House Prices, Occupations - list of neighborhoods. 2018. *Baltimore, MD Neighborhood Map - Income, House Prices, Occupations - list of neighborhoods*. [ONLINE] Available at: <http://www.city-data.com/nbmaps/neigh-Baltimore-Maryland2.html>. [Accessed 11 November 2018].
3. Home | Baltimore Police Department. 2018. *Home | Baltimore Police Department.* [ONLINE] Available at: <https://www.baltimorepolice.org/>. [Accessed 12 November 2018].
4. Maryland State Commission on Criminal Sentencing Policy (MSCCSP) . 2018. *Maryland State Commission on Criminal Sentencing Policy (MSCCSP)* . [ONLINE] Available at: <http://www.msccsp.org/>. [Accessed 12 November 2018].
5. Webster, D.W., Whitehill, J.M., Vernick, J.S. and Parker, E.M., 2012. *Evaluation of Baltimore’s Safe Streets Program: effects on attitudes, participants’ experiences, and gun violence. Baltimore, MD: Johns Hopkins Center for the Prevention of Youth Violence.*
6. Sci-Hub | Child maltreatment in an incarcerated sample in China: Prediction for crime types in adulthood | 10.1016/j.childyouth.2012.04.015. 2018. *Sci-Hub | Child maltreatment in an incarcerated sample in China: Prediction for crime types in adulthood | 10.1016/j.childyouth.2012.04.015.* [ONLINE] Available at: [https://sci-hub.tw/https://www.sciencedirect.com/science/article/pii/S0190740912001739](https://sci-hub.tw/https:/www.sciencedirect.com/science/article/pii/S0190740912001739). [Accessed 01 November 2018].
7. Sci-Hub | Crime Pattern Detection Using Data Mining | 10.1109/WI-IATW.2006.55. 2018. *Sci-Hub | Crime Pattern Detection Using Data Mining | 10.1109/WI-IATW.2006.55.* [ONLINE] Available at: [https://sci-hub.tw/https://ieeexplore.ieee.org/abstract/document/4053200/](https://sci-hub.tw/https:/ieeexplore.ieee.org/abstract/document/4053200/). [Accessed 01 November 2018].
8. Sci-Hub | Crimes analysis software: ‘pins in maps’, clustering and Bayes net prediction | 10.1016/S0957-4174(03)00097-6. 2018. *Sci-Hub | Crimes analysis software: ‘pins in maps’, clustering and Bayes net prediction | 10.1016/S0957-4174(03)00097-6.* [ONLINE] Available at: [https://sci-hub.tw/https://www.sciencedirect.com/science/article/pii/S0957417403000976](https://sci-hub.tw/https:/www.sciencedirect.com/science/article/pii/S0957417403000976). [Accessed 01 November 2018].