# Crime Analysis and Forecasting: Integrating Machine Learning with Population Density and Economy

By


**Mahabub Alam Khan**
**011181216**
**Suriya Najnine Nisamoni**
**011181131**
**Ahm Emramul Pasha**
**011181243**
**Sumiya Tabussum Dristy**
**011181237**
**Raj Shekhar Karmaker**
**011193149**
**Fatema**
**011193055**


Submitted in partial fulfilment of the requirements
of the degree of Bachelor of Science in Computer Science and Engineering

May 31, 2024


Department of Computer Science and Engineering
United International University

# Abstract

This paper reviews the application of machine learning techniques for analyzing and predicting crime rates in view of socioeconomic characteristics such as economic indicators and population density. The primary aim of this research project is the development of a prediction model that could be used by policy makers or law enforcement agencies in order to guide allocation of resources as well as crime prevention program planning. However extensively researched, the sphere of crime analysis, As of yet no research has been done on how the two variables of population type and size link with crime forecast. This paper seeks to establish this relationship by looking at the connection between crime rates and socio-economic status indicators. This information is sourced from different platforms such as World Bank or Macrotrends. Therefore our main aim is to fill this void with full understanding on how human concentration in an area tends to affect its economic standing as reflected by statistical evidences obtained from various sources including World Bank or Macrotrends in order to provide more comprehensive details on the causes of criminal activities. We are using ML techniques in analyzing crime rate and predicting it based on the traditional approach in criminology. We apply regression analysis to understand the relationships between crime rate and those social economic forces. In this way prediction models can be evolved which are able to give a clear insight on how crime is evolving in its connection with such factors like density or economic condition. In various ways this study could noticeably promote society. A more complex link between density of population and economic position achievement as well as crime is also understood better. Therefore, this will lead to better distribution of resources, increase in the efficiency of preventing strategies for crime and, lastly make communities safe from harm.

**Index Terms:** Bangladesh, crime prediction, analysis, Bangladesh police, ACLED, crime report

# Acknowledgements

Thanks to Almighty Allah. This work would not have been possible without the input and support of many people over the last trimesters. We would like to express our gratitude to everyone who contributed to it in some way or another. First, we would like to thank our academic advisor, Iftekharul Abedeen. Our sincere gratitude goes to our honorable class teacher Prof. Dr. Al-Sakib Khan Pathan. We also thank our group members, seniors, and classmates. Last but not least, we owe our family, including our parents, for their unconditional love and immense emotional support.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

In a rapidly changing world, studying and reducing crime contribute to the enhancement of public safety and improvement of society. This paper is mainly concerned with the issues of crime analysis as predicted by machine learning in conjunction with two significant factors; namely population dynamics and economic development. The main objective in our homeland is therefore; to observe, study it fully; even foresee its trends. This research seeks to provide intelligence on how societies can be better secured through convergence of technology, demographic data and economic indices; thus enabling policymakers shape up responsive mechanisms against crime onset and haphazardness in neighborhoods in conjunction with members constituting such establishment.

## 1.1 Project Overview

This project tries to look into how to analyse and forecast crime, with a focus on how to use machine learning techniques along with population change and economic growth as important factors in our country. We want to use complex algorithms to find hidden patterns in crime data and make prediction models that show how changes in population, the economy, and criminal activity are all connected in complex ways. Our goal is to give law enforcement and policymakers a full picture of crime patterns through time series analysis and creative data analysis. This will allow them to take proactive steps for public safety and community development. Our project aims to provide useful tools for dealing with and reducing crime problems in our specific social and economic setting by combining ideas from different fields.

## 1.2 Motivation

This study was inspired by a profound insight into the importance of coming to terms to deal with the sticky issues of crime in our society. As technology changes With more crime to record than ever, as well as the fact that the layout of streets and the way cities are patrolled are updated rapidly, the old criteria for appraising crime become obsolete. Using a predictive system based around crime typelikelihood, population change and economic growth as inputs using machine learning as main drivers can create a new era in attracting crime prevention and community development efforts to act proactively rather than reactively. We pick up on thoughts on how to keep the public safer from the way predictive modelling are going to allow us to prevent crime in advance. If crime patterns are studied other side of how the population and economy are changing so much so fast, and that we offer to information valuable to authorities and public policy. We are undertaking this study with the aim of not only enhancing scholarly endeavors but also, and of greater importance, creating functional instruments for our communities to address and lessen the menace of crime. It is our expectation that such an approach will yield tangible utility in strengthening safety and resilience in different societies.

## 1.3 Objectives

There are clear goals that this project aims to accomplish, which are:
**1. Prediction of Crime Patterns:** - Applying population and economic data in the development of machine learning models that can forecast crime occurrences in the future.

**2. Temporal Trend Analysis:** - Employing regression modeling techniques that look into crimes across different timespans.

**3. Visualization of Key Statistics:** - Consequently conducting exploratory data analysis aimed at making city-specific essential crime stats visually appealing when displayed on a screen.

**4. Collection of Decision-making Information:** Making wise decisions on how crime can be stopped by law enforcers and policy makers prompted for provision of important knowledge.

**5. Examination of Crime and Economic Development Relationship:** Influence of crime patterns, population transformations, and economic developments to society progress. The objectives of these objectives are to provide a deeper insight into crime patterns and to offer people interested in community safety measures valuable information to improve on the situation.

## 1.4 Methodology

This research project explores crime patterns in Bangladesh through a systemic approach involving use of varied machine-learning techniques. The main steps include:

**Collection of Data:** Acquire a range of datasets from the Bangladesh Police website and ACLED which consist of crime records, population data or any other economic indices that are within the regions chosen.

**Conducting Data Preprocessing:** The gathered data should be cleaned and preprocessed in order to be consistent and dependable. This implies correcting any missing values or anomalies that could potentially affect how effectively the model performs.

**Feature Engineering:** Select pertinent parts of the dataset while excluding superfluous ones, concentrating on factors like crime rates, population growth rates and some indicators of Bangladeshi economic performance.

**Train the models ready:** Take the processed data set and train each selected model, ensuring that the tunable parameters are good enough for performance and accuracy.

**Review the models:** Use indexes such as R-square, Mean Square Error(MSE) and Mean Absolute Error(MAE) for measuring how each model works, and evaluate its ability to identify crime trends within Bangladesh with some confidence.

## 1.5 Project Outcome

Upon completion of our study, we have generated valuable findings which are useful for law enforcement as well as for people living in the developing countries who want to know why it matters so much to examine crime. Here are the expected results:

**Advanced Predictive Models:** Utilize sophisticated machine learning algorithms to investigate crimes the way they involve demographic movements together with economic changes. The function of these models is to empower police departments with advance information that will enable them plan strategies as well as allocate resources.

**Informing Law-Enforcement Systems:** Providing them with intelligence-driven actions which enable them prevent any form of crime. This is due to the fact that when machine learning is combined with demographic and economic factors in the law enforcement strategies, it works better.

**Insights on community development:** this aspects are such as crime patterns, population changes, and how they relate to the level of economic growth in the third world countries. The intention of this work is to offer practical data such as policy makers can apply in their community development processes which will eventually lead to the betterment of society's wellbeing.

**Enhancing the security of members:** Through application of machine learning methods, the project is targeted at advancing the safety of the community by proactively identifying criminality trends. This finding is of great significance in ensuring safety and improving people's living standards.

**Fostering human capital in underdeveloped economies:** Concentration on the development of crime analysis skills, by sharing wisdom gained regarding crime investigation. Consequently, the project aims at enabling these Nations adopt innovative methods of curbing crime through machine learning techniques by also presenting the benefits it get from combining this with demographical and economic information.

To put it briefly, the project conforms with the broader quest of employing machine learning for human welfare. The research wants to significantly and permanently change our perception and approach towards criminal activities. This is possible by educating developing nations on the potential changes brought by insights derived from statistics and assisting to prevent crimes in advance by the police.

## 1.6 Organization of the Report

There is a title page. Then there is the abstract and the acknowledgment. Then, there is a table of contents. In the table of contents, the first chapter, Introduction, has a section named Project overview, a section called Motivation, a section called Objectives, a section called Methodology, Chapter two is Background, and there is a Preliminaries section. Then, the literature review section. Inside the literature review section. The next chapter, the name is Dataset. Then there is the Complex engineering problem chapter. The last chapter is the conclusion chapter. Two sections are there which are a Summary and future work.

# Chapter 2

# Background

To fully understand the context and impact of this undertaking, preliminary study is required. The goal of this study is to look at historical crime and violence trends and patterns in order to better assess current data. When we investigate details such as historical events and socioeconomic conditions, we can better understand the results of regression analysis, showing their underlying causes. This basic information is necessary when developing crime-prevention tactics and policies.

## 2.1 Preliminaries

**Regression:** regression analysis helps to understand how changes in independent variables influence the dependent variable, enabling both explanation and prediction of outcomes.

**Confusion Matrix:** A confusion matrix is a table used to evaluate the performance of a classification algorithm. It summarizes the outcomes of the classification process by comparing the actual (true) labels with the predicted labels. Each entry in the matrix represents the number of occurrences of a specific combination of predicted and actual classifications. The confusion matrix provides a detailed breakdown of the performance of a classification model, allowing for the identification of specific types of prediction errors and guiding improvements in model accuracy. Here we use R-Square, MSE, MAE.

**Population:** A population is the comprehensive group that researchers are interested in studying and about which they aim to make generalizations or gather insights.

**GDP:** GDP is a crucial metric in economics that provides a comprehensive snapshot of a country's economic activity and health.

## 2.2 Literature Review

For our research, there are some related literature reviews which include the Author's Year Site, Study Description, Method Adopted, and Results.

| Authors Year Site | Methodology | Dataset | Result |
|---|---|---|---|
| Bidun Christiana Obagbuwa and Ademola P. Abidoye Published 9 June 2021 [1] | The study uses the CRISP-DM methodology for data mining, including business understanding, data preparation, modeling, evaluation, and deployment | Crime data from South Africa's provinces, covering 27 crime categories, is utilized. Population and density information are included. | A linear regression model predicts crime occurrence based on population and density. The model explains 84. 7% of the variability in crime rates (R-squared = 0.847). |
| Luiz G.A. Alves a,, Haroldo V. Ribeiro b, Francisco A. Rodrigues 2018 [2] | The study employs the random forest algorithm for predicting crime and quantifying the influence of urban indicators on homicides. This machine learning approach is chosen due to its ability to handle non-linear relationships and multicollinearity in urban data. | The dataset includes urban indicators from all Brazilian cities, such as child labor, elderly population, GDP, illiteracy, family income, population, sanitation, unemployment, traffic accidents, suicides, and the number of homicides in the year 2000. These indicators are used to predict the number of homicides 10 years later. | The random forest model achieved up to 97% accuracy in crime prediction. Unemployment and illiteracy were found to be the most important variables for describing homicides in Brazilian cities. The model's robustness was confirmed as the importance of urban indicators remained stable under slight changes in the dataset. |
| Sridharan S1, Srish N2, Vigneswaran S3 and Santhi P4 15 February 2024[3] | Statistical and machine learning models (K-NN, Naive Bayes, Regression) analyze crime patterns using demographic, economic, social, victim, and geographic variables. | Indian crime data (2001-2016) includes over 500 entries. Focus on women, children, and IPC cases. | "Indian Crime Analysis" predicts crime ratios (2017-2020) with K-NN accuracy, identifying trend-changing years. |
| Rabia Musheer Aziz, Aftab Hussain, Prajwal Sharma, Pavan Kumar 2022[4] | The paper proposes a machine learning-based soft computing regression analysis approach for predicting crime data in India1. It utilizes various regression algorithms like Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Decision Tree Regression (DTR), Support Vector Regression (SVR), and Random Forest Regression (RFR) to build predictive models23. | The study used district-wise spatial-temporal crime data from 2001 to 2012, collected from the official website of the National Crime Records Bureau (NCRB)5. The data included various types of crimes such as murder, rape, kidnapping, abduction, riots, and more, across different regions and states in India. | The Random Forest Regression (RFR) model yielded the best fit for region-wise total Indian Penal Code (IPC) crime prediction with an adjusted R squared value of 0.9631551 and a Mean Absolute Percentage Error (MAPE) of 0.20274374. For region-wise theft crime count prediction, the RFR model also performed best with an adjusted R squared value of 0.966604 and a MAPE of 0.16571. |

| Authors Year Site | Methodology | Dataset | Result |
|---|---|---|---|
| Seema Aggarwal, Geeta Aggarwal and Manisha Bans2023 [5] | The study analyzes real-life data on crimes against women in India using statistical tests and visualization techniques. | The study analyzes real-life data on crimes against women in India using statistical tests and visualization techniques. | The study analyzes real-life data on crimes against women in India using statistical tests and visualization techniques. |
| Md Pavel Rahman, A.K.M Ifranul Hoque, Md. Faysal Ahmed, Iftekhirul, Ashraful Alam, Nahid Hossain 2022 [6] | Machine learning algorithms analyze crime patterns in Bangladesh using data from the Bangladesh Police website and ACLED. Decision Trees, Random Forest, and MLP models are used. | Crime statistics from 2010 to 2018 are collected from the Bangladesh Police website. ACLED data provides event types and geolocations | Successful crime trend predictions for 2019 using the Bangladesh Police dataset. Random Forest performs best for ACLED predictions in 2021. Law enforcement can benefit from this study. |
| MA Awal, J Rabbi, SI Hossain, MMA Hashem[7] | The paper provides an overview of the latest advancements in sentiment analysis. It covers preprocessing techniques, feature extraction methods, and classification techniques. The authors discuss machine learning, deep learning, and ensemble learning approaches. | The paper discusses widely used sentiment analysis datasets. Unfortunately, the specific datasets are not mentioned in the abstract. However, sentiment analysis datasets are commonly used for training and evaluating models. | The proposed hybrid semi-supervised method achieved the highest accuracy of 87.3% on the English SemEval 2017 dataset. The method was evaluated on three benchmark datasets using multiple classifiers, including machine learning, neural network, and ensemble learning. Overall, this paper provides valuable insights into sentiment analysis and serves as a resource for researchers and practitioners in the field |
| Amin Biswas ,Sarnali Basak (2019)[8] | The study employs a quantitative research approach. This means that it relies on statistical analysis and computational models to test hypotheses and predict outcomes.<br>Researchers collect data from various sources, ensuring a diverse and comprehensive sample for robust findings. | The paper analyzes a large-scale dataset. This dataset likely contains information relevant to the research question.<br>The dataset's size and diversity contribute to the study's validity. | The study concludes with significant findings. These findings could be related to accuracy, performance, or other relevant metrics. |

| Authors Year Site | Methodology | Dataset | Result |
|---|---|---|---|
| Md. Abdul Awal, Jakaria Rabbi, Sk. Imran Hossain, and M. M. A. Hashem (2016)[7] | The study uses linear regression to forecast crime trends in Bangladesh12. It involves collecting crime data, training the linear regression model, and then using the model to predict future crime occurrences. | The dataset is sourced from the Bangladesh police website and contains aggregated counts of various crimes4. It is divided into metropolitan and divisional region data, with 840 instances across the country. | The linear regression model was trained and used to forecast crimes like dacoit, robbery, murder, and others for 2016 The paper presents the accuracy of these forecasts and discusses the increase in crime rates alongside population growth. |
| Saqueeb Abdullah1, Farah Idid Nibir2, Suraiya Salam3, Akash Dey4, Md Ashraful Alam5 and Md Tanzim Reza (2020)[9] | The study applied machine learning algorithms (KNN, LR, RFC, DTC) to crime and victim data. It predicted criminal attributes like age, sex, race, and crime method. | Data were collected from the Bangladesh Police, including crime type, victim details, and area. Preprocessing handled missing values and encoded categorical data. | RFC achieved the highest accuracy (83.3%) for crime method prediction, while KNN had the lowest (58.2%) for criminal age range prediction. Further improvements are needed for a reliable system. |
| WAJIHA SAFAT, SOHAIL ASGHAR, (Member, IEEE), AND SAIRA ANDLEEB GILLANI (2021)[10] | The study applied various machine learning algorithms (e.g., logistic regression, SVM, Naïve Bayes, KNN, decision tree, MLP, random forest, XGBoost) and deep learning techniques (e.g., LSTM, ARIMA) for crime data analysis and forecasting. | Utilized crime datasets from Chicago and Los Angeles, focusing on 35 different crime types and various attributes (e.g., date, location, type of crime). Data preprocessing handled cleaning and transformation. | Algorithms were evaluated using metrics like accuracy, precision, recall, and F1-score. XGBoost performed best for Chicago, while KNN excelled for Los Angeles. LSTM provided |

## 2.3 Summary

Crime analysis and prediction using regression models involve understanding and forecasting crime patterns by analyzing the influence of various factors on criminal activities. Basically Regression analysis helps explain and predict outcomes by understanding how independent variables influence a dependent variable. A confusion matrix evaluates classification algorithms by comparing actual and predicted labels, providing detailed model performance and guiding improvements using metrics like R-Square, MSE, and MAE. Population refers to the comprehensive group of interest in a study, while GDP measures a country's economic activity and health. we have studied 15 papers and we can get see that Several studies have used machine learning to predict crime patterns. Obagbuwa and Abidoye (2021) achieved an R-squared of 0.847 using linear regression on South African data. Alves et al. (2018) used random forest on Brazilian data, reaching 97% accuracy in homicide predictions, with unemployment and illiteracy as key factors. Sridharan et al. (2024) applied K-NN, Naive Bayes, and regression to Indian data, effectively predicting crime ratios. Aziz et al. (2022) found Random Forest best for Indian Penal Code crime predictions with an adjusted R-squared of 0.966. Aggarwal et al. (2023) identified high-crime states in India. Rahman et al. (2022) used various models on Bangladeshi data, with Random Forest excelling. Awal et al. (2016) forecasted Bangladeshi crime rates using linear regression. Abdullah et al. (2020) achieved high accuracy with RFC on Bangladeshi data. Safat et al. (2021) used diverse techniques on US data, with XGBoost and KNN excelling for Chicago and Los Angeles, and LSTM and ARIMA providing trend insights.

# Chapter 3

# Project Design

The project's design is crucial since it establishes the methodology and specifies how the research will be carried out (systematically). A well-designed project will have well-defined objectives, a defined scope, appropriate data sources identified, and chosen analytical techniques. As a result, by carefully planning stronger controls, dependable and accurate data can be ensured, allowing for a more thorough analysis of relationships involving socioeconomic position and demographics, even though violent crimes are still classified as such.

## 3.1 Detailed Methodology and Design

**Dataset:** Machine learning depends on historical data for understanding past trends, patterns and relationships. In the case of crime analysis, predictions generated by different machine learning models are primarily based on past data related to crimes committed among other factors such as population demographics as well as economic factors. Past data is used in developing predictive models which are more precise compared to others when it comes to machine learning technologies. The utilization of historical data has enabled machine learning algorithms to unveil complex relationships and patterns within data. Through the examination of previous crime trends alongside socio-economic factors and population changes, patterns that have occurred before with potential impacts on future occurrences are identified using algorithms. We gathered historical crime data from the official website of the Bangladesh Police from 2010-2019. Demographics on Bangladesh came from "Worldometer" which also estimated population figures for the same years as mentioned earlier on demographic information above. World Bank database provided us with economic details such as GDP estimates. We also obtained information from the ACLED website. The dataset known as the Armed Conflict Location & Event Data Project(ACLED) is an itemized and regularly-employed one to the world for political violence and protest events. Bangladesh's homicide trends from 2001 up to 2020 have been scrutinized through using ACLED for this research. It has substantial coverage in terms of geographical area and specific types of incidents that occur, hence making it important in analyzing of crime data. The ACLED dataset is obtained from several credible sources, which include news articles, NGO reports and government security briefs. The subset that was used for the research is only about one country Bangladesh, with a total registered count of 30,645 cases. Hence there is enough data for study purposes. Once we have put together the datasets, we are given a total of fifteen separate crime datasets which run from 2010 through 2019 where the dataset features are years, population size and GDP, accompanied by the target variable being crime. Machine learning depends on historical data for understanding past trends, patterns and relationships. In the case of crime analysis, predictions generated by different machine learning models are primarily based on past data related to crimes committed among other factors such as population demographics as well as economic factors. Past data is used in developing predictive models which are more precise compared to others when it comes to machine learning technologies. The utilization of historical data has enabled machine learning algorithms to unveil complex relationships and patterns within data. Through the examination of previous crime trends alongside socio-economic factors and population changes, patterns that have occurred before with potential impacts on future occurrences are identified using algorithms. We gathered historical crime data from the official website of the Bangladesh Police from 2010-2019. Demographics on Bangladesh came from "Worldometer" which also estimated population figures for the same years as mentioned earlier on demographic information above. World Bank database provided us with economic details such as GDP estimates. We also obtained information from the ACLED website. The dataset known as the Armed Conflict Location & Event Data Project(ACLED) is an itemized and regularly-employed one to the world for political violence and protest events. Bangladesh's homicide trends from 2001 up to 2020 have been scrutinized through using ACLED for this research. It has substantial coverage in terms of geographical area and specific types of incidents that occur, hence making it important in analyzing of crime data. The ACLED dataset is obtained from several credible sources, which include news articles, NGO reports and government security briefs. The subset that was used for the research is only about one country Bangladesh, with a total registered count of 30,645 cases. Hence there is enough data for study purposes. Once we have put together the datasets, we are given a total of fifteen separate crime datasets which run from 2010 through 2019 where the dataset features are years, population size and GDP, accompanied by the target variable being crime.

**Dataset-1[Bangladesh Police Website]:**

| Variable Name | Description |
|---|---|
| Year | Time interval yearly (int) |
| Population | Total population of a country (int) |
| GDP | GDP per head (int) |
| Crime | Crime Crime rate of a specific crime(int) |

**Dataset-2[the Armed Conflict Location & Event Data Project (ACLED)]**

| Variable Name | Description |
|---|---|
| Year | The year in which the event occurred, ranging from 2001 to 2020. |
| BD_GDP | Bangladesh's Gross Domestic Product (GDP) for the corresponding year, providing an economic context for the analysis. |
| BD_GDP_GR | The growth rate of Bangladesh's GDP, which can indicate economic fluctuations and their potential impact on crime rates. |
| AvgPopulationdensity | The average population density, offering insights into the relationship between population concentration and crime incidents. |
| EVENT_TYPE | Event type, which categorizes the nature of each event. |

The section serves as a foundation for precise assessment and construction of prediction models in criminology through the emphasis of historical data in machine learning and detailed overview of the corpus assembled. After creating and setting up the data we split it into two parts so as to gauge how well your models function when analyzing different datasets this way there is some part of information that could still remain fresh. In our research we have followed the standard division for any machine learning dataset where 80% is allocated as training dataset while 20% is for testing dataset

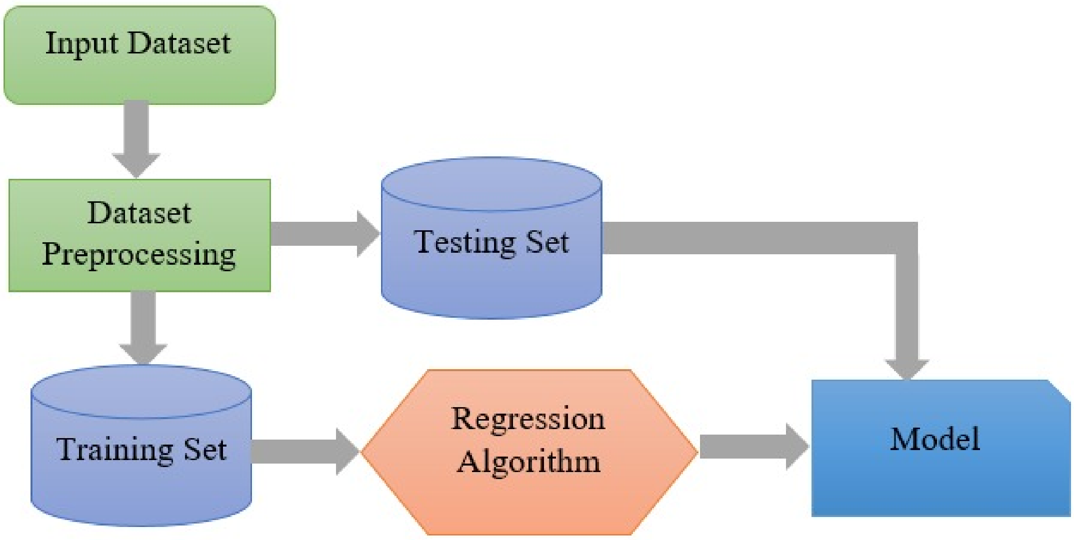


Figure 3.1: Crime Analysis Overview.

Machine learning models are trained using historical data which is the training set enabling them understand patterns and relationships between input features (for example crime events, population demographics, GDP) and the target variable such as crime rates. The test set is used to evaluate the generalization performance of trained models in isolation. Independent researchers can evaluate whether their models have the potential to make correct predictions from previously unknown data.

**Proposed Model (Regression):** In machine learning, multiple regression is a statistical technique that is used to establish links between more than one predictor factor and one target variable. This paper on crime analysis in Bangladesh has employed multiple regression in modeling relationships between crime rates and explanatory variables including population, GDP, and year. We selected Multiple regression for several reasons:

**Numeric Data:** The dataset consists of numeric variables (population, GDP, year) and aims to predict a numeric target (crime rate), making multiple regression an appropriate choice.

**Establishing Relationships:** Multiple regression allows for the examination of how changes in one or more predictor variables (e.g., population, GDP) are associated with changes in the target variable (crime rate). This helps in understanding the influence of socio-economic factors on crime dynamics.

**Interpretability:** The coefficients of multiple regression provide insights into the strength and direction of the relationships between predictor variables and the target variable, enabling interpretability of the model.

In multiple regression, the relationship between the predictor variables (X) and the target variable (Y) is represented by the following hyperplane equation:

$$\begin{aligned} Y &= \omega_0 + \omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n \\ &= \omega_0 + \sum_{i=1}^{n} \omega_i \cdot X_i \end{aligned} \tag{3.1}$$

**Where:**

- $Y$ is the target variable (crime rate).
- $X_1, X_2, \ldots, X_n$ are the predictor variables (year, population, GDP).
- $\omega_0$ is the intercept term (constant).
- $\omega_1, \omega_2, \ldots, \omega_n$ are the coefficients associated with each predictor variable, representing the change in the target variable for a one-unit change in the predictor variable, holding other variables constant.

## 3.2 Project Plan

The duration of Fydp II is 17 weeks in total. Therefore, over these 17 weeks, we will primarily be working on six tasks: reviewing the literature; gathering relevant data sets for our model; pre-processing the data sets in order to train the model; developing the model; assessing the obtained results; and, finally, preparing the report. Therefore, in order to properly understand what has already been done in this sector, we must first read a large number of related articles that are similar to our own. Here, we spend four weeks, and everyone in our group contributes. We must gather our data for the dataset in order to proceed to the second phase. Meanwhile, to obtain further ideas, a few members of our team continue to analyze relevant works in this sector. Collecting data from different sources takes about 3 weeks. We then have to process the datasets we have collected. It requires around two weeks. Between weeks seven and eight, we pre-processed our data set. Next, we concentrated on building a model. One of the primary roles is this. It takes around five weeks to create the models for our data sets. Most of the time is spent on this step. Next, we put our model-derived results to the test. Between weeks 14 and 16, test results are completed in three weeks. Finally, after testing every outcome, we began writing the research report. Every member of our team participates in report authoring. We schedule many meetings with our helpful mentor along our whole journey to address any questions or concerns.

## 3.3 Task Allocation

We follow this gantt chart to complete our research properly.

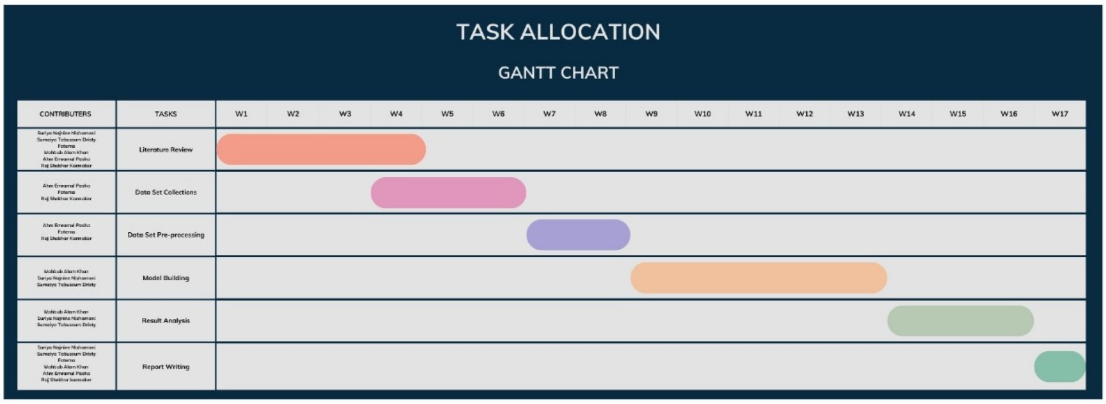


Figure 3.2: Task Allocation.

## 3.4 Summary

The duration of Fydp II is 17 weeks in total. Therefore, over these 17 weeks, we will primarily be working on six tasks: reviewing the literature; gathering relevant data sets for our model; pre-processing the data sets in order to train the model; developing the model; assessing the obtained results; and, finally, preparing the report. Therefore, in order to properly understand what has already been done in this sector, we must first read a large number of related articles that are similar to our own. Here, we spend four weeks, and everyone in our group contributes. We must gather our data for the dataset in order to proceed to the second phase. Meanwhile, to obtain further ideas, a few members of our team continue to analyze relevant works in this sector. Collecting data from different sources takes about 3 weeks. We then have to process the datasets we have collected. It requires around two weeks. Between weeks seven and eight, we pre-processed our data set. Next, we concentrated on building a model. One of the primary roles is this. It takes around five weeks to create the models for our data sets. Most of the time is spent on this step. Next, we put our model-derived results to the test. Between weeks 14 and 16, test results are completed in three weeks. Finally, after testing every outcome, we began writing the research report. Every member of our team participates in report authoring. We schedule many meetings with our helpful mentor along our whole journey to address any questions or concerns.

# Chapter 4

# Data Set

## 4.1 Dataset Analysis.

Our dataset was compiled from the websites of the Bangladesh Police and ACLED. The data for the Bangladesh Police website spans the years 2010 to 2019 and includes 15 attributes. For ACLED, we obtained 30,000 samples.

### 4.1.1 Year vs Population

| Year | Total Population |
|---|---|
| 2000 | 129193327 |
| 2001 | 131670484 |
| 2002 | 134139826 |
| 2003 | 136503206 |
| 2004 | 138789725 |
| 2005 | 140912590 |
| 2006 | 142628831 |
| 2007 | 144135934 |
| 2008 | 145421318 |
| 2009 | 146706810 |
| 2010 | 148391139 |
| 2011 | 150211005 |
| 2012 | 152090649 |
| 2013 | 154030139 |
| 2014 | 155961299 |
| 2015 | 157830000 |
| 2016 | 159784568 |
| 2017 | 161793964 |
| 2018 | 163683958 |
| 2019 | 165516222 |
| 2020 | 167420951 |
| 2021 | 169356251 |
| 2022 | 171186372 |

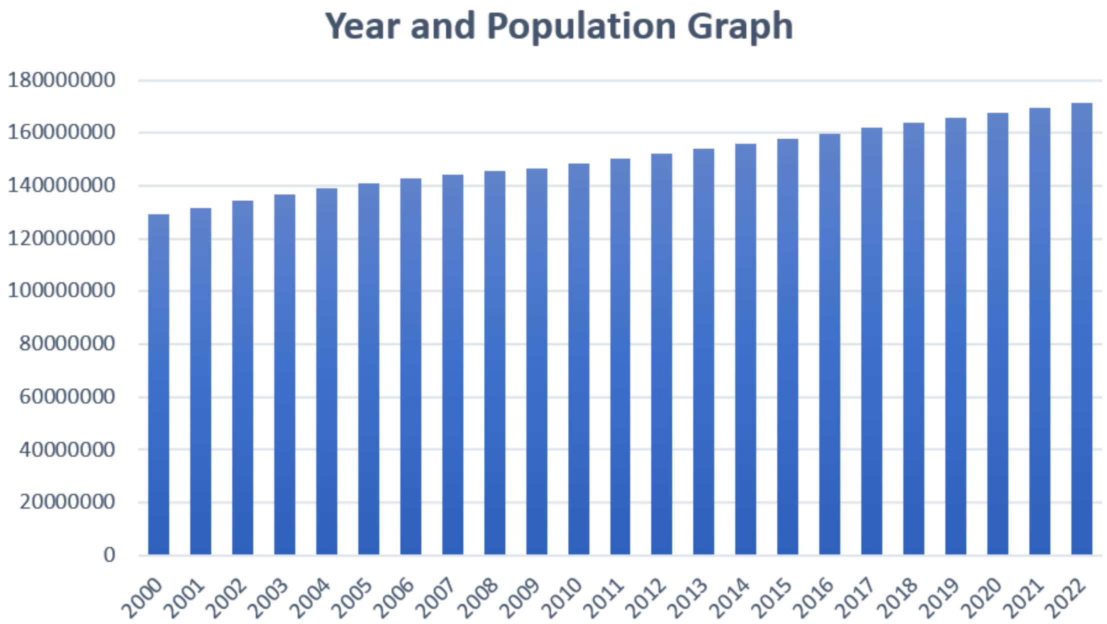


Figure 4.1: Year vs Population.

### 4.1.2 Year vs GDP Per Capital

| Year | GDP Per Capital |
|---|---|
| 2000 | 330 |
| 2001 | 310 |
| 2002 | 190 |
| 2003 | 290 |
| 2004 | 350 |
| 2005 | 490 |
| 2006 | 540 |
| 2007 | 590 |
| 2008 | 510 |
| 2009 | 410 |
| 2010 | 440 |
| 2011 | 520 |
| 2012 | 520 |
| 2013 | 470 |
| 2014 | 470 |
| 2015 | 530 |
| 2016 | 580 |
| 2017 | 530 |
| 2018 | 610 |
| 2019 | 670 |
| 2020 | 230 |
| 2021 | 570 |
| 2022 | 600 |

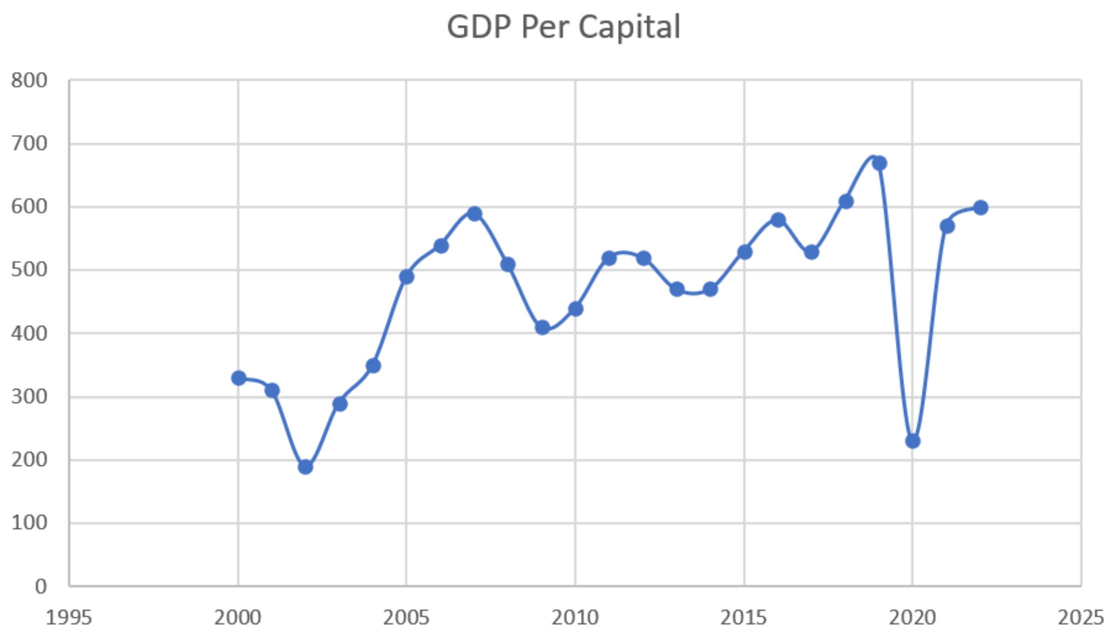


Figure 4.2: Year vs GDP Per Capital.

### 4.1.3 Year vs Dacoity

| Year | Dacoity |
|---|---|
| 2010 | 656 |
| 2011 | 650 |
| 2012 | 593 |
| 2013 | 613 |
| 2014 | 651 |
| 2015 | 492 |
| 2016 | 408 |
| 2017 | 336 |
| 2018 | 262 |
| 2019 | 32 |

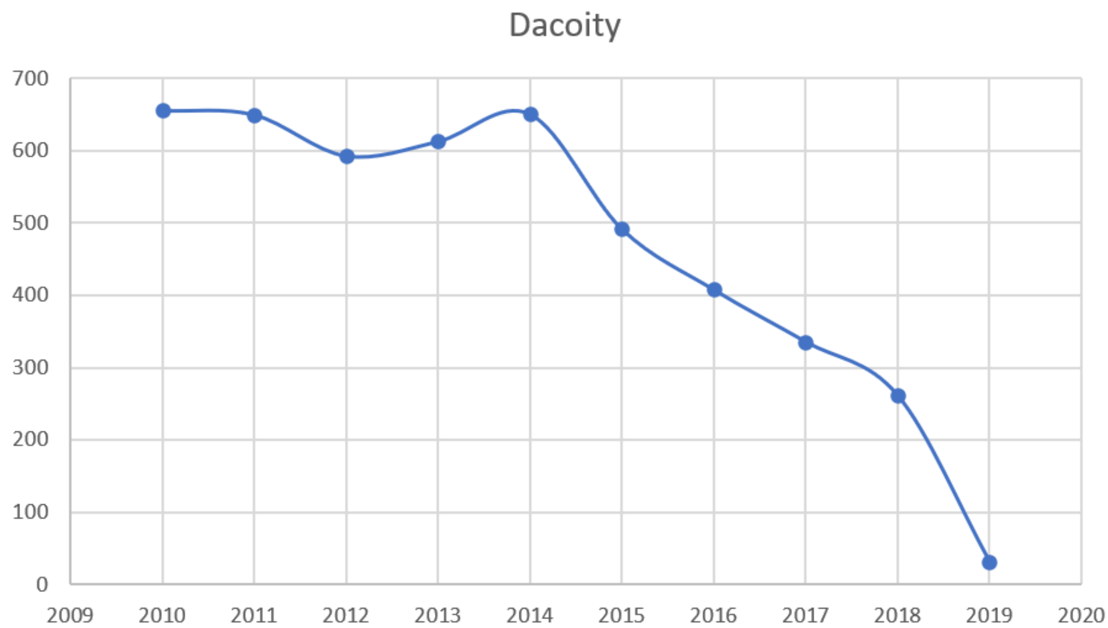


Figure 4.3: Year vs Dacoity.

### 4.1.4 Year vs Robbery

| Year | Robbery |
|---|---|
| 2010 | 1059 |
| 2011 | 1069 |
| 2012 | 964 |
| 2013 | 1021 |
| 2014 | 1155 |
| 2015 | 933 |
| 2016 | 722 |
| 2017 | 657 |
| 2018 | 562 |
| 2019 | 68 |

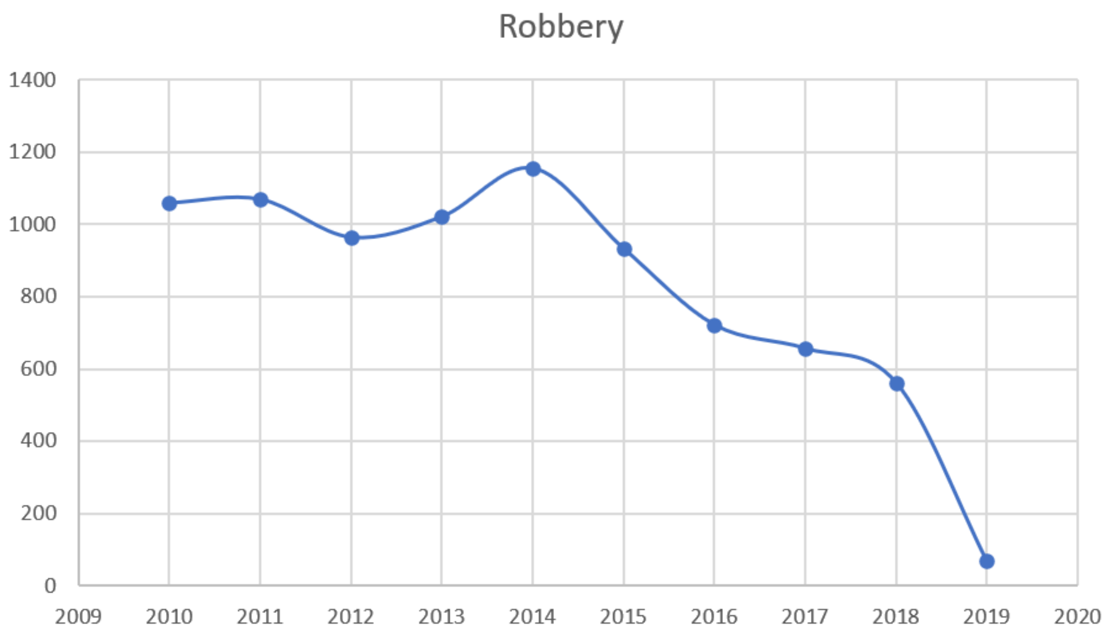


Figure 4.4: Year vs Robbery.

### 4.1.5 Year vs Speed Trial

| Year | Speed Trial |
|---|---|
| 2010 | 1666 |
| 2011 | 1863 |
| 2012 | 1907 |
| 2013 | 1896 |
| 2014 | 1716 |
| 2015 | 1549 |
| 2016 | 1052 |
| 2017 | 1045 |
| 2018 | 922 |
| 2019 | 48 |

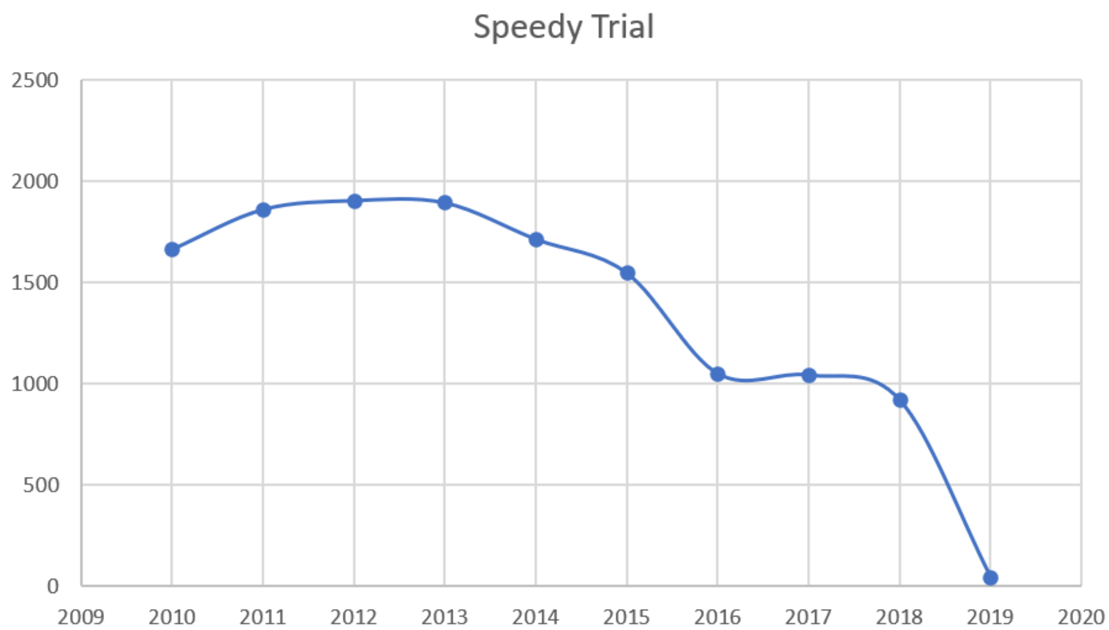


Figure 4.5: Year vs Speed Trial.

### 4.1.6 Year vs Riot

| Year | Riot |
|---|---|
| 2010 | 130 |
| 2011 | 109 |
| 2012 | 94 |
| 2013 | 172 |
| 2014 | 79 |
| 2015 | 93 |
| 2016 | 53 |
| 2017 | 23 |
| 2018 | 26 |
| 2019 | 1 |

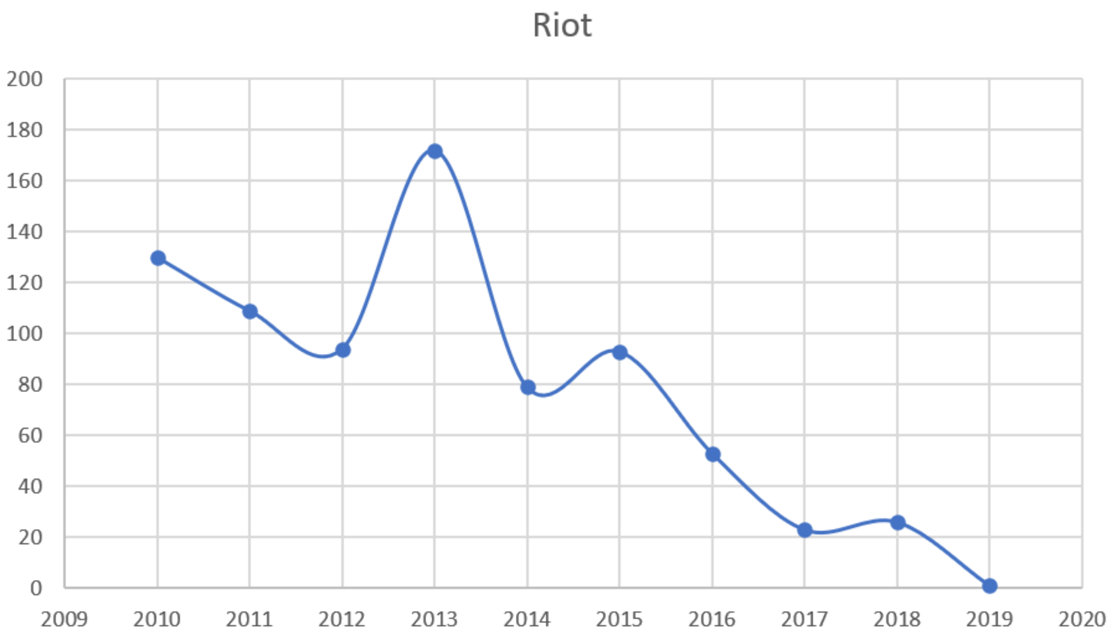


Figure 4.6: Year vs Riot.

### 4.1.7 Year vs Women Abuse

| Year | Women Abuse |
|---|---|
| 2010 | 17752 |
| 2011 | 21389 |
| 2012 | 20947 |
| 2013 | 19601 |
| 2014 | 21291 |
| 2015 | 21210 |
| 2016 | 18446 |
| 2017 | 17073 |
| 2018 | 16253 |
| 2019 | 1139 |

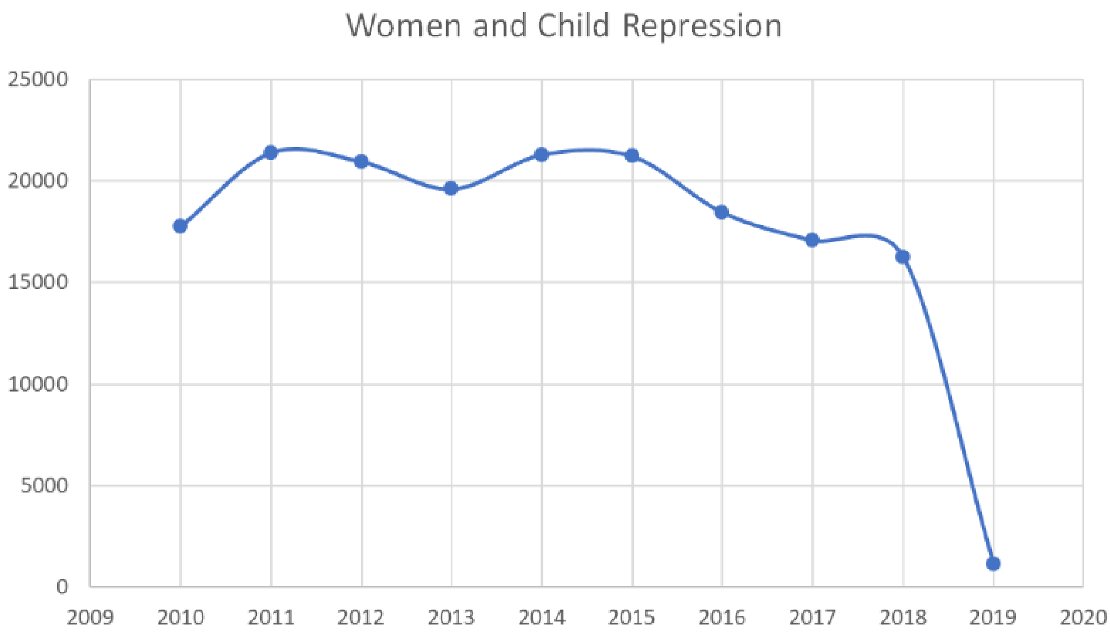


Figure 4.7: Year vs Women Abuse.

### 4.1.8 Year vs Kidnapping

| Year | Kidnapping |
|---|---|
| 2010 | 870 |
| 2011 | 792 |
| 2012 | 850 |
| 2013 | 879 |
| 2014 | 920 |
| 2015 | 805 |
| 2016 | 639 |
| 2017 | 509 |
| 2018 | 444 |
| 2019 | 46 |

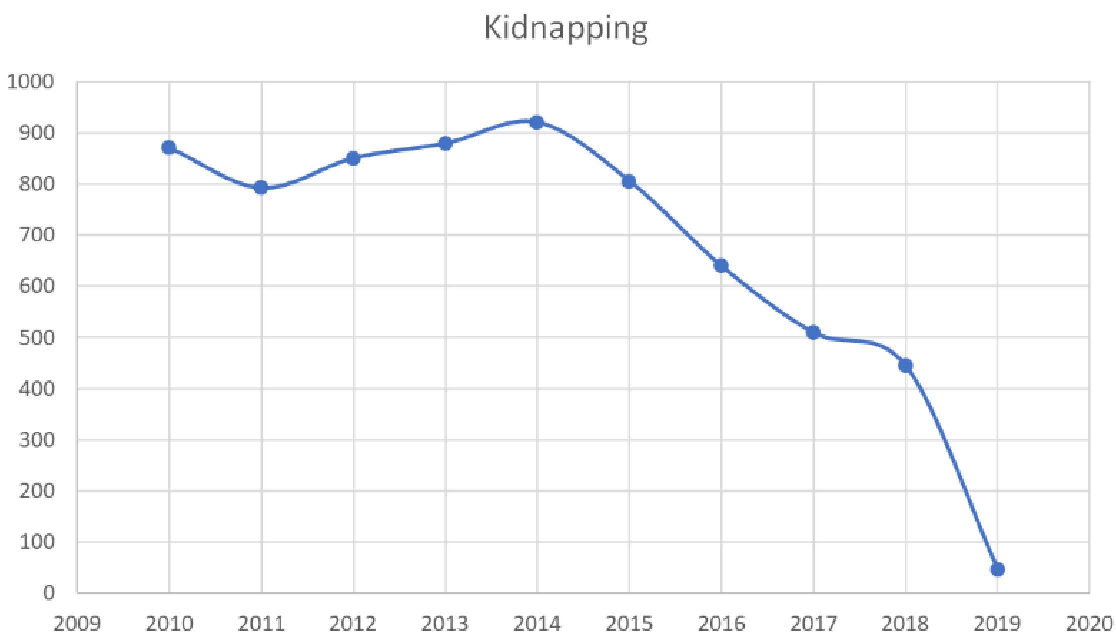


Figure 4.8: Year vs Kidnapping.

### 4.1.9 Year vs Police Assault

| Year | Police Assault |
|---|---|
| 2010 | 473 |
| 2011 | 581 |
| 2012 | 659 |
| 2013 | 1257 |
| 2014 | 702 |
| 2015 | 634 |
| 2016 | 521 |
| 2017 | 543 |
| 2018 | 811 |
| 2019 | 69 |

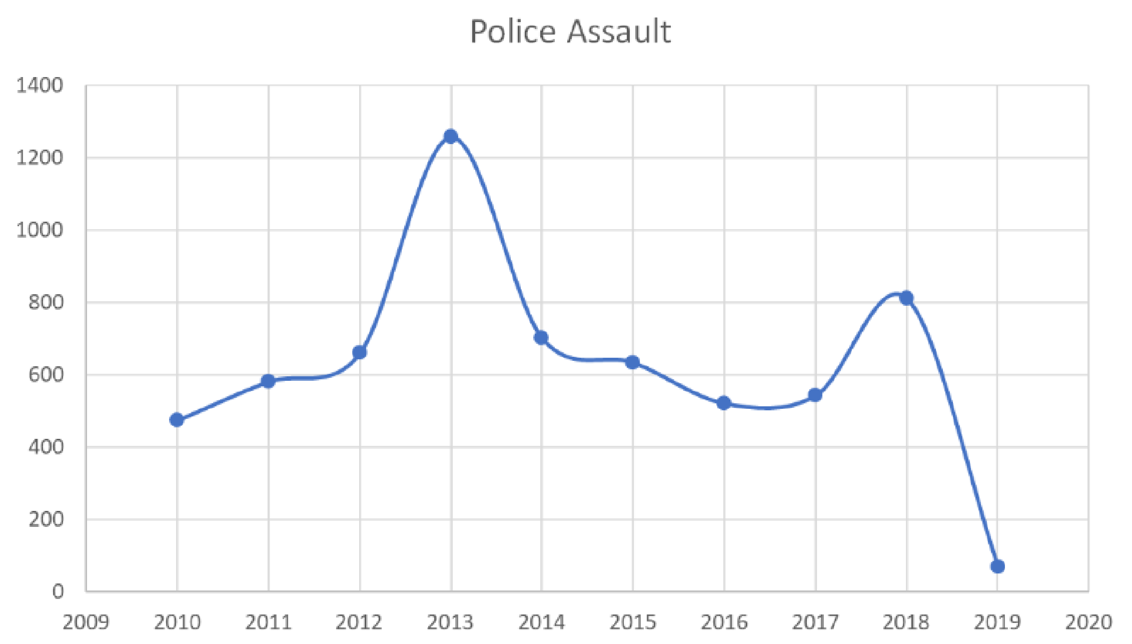


Figure 4.9: Year vs Police Assault

### 4.1.10 Year vs Burglary

| Year | Robbery |
|---|---|
| 2010 | 3101 |
| 2011 | 3134 |
| 2012 | 2927 |
| 2013 | 2762 |
| 2014 | 2809 |
| 2015 | 2495 |
| 2016 | 2213 |
| 2017 | 2163 |
| 2018 | 2137 |
| 2019 | 174 |

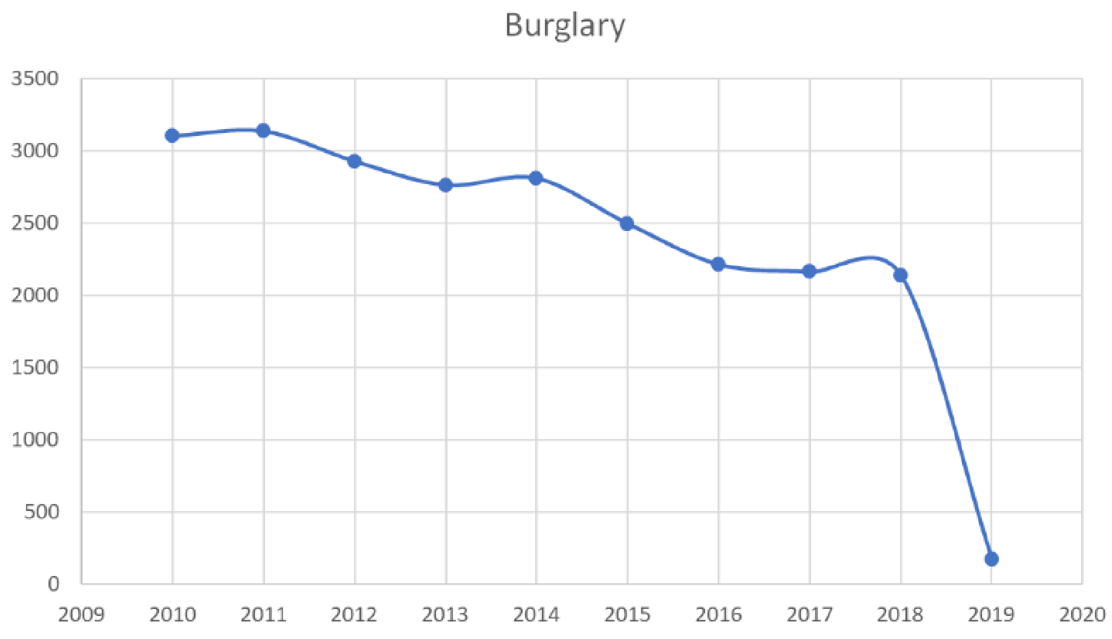


Figure 4.10: Year vs Burglary

### 4.1.11 Year vs Theft

| Year | Theft |
|---|---|
| 2010 | 8529 |
| 2011 | 8873 |
| 2012 | 8598 |
| 2013 | 7882 |
| 2014 | 7660 |
| 2015 | 6821 |
| 2016 | 6110 |
| 2017 | 5833 |
| 2018 | 5561 |
| 2019 | 494 |

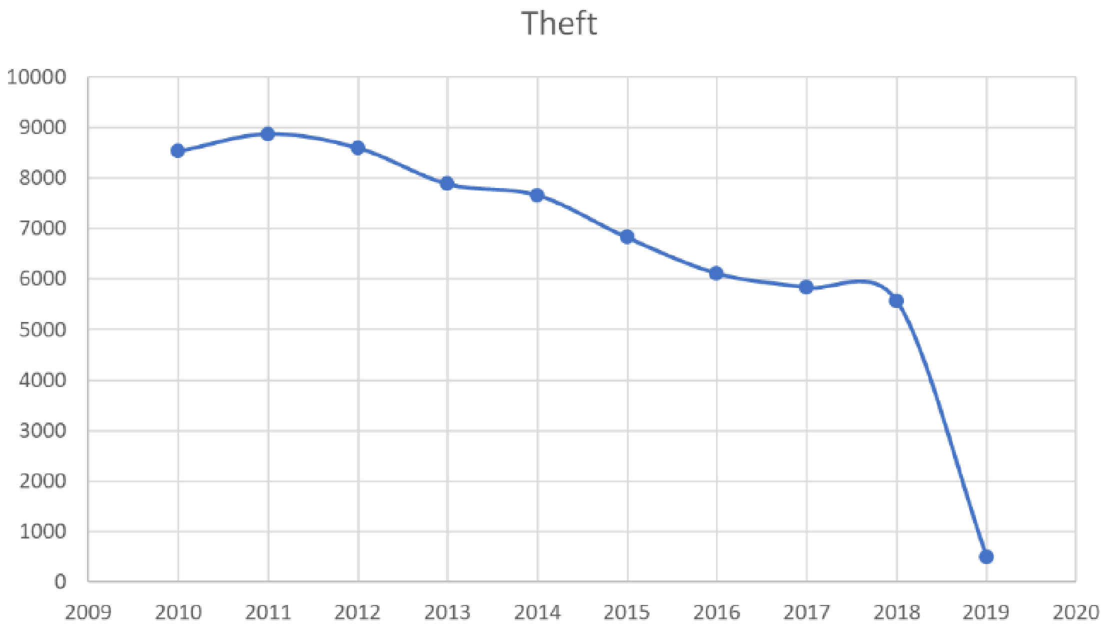


Figure 4.11: Year vs Theft

### 4.1.12 Year vs Other Cases

| Year | Other Cases |
|---|---|
| 2010 | 87139 |
| 2011 | 88355 |
| 2012 | 96112 |
| 2013 | 93930 |
| 2014 | 90400 |
| 2015 | 84117 |
| 2016 | 77747 |
| 2017 | 74645 |
| 2018 | 69736 |
| 2019 | 5428 |

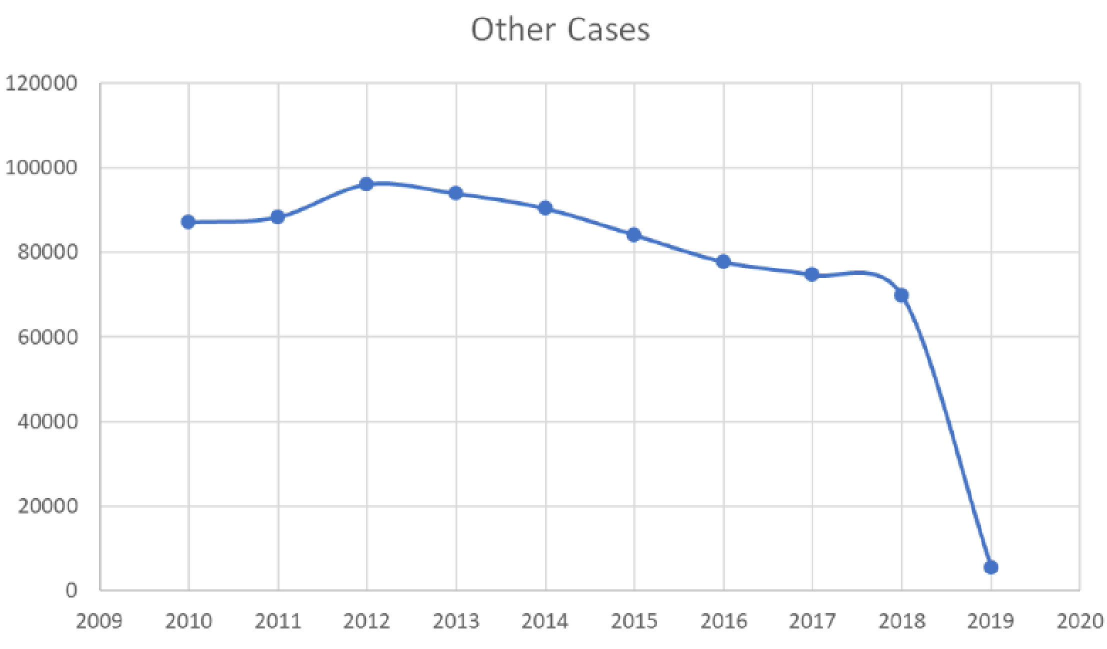


Figure 4.12: Year vs Other Cases

### 4.1.13 Year vs Arms Act.

| Year | Arms Act |
|---|---|
| 2010 | 1575 |
| 2011 | 1269 |
| 2012 | 1511 |
| 2013 | 1517 |
| 2014 | 2023 |
| 2015 | 2079 |
| 2016 | 2291 |
| 2017 | 2208 |
| 2018 | 2515 |
| 2019 | 174 |

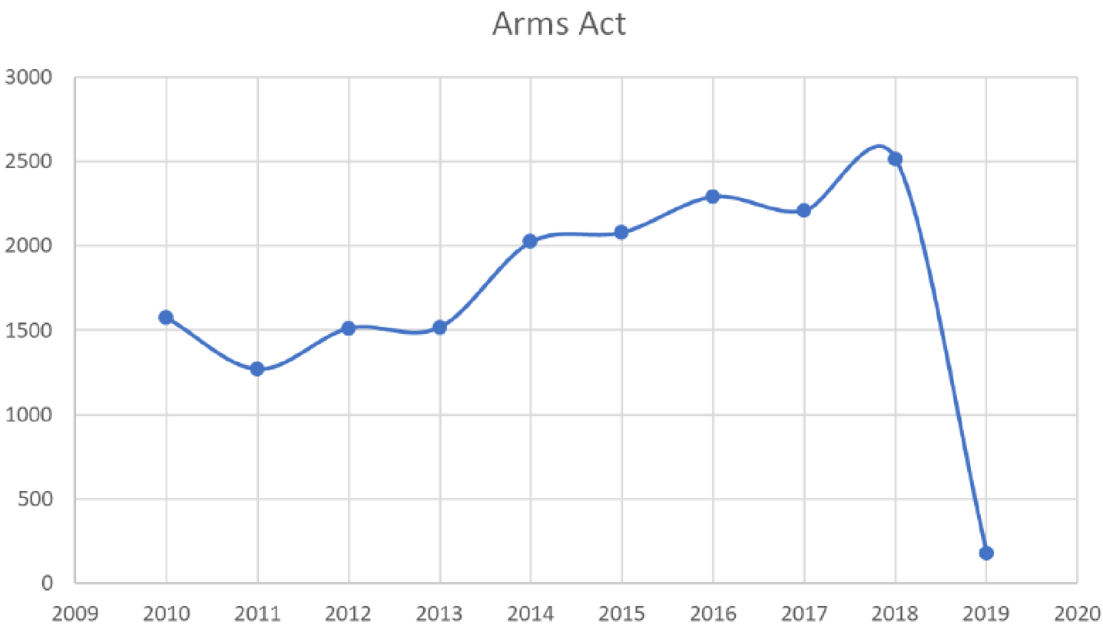


Figure 4.13: Year vs Arms Act

### 4.1.14 Year vs Explosive

| Year | Explosive |
|---|---|
| 2010 | 253 |
| 2011 | 207 |
| 2012 | 289 |
| 2013 | 1007 |
| 2014 | 520 |
| 2015 | 725 |
| 2016 | 487 |
| 2017 | 362 |
| 2018 | 1310 |
| 2019 | 30 |

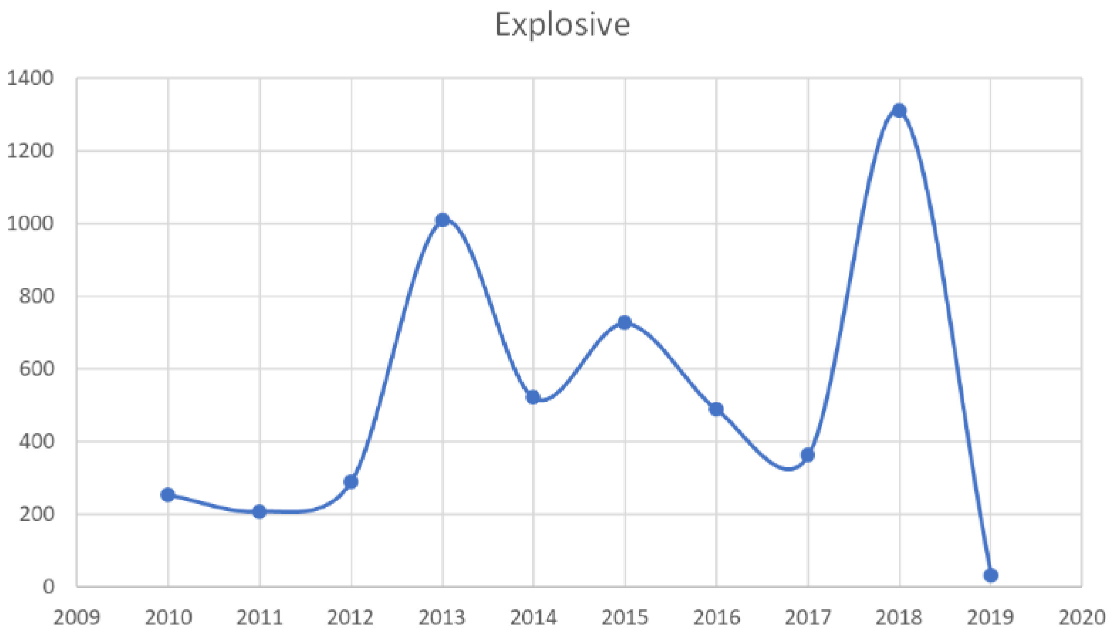


Figure 4.14: Year vs Explosive

### 4.1.15 Year vs Narcotics

| Year | Narcotics |
|---|---|
| 2010 | 29344 |
| 2011 | 31696 |
| 2012 | 37264 |
| 2013 | 35832 |
| 2014 | 42501 |
| 2015 | 47666 |
| 2016 | 62208 |
| 2017 | 98984 |
| 2018 | 112549 |
| 2019 | 9069 |

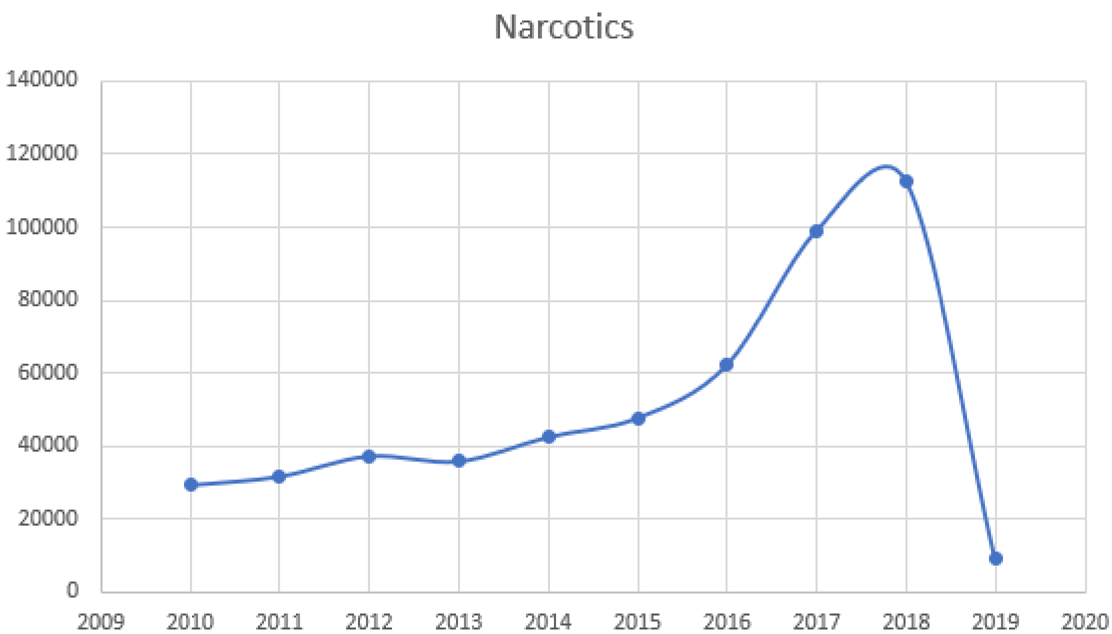


Figure 4.15: Year vs Narcotics

### 4.1.16 Year vs Smuggling

| Year | Smuggling |
|---|---|
| 2010 | 6363 |
| 2011 | 5714 |
| 2012 | 6578 |
| 2013 | 6437 |
| 2014 | 6788 |
| 2015 | 6179 |
| 2016 | 4680 |
| 2017 | 5599 |
| 2018 | 4501 |
| 2019 | 361 |

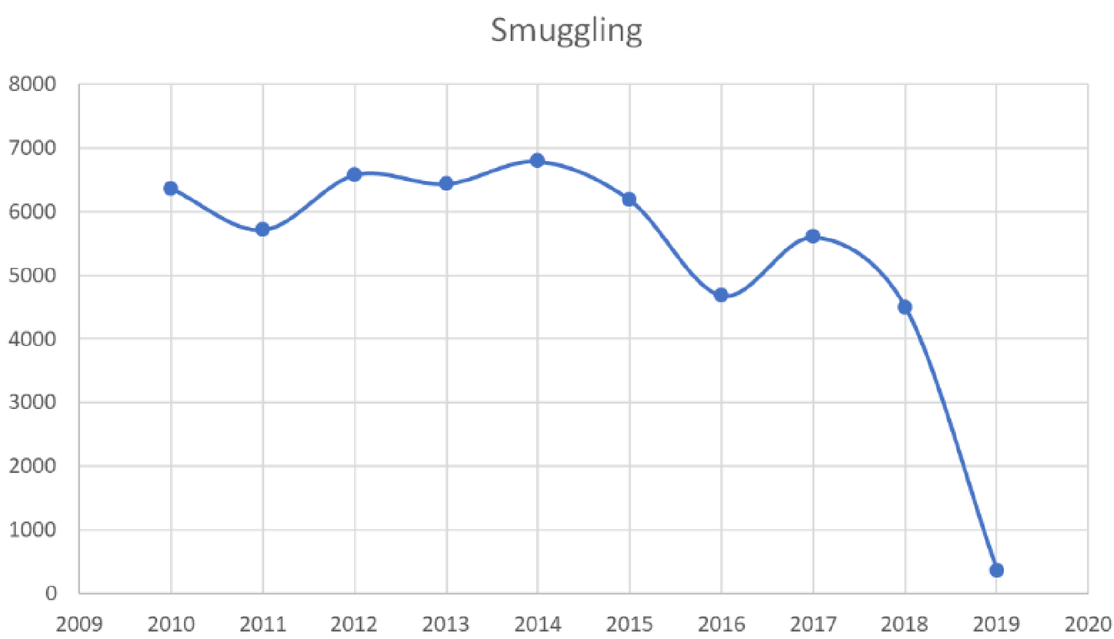


Figure 4.16: Year vs Smuggling

## 4.2 Dataset Preprocessing.

We split up the data and created a subset to determine the relationship between crime and population.

### 4.2.1 Arms Act

| Year | Total Population | GDP Per Capital | Arms Act |
|---|---|---|---|
| 2010 | 148391139 | 440 | 1575 |
| 2011 | 150211005 | 520 | 1269 |
| 2012 | 152090649 | 520 | 1511 |
| 2013 | 154030139 | 470 | 1517 |
| 2014 | 155961299 | 470 | 2023 |
| 2015 | 157830000 | 530 | 2079 |
| 2016 | 159784568 | 580 | 2291 |
| 2017 | 161793964 | 530 | 2208 |
| 2018 | 163683958 | 610 | 2515 |
| 2019 | 165516222 | 670 | 174 |

### 4.2.2 Burglary

| Year | Total Population | GDP Per Capital | Burglary |
|---|---|---|---|
| 2010 | 148391139 | 440 | 3101 |
| 2011 | 150211005 | 520 | 3134 |
| 2012 | 152090649 | 520 | 2927 |
| 2013 | 154030139 | 470 | 2762 |
| 2014 | 155961299 | 470 | 2809 |
| 2015 | 157830000 | 530 | 2495 |
| 2016 | 159784568 | 580 | 2213 |
| 2017 | 161793964 | 530 | 2163 |
| 2018 | 163683958 | 610 | 2137 |
| 2019 | 165516222 | 670 | 174 |

### 4.2.3 Dacoity

| Year | Total Population | GDP Per Capital | Dacoity |
|---|---|---|---|
| 2010 | 148391139 | 440 | 656 |
| 2011 | 150211005 | 520 | 650 |
| 2012 | 152090649 | 520 | 593 |
| 2013 | 154030139 | 470 | 613 |
| 2014 | 155961299 | 470 | 651 |
| 2015 | 157830000 | 530 | 492 |
| 2016 | 159784568 | 580 | 408 |
| 2017 | 161793964 | 530 | 336 |
| 2018 | 163683958 | 610 | 262 |
| 2019 | 165516222 | 670 | 32 |

### 4.2.4 Explosive

| Year | Total Population | GDP Per Capital | Explosive |
|---|---|---|---|
| 2010 | 148391139 | 440 | 253 |
| 2011 | 150211005 | 520 | 207 |
| 2012 | 152090649 | 520 | 289 |
| 2013 | 154030139 | 470 | 1007 |
| 2014 | 155961299 | 470 | 520 |
| 2015 | 157830000 | 530 | 725 |
| 2016 | 159784568 | 580 | 487 |
| 2017 | 161793964 | 530 | 362 |
| 2018 | 163683958 | 610 | 1310 |
| 2019 | 165516222 | 670 | 30 |

### 4.2.5 Kidnapping

| Year | Total Population | GDP Per Capital | Kidnapping |
|---|---|---|---|
| 2010 | 148391139 | 440 | 870 |
| 2011 | 150211005 | 520 | 792 |
| 2012 | 152090649 | 520 | 850 |
| 2013 | 154030139 | 470 | 879 |
| 2014 | 155961299 | 470 | 920 |
| 2015 | 157830000 | 530 | 805 |
| 2016 | 159784568 | 580 | 639 |
| 2017 | 161793964 | 530 | 509 |
| 2018 | 163683958 | 610 | 444 |
| 2019 | 165516222 | 670 | 46 |

### 4.2.6 Murder

| Year | Total Population | GDP Per Capital | Murder |
|---|---|---|---|
| 2010 | 148391139 | 440 | 3988 |
| 2011 | 150211005 | 520 | 3966 |
| 2012 | 152090649 | 520 | 4114 |
| 2013 | 154030139 | 470 | 4393 |
| 2014 | 155961299 | 470 | 4514 |
| 2015 | 157830000 | 530 | 4037 |
| 2016 | 159784568 | 580 | 3591 |
| 2017 | 161793964 | 530 | 3549 |
| 2018 | 163683958 | 610 | 3830 |
| 2019 | 165516222 | 670 | 351 |

### 4.2.7 Narcotics

| Year | Total Population | GDP Per Capital | Narcotics |
|---|---|---|---|
| 2010 | 148391139 | 440 | 29344 |
| 2011 | 150211005 | 520 | 31696 |
| 2012 | 152090649 | 520 | 37264 |
| 2013 | 154030139 | 470 | 35832 |
| 2014 | 155961299 | 470 | 42501 |
| 2015 | 157830000 | 530 | 47666 |
| 2016 | 159784568 | 580 | 62208 |
| 2017 | 161793964 | 530 | 98984 |
| 2018 | 163683958 | 610 | 112549 |
| 2019 | 165516222 | 670 | 9069 |

### 4.2.8 Other Cases

| Year | Total Population | GDP Per Capital | Other Cases |
|---|---|---|---|
| 2010 | 148391139 | 440 | 87139 |
| 2011 | 150211005 | 520 | 88355 |
| 2012 | 152090649 | 520 | 96112 |
| 2013 | 154030139 | 470 | 93930 |
| 2014 | 155961299 | 470 | 90400 |
| 2015 | 157830000 | 530 | 84117 |
| 2016 | 159784568 | 580 | 77747 |
| 2017 | 161793964 | 530 | 74645 |
| 2018 | 163683958 | 610 | 69736 |
| 2019 | 165516222 | 670 | 5428 |

### 4.2.9 Police Assault

| Year | Total Population | GDP Per Capital | Police Assault |
|---|---|---|---|
| 2010 | 148391139 | 440 | 473 |
| 2011 | 150211005 | 520 | 581 |
| 2012 | 152090649 | 520 | 659 |
| 2013 | 154030139 | 470 | 1257 |
| 2014 | 155961299 | 470 | 702 |
| 2015 | 157830000 | 530 | 634 |
| 2016 | 159784568 | 580 | 521 |
| 2017 | 161793964 | 530 | 543 |
| 2018 | 163683958 | 610 | 811 |
| 2019 | 165516222 | 670 | 69 |

### 4.2.10 Riot

| Year | Total Population | GDP Per Capital | Riot |
|---|---|---|---|
| 2010 | 148391139 | 440 | 130 |
| 2011 | 150211005 | 520 | 109 |
| 2012 | 152090649 | 520 | 94 |
| 2013 | 154030139 | 470 | 172 |
| 2014 | 155961299 | 470 | 79 |
| 2015 | 157830000 | 530 | 93 |
| 2016 | 159784568 | 580 | 53 |
| 2017 | 161793964 | 530 | 23 |
| 2018 | 163683958 | 610 | 26 |
| 2019 | 165516222 | 670 | 1 |

### 4.2.11 Robbery

| Year | Total Population | GDP Per Capital | Robbery |
|---|---|---|---|
| 2010 | 148391139 | 440 | 1059 |
| 2011 | 150211005 | 520 | 1069 |
| 2012 | 152090649 | 520 | 964 |
| 2013 | 154030139 | 470 | 1021 |
| 2014 | 155961299 | 470 | 1155 |
| 2015 | 157830000 | 530 | 1155 |
| 2016 | 159784568 | 580 | 722 |
| 2017 | 161793964 | 530 | 657 |
| 2018 | 163683958 | 610 | 562 |
| 2019 | 165516222 | 670 | 68 |

### 4.2.12 Smuggling

| Year | Total Population | GDP Per Capital | Smuggling |
|---|---|---|---|
| 2010 | 148391139 | 440 | 6363 |
| 2011 | 150211005 | 520 | 5714 |
| 2012 | 152090649 | 520 | 6578 |
| 2013 | 154030139 | 470 | 6437 |
| 2014 | 155961299 | 470 | 6788 |
| 2015 | 157830000 | 530 | 6179 |
| 2016 | 159784568 | 580 | 4680 |
| 2017 | 161793964 | 530 | 5599 |
| 2018 | 163683958 | 610 | 4501 |
| 2019 | 165516222 | 670 | 361 |

### 4.2.13 Speed Trial

| Year | Total Population | GDP Per Capital | Speed Trial |
|---|---|---|---|
| 2010 | 148391139 | 440 | 1666 |
| 2011 | 150211005 | 520 | 1863 |
| 2012 | 152090649 | 520 | 1907 |
| 2013 | 154030139 | 470 | 1896 |
| 2014 | 155961299 | 470 | 1716 |
| 2015 | 157830000 | 530 | 1549 |
| 2016 | 159784568 | 580 | 1052 |
| 2017 | 161793964 | 530 | 1045 |
| 2018 | 163683958 | 610 | 922 |
| 2019 | 165516222 | 670 | 48 |

### 4.2.14 Theft

| Year | Total Population | GDP Per Capital | Arms Act |
|---|---|---|---|
| 2010 | 148391139 | 440 | 8529 |
| 2011 | 150211005 | 520 | 8873 |
| 2012 | 152090649 | 520 | 8598 |
| 2013 | 154030139 | 470 | 7882 |
| 2014 | 155961299 | 470 | 7660 |
| 2015 | 157830000 | 530 | 6821 |
| 2016 | 159784568 | 580 | 6110 |
| 2017 | 161793964 | 530 | 5833 |
| 2018 | 163683958 | 610 | 5561 |
| 2019 | 165516222 | 670 | 494 |

### 4.2.15 Women and child Repression

| Year | Total Population | GDP Per Capital | Women and child Repression |
|---|---|---|---|
| 2010 | 148391139 | 440 | 17752 |
| 2011 | 150211005 | 520 | 21389 |
| 2012 | 152090649 | 520 | 20947 |
| 2013 | 154030139 | 470 | 19601 |
| 2014 | 155961299 | 470 | 21291 |
| 2015 | 157830000 | 530 | 21210 |
| 2016 | 159784568 | 580 | 18446 |
| 2017 | 161793964 | 530 | 17073 |
| 2018 | 163683958 | 610 | 16253 |
| 2019 | 165516222 | 670 | 1139 |

# Chapter 5

# Implementation and Results

The implementation phase is so important because it includes the use of statistical procedures and machine learning systems to examine the datasets for which they were intended to convert theories into intelligence that is tangible and can be acted upon. All of our work has been uploaded in here: `https://github.com/KhanStack/Final-Year-Design-Project` By doing regression analysis we are able to determine how much casual relationships there may be between different factors affecting crimes like GDP (Gross Domestic Product) rates or even population changes in certain areas.

## 5.1 Environment Setup

Using Google Colab, a cloud platform, this Crime analysis machine learning project was developed and implemented, being a cloud-based that provides the best and most flexible environment for doing data analysis and machine learning tasks. These are the steps for setting up and configuring the system:

**Google Colab:** Google Colab (Collaboratory) is a free, online-based Jupyter Notebook platform that permits its users to compose as well as to run Python programs within the confines of a web browser. It stands out in support of jobs on machine learning as a result of its compatibility with the major libraries, the capability of utilizing graphic processing units and a direct link to Google Drive so that users can store their data on the same platform.

**Dataset Upload and Access:** After we accomplish this step, our dataset will be safely stored in a special folder on Google Drive for future reference. Subsequent mounting of the Google Drive to the Google Colab environment was done as follows:

from google. colab import drive
drive.mount('/content/drive')

When Google Drive has been successfully entered into, read all the files from the dataset into Colab. We are going to load the dataset.

data_path1='/content/gdrive/MyDrive/FYDP2_DATACode/Separate_Data/Burglary.xlsx'
df = pd.read_excel(data_path1, header=None)

**Libraries and Dependencies:** Data analysis, preprocessing, and machine learning require many libraries are used. At the beginning of the notebook, all these were imported and installed. The key libraries were as follows:
**Pandas:** For data manipulation and analysis.
**NumPy:** For numerical computations.
**Matplotlib and Seaborn:** For data visualization.
**Scikit-learn:** For machine learning algorithms and evaluation metrics.

## 5.2 Testing and Evaluation

**Evaluation Matrix for Regression Model:**
The necessary metrics and measures have been calculated. These figures include:

**R-square:**
R-squared is a way to measure statistics, which tells us how well dependent variables are predicted using independent variables. Specifically, R-squared ranges from zero to one, zero implying that target variables cannot be explained by any model, while one suggests total explanation for any given model.

$$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{total sum of squares}} = 1 - \frac{\sum_i (y_i - \bar{y})^2}{\sum_i e_i^2} \tag{5.1}$$

**Mean Squared Error (MSE):**
The mean squared error (MSE) is the average of the squared differences between the actual values of the target variable and the projected values of the regression model. It provided quantifiable assessment of the model's accuracy with smaller numbers indicating more accurate predictions. This is derived as the average of squared residuals (difference between the actual and predicted values) in the dataset.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{5.2}$$

**Mean Absolute Error (MAE):**
Mean absolute error (MAE) is a close measure to MSE as it finds the average absolute gap between actual value and its predicted number, rather than squaring the difference. This measure will give one estimation average size of errors one can expect in his predictions irrespective of their accuracy. Lesser values mean better predictions in both cases with MAE having an edge over MSE due to its lower sensitivity on extreme cases.

$$\begin{aligned} \text{MAE} &= \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \\ &= \frac{\sum_{i=1}^{n} |e_i|}{n} \end{aligned} \tag{5.3}$$

After performing the regression we obtain with these values for confusion matrix. These following values are only for Bangladesh Police website dataset.

| Crime | R-Square | MSE | MAE |
|---|---|---|---|
| Dacoity | 0.76 | 5957.65 | 77.04 |
| Robbery | 0.34 | 28111.87 | 167.24 |
| Murder | -3.28 | 186105.60 | 388.20 |
| Speedy Trial | 0.94 | 9208.67 | 70.07 |
| Riot | 0.62 | 707.71 | 21.30 |
| Woman and Child Repression | 0.54 | 2131812.04 | 1146.26 |
| Kidnapping | 0.74 | 5260.92 | 56.75 |
| Police Assault | -135.37 | 49227.99 | 200.78 |
| Burglary | 0.85 | 34438.09 | 182.11 |
| Theft | 0.97 | 63116.31 | 197.15 |
| Other Cases | 0.79 | 9705051.01 | 3099.81 |
| Arms Act | 0.69 | 67980.86 | 256.30 |
| Explosive | -62.35 | 380488.02 | 478.94 |
| Narcotics | 0.97 | 39488639.12 | 5028.52 |
| Smuggling | -53.26 | 179395.33 | 418.80 |

After performing the regression we obtain with these values for confusion matrix. These following values are only for ACLED dataset.

## 5.3 Results and Discussion

We applied two datasets in our study, one from the web portal of the Bangladesh Police and another one from the Armed Conflict Location & Event Data Project (ACLED). Regression analysis was carried out on both data to find out the trends and patterns in

| EVENT_TYPE | R-Square | MSE | MAE |
|---|---|---|---|
| Battles | -1.33 | 2924.85 | 50.44 |
| Riot | -0.66 | 501012.54 | 561.07 |
| Protests | -23.36 | 129848.08 | 325.19 |
| Remote Violence | 0.00 | 1.85 | 1.28 |
| Explosion Remote Violence | -5.14 | 304.94 | 14.48 |
| Strategic Development | 0.00 | 6.96 | 2.47 |
| Violence Against Civilians | 0.00 | 7343.37 | 80.41 |

different criminal offenses and occurrences throughout time.

**Analysis of Dataset of Police Bangladesh (2010-2019):** We have been analyzing the Bangladeshi Police data and concentrating on many categories of crimes such as Arms act, explosives, murder, narcotics, police assault, burglary, dacoity, kidnapping, other cases, riot, robbery, smuggling, speedy trial, theft and woman_and_child_repression in the country. From that research, it became apparent that trends are as follows:

**Crime Rate Has Been Increasing:** From 2010 to 2019, there was an upward trajectory in offenses such as arms act, explosives, murders, narcotics and police assault an indication that even though they have been made better in some spheres; it is only these kind of offences that have increased steadily.

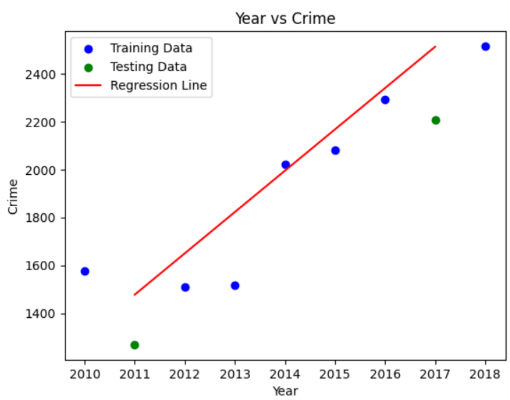


Figure 5.1: Arms Act

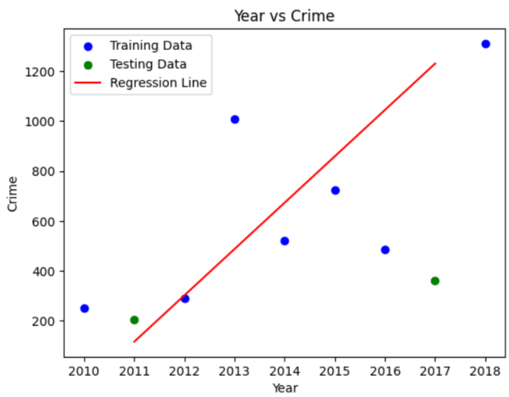


Figure 5.2: Explosives

**Decreasing crimes:**Burglary, Kidnapping, other cases, riot, robbery, smuggling, speedy trial, theft, dacoity, woman and child repression—these have had decreasing trends across similar periods. It is worth noting that such offenses are usually done for money purposes. The dramatic fall in financial-driven crimes between 2010 and 2019 has been linked to an exceptional growth in GDP. People experienced higher living standards and economic stability as they became wealthier, which lowered their desire for financial vices. This decrease could also be due to increased reading and comprehension capabilities.

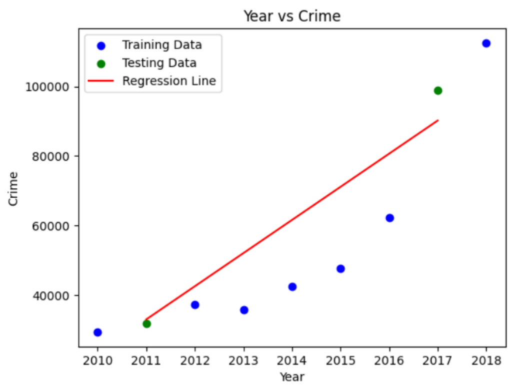


Figure 5.3: Narcotics

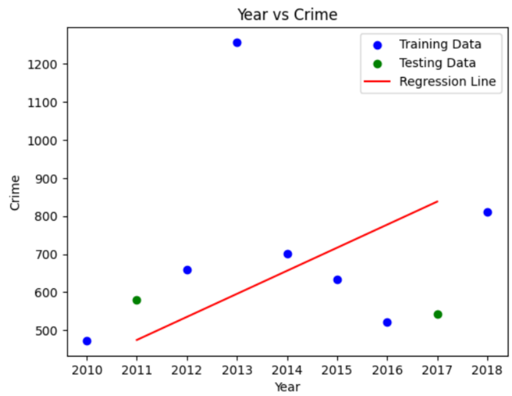


Figure 5.4: Police Assault

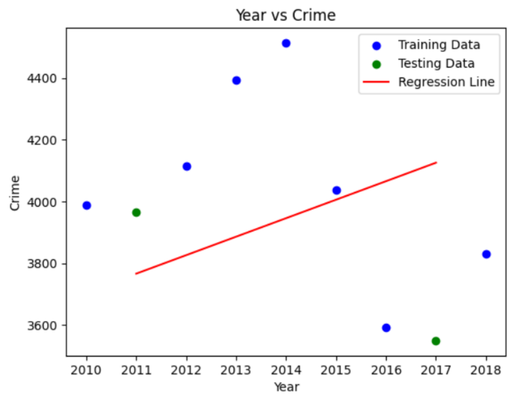


Figure 5.5: Murder

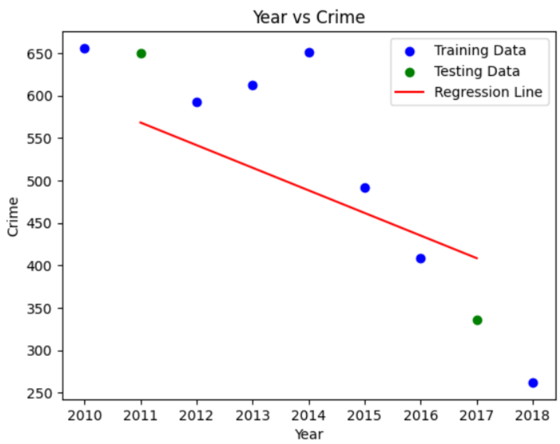


Figure 5.6: Dacoity

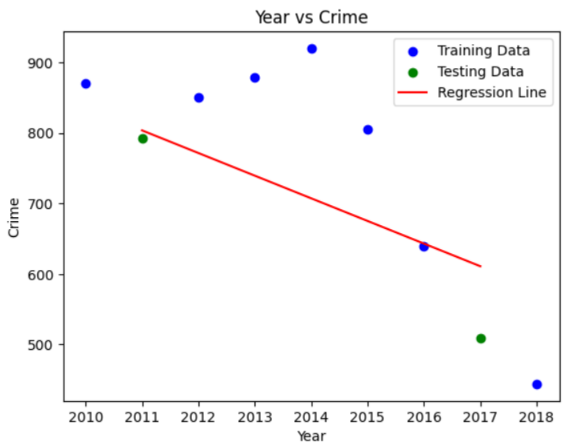


Figure 5.7: Kidnapping

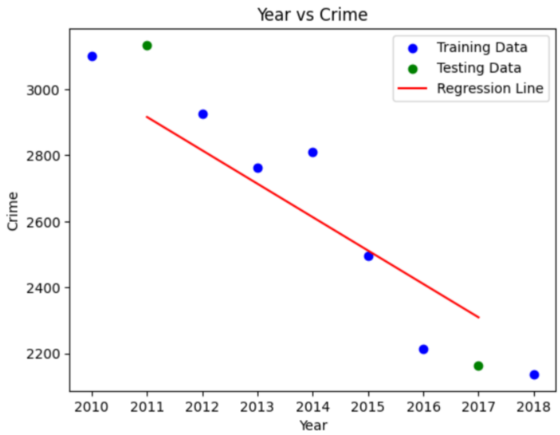


Figure 5.8: Burglary

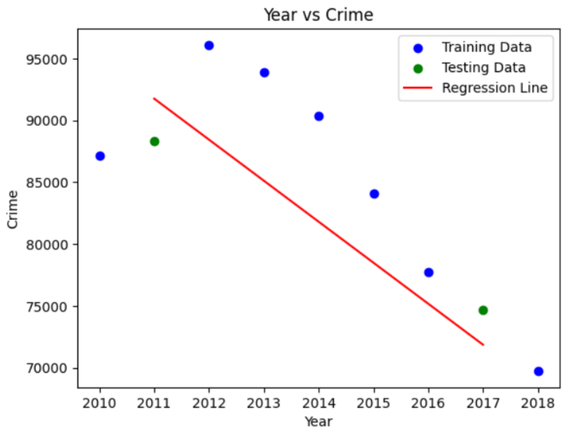


Figure 5.9: Other Cases

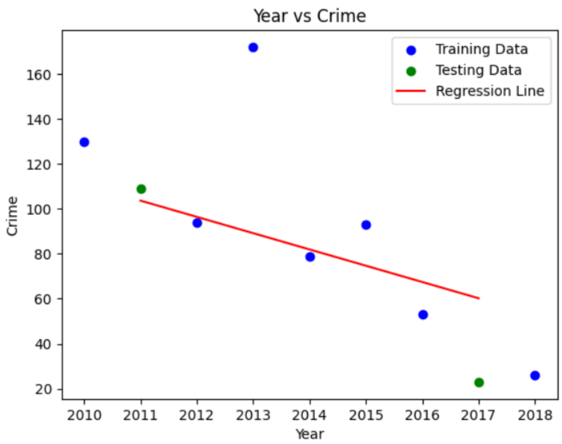


Figure 5.10: Riot

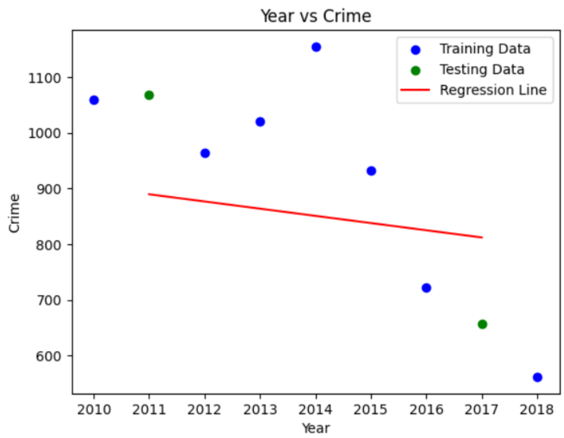


Figure 5.11: Robbery

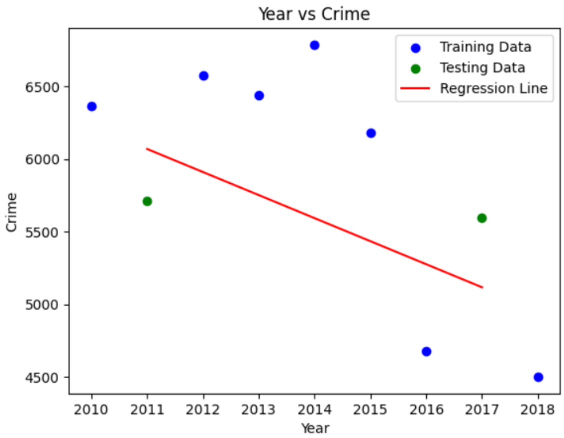


Figure 5.12: Smuggling

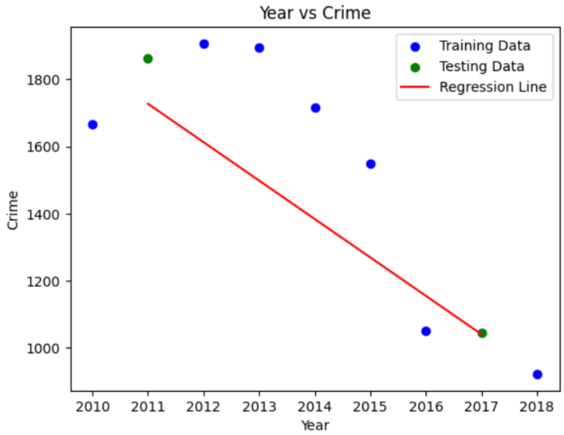


Figure 5.13: SpeedyTrial

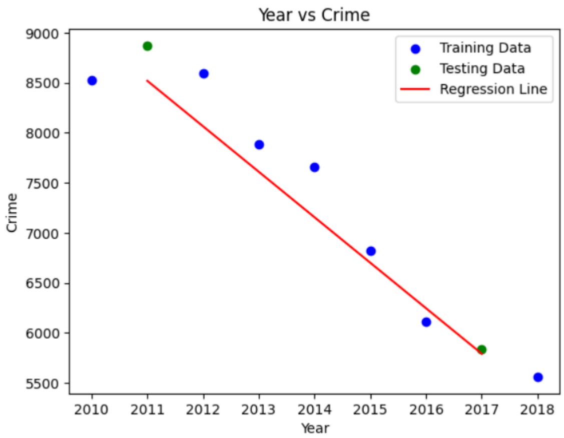


Figure 5.14: Theft

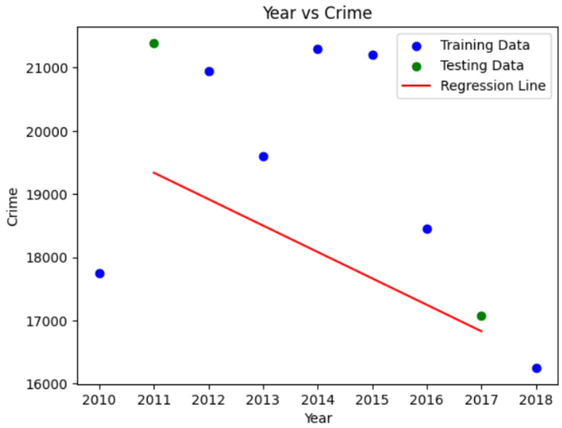


Figure 5.15: Decreasing crimes/Woman and Child Repression.

**ACLED Dataset Analysis (2001-2020):** The ACLED dataset gave an overview on different violent events including battles, riots, violence against civilians, remote violence, explosion remote violence, and protests. Our regression analysis led to the following final results.

**Increasing event:** Protests have increased significantly throughout this period. The red dots on the geospatial map indicate areas where protests have been more common. Growing literacy rates and awareness among the population contribute to the rise in protests. When they are aware of their rights and become more educated, these people tend to take part in protests in order to ask for change or fair treatment. What emerges from this change is how crucial it is to engage politically and promote civic education thereby making it possible for citizens to fight for their own rights.

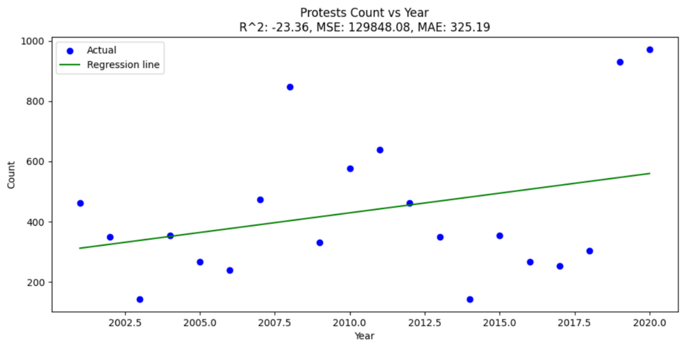


Figure 5.16: Protest.

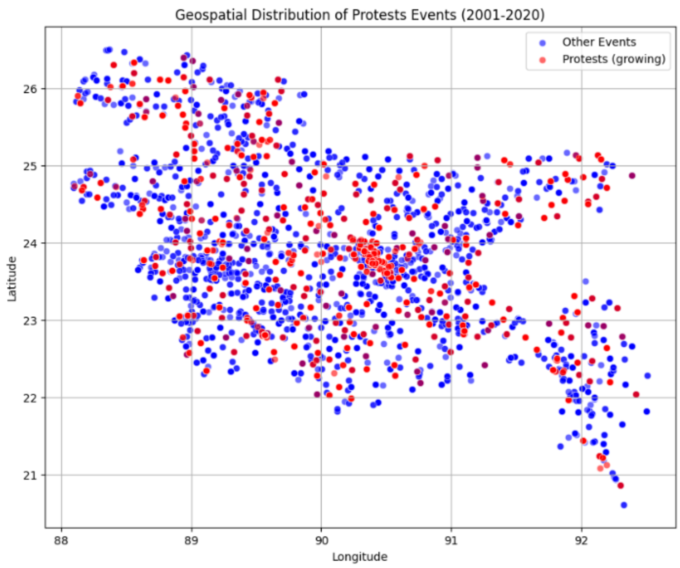


Figure 5.17: Protest_Event.

**Decreasing events:** Battles, riots, violence against civilians, remote violence, and explosions. Remote Violence: From 2001 to 2020, these events exhibited a declining tendency. Green dots on the geospatial map indicate locations where violent occurrences have decreased in frequency.

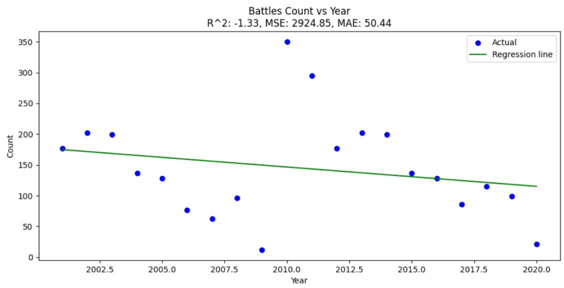


Figure 5.18: Battles

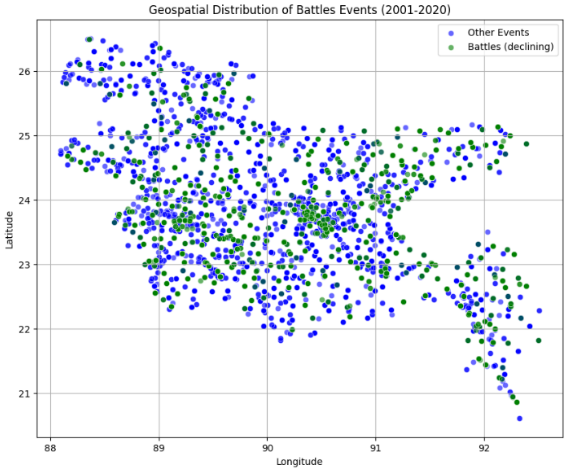


Figure 5.19: BattlEvent

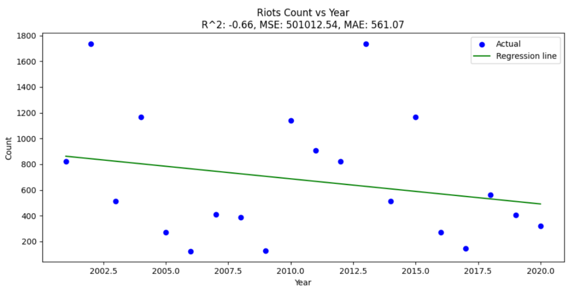


Figure 5.20: Riots

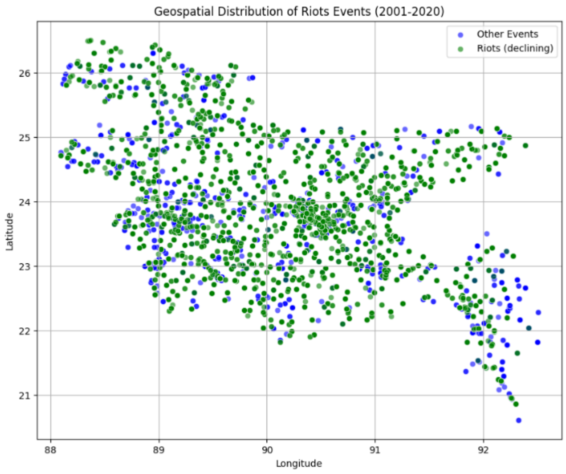


Figure 5.21: RiotEvent

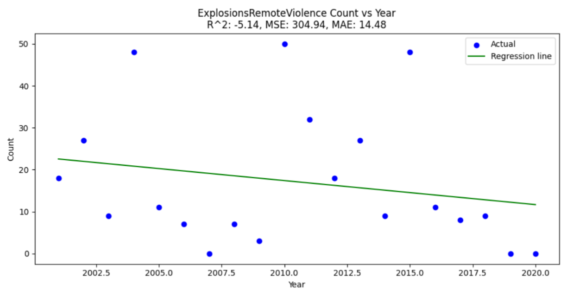


Figure 5.22: ExplosionRemoteViolence

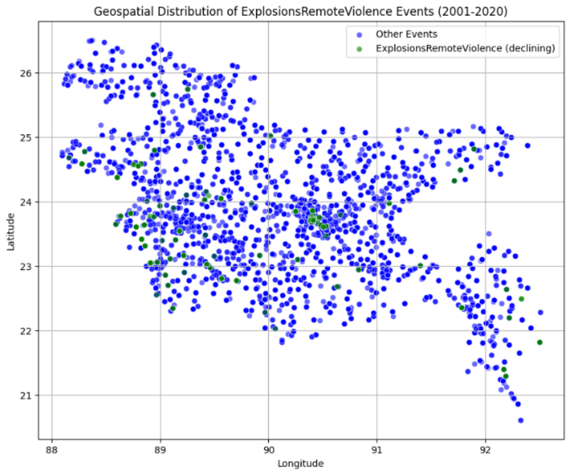


Figure 5.23: ExplosionRemoteViolentEvent

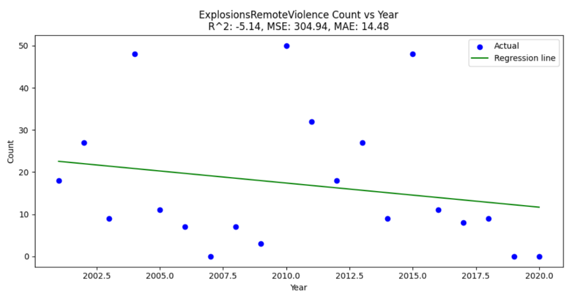


Figure 5.24: RemoteViolence

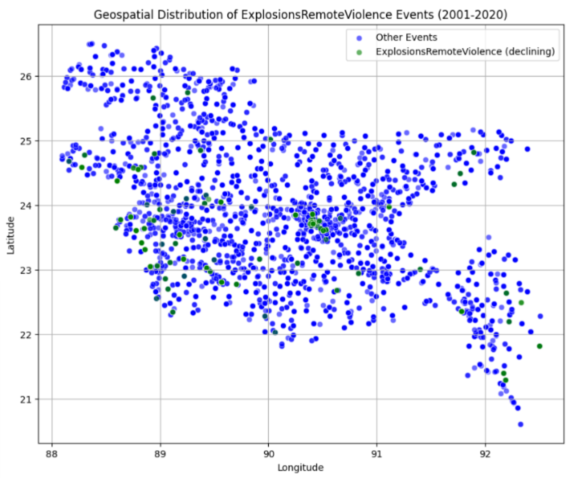


Figure 5.25: RemoteViolenceEvent

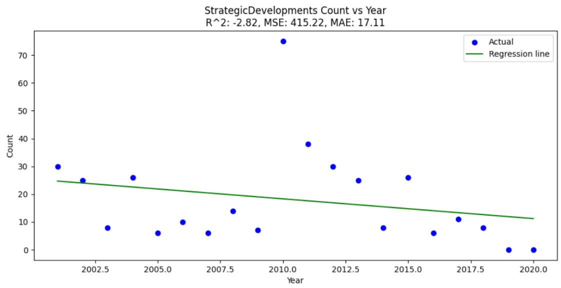


Figure 5.26: Strategic Development

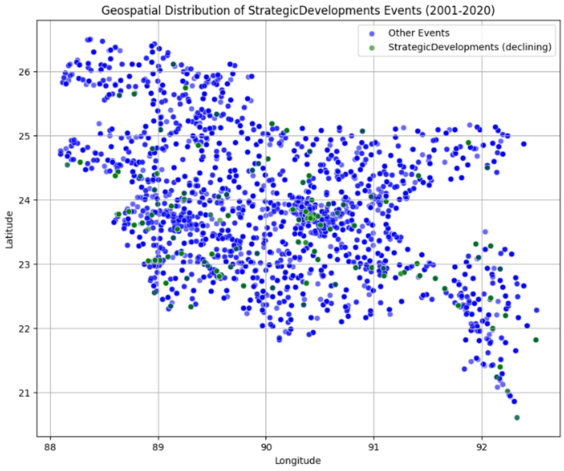


Figure 5.27: StrategicDevelopmentEvent

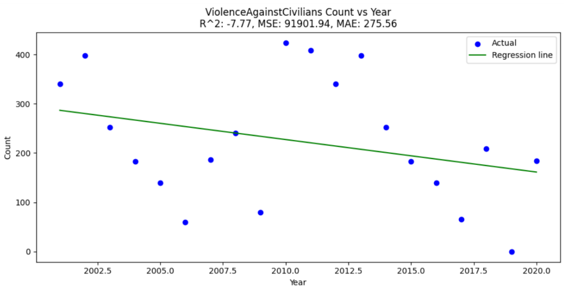


Figure 5.28: Violence Against Civilians

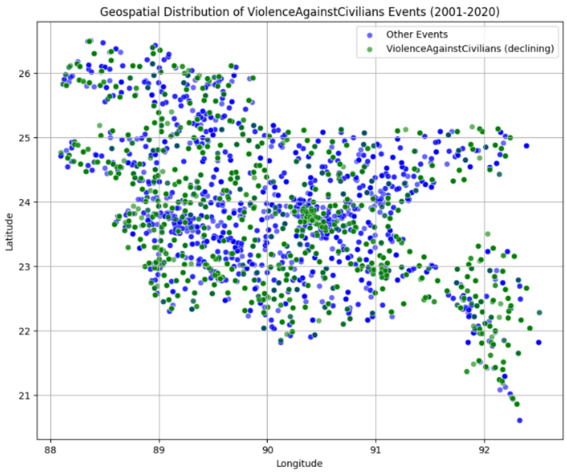


Figure 5.29: ViolenceAgainstCivilianseEvent

**Heatmap:** The heatmap demonstrates how various types of events have been spread throughout the years between 2001 and 2020 regarding their prevalence. The number of times a given event occurred within a given year is what each cell in a heatmap symbolizes. More frequent events are represented by darker colors within cells whose color intensities correspond to how frequently they have taken place.

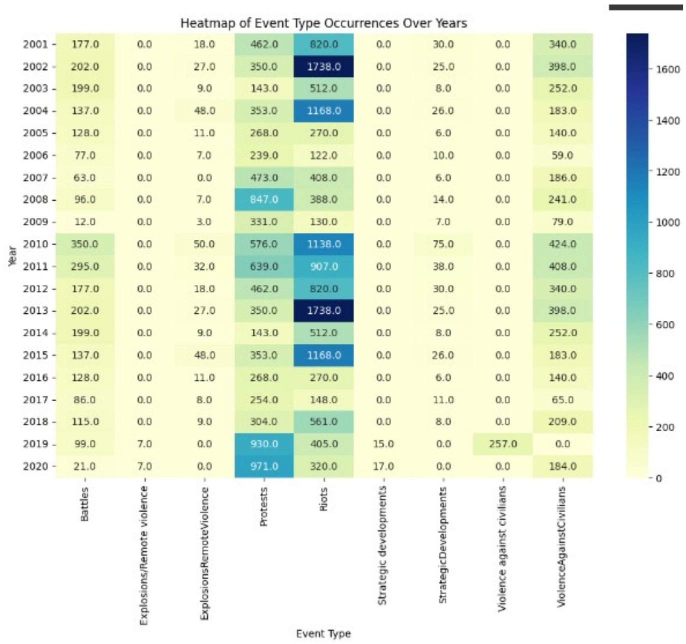


Figure 5.30: Heat Map.

## 5.4 Summary

The regression analysis of both datasets reveals substantial trends and patterns in criminal and violent activity during the study periods. The findings show that economic development and increasing literacy rates are critical in reducing financially motivated crimes. Conversely, the increase in protests demonstrates a more aware and active population fighting for their rights. These findings can help policymakers and law enforcement authorities focus on regions that need attention and promote a safer, more fair society.

# Chapter 6

# Complex Engineering

### 6.0.1 Knowledge Profile

This section will concentrate on the knowledge profile process, which compiles information on a project's qualities.

| Profile | Description | Status |
|---|---|---|
| K1 | Natural Sciences applicable to our project | No |
| K2 | Required Mathematical or Computational analysis | Yes |
| K3 | Engineering Fundamentals that need to be sustained | Yes |
| K4 | Specialist in-depth knowledge required | Yes |
| K5 | Engineering Design and Front-end work | Yes |
| K6 | Engineering practices required to be upheld | Yes |
| K7 | Comprehensive Engineering Ethics and real-life impacts and Implications | Yes |
| K8 | Literature and website research on related projects | Yes |

### 6.0.2 Complex Problem Solving

Table 6.1: Mapping with complex problem-solving.

| Attribute | Status | Problem Description |
|---|---|---|
| P1 (Depth of Knowledge) | Yes | The Mathematical terms & data analysis are included(k2). Implementation of word matching and searching (k3,k4). Web-based front end, interface and database management (K5) and keeping the project safe and cost-efficient (K6). (K8) helps us find the limitation and need to improve our paper. |
| P2 (Range of Conflicting Requirements) | Yes | Wide-ranging or conflicting technical, engineering, and other issues are common in many fields and require careful consideration to find a solution that satisfies all stakeholders. It may involve trade-offs, compromises, and a deep understanding of the various factors to arrive at a satisfactory answer. |
| P3 (Depth of Analysis) | Yes | This project integrates multi-variable socio-economic information and models crime patterns and their association with GDP and population metrics using advanced regression analysis and machine learning algorithms, guaranteeing robust forecasting accuracy and actionable insights. |
| P4 (Familiarity of issues) | Yes | A thorough spatial-temporal analysis is provided by the application of geospatial data visualization techniques, such as heatmaps and scatter plots with regression lines, which improve the interpretability of the intricate relationships between socioeconomic factors and crime patterns. |
| P5 (Extent of applicable codes) | Yes | Professional engineering codes and maintain standards for their projects. They continuously evolve to keep pace with changes in industry and society, resulting in engineering projects that are safe, dependable, and socially responsible. |
| P6 (Extent of stakeholder involvement and conflicting requirements) | Yes | SThe research demonstrates the application of sophisticated data engineering techniques to generate significant correlations by methodically processing and analyzing large-scale datasets from various sources. This supports the development of data-driven policies and strategic planning in the field of crime prevention. |
| P7 (Interdependence) | Yes | In order to ensure a multifaceted study that takes into account temporal, spatial, and economic dimensions, the project integrates data from numerous sources in a comprehensive manner to examine the socio-economic drivers of crime. This integration enables the creation of evidence-based interventions and policy decisions by facilitating a comprehensive understanding of criminal dynamics. |

### 6.0.3 Engineering Activities

Table 6.2: Complex Engineering Activities.

| Attribute | Status | Complex Engineering Problems have characteristics P1 and some or all of P2 to P7: |
|---|---|---|
| A1 (Range of resources) | Yes | Define and identify the important socioeconomic factors (GDP, population, and crime statistics) to guarantee that the goals and scope of the project are understood. |
| A2 (Interaction Level) | Yes | Create a pipeline for data processing that will clean, preprocess, and merge datasets from ACLED and Bangladesh Police while preserving consistency and integrity of the data. |
| A3 (Innovation) | Yes | Establish relationships between socioeconomic characteristics and crime rates using regression analysis and machine learning models; validate the models using pertinent statistical indicators. |
| P4 (Familiarity of issues) | Yes | Use heatmaps, regression lines, and scatter plots to visually represent the data so that trends in crime patterns and findings may be communicated clearly. |
| A5 (Consequences for society and the environment) | Yes | To make sure the predictive insights obtained from the study are accurate and reliable, assess the models' performance using R-square, MSE,and MAE. |
| A6 (Familiarity) | Yes | Provide thorough reports and graphical summaries that capture the research findings and provide stakeholders and policymakers with useful information to guide crime prevention initiatives. |

## 6.1 Summary

In order to simulate the relationship between crime trends and socioeconomic variables like GDP and population, this project makes use of sophisticated regression analysis and machine learning techniques. It combines multi-variable data with extensive datasets from Bangladesh Police and ACLED to produce a strong forecasting framework. In order to provide a thorough spatial-temporal study of crime patterns, the project applies geospatial data visualization techniques including heatmaps and scatter plots. These approaches guarantee a sophisticated comprehension of the variables impacting criminal activity, facilitating data-driven policy development and tactical planning for successful crime prevention.

# Chapter 7

# Conclusion

Our primary objective was to determine the connections between GDP, crime, and population, and we were successful in visualizing these relationships in our research. The data revealed that while crimes like protests have grown due to more understanding of rights, economic progress and increased literacy are connected with a drop in financially motivated crimes. The heatmap, which showed the temporal and spatial variations in crime occurrences, provided an additional level of comprehension. Policymakers and other stakeholders can benefit greatly from these findings, which can be used to help build focused measures to reduce crime and enhance public safety.

## 7.1 Summary

In our study, we used statistics from the Bangladesh Police and ACLED to investigate the relationships between population, GDP, and crime rates. Regression analysis was used to find trends in different types of crimes and how they related to socioeconomic variables. According to our data, some crimes with financial motivations have decreased, but others have increased as a result of socioeconomic shifts. Additionally, we used heatmaps, regression lines, and scatter plots to illustrate our findings, giving us a thorough insight of the patterns of crime over time and across various geographies.

## 7.2 Limitation

The project is mostly about crime patterns that are related to general illegal activities. It doesn't look into specific areas like money laundering or trafficking people. The analysis is not as complete because these specific crimes were not included. The project doesn't look at digital crimes like cyberbullying and other cyber offences. Because these crimes are so complicated, they need more in-depth analysis than what the project is doing now, which is mainly looking at common crime patterns.

## 7.3 Future Work

To make room for future research, studying digital crimes like cybercrime and online bullying is still uncharted territory that needs special care. The accuracy and reliability of crime predictions could be improved by adding more socioeconomic variables to predictive models and using a larger dataset. Additionally, working with law enforcement to use machine learning in the real world and keep testing it against changing crime patterns is a promising way to make it a practical part of crime prevention strategies.

# References


[1] Ibidun Christiana Obagbuwa and Ademola P Abidoye. South africa crime visualization, trends analysis, and prediction using machine learning linear regression technique. *Applied Computational Intelligence and Soft Computing*, 2021:1–14, 2021.

[2] Luiz GA Alves, Haroldo V Ribeiro, and Francisco A Rodrigues. Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and its Applications*, 505:435–443, 2018.

[3] S Sridharan, N Srish, S Vigneswaran, and P Santhi. Crime prediction using machine learning. *EAI Endorsed Transactions on Internet of Things*, 10, 2024.

[4] Rabia Musheer Aziz, Aftab Hussain, Prajwal Sharma, and Pavan Kumar. Machine learning-based soft computing regression analysis approach for crime data prediction. *Karbala International Journal of Modern Science*, 8(1):1–19, 2022.

[5] Seema Aggarwal, Geeta Aggarwal, and Manisha Bansal. Crime against women and girls in indian states: Analysis, visualization and prediction.

[6] Md Pavel Rahman, AK M Ifranul Hoque, Md Faysal Ahmed, Iftekhirul Iftekhirul, Ashraful Alam, and Nahid Hossain. Bangladesh crime reports analysis and prediction. In *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*, pages 453–458. IEEE, 2021.

[7] Md Abdul Awal, Jakaria Rabbi, Sk Imran Hossain, and MMA Hashem. Using linear regression to forecast future trends in crime of bangladesh. In *2016 5th international conference on informatics, electronics and vision (ICIEV)*, pages 333–338. IEEE, 2016.

[8] Al Amin Biswas and Sarnali Basak. Forecasting the trends and patterns of crime in bangladesh using machine learning model. In *2019 2nd international conference on intelligent communication and computational techniques (ICCT)*, pages 114–118. IEEE, 2019.

[9] Saqueeb Abdullah, Farah Idid Nibir, Suraiya Salam, Akash Dey, Md Ashraful Alam, and Md Tanzim Reza. Intelligent crime investigation assistance using machine learn-

ing classifiers on crime and victim information. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pages 1–4. IEEE, 2020.

[10] Wajiha Safat, Sohail Asghar, and Saira Andleeb Gillani. Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE access*, 9:70080–70094, 2021.