

Crime Rate Analysis & Visualization

Submitted in partial fulfillment of the requirements of the
degree

BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING

By

Kritarth Kambli (Roll no. 28)

Rahul Patil (Roll no. 50)

Pooja Patkar (Roll no. 52)

Vishal Shirke (Roll no. 64)

Name of the Mentor

Prof. Avinash Gondal



Department of Computer Engineering

**Watumull Institute of Electronics Engineering and
Computer Technology**

Ulhasnagar-421003

University of Mumbai

(AY 2022-23)

CERTIFICATE

This is to certify that the Project entitled “**Crime Rate Analysis & Visualization**” is a bonafide work of

Kritarth Kambli (Roll no. 28)

Rahul Patil (Roll no. 50)

Pooja Patkar (Roll no. 52)

Vishal Shirke (Roll no. 64)

submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of “**Bachelor of Engineering**” in “**Computer Engineering**”.

(Prof. Avinash Gondal)

Mentor

(Prof. Rahul Jintunkar)

Head of Department

(Prof. Sunita Sharma)

Principal

PROJECT APPROVAL

This Project entitled “Crime Rate Analysis & Visualization” by

Kritarth Kambli (28)

Rahul Patil (50)

Pooja Patkar (52)

Vishal Shirke (64)

is approved for the degree of **Bachelor of Engineering** in **Computer Engineering**.

Examiners

Prof. Avinash Gondal

(Internal Examiner Name & Sign)

(External Examiner name & Sign)

Date:

Place:

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Kritarth Kambli (Roll no. 28)

Rahul Patil (Roll no. 50)

Pooja Patkar (Roll no. 52)

Vishal Shirke (Roll no. 64)

Date:

ACKNOWLEDGEMENT

We would like to extend our sincere and heartfelt gratitude to our professor Mr. Avinash Gondal who has helped us in this endeavor and has always been very cooperative and without his help, cooperation, guidance and encouragement, the project couldn't have been what it evolved to be.

We extend our heartfelt thanks to our faculty for their guidance and constant supervision, as well as, for providing us the necessary information regarding the project. We must thanks to our classmates for their timely help& support for compilation of this project.

Last but not least, we would like to thank all those who helped us (directly or indirect) towards the completion of this project within a limited time frame.

Kritarth Kambli (28)

Rahul Patil (50)

Pooja Patkar (52)

Vishal Shirke (64)

(B.E.)

ABSTRACT

To be better prepared to respond to criminal activity, it is important to understand patterns in crime. In our project, we analyse crime data from Maharashtra state, scraped from publicly available website of NCRB.

Growing crime rate is a major concern and it is essential to find techniques to reduce it. Criminal data is available for public but it is mostly focused on combined data about a state. The data for individual cities is not easily accessible and is mostly in the form of FIR/charge sheets in local languages. Crime analysis gives us better understanding of crime trends which is highly beneficial. In this project, we analyse crime data and use it to produce simpler data in the form of graphs and crime maps which are easy to understand. The application will provide a brief overview of current state of crime. This can be solved by analysing, classifying and creating a GIS environment for individual cities. Simple crime maps and statistical data in the form of graphs can help the police as well as the citizens in many ways.

Government of Maharashtra has sanctioned the proposal of Mumbai Police to implement GIS/GPS based systems. The current available systems are in initial phase and still being developed. We aim to build a GIS based system model which will be available to public and can be used as a starting point to build more complex prediction systems in future. Improvements in crime prevention technology will likely spur increased total spending on this technology which will result in better measures to tackle increasing crime rate problems and provide a healthy and crime free neighbourhood for the people.

TABLE OF CONTENTS

Certificate	ii
Project Approval	iii
Declaration	iv
Acknowledgements	v
Abstract	vi
1. Introduction	01
1.1. Introduction	01
1.2. Product Scope	01
1.3. Problem Statement & Objectives	01
2. Literature Survey	02
2.1. A Crime Analysis Tool	02
2.2. Current Scenario	03
2.3. Data Dissemination, Data secrecy and Fear of Crime	03
2.4. Software Requirements Analysis (SRS)	04
2.5. Limitation of Existing Systems	09
2.6. Project Contribution	10
3. Proposed System	11
3.1. Architecture/Framework	12
3.2. Algorithm & Process Design	12
3.2.1. Prediction	13
3.2.2. Clustering	14
3.2.3. Dataset Information	17
3.3. Details of Hardware & Software	18
3.3.1. Hardware requirements	18
3.3.2. Software requirements	18
4. Visualization and Results	19
4.1. Crime Rate Analysis (CRA) Website	19
4.1.1. Home Page	19
4.1.2. Crime Charts	20
4.1.3. Crime Data	20
4.1.4. Crime Maps	21
4.1.5. Crime News	22
4.2. Visualization	22
4.2.1. First Method	22
4.2.2. Second Method	27
4.2.3. Results	29
5. References	30

LIST OF FIGURES

1. Introduction	01
2. Literature Survey	02
2.1. Figure 2.1	02
3. Proposed System	11
3.1. Figure 3.1	12
3.2. Figure 3.2	14
3.3. Figure 3.3	14
3.4. Figure 3.4	15
3.5. Figure 3.5	15
3.6. Figure 3.6	15
3.7. Figure 3.7	16
3.8. Figure 3.8	16
3.9. Figure 3.9	17
4. Visualization and Results	19
4.1. Figure 4.1	19
4.2. Figure 4.2	20
4.3. Figure 4.3	20
4.4. Figure 4.4	21
4.5. Figure 4.5	21
4.6. Figure 4.6	22
4.7. Figure 4.7	24
4.8. Figure 4.8	25
4.9. Figure 4.9	26
4.10. Figure 4.10	27
4.11. Figure 4.11	28
4.12. Figure 4.12	29
5. References	30

Chapter 1

INTRODUCTION

1.1 Introduction

Since crime is a growing concern in every part of the world, it is very essential to find techniques to reduce it. Crime analysis is a well-organized way of detecting and examining patterns, trends in crime and a better understanding of crime is beneficial. With the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime is an active field of research. In India, the Police agencies of Mumbai, Delhi, Bangalore, etc. use GIS (A geographic information system that creates, manages, analyzes, and maps all types of data.) for crime mapping. But applications for GIS in policing is in initial stages and GIS based data analysis can be useful to police as well as citizens across major cities to tackle the increasing crime scenario.

1.2 Product Scope

There is a change in the trends of crimes and it is very challenging to find the new trends and patterns in crime. Criminal data is available for public but it is mostly focused on combined data about a state. The data for individual cities is not easily accessible and is mostly in the form of FIR/charge sheets in local languages. This can be solved by analyzing, classifying and creating a GIS environment for individual districts. Simple crime maps and statistical data in the form of graphs can help the police as well as the citizens in many ways.

1.3 Problem Statement and Objectives

We propose to build a mobile application/website which will provide a brief overview of current state of crime. The frontend of the application will focus on statistical data visualization and crime mapping. The backend system will focus on generating accurate statistical data and intensity of crime for GIS mapping.

Objective are as follows:

- Predicting surges and hotspots of crime
- Understanding crime pattern.
- Classify crime based on location (Districts)
- Analysis of crime in Maharashtra State
- Easy accessibility of application/website available for public and private use
- Understanding patterns of criminal behaviour that could help in solving criminal investigations.

Chapter 2

LITERATURE SURVEY

2.1. A Crime Analysis Tool

In North America a company named Central Square, a technologies foundation has been developing advanced software for public sector. One which includes a website called ***CrimeMapping.com*** to help law enforcement agencies throughout North America provide the public with valuable information about recent crime activity in their neighborhood.

Their goal is to assist police departments in reducing crime through a better-informed citizenry. Creating more self-reliance among community members is a great benefit to community-oriented policing efforts everywhere and has been proven effective in combating crime.

CrimeMapping.com utilizes an advanced mapping engine, which helps them provide a high level of functionality as well as flexibility to the agencies they serve. Crime data is extracted on a regular basis from each department's records system so that the information being viewed through a Web browser is the most recently collected.

This data is always verified for accuracy and all address information is generalized by block in order to help ensure privacy is protected. The data provided through departments plays a crucial role for visualizing crime statistics. It provides type and date with the exact time and location where the actual crime took place.

Fig.2.1 shows the visualization of an area's crime according to the departments data.

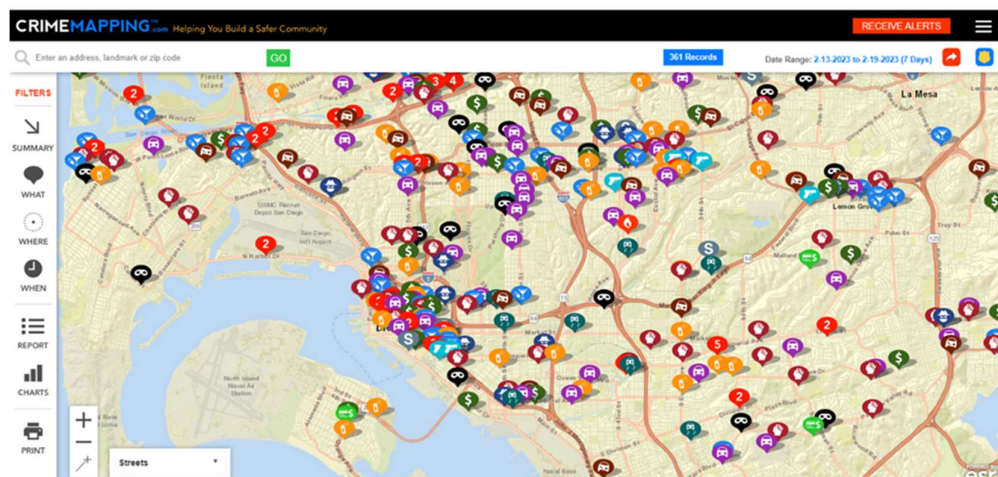


Figure 2.1

2.2. Current Scenario

In India, the individual police departments, both rural and urban, collect crime data. Most of the data is collected from the complaints from victims or informants in the form of First Information Report (FIR). Later they are sent to the district crime records bureau and later sent to State crime records bureau.

State crime records bureau compiles and sends the data to National crime records bureau. With the data, the National Crime Records Bureau (NCRB) publishes a report called "Crime in India" every year. This publication has a graphical and map representations but no significant spatial analysis output is presented.

State Crime records bureaus produce only graphical representation on crime scenario and they have never thought of serious analysis. Moreover, the conversion of data to information is insufficient. State crime records bureaus are data rich, but they are information poor.

On the other hand, computerization of records in those bureaus have been done long back and NCRB has its own software (Crime and Criminal Information System (CCIS)) and it gives training in this software for the personnel of all State crime records bureau. Unfortunately, hot technologies like GIS and GPS are not given enough priority. GIS technology is still beyond the limit of records bureaus, due to lack of awareness and ignorance of its importance.

2.3. Data Dissemination, Data secrecy and Fear of Crime

Indian police do data dissemination of crime efficiently. A network to the state headquarters connects most of the districts. This makes the police to collect the data from all the districts. Nevertheless, the problem here is the lack of trained officers to take care of the network.

In India, always crime data are maintained secretly. Even after the growth of the Internet. The reason is that, Indian police feel that showing the crime data publicly will create fear among the public, by publishing crime data and maps in the internet.

2.4. System Requirement Specification (SRS)

Software Requirements Specification For

CRIME RATE ANALYSIS

Prepared by

- 1. Kritarth Kambli (20)**
- 2. Rahul Patil (37)**
- 3. Pooja Patkar (39)**
- 4. Vishal Shirke (48)**

Table of Contents

1. Introduction	4
1.1. Purpose	4
1.2. Document Conventions	4
1.3. Intended Audience and Reading Suggestions	4
1.4. Product Scope	4
1.5. References	4
2. Overall Description	5
2.1. Product Perspective	5
2.2. Product Functions	5
2.2.1. Crime Rate Analysis	5
2.3. User Classes and Characteristics	5
2.4. Operating Environment	5
3. External Interface Requirements	6
3.1. User Interface	6
4. System Features	6
4.1. Crime Rate Analysis	6
4.1.1. Description and Priority	6
4.1.2. Stimulus	6
4.2. Functional Requirements	6
5. Other Non-functional Requirements	7
5.1. Software Quality Attributes	7
5.1.1. Availability	7
5.1.2. Security	7
5.1.3. Maintainability	7

1. Introduction

1.1 Purpose

The purpose of this Software Requirements Specifications (SRS) is to fully document the specifications and requirements for the Crime Rate Analysis System.

1.2 Document Conventions

- CRA: CRIME RATE ANALYSIS.
- This document will use IEEE format. For clarity, acronyms and technical jargon, deemed uncommon by the author, will be annotated and included in the glossary. The format for headings is as followed: Major headings are in bold 16pt font, and concurrent headings in bold 14 pt. font and the rest of the font in 12 pt.

1.3 Intended Audience and Reading Suggestions

The audience of this SRS will be the examiners who will evaluate the work and the students developing the software.

The overall SRS is organized as follows:

- Section 1(current) Introduction of project.
- Section 2 is about overall description of the project.
- Section 3 cites the External Interface Requirements.
- Section 4 about various System Features.
- Section 5 about Other Non-functional Requirements.

1.4 Product Scope

The objective of this project is to create and implement a mobile application/website which will be available for public and private use. The application will provide a brief overview of current state of crime. The frontend of the application will focus on statistical data visualization and crime mapping. The backend system will focus on generating accurate statistical data and intensity of crime for GIS mapping.

2. Overall Description

2.1 Product Perspective

This product is an entirely new product. It is not a component of a larger system. The Crime Rate Analysis System will provide users with simple crime maps and statistical data in the form of graphs which will be interactive. The application will provide a brief overview of current state of crime.

2.2 Product Functions

2.2.1 Crime Analysis Module

The Crime Analysis function shall allow users to view and interact with crime maps and other graphs providing the criminal data. On the website the user will click on different maps and graphs for different districts and view crime data in graphical form. Rationale: This provides a method for user to get a view of the state of crime in a particular area or in a various district.

2.3 User Classes and Characteristics

The types of user classes are normal people, police, students, security personals, tourists etc. depending on the user the minimum criteria changes as follows:

A user should be minimum 8 years of age. Users of the website must possess a minimal educational level and knowledge to use a website. Users of the application must know how to navigate in a website.

2.4 Operating Environment

Hardware Requirements:

- Hardware i3 6th gen, 4 GB ram

Software Requirements:

- Windows 7 operating system or higher.
- Jupyter notebook or VS Code.
- Flask, Heroku
- SKlearn libraries.

3. External Interface Requirements

3.1 User Interfaces

The system will provide a web platform for user to get crime data in the form of statistical data and GIS maps.

- The application will be used to get an overview of the crime state/data using GIS maps and graphical data.

4. System Feature

4.1 Crime Rate Analysis:

4.1.1 Description and Priority

User has to click on the different crime maps and other graph data tabs.

4.1.2 Stimulus: Click "Crime Maps" Button: GIS Maps of district wise crime data.

- The system shall allow any user to access the website and get crime data.
- User has to select from different crime maps and graphs to view.
- The system will process the user's request and provide a GIS Map or statistical data on crime.

4.2 Functional Requirements:

Client/Server System:

A client system which will be used to collect user information as the front end application, and a server system to process and run the crime analysis module i.e. A Crime Rate Analysis and Mapping system working as the back end of the application.

Data Model: The input from the user will be taken as a String or Click on tab and the output in the form of statistical data or graphs and GIS maps.

5. System Feature

5.1 Software Quality Attributes

5.1.1 Availability

This system is an online system which does require internet, so as long as the user's system supports web services like a web browser, the system should function.

5.1.2 Security

Users will be able to access all the data on the site as it is free to use and uses the publicly available criminal data from government websites.

5.1.3 Maintainability

Any updates or defect fixes shall be able to be made on server-side Computers only without any patches required by the user.

2.5. Limitation of Existing Systems

The current available systems are in initial phase and still being developed. For Chennai Police, ESRI India has sold two licenses of Arc view GIS software and has imparted training. GIS based systems are being used by New Delhi, Hyderabad and Assam Police Departments. Government of Maharashtra has sanctioned the proposal of Mumbai Police to implement GIS/GPS based systems. Technologies like GIS and GPS are not given enough priority. GIS technology is still beyond the limits of crime records bureaus, due to lack of awareness and ignorance of its importance.

In India, the individual police departments, both rural and urban, collect crime data. Most of the data are collected from the complaints from victims or informants in the form of First Information Report (FIR). With the data, the National Crime Records Bureau (NCRB) publish a report called "Crime in India" every year. This publication has graphical and map representations but no significant spatial analysis output is presented. State Crime records bureaus produce only graphical representation on crime scenario and they have never thought of serious analysis.

Lack of training in GIS, spatial analysis, and computer mapping. Police officers lack training in these new technologies and they might even be computer illiterates. For this reason, any GIS software product developed for police departments needs to be easy to use, preferably with a point and click interface. The cost of training is also very high.

Properly mapped and statistical information is not available for public. Criminal data is available for public but it is mostly focused on combined data about a state. The data for individual cities is not easily accessible

Currently available systems are only capable of mapping past data and not for prediction and other analytical use. Such GIS systems are recently being considered and it might take a few years to develop properly working systems which even public can access.

2.6. Project Contribution

The use of AI/ML in predicting crimes or an individual's likelihood for committing a crime has promise but is still more of an unknown. The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative. Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology.

Possible avenues through which to extend this work include time-series modeling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories of crimes.

The idea behind this project is that crimes are relatively predictable; it just requires being able to sort through a massive volume of data to find patterns that are useful to law enforcement. This kind of data analysis was technologically impossible a few decades ago, but the hope is that recent developments in machine learning are up to the task.

Public safety and protection relate to crime, and a better understanding of crime is beneficial in multiple ways: it can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime, and more concerted efforts by citizens and authorities to create healthy neighborhood environments. With the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis, understanding patterns in crime from data is an active and growing field of research.

Chapter 3

PROPOSED SYSTEM

Past decades have experienced the rage of GIS technology in a never-ending scope. In today's era, visualization of high-dimensional hyperspectral data is an indispensable task and GIS is simply a platform to practically experience the visualization. Furthermore, crime is an unprecedented event and to analyze crime there exist many technologies but to visualize it we are limited to adopt a few technologies and GIS is one of those few.

With the recent development of the Internet and satellites, the scope of utilization of the Geographic Information System (GIS)-based technology has been increasing. GIS refers to a system that integrates and manages spatial data and attribute data for objects with geographical locations to provide various forms of information, such as maps, diagrams and drawings.

It is expected that such crime map visualization research will identify the distribution of spatial characteristics of crime and further combine them with urban-engineered, socio-economic, and demographic characteristics to be used to establish policing policies that consider regional characteristics.

With the rapid urbanization and development of big cities and towns, the graph of crimes is also on the increase. This phenomenal rise in offences and crime in cities is a matter of great concern and alarm to all of us. For easy accessibility and use we propose to build a mobile application/website which will be available for public and private use.

The application will provide a brief overview of the current state of local crime. The frontend of the application will focus on statistical data visualization and crime mapping. The backend system will focus on generating accurate statistical data and intensity of crime for GIS mapping.

3.1. Architecture/Framework

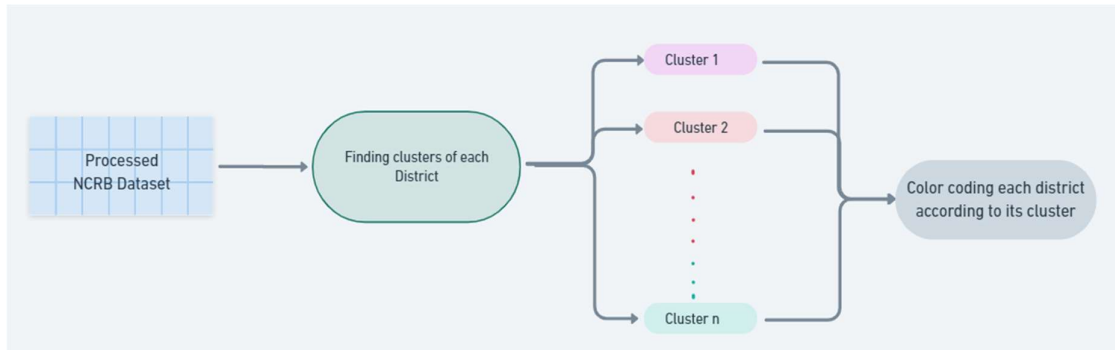


Figure 3.1

The above flowchart depicts the methodology followed for the analysis of time series crime data set.

1. Collection of datasets from the National Crime Records Bureau.
2. Preprocessing of dataset (data cleaning, data selection, data transformation)
3. Analysis of k-means using clustering tool
 - a. Identification of k using silhouette measure/elbow method.
 - b. Inputting the data into k-means clustering tool.
4. Cluster 0 to cluster 4 obtained using k-means
5. Analysis of clusters obtained using k-means and case study of crime at various locations.

3.2. Algorithm and Process Design

Data Mining is an operation that analyses data from various perspectives and summarizes or generalizes it into useful information or relationships. It is also characterized as the process of correlation discovery or discovery of patterns among fields in relational databases. Data mining is defined by Larose as an “interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases and visualization.”

Data mining is a very illustrious and imperative method to determine any crime. Data mining is made with the addition of two words ‘data’ and ‘mining’ in which mining includes the relation between the values of data of historical and current stipulation. Data mining is being used in many different fields to better understand and visualize data while also finding patterns and trends to enhance current information. Crime data mining is being studied by federal, state, local,

commercial and academic agencies. It aims to find patterns and trends using data collected by local law enforcement. One such system was theorized in the “Crime Data Mining:

A General Framework & Some Examples” paper. In this paper they discussed the framework that they have implemented. “The framework shows relationships between data mining techniques applied in criminal and intelligence analysis and the crime types listed.” They applied several data mining techniques on criminal data using their framework, such as “Entity extraction, Clustering, Association, Sequential pattern mining and others” in order to achieve their goal.

According to their paper, different types of data mining techniques were used for different crime types in order to achieve the optimal result. Each technique is employed for a specific crime. They conducted three different data mining tasks on the COPLINK project dataset. There are also other systems that are implementing criminal data mining tasks not only for law enforcement, but also for public use such as real estate agents, home owners etc.

3.2.1. Prediction

The use of predictive analytics is to predict future outcomes based on past data. The predictive algorithm can be used in many ways to help companies gain a competitive advantage or create better products, such as medicine, finance, marketing, and military operations. In Crime analysis, prediction based on historical data is primary source to predict or find patterns in criminal activity. Based on the data collected the best option is to use a Time-series based model like ARIMA.

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

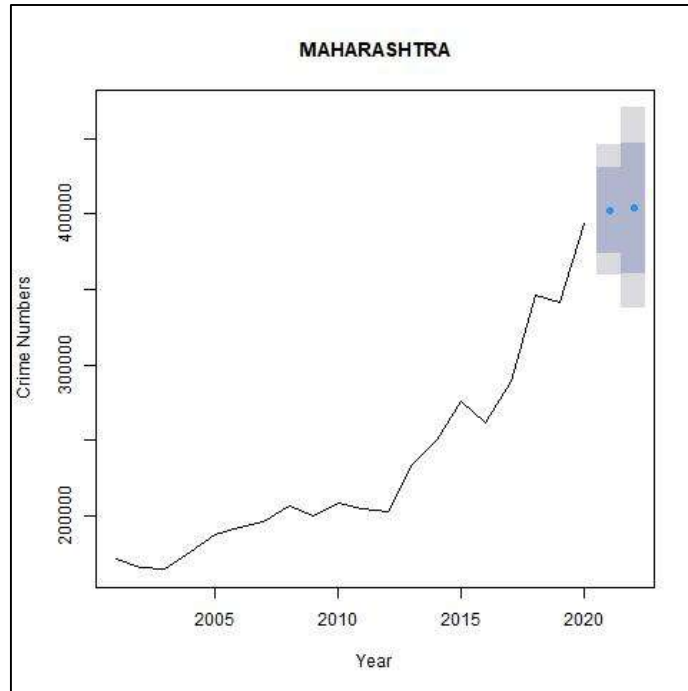


Figure 3.2

3.2.2. Clustering

In clustering, the data items are clustered according to their logical relationships or natural groupings and a structure as a whole is generated. There are no pre-defined groups, thus, clustering comes in the group of undirected Data Mining techniques. Each cluster is a collection of homogeneous elements, which may be exclusive to that group, but are similar to each other. K-means clustering is a simple clustering method that has been used in similar research.

We have to find the optimal K value for clustering the data. Now we are using the Elbow method to find the optimal K value.

```
1 from sklearn.cluster import KMeans
2
3 #create a list for the wcss parameter
4 wcss = []
5 #test with 14 clusters
6 for i in range(1, 15):
7     kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 0)
8     kmeans.fit(df1)
9     wcss.append(kmeans.inertia_)
```

Figure 3.3

The point at which the elbow shape is created is 4, that is, our K value or an optimal number of clusters is 4. Now let's train the model on the dataset with a number of clusters 4.

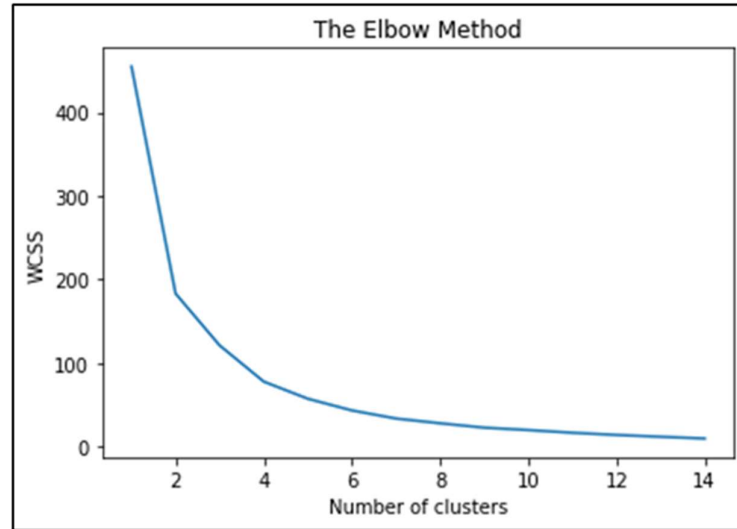


Figure 3.4

```

: #K-Means Clustering using k=6
km=KMeans(n_clusters=4)
y_kmeans=km.fit_predict(df2)

C:\Users\Patil\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(

: y_kmeans
: array([3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 3, 0, 0, 3, 0,
        0, 0, 1, 0, 0, 0, 0, 0, 3, 1, 0, 0, 0])

```

Figure 3.5

kmeans give us different clusters corresponding to X. Now let's plot all the clusters using seaborn.

```

1 km=KMeans(n_clusters=6)
2 y_pred=km.fit_predict(df1)

1 y_pred
array([5, 3, 0, 0, 0, 3, 0, 3, 3, 3, 3, 3, 0, 0, 0, 3, 2, 5, 0, 3, 5, 3,
       3, 3, 1, 3, 3, 0, 0, 3, 0, 4, 3, 3, 3])

```

Figure 3.6

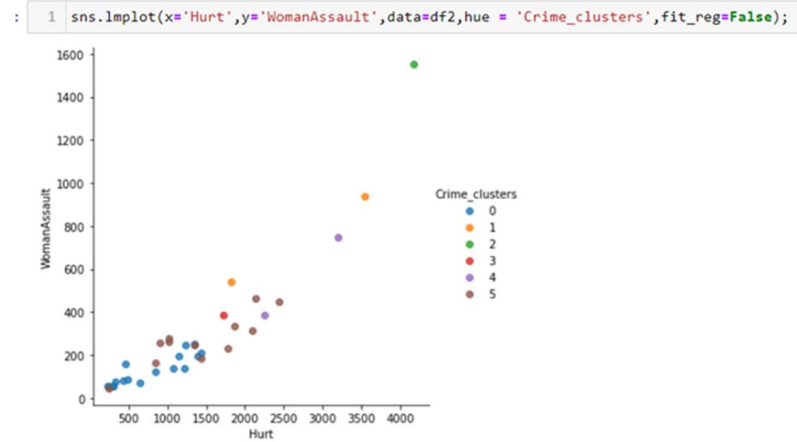


Figure 3.7

Similar to the Elbow Method as shown in Fig 3.4 Silhouette Measure is another way to evaluate the quality of clusters created using clustering algorithms such as K-Means in terms of how well samples are clustered with other samples that are similar to each other. The Silhouette score is calculated for each sample of different clusters.

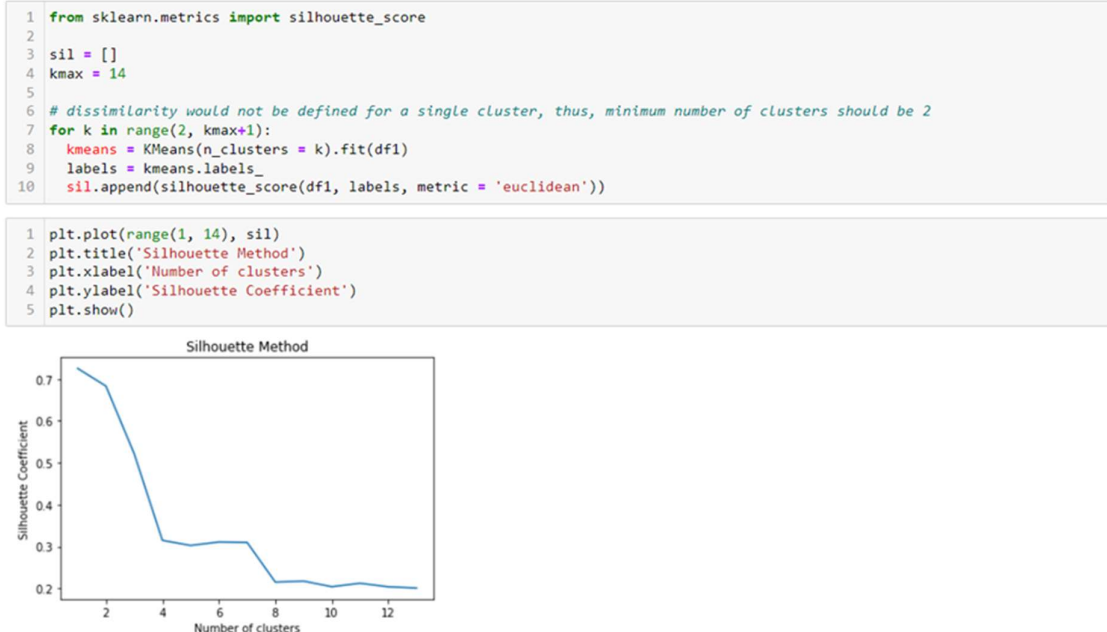


Figure 3.8

The results of clustering the datasets for each year have been added as an additional column to each year's dataset for further analysis. Each color represents

different cluster based on similarity between different crimes mentioned below. A good way for finding prominent features from which the similarities between datapoints were found is to plot graphs and visualize the behavior of each cluster. In sections where there are clear separations between different clusters and minimal space within clusters will be a prominent feature similar in a particular cluster.

3.2.3. Dataset Information

NCRB was set-up in 1986 to function as a repository of information on crime and criminals so as to assist the investigators in linking crime to the perpetrators based on the recommendations of the Tandon Committee, National Police Commission (1977-1981) and the MHA's Task force (1985).

Subsequently, NCRB was entrusted with the responsibility for monitoring, coordinating and implementing the Crime and Criminal Tracking Network & Systems (CCTNS) project in the year 2009. The project has connected 15000+ police stations and 6000 higher police offices in the country.

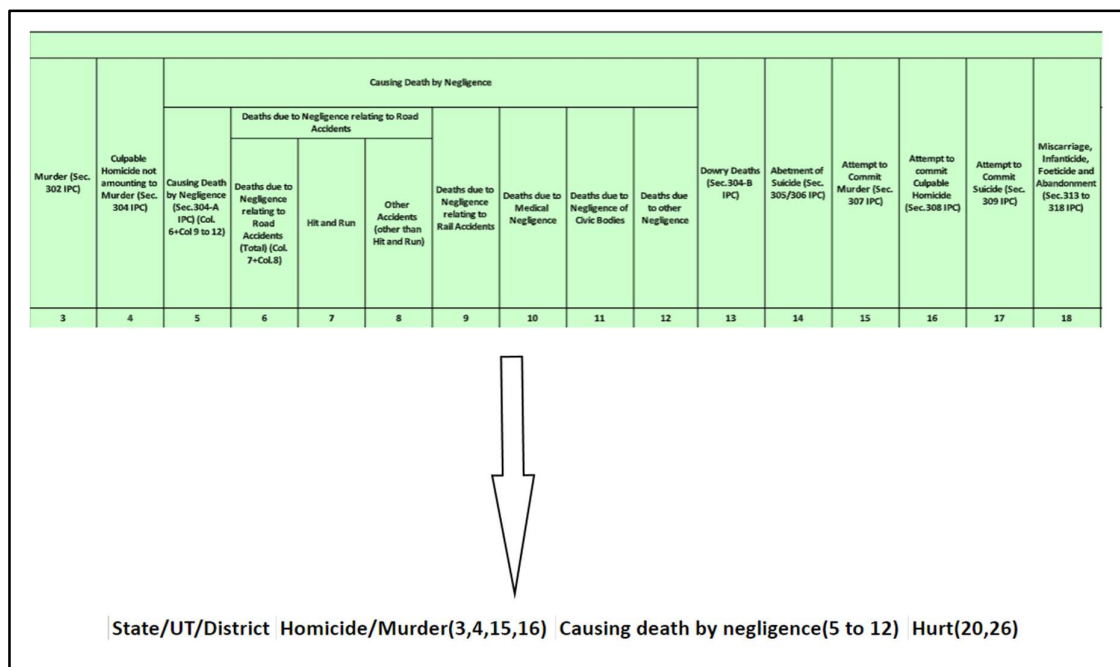


Figure 3.9

A column-oriented dataset like shown in the upper part of Fig 3.9 with several dependent variables can prove difficult for Data Analysis. Hence, we aggregate several features and rows of the dataset to form a simpler database.

3.3. Details of Hardware and Software Requirements

3.3.1. Hardware Requirements

Laptop or Desktop with minimum of intel core i5 Processor with At least 8gb of Ram and 500gb of storage.

3.3.2. Software Requirements

- Any IDE (Visual Studio Code Recommended)
- Latest version of Python (3.9)
- Python Libraries like NumPy, Pandas, Flask
- Jupyter Notebook

Chapter 4

Visualization and Results

4.1 Crime Rate Analysis (CRA) Website:

A website designed for analyzing crime statistics and viewing data via graph plotting and data mapping across a geographical area. Includes several types of charts, general statistics of the areas such as crime rate, population, total Crime, Area rankings etc. This website is designed for public as well as for commercial purpose for private organizations such as the police department. This website contains several methods of visualizing current criminal statistics over a specific area. It consists of following sections:

4.1.1 Home Page

The Home Page is the index page of our website which consists of a general information about Crime Analysis and purpose of the website.

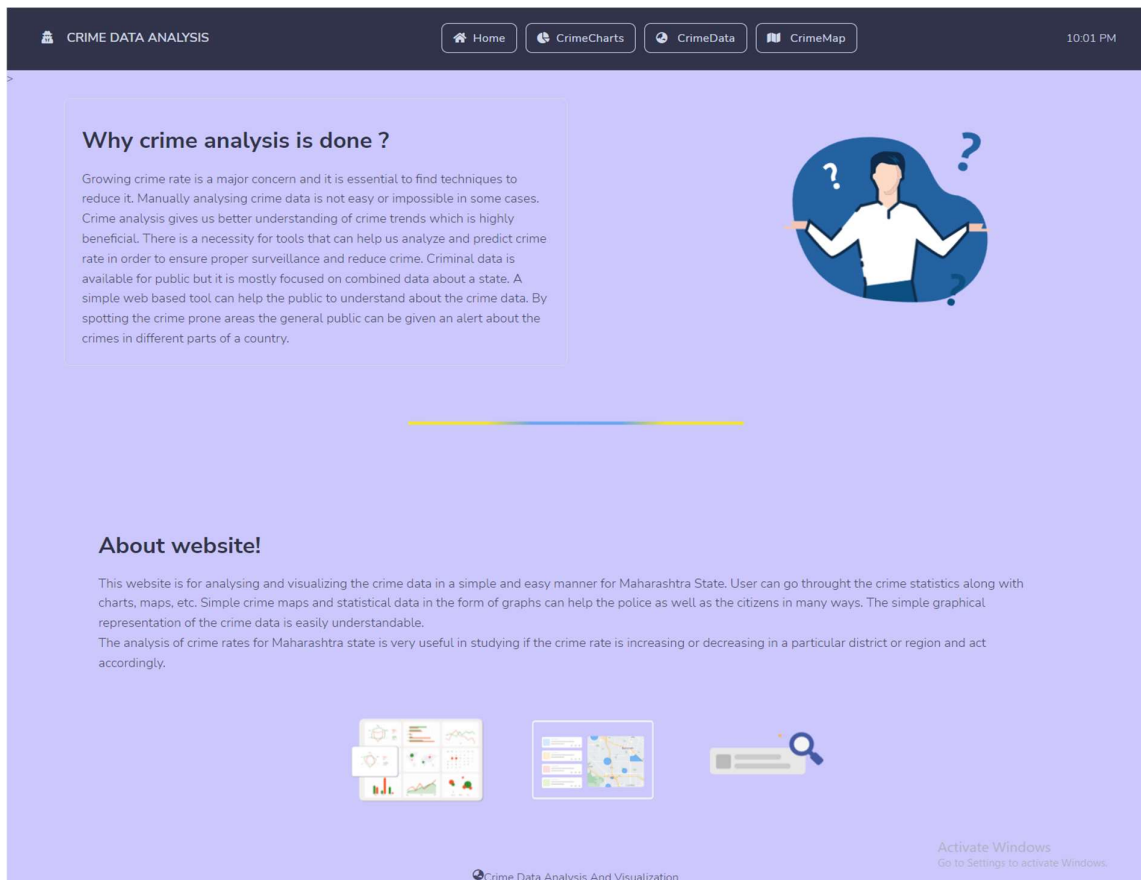


Figure 4.1

4.1.2 Crime Charts

This page is used to visualize the statistical data with the use of 3 types of charts such as bar charts for crime numbers within one year, pie charts for percentage share of a crime across different areas and Area charts for crime numbers across time.

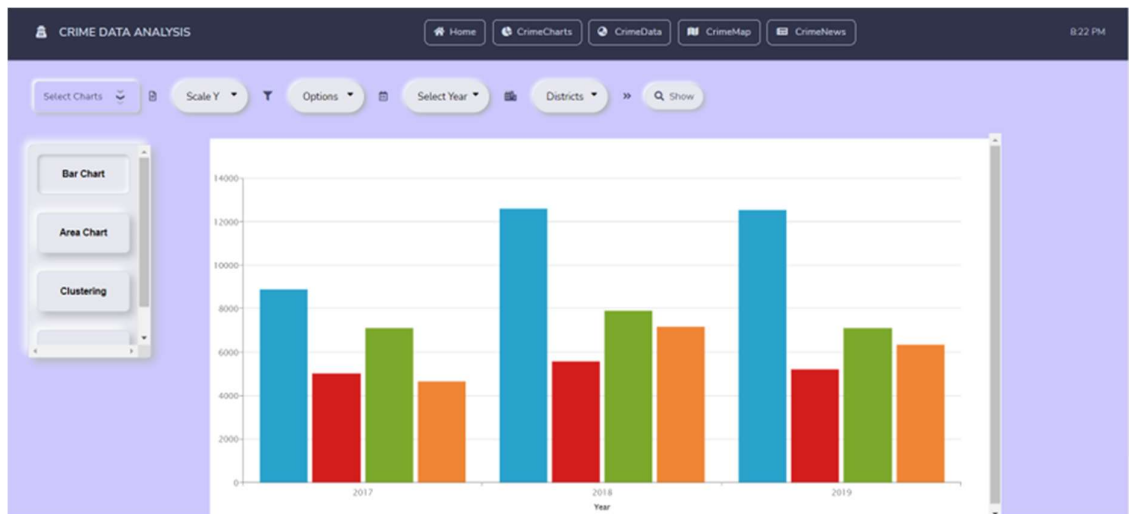


Figure 4.2

4.1.3 Crime Data

This section displays the Crime Data from the years 2017-2020 of the Maharashtra State. The search bar takes the input from the user as district name or pin code. The Crime Data is represented in a tabular format. It displays the name of the place i.e., District and all the different types of crimes and their statistical data. A map is displayed here showing the region for which the crime data is searched.

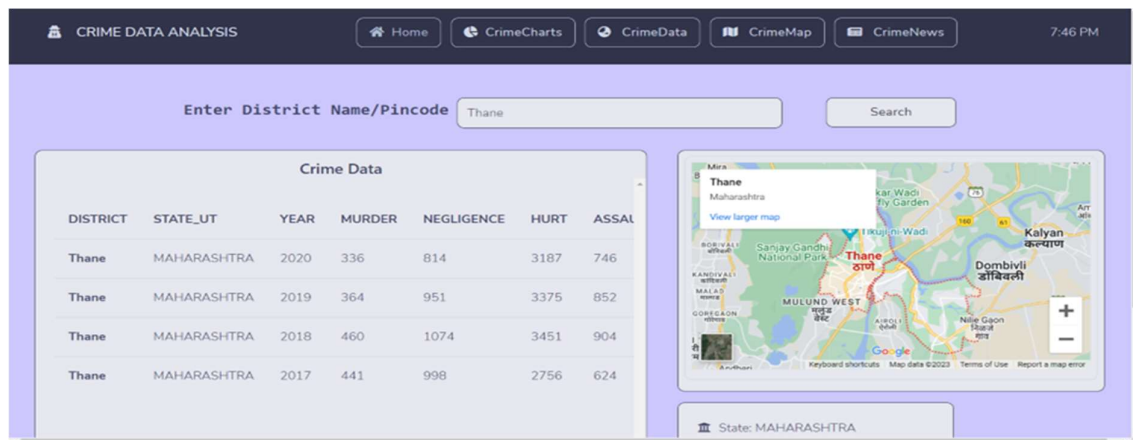


Figure 4.3

4.1.4 Crime Maps

It displays the detailed information about the Crime Rate Analysis of the Maharashtra State. For the crime maps SVG file is used. The Crime Data is displayed district wise in this section. There is a drop-down list which has a list of years. After selecting a particular year, a district is selected by clicking on the district in the map.

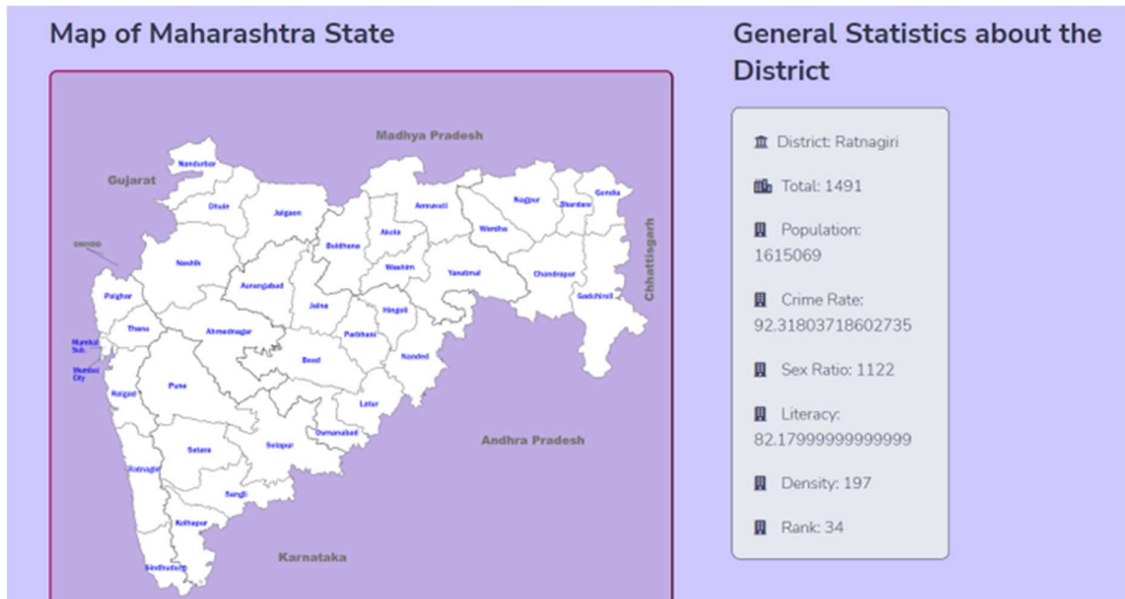


Figure 4.4

Heat maps are used to display the crime data. Different colors are used to denote the crime rate in each region. Red denotes maximum crime rate.

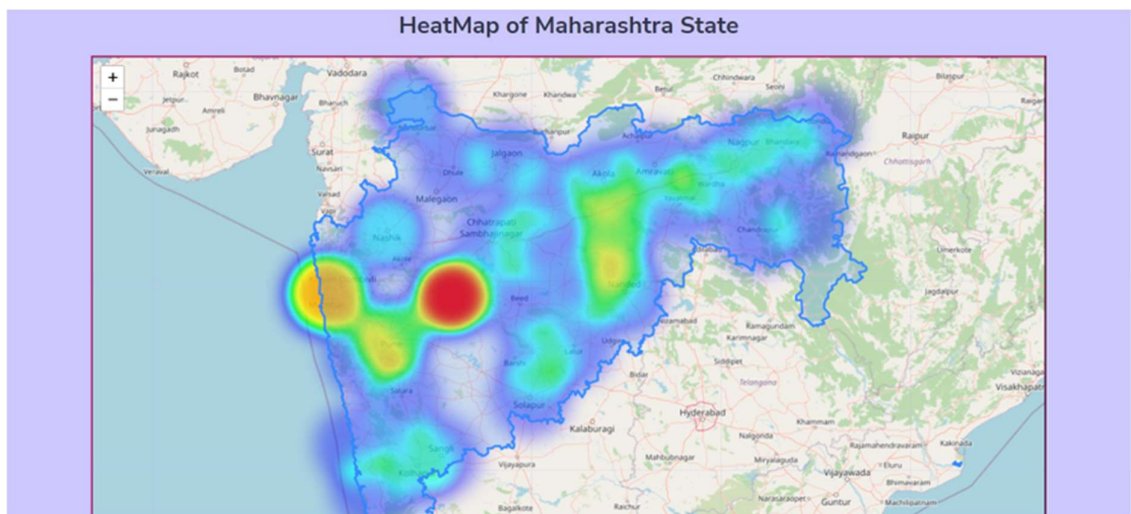


Figure 4.5

4.1.5 Crime News

In this Section a free API called News API is used to display current news across Maharashtra collected from several different News platforms. User can search a particular by typing the keyword in the search bar.

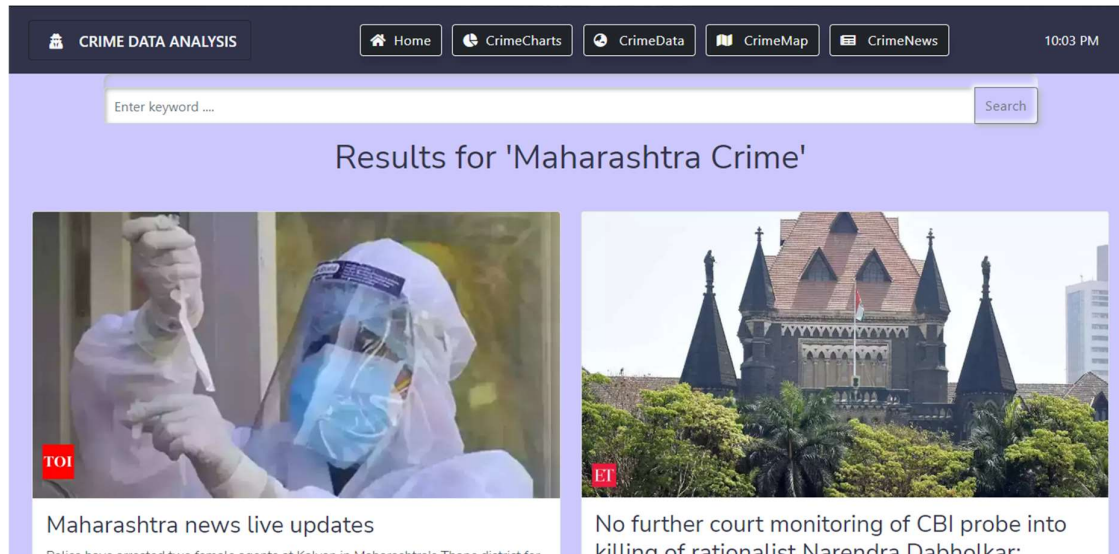


Figure 4.6

4.2 Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

Visualization of the crime data is done through plotting Graphs in two methods:

- First Method: Using metrics which have a significant impact on the Crime Rate of the society.
- Second Method: Visualizing Crime Against Crime.

4.2.1 First Method

Historically, the causes and origins of crime have been the subjects of investigation by many disciplines. Some factors that are known to affect the volume and type of crime occurring from place to place are:

- Population density and degree of urbanization.
- Variations in composition of the population, particularly youth concentration.
- Stability of the population with respect to residents' mobility, commuting patterns, and transient factors.
- Modes of transportation and highway system.
- Economic conditions, including median income, poverty level, and job availability.
- Cultural factors and educational, recreational, and religious characteristics.
- Family conditions with respect to divorce and family cohesiveness.
- Climate.
- Effective strength of law enforcement agencies.
- Administrative and investigative emphases of law enforcement.
- Policies of other components of the criminal justice system (i.e., prosecutorial, judicial, correctional, and probational).
- Citizen's attitudes toward crime.
- Crime reporting practices of citizens

For Visualizing the crime rate changes, we have chosen few of the categories mentioned above:

- **Population:**

Population and Crime rate have a linear relationship between each other as the density of the society increase the need of self-sustenance increases. A densely populated city has a higher rate growth than the other locations which signifies that a person driven by their own growth can easily be swayed to commit a crime due to the increase in competition, living conditions, individuals struggle and many other factors.

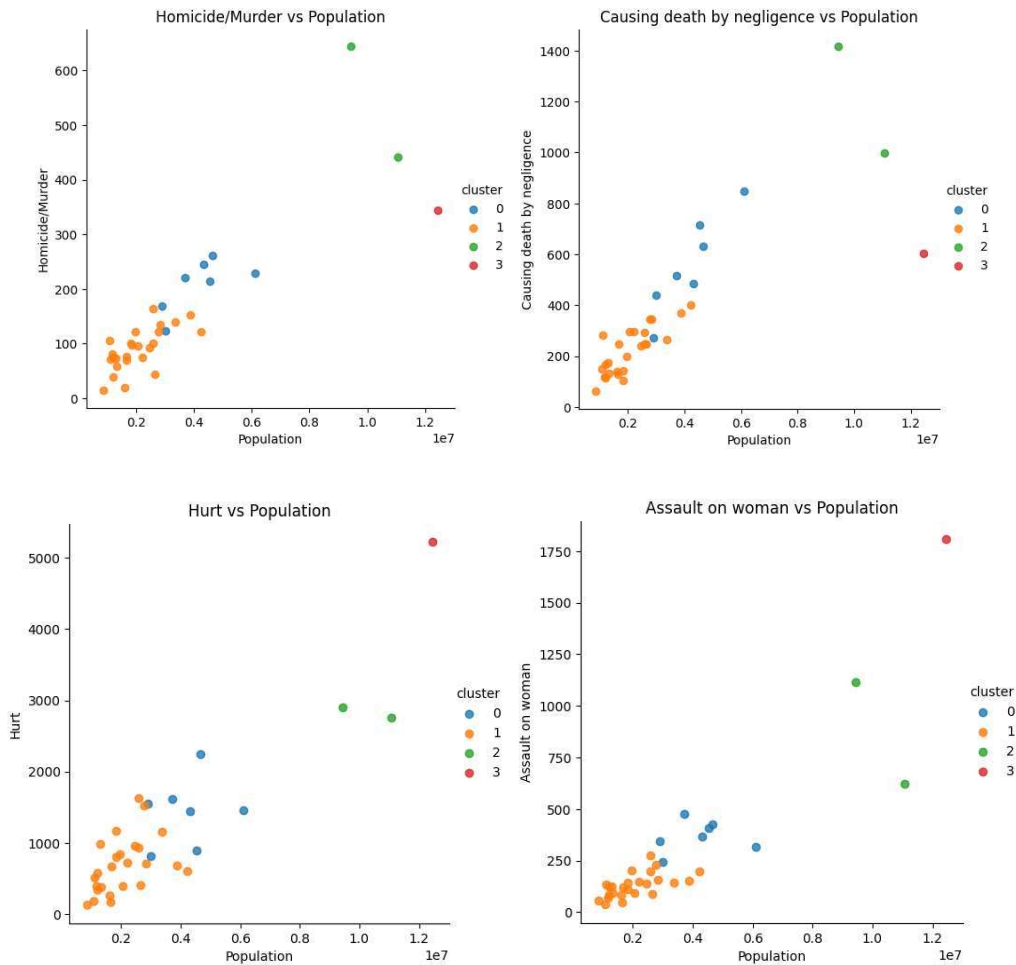


Figure 4.7

- **Sex ratio:**

Recent evidence suggests that a higher male sex ratio has a positive effect on reported crime (Dreze & Khera, 2000). Imbalances in sex ratios are found to have accounted for a one seventh increase in crime in China (Edlund, Li, Yi & Zhang, 2013). In India, South, Trent and Bose (2014) estimate an association between male-female sex ratio and selfreported victimization cases of theft, breaking, entering and assault. The findings also suggest that a higher male sex ratio is associated with frequent harassment of unmarried females. Using district level data from India, spanning across five census years since 1961, we study if an improvement in the sex ratio in favor of women, 1 i.e. an increase in the number of women per 1000 men, causes a reduction in crime.

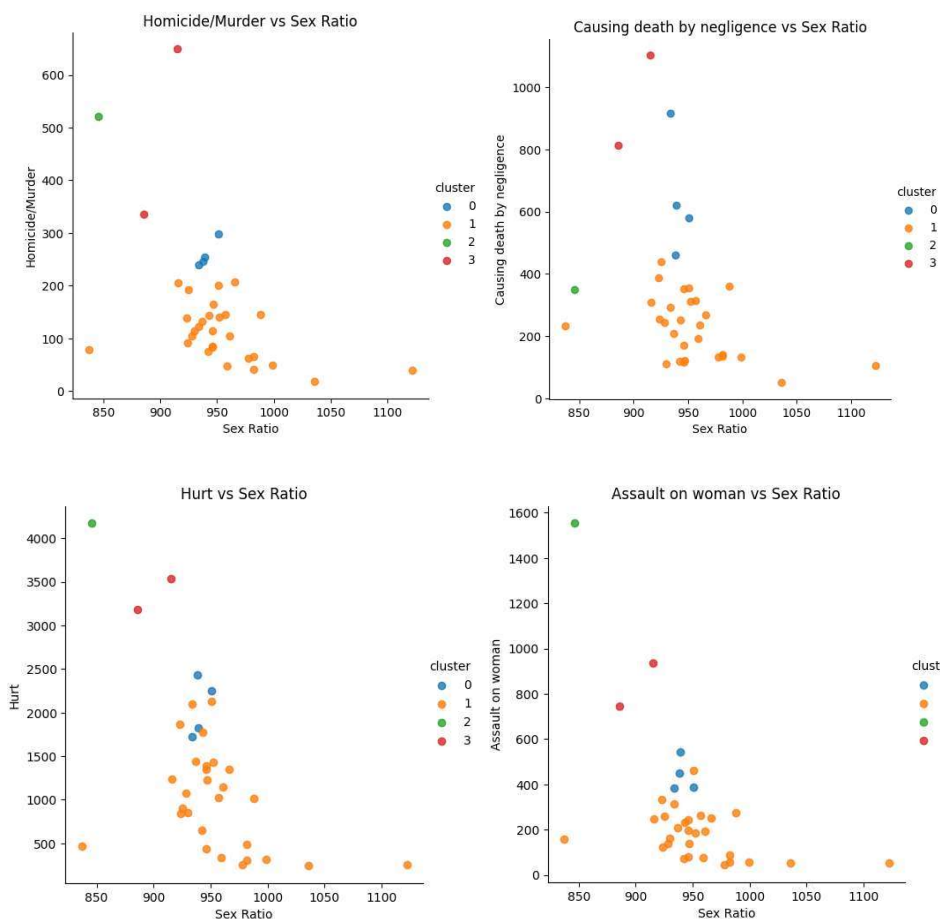


Figure 4.8

- **Literacy**

It is expected that more the literacy rate of a state, the less the crime would be. Though crime has more to do with psychological factors, a negative relationship can be expected in literacy rate and crime rate. It transcends race, gender, age, and is perhaps the strongest common thread between the majority of criminals: a lack of literacy. Some argue this lack of literacy leads to high rates of recidivism, and increased crime instead of being able to help rehabilitate an offender. Recidivism, or the tendency of a criminal to reoffend or relapse into crime, places a huge burden on the Criminal Justice System.

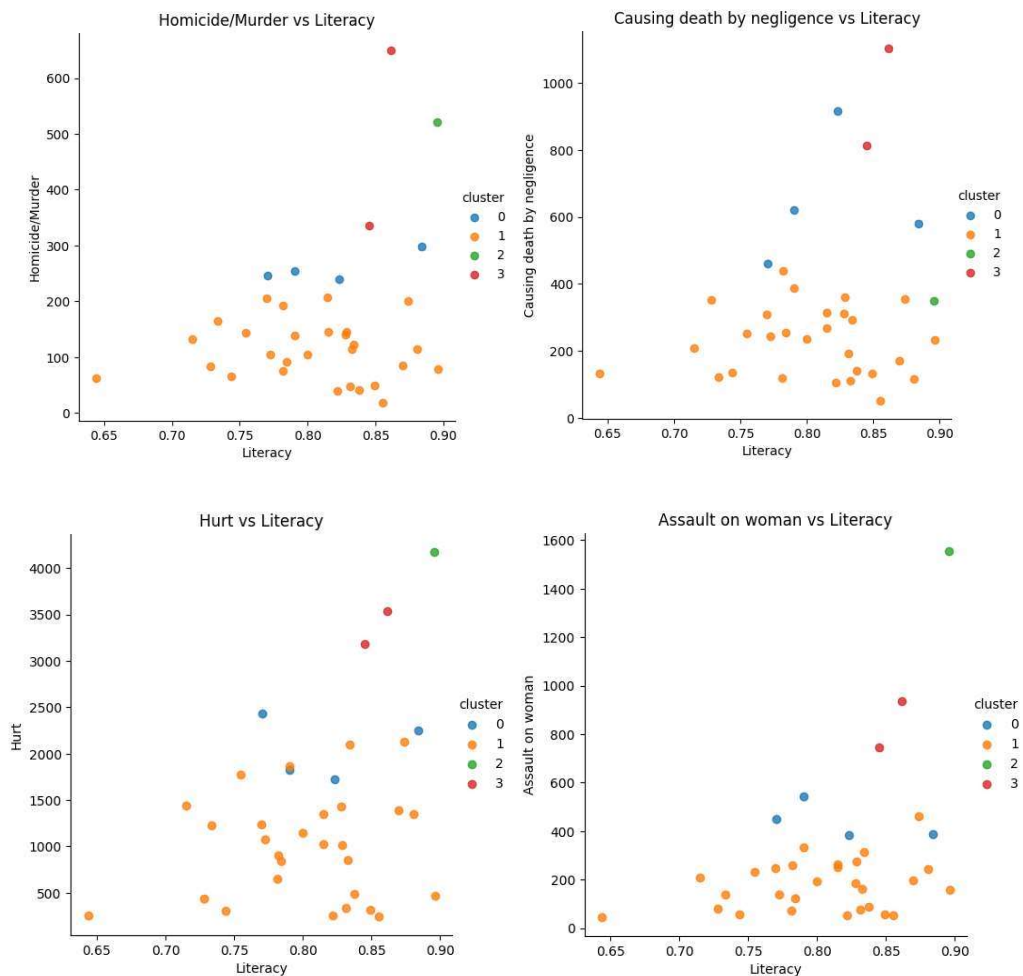


Figure 4.9

4.2.2 Second Method

The Figure below shows a cluster correlation between different crimes which indicates the interrelation associated with one crime towards the other.

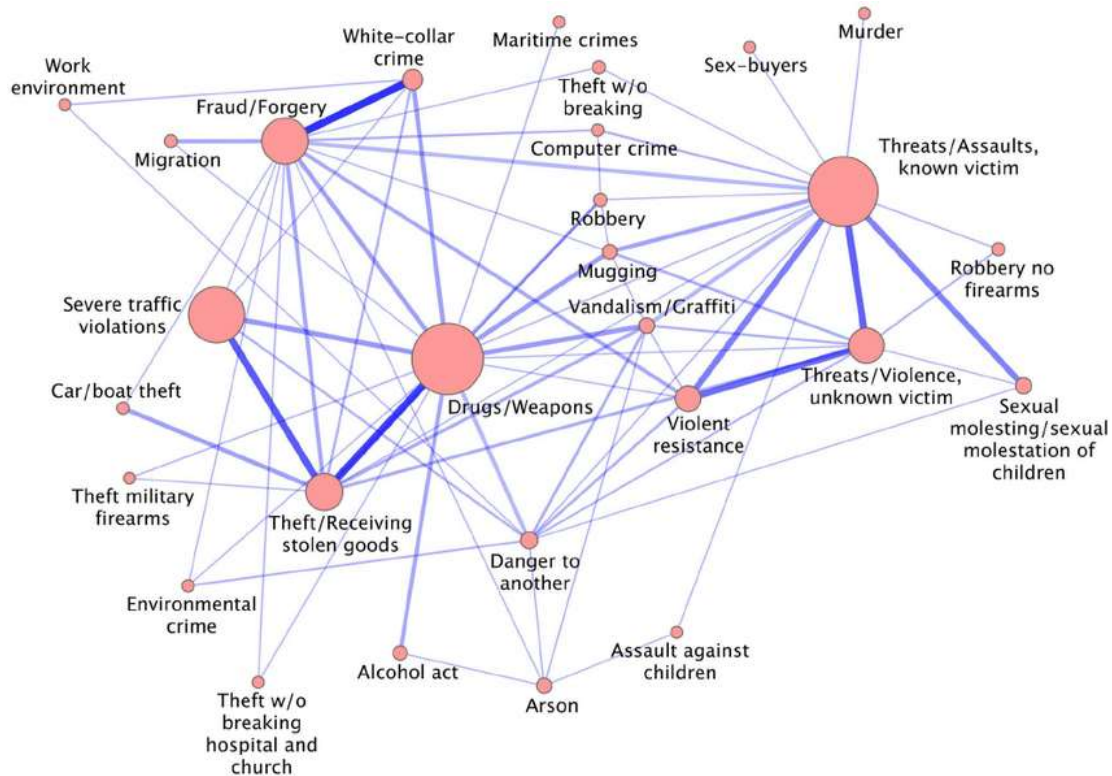


Figure 4.10

A tight connection is observed between the cluster Threats/Assaults, known victim and Sexual molestation of children/Sexual molestation, indicating the presence of a pattern that relates types of crimes of assault in which the suspect is acquainted with the victim to sexual offenses against underage victims. Cluster Threats/Assaults, known victim is also strongly connected to the cluster Violent resistance, which is the cluster of violence against public servants, and cluster Violent resistance, in turn, is connected with cluster Threats/Violence, unknown victim, which includes types of crimes of assault in which the suspect is not acquainted with the victim. The loop is finally closed by the connection between cluster Threats/Violence, unknown victim, and the cluster Threats/Assaults, known victim. An expected connection is observed between the cluster Fraud/Forgery and the cluster White collar crime. Another interesting loop cycle is the one that involves the clusters Drugs/Weapons, Theft and Severe traffic violations.

Crime against crime visualization can supply correlations between different crimes. Here are some of the results

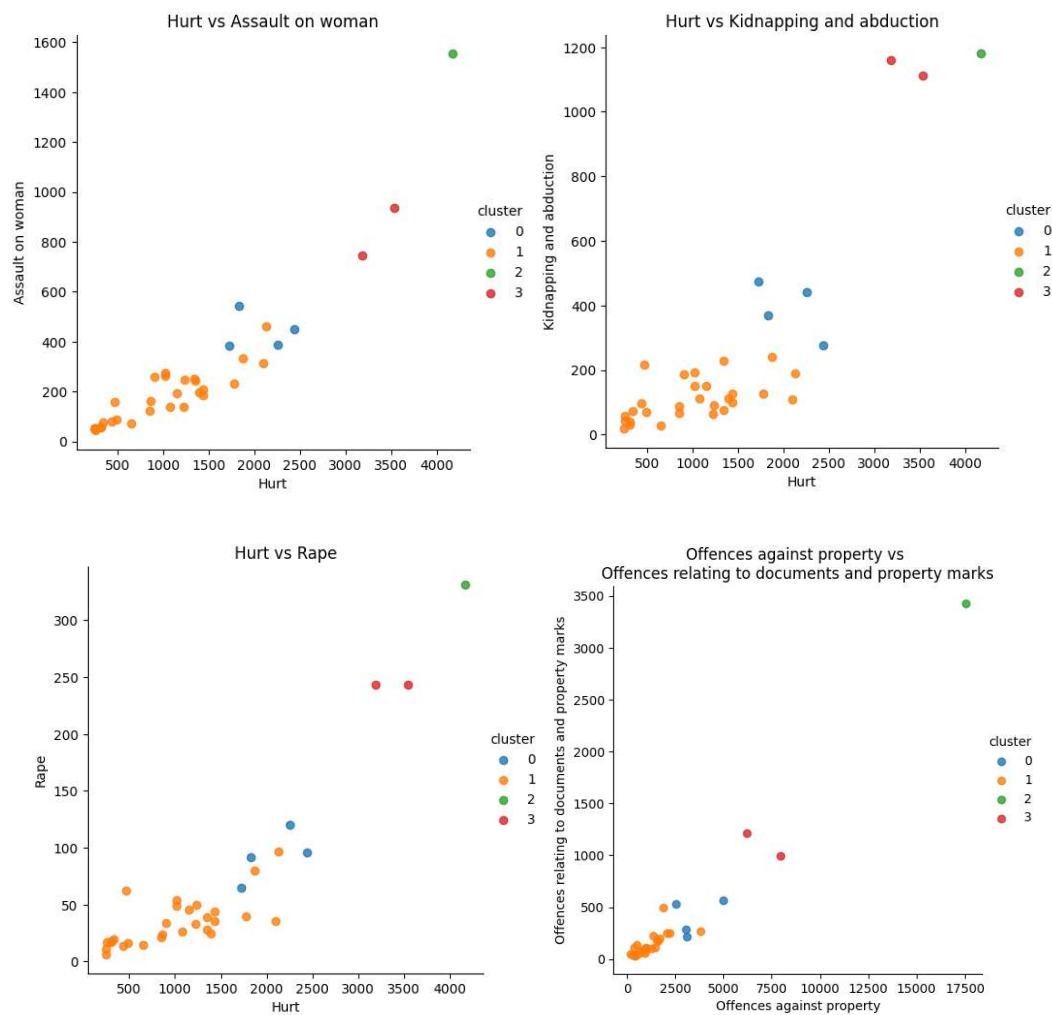


Figure 4.11

4.2.3 Results

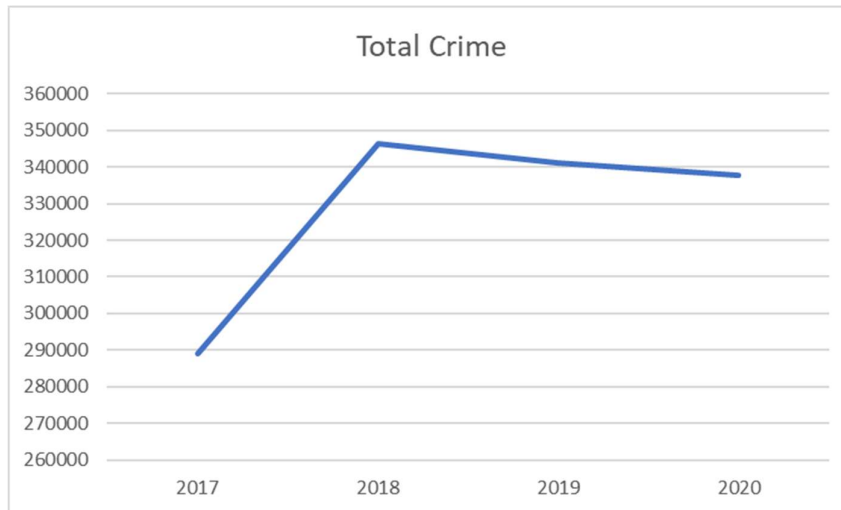


Figure 4.12

As it seen the total crime has exponentially increased from 2017-2018 after which there is a steady decline. This decline is majorly caused due the COVID-19 pandemic but apparently the crimes numbers are still on the rise. Despite the steady increase over these years the rate at which crime is occurring is reduced.

- India crime rate & statistics for 2020 was 2.95, a 0.56% decline from 2019.
- India crime rate & statistics for 2019 was 2.96, a 2.23% decline from 2018.
- India crime rate & statistics for 2018 was 3.03, a 1.11% decline from 2017.
- India crime rate & statistics for 2017 was 3.06, a 4.1% decline from 2016.

There are many possible reasons majorly being the rise in literacy and technological advancements implemented in the society. Health and lifestyle improvement is still a concern but it's improving as the attention towards poverty and hunger is increased. There are many Ways to reduce crime which involve different approaches, each of which focuses on technology and lifestyle improvement. Ultimately the urge of committing a crime comes down towards individual needs.

Chapter 4

References

- [1] ‘Crime Mapping in India: A GIS Implementation in Chennai City Policing’.
- [2] By K. Jaishankar, S. SHanmugapriya, and V. Balmurugan, June 2004.
- [3] ‘Analysis and mapping of crime perception: A quantitative approach of sketch maps’ by Mariana Vallejo Velazquez , Ourania Kounadi, and Andrea Podor.
- [4] ‘Crime Prediction and Analysis Using Machine Learning’ Alkesh Bharati1, Dr Sarvanaguru, 09 Sep 2018.
- [5] Shivani Dandir(2019).Indian Crime Data Analysis.
<https://towardsdatascience.com/indian-crime-data-analysis-85d3afdc0ceb>.
- [6] Tapan Chakraborty.An Alternative to Crime Trend Analysis in India.
<https://www.satp.org/satporgtp/publication/faultlines/volume14/article5.htm>
- [7] Vikramaditya Singh Bhatia,Sourabh Tiwarib, Jasvant Mandloic. Machine Learning Model to Predict, Classify and Analyze Crime in Indore City.
- [8] Prajakta Yerpude and Vaishnavi Gudur, “Predictive Modelling of Crime Dataset Using Data Mining”, International Journal of Data Mining & Knowledge Management Process (IJDKP), Vol.7, No.4, July 2017, DOI: 10.5121/ijdkp.2017.7404.
- [9] Skogan, W. G. (1984). Reporting crimes to the police: The status of world research. Journal of Research in Crime and Delinquency, 21, 113–137
- [10] Verma, A. (2000). Lies, damn lies and police statistics. Indian Police Journal, 2, 29–36. S N Mishra, Indian Penal Code 233 (Central Law Publication 19th ed. 2013).
- [11] Chaudhuri, K., Chowdhury, P., Reilly, K., (2013). A new perspective on violent crime burden index: evidence from Indian districts. Social Indicators Research. 110, 771-789.