Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 9179 - 9191

Research Article

Implementation of Exploratory Data Analysis(EDA) Concepts on Crime Data With Juvenile Delinquency

Dr. Praveen Kumar Misra,

Department of Mathematics & Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, praveenkumarmisra@gmail.com

Dr. Anuradha Misra*,

Amity School of Engineering & Technology, Amity University, Lucknow Campus, amisra@lko.amity.edu,

ORC ID: 0000-0002-3790-8798

Shivanshi Yadav,

Department of Mathematics & Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow

Abstract:
Juvenile delinquency is a serious offence and it is detrimental for the social order in any country. There is a trend of increase in juvenile crimes world-over, with more and more involvement of the youth in violent crimes. India shows similar trends of increasing rate of violent crimes committed by the juveniles. It is a very serious concern for the nation and solutions to end the problem need to be sought very carefully. Indian legal system and judiciary has responded to these trends and has brought some amendments in the laws pertaining to juvenile justice in India. This paper aims at looking at the causes of juvenile delinquency and explanations given by scholars from various fields to explain the problem. The analysis of statistical data available at official sites indicates increasing involvement of the juveniles in heinous crimes. To contain the problem of juvenile delinquency in India, the Act pertaining to Juvenile Delinquency has been amended and now trial of juveniles involved in heinous crimes is held as adults.
Keywords: Juvenile Delinquency, EDA, Classification

I. Introduction to Crime

Children are the rock of any nation on which its future is built. They become the leaders of the country, the creators of national wealth, who care for and protect the human community of the land to which they are rooted. These children across the world develop at different rate and develop different world- view. They increase their ability to think abstractly and develop their own views regarding social and political issues. They develop ability to indulge in long- term – planning and goal setting. There is also a tendency of making comparison of self with others. They yearn for separate identity and independence from

parents. This is the age when peer influence and acceptance becomes very important. They also develop strong romantic/ sexual ideas, and tend to show indulgence in Love and long- term relationships.
However, these are normal changes and there are no anomalies generally. Problems arise when these juveniles develop delinquent tendencies, and get into law and order problems. There occurs to be a very strong relationship of crime/ deviance with age- according to Hirschi and Gottfredson (1983), the age-crime relationship is universal.

Under the Indian Laws, Section 2 (k) of the Juvenile Justice (Care and Protection of Children) Act, 2015 (referred generally as JJ Act) juvenile is a person who below 16 years. Prior to JJ Act of 2015, the age bar for juveniles was 18 years (Juvenile Justice (Care and Protection of Children) Act, 2000, 2006, 2012). In fact, the age of the juvenile under the Indian legislations has taken variation in temporal and spatial perspectives. It varies from 14 to 18 years under different laws and different Indian states[1].

Data mining is a powerful tool that enables criminal investigators who may lack extensive training as data analysts to explore large databases quickly and efficiently[2]. Computers can process thousands of instructions in seconds, saving precious time. In addition, installing and running software often costs less than hiring and training personnel. Computers are also less prone to errors than human investigators, especially those who work long hours. We present a general framework for crime data mining that draws on experience gained with the Copylink project (http://ai.bpa.arizona.edu/coplink), which researchers at the University of Arizona have been conducting in collaboration with the Tucson and Phoenix police departments since 1997[3][4][10].

II. Who is a Juvenile?

Juvenile means anyone who has not yet reached the age of adults in terms of childishness or immaturity. In the Legal sense, a juvenile can be defined as a child who has not attained a certain age at which he can be held liable for his criminal acts like an adult person under the law of the country. "Juvenile" or Child is a person who has not completed eighteen years of age.

III. Juvenile Delinquency

William Coxton in the year 1484 used the word delinquent to refer a person who was found guilty. Juvenile delinquency means the involvement by the teenagers in an unlawful behavior who is basically under the age of 18 and commits an act which is considered as a crime. A child is known as a delinquent when he/she perpetrate a mistake which is against the law and is not accepted by the society. A child is known as a delinquent when he/she commits a mistake which is against the law and which is not accepted by the society.

IV. Juvenile Delinquency in India- Current Trends

The legal definition of child affects how the courts in a country deal with offenders. As per the international norms, and also under the juvenile Justice System in India, a minor or a child cannot be tried in the same manner as an adult. A child is treated as doliincapax, with no mensrea- he/ she is not capable of understanding consequences of his/ her actions[5].

Keeping this logic in mind, children are dealt under juvenile justice system, and not under the adult criminal justice system. They can never be given imprisonment or death penalty. Hence, under the Indian legal system, Art. 40 (3) (a) of CRC requires State Parties to promote establishment of minimum age below which child is presumed not to have capacity to deviate the penal law. Age of criminal responsibility is held to be 7 years- child below 7 years cannot be considered a child in conflict with law – section 82 of IPC, 1860. Thus, nothing is an offence done by a child between 7 and 12 years, who has not attained sufficient maturity to judge the nature and consequences of his/ her conduct and did not know that what she was doing was wrong - Section 83 of IPC, 1860.

V. DATA DESCRIPTION

The source of the data is **"Rajkiya Balgrah (Balak)"** Lucknow. We have considered two groups of children with respect to age, first group consist of children below 16 years and second group consist of children of age 16 to 18 years old. We have considered three places (lko, unnao, raibareli) in Lucknow. We define six crime(variable) which are as follows:

1. **Petty Crime** is a type of crime that is not considered serious when compared with some other crime: such as shoplifting.
2. **Serious Crime** such as murder aggravated assault various forms of fraud.
3. **Heinous Crime** is the crimes which are depending on the jurisdiction violent crimes may include homicide, murder, assault, rape etc. A heinous crime is very evil or wicked.
4. **POCSO (Protection of children from sexual offences)** the act define different forms of sexual abuse which include penetrative and non-penetrative assault. It also involves sexual harassment, pornography etc.
5. **Under IPC (Indian penal code) 302:** Section 302 of IPC deals with punishment for murder. It states the following "whoever commits murder shall be punished death or imprisonment for life and shall also be liable to fine".
6. **Under IPC (Indian penal code) 376:** Whoever commit the rape, shall be punished with rigorous imprisonment of either description for a term which shall not be less than 7 years, bail which extend to imprisonment for life, and shall be liable to fine.[6][7][8]

Month	District	No. of children below 16	No. of children age 16-18	Inhabitant with petty offence	Inhabitant with serious offence	Inhabitant with heinous offence	Under POCSO	Under IPC 302	Under 376 IPC
Aug-17	Lko	20	21	28	2	22	7	3	6
Aug-17	Raibarelly	4	6	3	1	8	4	2	3
Aug-17	Unnao	5	22	7	2	24	14	1	13
Sep-17	Lko	15	16	20	9	19	11	7	8
Sep-17	Raibarelly	4	6	3	1	7	3	2	3

Sep-17	Unnao	11	14	7	3	17	10	2	11
Oct-17	Lko	17	18	13	7	22	14	9	10
Oct-17	Raibarelly	5	6	2	1	9	4	2	5
Oct-17	Unnao	15	10	6	3	19	13	2	13
Nov-17	Lko	16	15	9	4	27	15	9	13
Nov-17	Raibarelly	6	5	4	1	10	6	2	4
Nov-17	Unnao	3	9	5	4	15	12	1	11
Dec-17	Lko	18	16	20	7	26	13	17	10
Dec-17	Raibarelly	6	7	4	3	7	15	14	4
Dec-17	Unnao	10	6	4	1	13	6	10	0
Feb-18	Lko	23	18	21	15	16	14	14	2
Feb-18	Raibarelly	5	3	1	4	3	3	3	1
Feb-18	Unnao	13	5	7	6	5	5	5	1
Apr-18	Lko	22	28	15	15	16	16	15	16
Apr-18	Raibarelly	2	3	0	3	1	2	3	4
Apr-18	Unnao	9	7	4	0	12	5	2	6
May-18	Lko	23	18	21	15	16	14	14	2
May-18	Raibarelly	5	3	1	4	3	3	3	1
May-18	Unnao	13	5	7	6	5	5	5	1
Jun-18	Lko	28	18	29	17	19	17	10	2
Jun-18	Raibarelly	5	3	0	4	5	5	3	0
Jun-18	Unnao	14	8	9	5	11	10	1	1
Jul-18	Lko	36	22	36	14	28	23	12	3
Jul-18	Raibarelly	6	4	0	5	7	6	3	0
Jul-18	Unnao	20	9	13	17	13	12	1	1

Aug-18	Lko	43	27	32	12	26	20	12	6
Aug-18	Raibarelly	7	1	3	4	11	4	4	7
Aug-18	Unnao	15	12	2	4	16	16	4	2
Sep-18	Lko	55	15	20	20	32	32	6	1
Sep-18	Raibarelly	10	15	3	12	12	8	6	1
Sep-18	Unnao	21	16	7	8	19	19	7	1
Oct-18	Lko	45	30	30	32	13	32	6	1
Oct-18	Raibarelly	9	15	6	12	6	8	6	1
Oct-18	Unnao	15	19	6	19	9	19	7	1
Nov-18	Lko	45	30	30	13	32	32	6	1
Nov-18	Raibarelly	9	15	6	6	12	8	6	1
Nov-18	Unnao	15	19	6	9	19	19	7	1
Dec-18	Lko	55	15	20	20	28	20	8	1
Dec-18	Raibarelly	10	15	3	12	10	10	6	0
Dec-18	Unnao	21	16	7	8	22	16	6	0
Jan-19	Lko	50	21	26	16	29	29	8	0
Jan-19	Raibarelly	12	8	2	7	11	11	3	1
Jan-19	Unnao	27	13	7	9	24	22	6	1

VI. Classification of the data

Measurement lies at heart of statistics. Indeed, no statistic would be possible without the concept of measurement. Measurement is also an integral part of our everyday lives. We routinely classify and assign values to people and objects without giving much thought to the processes that underlie our decisions and evaluations. In statistics, such classification and ordering of values must be done in a systematic way. There are clear rules for developing different types of measures and defined criteria for deciding which are most appropriate for answering a specific research question[9S].

VII. Science and Measurement

Classification

Criminal justice research is a scientific enterprise that seeks to develop knowledge about the nature of crimes, criminals, and the criminal justice system. The development of knowledge can, of course, be carried out in a number of different ways.

Measurement in science begins with the activity of distinguishing groups or phenomena from one another. This process, which is generally termed **classification,** implies that we can place units of scientific study such as victims, offenders, crimes, or crime places in clearly defined categories. The classification process leads to the creation of **variables**. A variable is a trait, characteristic, or attribute that can be measured. What differentiates measurement in science from measurement in our everyday lives is that there must be systematic criteria for determining both what each category of a variable represents and the boundaries between categories. We now turn to a discussion of these criteria as they relate to different **levels of measurement.** There are a number of different ways we can classify the people, places, or phenomena we wish to study. We may be content to simply distinguish one category from another. But we may also be interested in how those categories relate to one another. Do some represent more serious crime or less serious crime?

Can we rank how serious various crimes are in a clear and defined order?

Is it possible to define exactly how serious one crime is relative to another?

Recognizing this complexity, statisticians have defined four basic groups of measures, or **scales of measurement,** based on the amount of information that each takes advantage of. The four are generally seen as occupying different positions, or levels.

Nominal Scales

Nominal scale variables simply distinguish one phenomenon from another. Suppose, for example, that you want to measure crime types. In your study, you are most interested in distinguishing between violent crime and other types of crime. To fulfill the requirements of a nominal scale, and thus the minimum requirements of measurement, you need to be able to take all of the crime events in your study and place them in one of two categories: either violent crime or other crime. There can be no overlap.

Nominal-Scale Variables Commonly Found in Criminal Justice Research

Variable	**Common Categories**
Gender	Male, Female
Race -Ethnicity	Non-Hispanic Black, Non-Hispanic White, Hispanic (any race)
Marital Status	Single, Married, Separated, Divorced, Widowed
Pretrial Release Status	Detained, Released
Type of Case Disposition	Dismissed, Acquitted, Diverted, Convicted
Method of Conviction	Negotiated guilty plea, Nonnegotiated guilty plea, Bench trial, Jury trial
Type of Punishment	Incarceration, Nonincarceration

Ordinal Scales

What distinguishes an ordinal from a nominal scale is the fact that we assign a clear order to the categories included. Now not only can we distinguish between one category and another; we also can place these categories on a continuum. This is a very important new piece of information;

it allows us to rank events and not just categorize them. In the case of crime, we might decide to rank in order of seriousness. For example, in a well-known London survey of victimization, fear of crime was

measured using a simple four-level ordinal scale. Researchers asked respondents: "Are you personally concerned about crime in London as a whole? Would you say you are

(1) very concerned
(2) quite concerned
(3) a little concerned, or
(4) not concerned at all?"

Ordinal Scale Variables Commonly Found in Criminal Justice Research

Variable Common Categories

Level of Education Less than high school, Some high school, High school graduation, Some college or trade school, College graduate, Graduate/ professional school
Severity of Injury in an Assault None, Minor—no medical attention, Minor—medical attention required, Major—medical attention required with no hospitalization, Major—medical attention required with hospitalization
Attitude and Opinion Survey Strongly disagree, Disagree, No opinion, Agree, Strongly agree;
Questions Very high, High, Moderate, Low, Very low
Bail-Release Decision Released on own recognizance, Released on bail, Detained— unable to post bail, Denied release
Type of Punishment Probation/community service, Jail incarceration, Prison incarceration, Death sentence

Interval and Ratio Scales

Interval scales not only classify and order people or events; they also define the exact differences between them. An interval scale requiresthat the intervals measured be equal for all of the categories of the scale examined. Thus, an interval-scale measure of prior record would not simply rank prior record by seriousness; it would allow us to say how much more serious one offender's record was than another's in a standard unit of measurement—for example, number of arrests, convictions, or prison stays. Most criminal justice variables that meet the criteria of an interval scale also meet the criteria of a **ratio scale**. A ratio scale has all of the characteristics of an interval scale but also requires that there be a non-arbitrary, or true, zero value. This means simply that zero represents the absence of the trait under study.

Variable Common Categories

Age Years
Education Years
Income or Salary Dollars, etc.
Number of Crimes in a
City/County State Nation Count

Crime Rates for a
City/County/State/Nation Count of crimes, adjusted for the size of the population
Self-Reported Delinquent Acts Count

There are three separate factors that affect the quality of a measure.
The researcher should strive for a measure that has (1) a high scale of measurement (one that uses the most information); (2) a high level of **validity** (one that provides an accurate reflection of the concept being studied); and (3) a high level of **reliability** (one that provides consistent results across subjects or units of study).

VIII. Representing and Displaying Data

The graphical representation of data is an important tool for presenting statistical results in such a way that the key features or characteristics of an analysis are highlighted. There are many different ways the same data might be displayed.

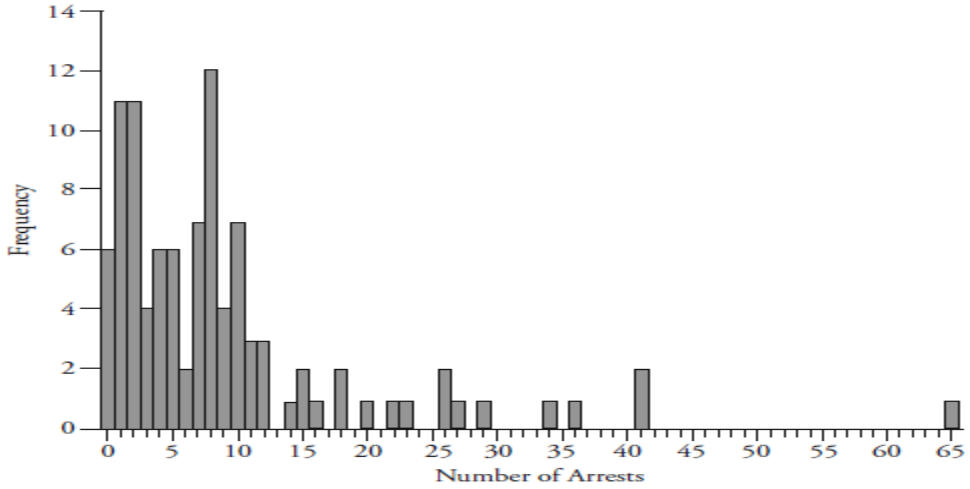
i. Frequency Distributions

When we array scores according to their value and **frequency,** we construct what is called a **frequency distribution.** Here**, frequency** means the number of times that a score or value occurs.
Let's take the following data on previous arrests of 100 known offenders as an example:14,0,34,8,7,22,12,12,2,8,6,18,1,18,8,1,10,10,2,12,26,8,7,9,9,3,2,7,16,8,65,8,2,4,2,4,0,7,2,1,2,11, 2,1,1,5,7,4,10,11,3,41,15,1,23,10,5,2,10,20,0,7,6,9,0,3,1,15,5,27
8,26,8,1,1,11,2,4,4,8,41,29,18,8,5,2,10,1,0,5,36 3,4,9,5,10,8,0,7.
We first group all of the cases with the same value together. Accordingly, we group together the cases with no prior arrests, one prior arrest, two prior arrests, and so forth, until we have covered all of the potential scores in the distribution. Then we arrange these scores in order of magnitude. Looking at the data in this way allows us to get a sense of the nature of the distribution of scores.
Frequency Distribution of Prior Arrests for 100 Known Offenders

VALUE	FREQUENCY
0	6
1	11
2	11
3	4
4	6
5	6
6	2
7	7
8	12
9	4
10	7
11	3
12	3
14	1
15	2
16	1
18	2
20	1
22	1
23	1
26	2
27	1
29	1
34	1
36	1
41	2
65	1
Total	100

ii. Histogram

Histogram means a bar graph used to represent a frequency distribution. To make a histogram, we take the scores and values from a frequency distribution and represent them in pictorial form. In this case, we use a bar to represent each value in the frequency distribution. The *x*-axis (the horizontal axis) of the histogram represents the values of the variable we are analyzing—here, the number of arrests. The *y*-axis (the vertical axis) captures the height of the bars and indicates the number of scores the frequency found in each category. A histogram of the data on prior arrests. The information presented in the histogram is identical to the information presented in the frequency distribution, but the histogram conveys to the reader an immediate sense of the range of values, the location of clusters of cases, and the overall shape of the distribution—information that is not as easily obtainable from a frequency distribution.



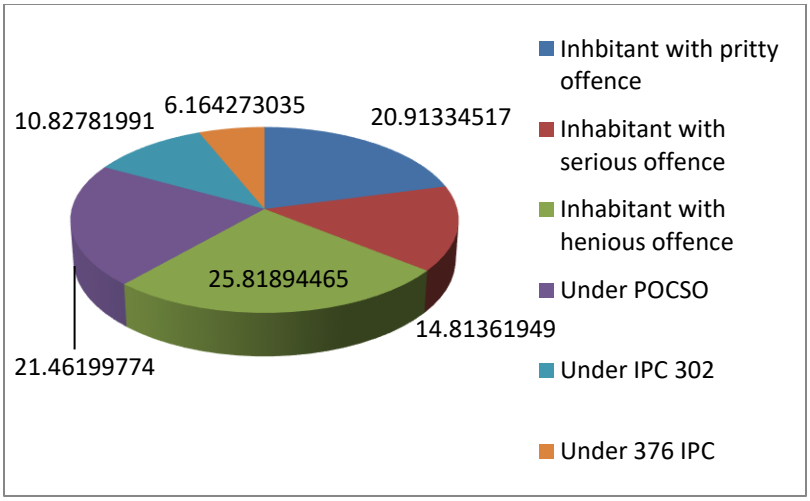
Interpretation:

According to this histogram, there is a large gap in the distribution from 41 arrests to 65 arrests. There are relatively few observations for increasingly greater numbers of arrests and that the distance between categories starts to increase. Histogram shows the right tailed distribution. If the bars were adjacent to

each other and correctly labeled, it would still be possible for the reader to discern the spread of observations. But when the values with an observed frequency of zero are portrayed as well, it is often easier to interpret the histogram.

iii. Pie chart

Pie chart means a graph in which a circle (called a pie) is cut into wedges to represent the relative size of each category's frequency count. When we present data on the percentage or proportion of cases in a pie chart, the information contained in a pie chart is identical to that presented in a bar chart.



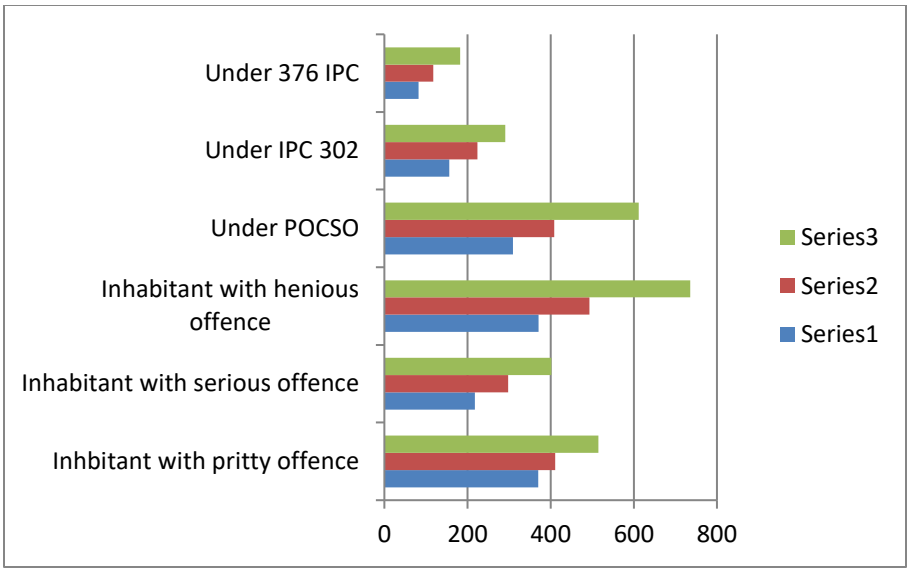
Interpretation:

Here pie chart shows the crime percent of the juvenile of Lucknow in 2018. In this graph maximum percent is 25.8% show the heinous crime. And the 6.16% is the minimum percent which shows crime under IPC 376.

iv. Bar chart

Bar chart means a graph in which bars represent frequencies, percentages, or proportions. A simple extension of the histogram to multiple groups makes use of the **bar chart.** Bar charts allow us to present information for multiple groups simultaneously. Bar charts are constructed in much the same way as histograms. The *x*-axis generally represents the values of the variable, and the *y*-axis the size of the bar for the categories or values of a variable.

Bar charts are also quite useful for visually representing nominally or ordinary measured variables.



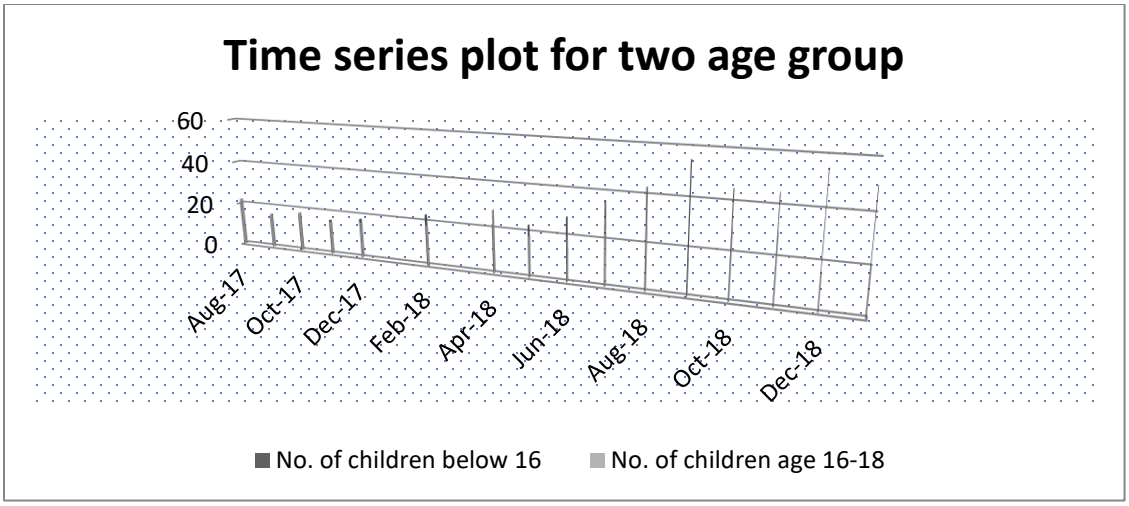
Interpretation:
In this bar graph Series1 represents crimes in Lucknow, Series2 represents crime in Raebareli and Series3 represents crimes in Unnao. We can see that the heinous crime most frequently occur in Unnao. In comparison to Unnao and Raebareli, Lucknow has the lowest crime rate and the crime which comes under IPC 376 is minimum.

Proportion
A relation between two numbers in which the whole is accorded a value of 1 and the other number is given a numerical value corresponding to its share of the whole.

v. Time Series Data

Time series data means repeated measures of the same variable over some regularly occurring time period, such as days, months, or years. Study of many important issues in crime and criminal justice requires the use of **time series data.** Time series data include measures on the same variable for the same unit of analysis at more than one point in time.

Time Series Plot
In a **time series plot,** data are graphed over time, with the measure of time along the *x*-axis. For the data, the measure of time is years, daily, weekly, monthly, or quarterly data.



Interpretation: According to the given time series plot, it is concluded that the crime rates are increasing with time i.e.., the trend line is increasing.

IX. Conclusion and Future Scope

Juvenile crime is a very serious issue of the society. It needs proper attention so that it can be controlled.

The analysis performed in this paper can be very helpful for the updating and proper maintenance of the IPC acts regarding juvenile crime. We can see that the heinous crime most frequently occur in Unnao. In comparison to Unnao and Raebareli, Lucknow has the lowest crime rate and the crime which comes under IPC 376 is minimum. Moreover, it can be seen that the crime rates are increasing with time.
Based on the analysis, we can further apply some more dimensions for the similar analysis using data mining tools to further modify the accuracy of the results. Such type of analysis are very beneficial to maintain proper law and order.

References

[1] U.M. Fayyad and R. Uthurusamy, Evolving Data Mining into Solutions for Insights, Comm. ACM, Aug. 2002, pp. 28-31.

[2] W. Chang et al., An International Perspective on Fighting Cybercrime, Proc. 1st NSF/NIJ Symp. Intelligence and Security Informatics, LNCS 2665, Springer-Verlag, 2003, pp. 379-384.

[3] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2001.

[4] H. Kargupta, K. Liu, and J. Ryan, Privacy-Sensitive Distributed Data Mining from Multi-Party Data, Proc. 1st NSF/NIJ Symp. Intelligence and Security Informatics, LNCS 2665, Springer-Verlag, 2003, pp. 336-342.

[5] M.Chau, J.J. Xu, and H. Chen, Extracting Mean- insightful Entities from Police Narrative Reports, Proc. Natl Conf. Digital Government Research, Digital Government Research Center, 2002, pp. 271-275.

[6] A. Gray, P. Sallis, and S. MacDonell, Software Forensics: Extending Authorship Analysis Techniques to Computer Programs, Proc. 3rd Biannual Conf. Intl Assoc. Forensic Linguistics, Intl Assoc. Forensic Linguistics, 1997, pp. 1-8.

[7] R.V. Hauck et al., Using Coplink to Analyze Criminal-Justice Data, Computer, Mar. 2002, pp. 30-37.
[8] T. Senator et al., The FinCEN Artificial Intelligence System: Identifying Potential Money Laundering from Reports of Large Cash Transactions, AI Magazine, vol.16, no. 4, 1995, pp. 21-39.

[9] W. Lee, S.J. Stolfo, and W. Mok, A Data Mining Framework for Building Intrusion Detection Models, Proc. 1999 IEEE Symp. Security and Privacy, IEEE CS Press, 1999, pp. 120-132.

[10]. Naila, Anuradha Sharma, 2016, "Data Mining Application in Diabetes Diagnosis using biomedical Records of Pathological Attribute", International Journal of Science and Research (IJSR),pp 1077-1083, **Volume 5 Issue 6 ,** ISSN (Online): 2319-7064