

A REVIEW WORK ON CRIME DATA ANALYSIS OF WESTERN & MIDDLE PART OF INDIA

A Project Submitted in Partial Fulfillment of the

Requirements For the Degree

Of

Bachelor of Science in

Statistics By Avisikta Das

Semester-6 Paper- DSE-B2

Registration No- 012-1212-

0625-20

Roll No- 203012-11-0235

Under the supervision of Prof. Dr Shirsendu
Mukherjee

DEPARTMENT OF STATISTICS, ASUTOSH COLLEGE , 92,
Shyama Prasad Mukherjee Road, Jatin Das Park,
Patuapara,Bhowanipore,Kolkata, West Bengal 700026
AFFILIATED TO THE UNIVERSITY OF CALCUTTA

CERTIFICATE

This is certified that the project paper entitled "A REVIEW WORK ON CRIMINAL DATA ANALYSIS ON WESTERN & MIDDLE PART OF INDIA" submitted by Avisikta Das in partial fulfillment of the requirement for the Bachelor degree of Statistics (Honours) is based upon the result of benefited research work carried out by the investigator under my guidance and supervision.

The results of the investigator reported in this project paper have not so far been submitted for any degree or diploma.

Date:

Prof. Dr Shirsendu

Mukherjee Assistant

Professor Department of

Statistics Asutosh

College

DECLARATION

I **Avisikta Das**, a student of B.Sc. Semester-6, Statistics Honours, of University of Calcutta, **Registration no-012-1212-0625-20, Roll no- 203012-11-0235**, hereby declare that I have done this piece of project work entitled as, "**A Review Work On Criminal Data Analysis Of Western & Middle Part Of India**" under the supervision of **Dr. Shirsendu Mukherjee** (Assistant professor, Department of Statistics, Asutosh College) as a part of B.Sc. Sem-6 examination according to the syllabus paper DSE-B2. I further declare that the piece of project work has not been published elsewhere for any degree or diploma or taken from any published project.

Signature:

Table Contents

Title	1
Certificate	2
Declaration	3
Abstract	6
Introduction	7-8
Purpose of The Project	8
Methodology	
Data collection	9
Analysis	9-18
1. Descriptive Analysis	9-12
2. Inferential Analysis	12-17
3. Crime Index	17-18
Data	19
Result	
1. Pie Charts	20-23
2. Multiple Divided Bar Diagram	24
3. Moving Average Trends	25-30
4. Chi-square Test	31-32
5. Two-way ANOVA Test	32-34
6. Crime Index	35-36

7. Overall Crime Scenario of India **37-40****Conclusion** **41****Bibliography** **42****Acknowledgement** **43****Appendix** **44-46**

ABSTRACT

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of inferential statistics we can observe the crime situation in a region and also predict future outcomes. Here we have an approach between computer science and criminal justice to develop a data driven procedure that can help show the present crime scenario and also project the future. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etc we are focusing mainly on crime factors of each year in various parts of western & middle India.

INTRODUCTION

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So for building such a powerful crime analytics tool we have to collect crime records and evaluate it. It is only within the last few decades that the technology made spatial data mining a practical solution for wide audiences of Law enforcement officials which is affordable and available. Since the availability of criminal data or records is limited we are collecting crime data from various sources like web sites, news sites, blogs, social media, RSS feeds etc. This huge data is used as a record for creating a crime record database. So the main challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively.



The present study is an effort in a direction of predicting both the present scenario and project the future by suitable statistical technique.

Study Area:

The Study area includes few western regions of India like Gujarat, Rajasthan, Haryana and middle regions of India like Madhya Pradesh, Chhattisgarh, Maharashtra .These

cities are some of the few most developed and densely populated cities in India and witnessing urbanization and industrialization at accelerated pace coupled with increased socio-economic activities. These areas are one of the most prime hotspots of crime in India and hence we are interested in the Statistical analysis of the western & middle region of India.

PURPOSE OF THE PROJECT

The criminal cases in India are rapidly increasing in India due to which numbers of cases pending are also piling up. Crime rate cannot be predicted since it is neither systematic nor random. Even though we cannot predict who all maybe the victims of crime but we may predict the place that has probability for its occurrence.

Here the main objective of our crime analysis project includes:

1. Extraction of crime patterns by analysis of available crime and criminal data
2. Prediction of crime based on spatial distribution of existing data anticipation of crime rate using different statistical techniques
3. Prediction of regions which have high probability for crime occurrence and visualize crime prone areas.

This crime data analysis on the basis of time, location and so on can help in providing useful information to law enforcement from strategically perspective. Although this predicted result cannot be assured of 100% accuracy but the result shows our application helps in reducing crime rate to a certain extent by providing more securities in crime sensitive areas.

METHODOLOGY

1. Data Collection:

Geographic and demographic diversity is perhaps the most striking feature of India, the second most populous country in the world. These are connected to the web-based system, and the data are open to access for the public (CPCB 2020a). These monitoring stations are maintained by the respective state Crime control boards. Considering the size of the country, the number of air quality monitoring stations is insufficient.

The major crimes considered in this study include Murder, Rape, Kidnapping, Robbery ,Counterfeiting. The period from the year 2001 to 2012 is considered as time of reference in this study. In this study we will observe the changes in the concentration of these major crimes in 6 western & middle states of India namely **Gujarat, Rajasthan, Haryana, Madhya Pradesh, Chhattisgarh, and Maharashtra.**

2. Analysis:

A. Descriptive Analysis:

The data that was collected from kaggle.com was extracted into a Microsoft Excel 2007 spreadsheet where descriptive analysis was conducted, and then we arrange them as a table in which major crimes are on column-heads and years on row-heads for specific states we are studying.

At first the Yearly data of western and middle states from 2001 to 2012 converted to single average data using average function in Excel. It was done for every 6 western & middle states of India.

Then we create a percentage table for respective data for 6 western and middle states and graphically represented percentage of type of crime in a whole, using **Pie Chart**. Also, to compare the difference of level of every single factor of crime between the states, we used **Divided bar diagram**.

Pie Diagram & Divided Bar Diagram:

In some situations, the values of a variable are available for a number of components and comparison among the different components or the relation between each part and the whole may be necessary in this context, the proportions or percentages of the various components are given more importance than the absolute values. So, we have to think of a diagrammatic representation of the percentages for different categories. Pie diagram and divided bar diagram are appropriate device for this purpose.

Pie diagram is used for exhibiting the relative sizes of the different parts of whole. In this case, a circle is partitioned into several sectors by drawing angles at the centre, the area of each sector indicating the corresponding percentage. In fact, the area enclosed by the circle is regarded as 100. Since, the total angle at the centre is 360° , the desired angle for some particular category will be 3.6 times the relevant percentage. The diagram, thus constructed, is termed as a pie diagram.

For divided bar diagram, a single bar of suitable length and width is taken and its area is considered as 100. The area of the bar is then divided into a number of parts, depending on the number of categories, with the help of lines drawn parallel to the base, so that the area of a part represents the percentage for the category concerned.

Here, we used pie diagram for each 6 western & middle states individually to compare the 6 crime aspects. And we used multiple divided bar diagram to represent the crimes aspect of each states in one frame.

Now, for a Time Series data it is important to identify and fit any systematic time patterns which may be present. Using the total data set to identify where possible reversals will take place, we draw **Trend Line**) of each factors at each stations. Trend lines are easily recognizable lines that traders draw on charts to connect a series of values together or show some data's best fit. The resulting line is then used to give the trader a good idea of the Crime at different western and middle states. A moving average trend line smoothes out fluctuations in data to show a pattern or trend more clearly.

We create Trend Line in the following way:

Moving Average:

Moving Average method is a very commonly used method for the computation of Trend in a time series. This method consists in measurement of Trend by smoothing out the fluctuations present in the data by means of a Moving Average.

Moving Average may be defined as an arithmetic mean of a given number of observations, each time in a time series. The period of it is adjusted successively by replacing the first observations of the previously averaged groups by the next observations below the relevant group in calculating such successive average. Thus a Moving Average of period 'k' is a series of successive arithmetic means of observations, each computed on the basis of 'k' observations. We start with the first 'k' observations. At the next step, we leave the first and include the $(k+1)^{th}$ observation. This process is repeated until we arrive at the last set of 'k' observations. The averages, so obtained, are placed against the middle of the time-interval used in the calculation of the Moving Average. If 'k', the period of Moving Average, is odd, for example – 3 years, 5 years, 7 years etc. then each Moving Average value will correspond to the tabulated time point. However, if 'k' is even, for example – 4 years, 6 years etc. the Moving Average value falls midway between two tabulated time points. In this case, we calculate a subsequent

2-item Moving Average to make the resultant Moving Average value correspond to the tabulated time points. This process is known as Centering.

These values of the Moving Averages, when plotted on a graph paper, will give the Trend values of the original time series.

The Moving Average employed on any series showing random fluctuations will produce a series much smoother than the original one. The greater the period of Moving Average, smoother will be the resulting series. It has also been observed that Moving Averages acting on a series showing periodic oscillations will produce a straight line, provided the extent period of Moving Average is equal to the period of oscillation or some multiple of it. If, however, the extent of Moving Average is different from the period of oscillation, then the resultant series will exhibit same type of oscillation as before, but with much diminished amplitude.

So, moving average method is flexible in the sense that any change in the Trend is faithfully reflected by the Moving Averages.

Here, we are going to plot 3-years moving average for the 6 western & middle states on the 6 crime aspects (Murder, Rape, Kidnapping, Robbery, Riots, and Counterfeiting). By, this we can smooth out the random fluctuations in trend of 6 crime aspects for the states, individually.

B. Inferential Analysis:

i. Chi-square Test:

The Formula of Chi-square is:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

c = Degrees of freedom

O = Observed value(s)

E = Expected value(s)

A chi-square (χ^2) statistic is a test that measures how a model compares to actual observed data. The data used in calculating a chi-square statistic must be random, raw, mutually exclusive, drawn from independent variables, and drawn from a large enough sample. For example, the results of tossing a fair coin meet these criteria.

Chi-square tests are often used in hypothesis testing. The chi-square statistic compares the size of any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship.

For these tests, degrees of freedom are utilized to determine if a certain null hypothesis can be rejected based on the total number of variables and samples within the experiment. As with any statistic, the larger the sample size, the more reliable the results.

ii. ANOVA:

We use Two-way ANOVA on our data to make appropriate claims and prove the truth of certain hypothesis.

ANOVA is helpful for testing three or more variables. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse source.

It is a way to find out if the experiment results are significant. It helps us to figure out if we need to reject the null hypothesis or accept the alternative hypothesis.

For the results of an ANOVA to be valid, the following assumptions should be met:

1) Normality – The crime observations are count data, so we can say that it follows Poisson distribution. So, we are using Central Limit Theorem. Thus we can say that our observations are normally distributed.

[Observations (Y_{ij})~Poisson(θ)]

Then by CLT theorem

$$\frac{Y_{ij} - \theta}{\sqrt{\theta}} \rightarrow N(0,1)$$

Thus, Y_{ij} will approximate normal distribution.

2) Equal Variances – The variances for each group should be roughly equal.

3) Independence – The observations in each group are independent of each other and the observations within groups were obtained by a random sample.

Two-way ANOVA:

A two-way ANOVA is used to estimate how the mean of a quantitative variable changes according to the levels of two categorical variables. It has two independent variables (i.e. years and states). Now, we want to test whether the crimes are equally effective over time and over states.

Let, Y_{ij} be the crime observations corresponding to the i_{th} year and j_{th} state.

$i=1(1)p, j=1(1)q$.

So, total number of observations are (n)= $p*q$

Model:

The appropriate mathematical model is given by—

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

$\forall i=1(1)p; j=1(1)q$.

Where,

μ : General mean effect

α_i : Additional fixed effect over the general mean effect due to i_{th} year, $i=1(1)p$.

β_j : Additional fixed effect over the general mean effect due to j_{th} state, $j=1(1)q$.

e_{ij} : Random error component, $i=1(1)p, j=1(1)q$.

Assumptions:

i) $e_{ij} \sim N(0, \sigma_e^2)$, σ_e^2 is unknown.

ii) $\sum_{i=1}^p \alpha_i = 0$ iii) $\sum_{j=1}^q \beta_j = 0$

Hypothesis to be tested:

Here we want to test whether there is any significant difference in the observations corresponding to the different years, i.e.

$H_{01}: \alpha_i = 0, \forall i = 1(1)p$ Against $H_{11}: \text{Not } H_{01}$

We also want to test whether there is any significant difference in the observations corresponding to the different states, i.e.

$H_{02}: \beta_j = 0, \forall j = 1(1)q$ Against $H_{12}: \text{Not } H_{02}$

Test Statistic:

The appropriate test statistic for testing H_{01} against H_{11} is given by,

$$F_1 = \frac{\text{MSY}}{\text{MSE}} \sim F_{(p-1), (p-1)(q-1)} \quad \text{Under } H_{01}$$

Where MSY: Mean Sum of Squares due to Years=Sum of Squares due to Years (SSY)/d.f. (degrees of freedom) of SSY.

Here the d.f. of SSY is (p-1)

MSE: Mean Sum of Squares due to Error=Sum of Squares due to Error (SSE)/d.f. (degrees of freedom) of SSE.

Here the d.f. of SSE is (p-1).(q-1)

Now,

The appropriate test statistic for testing H_{02} against H_{12} is given by,

$$F_2 = \frac{\text{MSS}}{\text{MSE}} \sim F_{(q-1), (p-1)(q-1)} \quad \text{Under } H_{02}$$

Where MSS: Mean Sum of Squares due to States=Sum of Squares due to States (SSS)/ d.f. (degrees of freedom) of SSS.

Here the d.f. of SSS is (q-1)

MSE: Mean Sum of Squares due to Error=Sum of Squares due to Error (SSE)/d.f. (degrees of freedom) of SSE.

Here the d.f. of SSE is (p-1).(q-1)

Critical Region:

We reject H_{01} against H_{11} at level α iff

$$\text{Observed } F_1 > F_{\alpha;(p-1),(p-1).(q-1)}$$

Here $F_{\alpha;(p-1),(p-1).(q-1)}$ is the upper α point of $F_{\alpha;(p-1),(p-1).(q-1)}$ distribution.

Otherwise, we will accept H_{01} .

We reject H_{02} against H_{12} at level α iff

$$\text{Observed } F_2 > F_{\alpha;(q-1),(p-1).(q-1)}$$

Here $F_{\alpha;(q-1),(p-1).(q-1)}$ is the upper α point of $F_{(q-1),(p-1).(q-1)}$ distribution.

Otherwise, we will accept H_{02} .

Steps for Computing Mean Sum of Squares:

1. Grand Total: $G = \sum_{i=1}^p \sum_{j=1}^q Y_{ij}$

2. Correction Factor: $C.F. = \frac{G^2}{n}$

3. Raw Total Sum of Squares: $RSS = \sum_{i=1}^p \sum_{j=1}^q Y_{ij}^2$

4. Total Sum of Squares: $TSS = RSS - C.F.$

5. Sum of Squares due to Years: $SSY = \frac{1}{q} \sum_{i=1}^p T_{i0}^2 - C.F.$ where, $T_{i0} = \sum_{j=1}^q Y_{ij}$

6. Sum of Squares due to States: $SSS = \frac{1}{p} \sum_{j=1}^q T_{0j}^2 - C.F.$ where, $T_{0j} = \sum_{i=1}^p Y_{ij}$

7. Sum of Squares due to Error: $SSE = TSS - SSY - SSS$

8. Mean Sum of Squares due to Years: $MSY = \frac{SSY}{(p-1)}$ as d.f. of $SSS = (p-1)$

9. Mean Sum of Squares due to States: $MSS = \frac{SSS}{(q-1)}$ as d.f. of $SSS = (q-1)$

10. Mean Sum of Squares due to Error: $MSE = \frac{SSE}{(p-1).(q-1)}$

as d.f. of $SSE = (p-1).(q-1)$

ANOVA Table:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	Observed F	Tabulated F
Years	(p-1)	SSY	MSY	F ₁	F _{α;(p-1),(p-1).(q-1)}
States	(q-1)	SSS	MSS	F ₂	F _{α;(q-1),(p-1).(q-1)}
Error	(p-1).(q-1)	SSE	MSE		
Total	p.q-1	TSS	-----	-----	-----

C. Crime Index:

Step 1: To find the index of every Crime aspect in each state, we selected 2001 year as the base year and 2002-2012 year as recent years.

Step 2: Then we need to find the Index Relatives of every Crime aspect for respective state for each year (2002-2012) based on the year 2001 using the formula:

$$I_{ijk} = \frac{Y_{ijk}}{Y_{0jk}} \times 100$$

Here, i = 11, j = 6, k = 6

Where I_{ijk} : Index number corresponding to i_{th} year, j_{th} state and k_{th} crime aspect.

Y_{ijk} : Crime observations corresponding to i_{th} year and j_{th} state and k_{th} crime aspect.

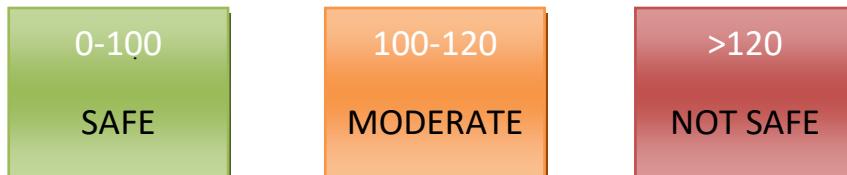
Y_{0jk} : Crime observations corresponding to 2001 year, j_{th} state and k_{th} crime aspect.

Step 3: After finding all the I_{ijk} , we will find Average Crime Index (CI_{ij}) for respective state for each year (2002-2012) based on the year 2001 using formula:

$$CI_{ij} = \frac{1}{6} \sum_{k=1}^6 I_{ijk}$$

Here, $i = 11, j = 6, k = 6$.

After calculate the Index Numbers we have to see in which group the quality index is falling.



DATA

We have 12 years data of six monitoring western & middle states from India.

Chhattisgarh:

YEAR	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
2001	880	959	207	338	871	8
2002	844	992	178	326	817	18
2003	776	898	177	331	703	24
2004	927	969	214	363	790	42
2005	1013	990	246	401	893	73
2006	1098	995	239	439	905	33
2007	1097	982	244	427	881	64
2008	1169	978	273	500	1144	71
2009	1083	976	286	554	957	68
2010	1065	1012	359	552	1092	59
2011	1110	1053	472	470	934	65
2012	998	1034	450	412	910	53

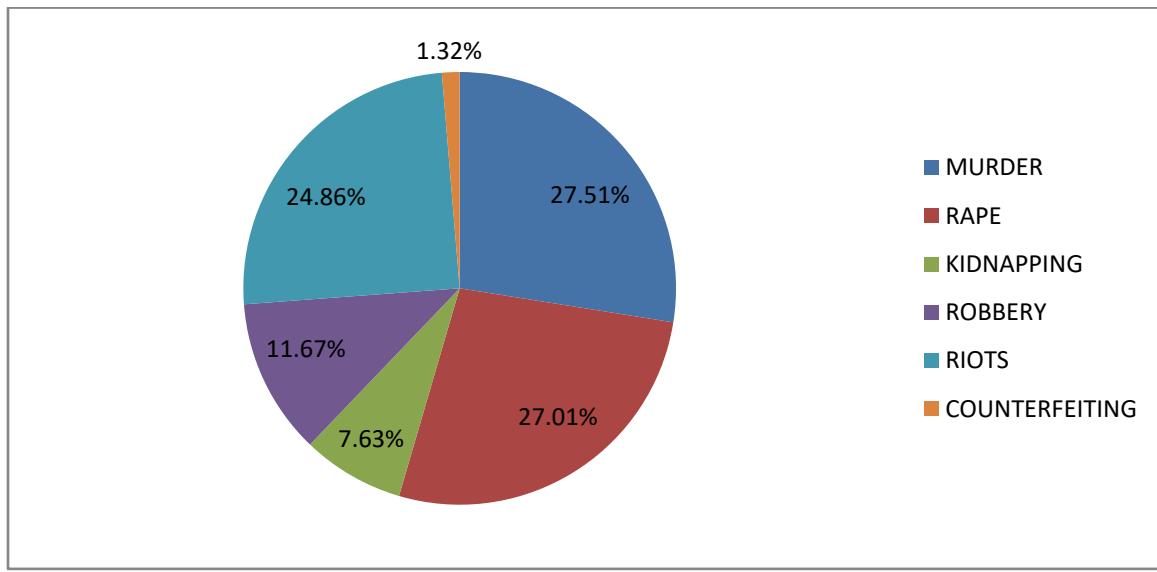
The data for other 5 states (Gujarat, Haryana, Madhya Pradesh, Maharashtra, and Rajasthan) are in the '**Appendix**' which is given in the last of the project.

RESULT

I. Pie Chart of Different crimes in western & middle states of India:

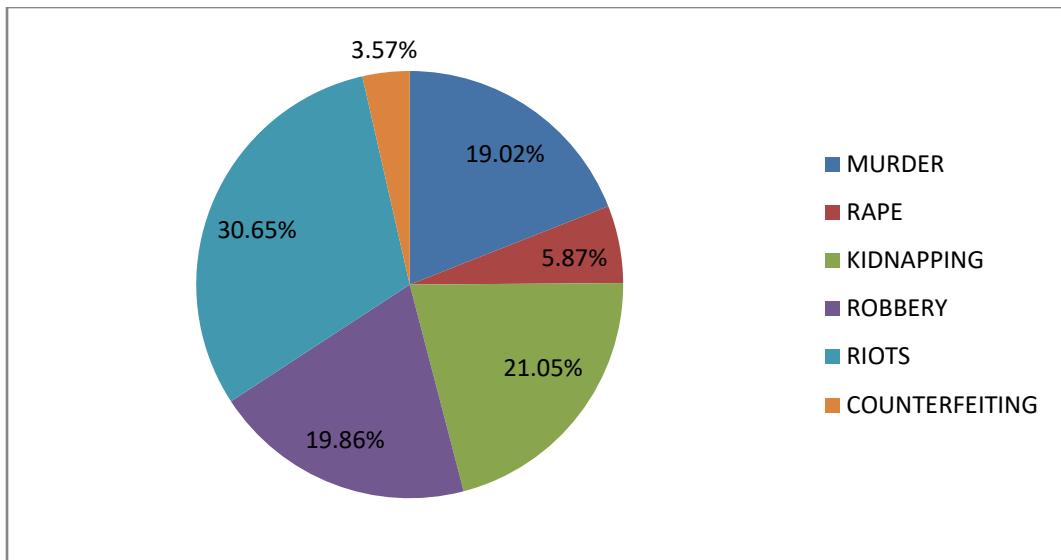
1. Chhattisgarh:

CRIME	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
AVERAGE	1005	986.5	278.75	426.08333	908.1	48.167
% TOTAL	27.51%	27.01%	7.63%	11.67%	24.86%	1.32%



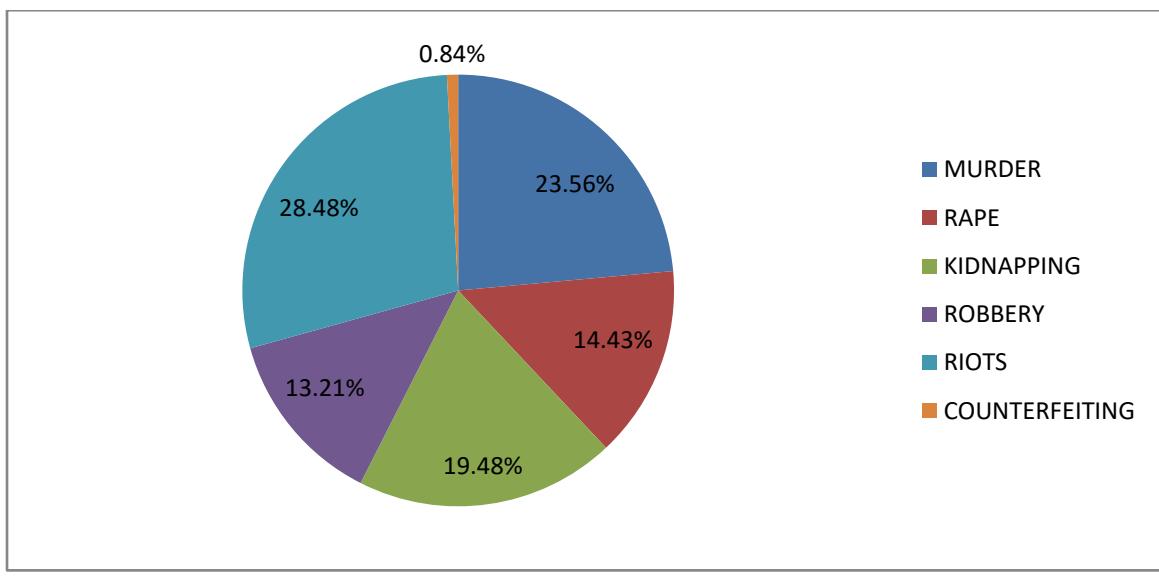
2. Gujarat:

CRIME	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
AVERAGE	1147.9167	354.08	1270.5	1198.5833	1849	215.17
%TOTAL	19.02%	5.87%	21.05%	19.86%	30.65%	3.57%



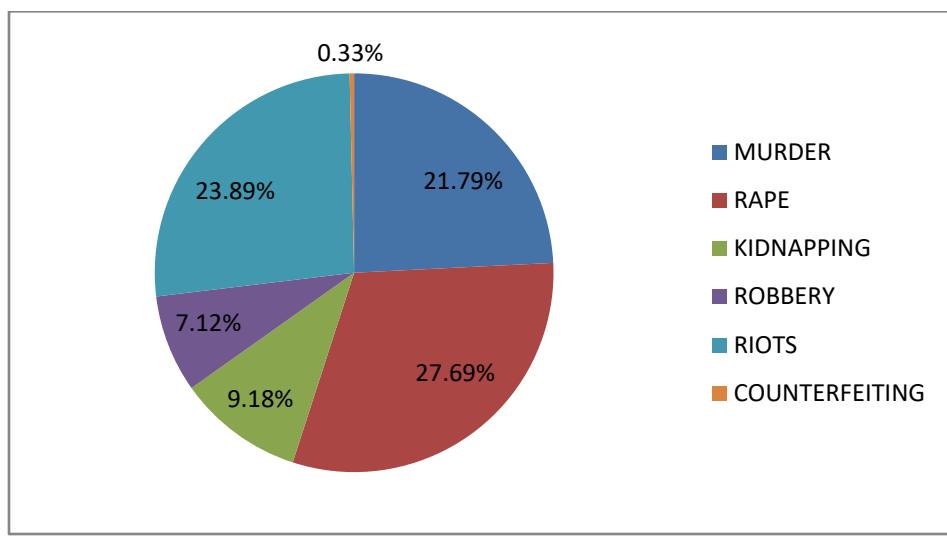
3. Haryana:

CRIME	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
AVERAGE	872.58333	534.17	721.3333333	489.25	1055	31
%TOTAL	23.56%	14.43%	19.48%	13.21%	28.48%	0.84%



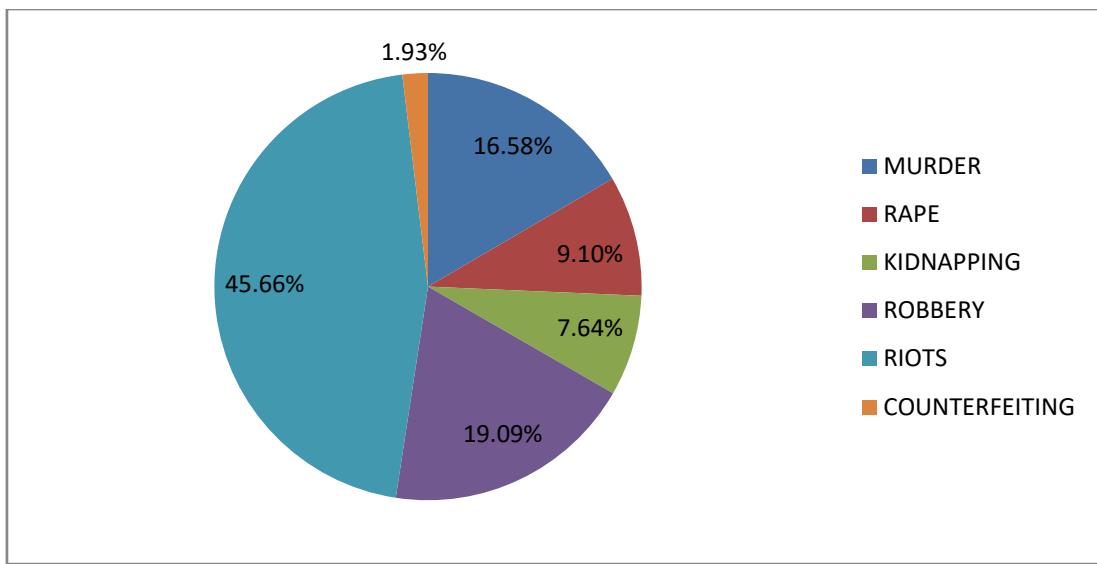
4. Madhya Pradesh:

CRIME	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
AVERAGE	2366.5833	3007.3	996.5	1859.5	2594	35.333
%TOTAL	21.79%	27.69%	9.18%	7.12%	23.89%	0.33%



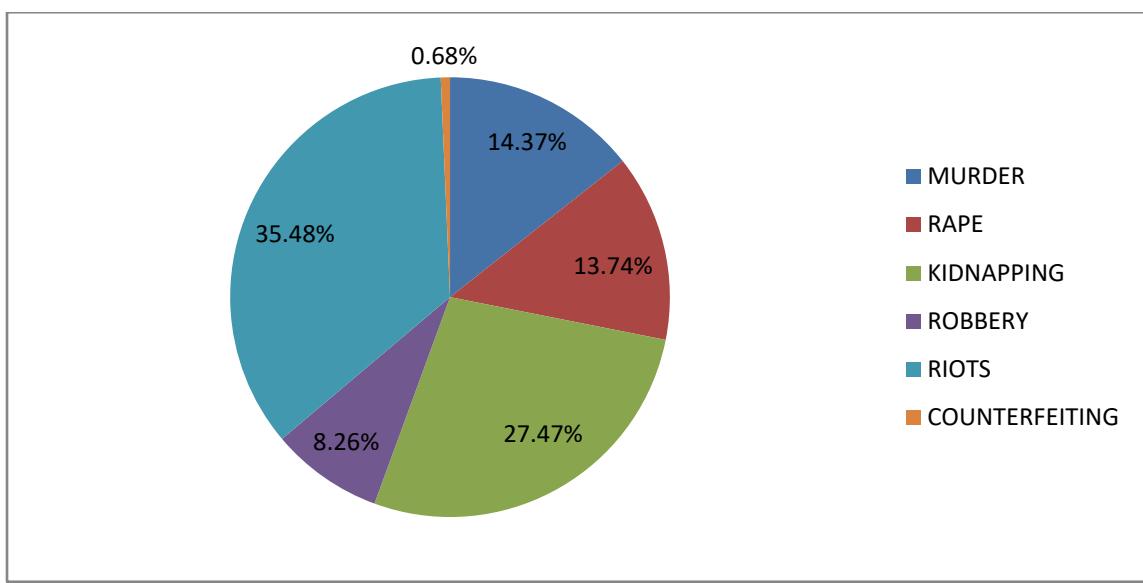
5. Maharashtra:

CRIME	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
AVERAGE	2730.5833	1498.8	1258.416667	3143.25	7520	317.17
%TOTAL	16.58%	9.10%	7.64%	19.09%	45.66%	1.93%



6. Rajasthan:

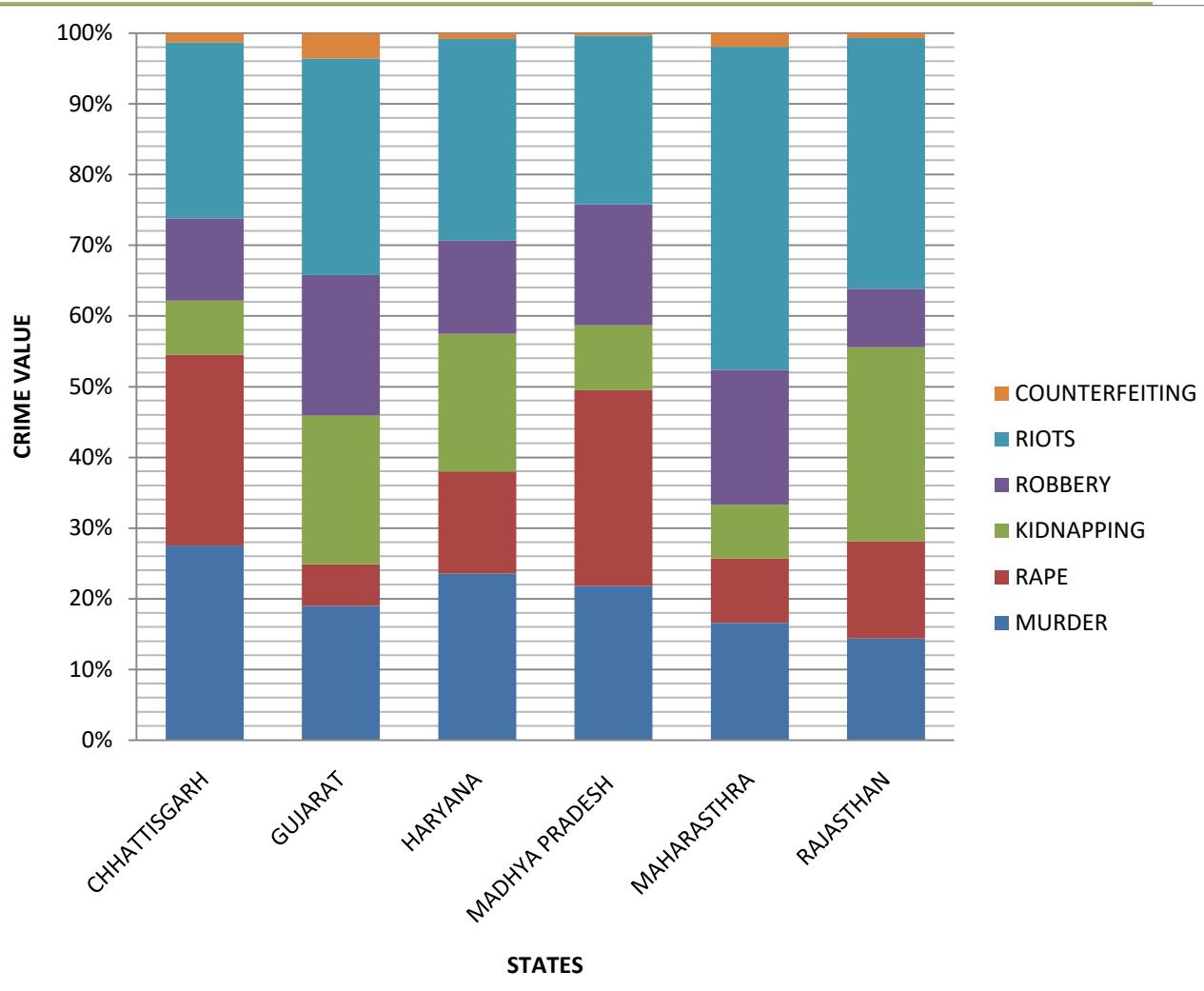
CRIME	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
AVERAGE	1307.5455	1249.9	2498.545455	751.27273	3227	61.545
%TOTAL	14.37%	13.74%	27.47%	8.26%	35.48%	0.68%



II. Multiple Divided Bar diagram: Comparison in levels of each Types of Crime for different Western & Middle States of India

Table 2.1 showing average crime values of different western and middle states over 12 year time frame.

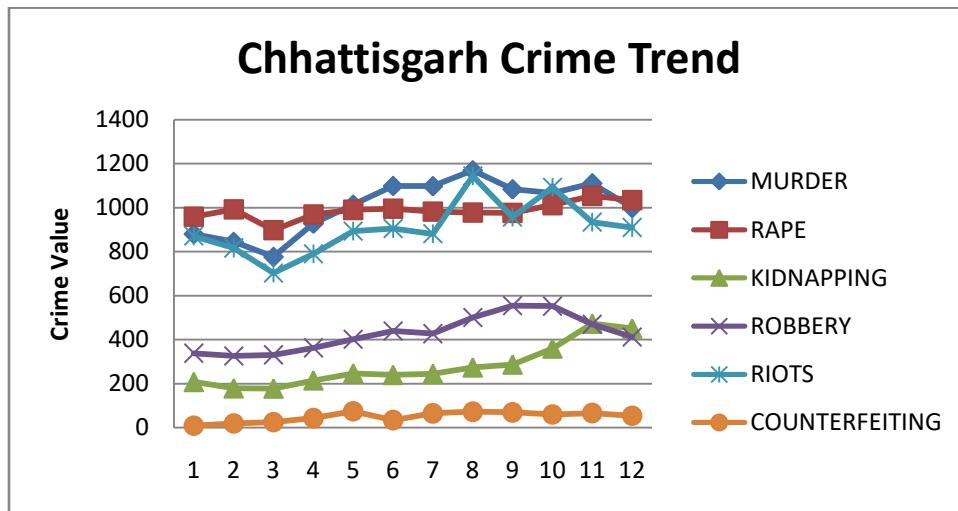
STATE	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
CHHATTISGARH	1005	986.5	278.75	426.0833	908.0833	48.16667
GUJARAT	1147.917	354.0833	1270.5	1198.583	1849.333	215.1667
HARYANA	872.5833	534.1667	721.3333333	489.25	1054.667	31
MADHYA PRADESH	2366.583	3007.25	996.5	1859.5	2593.667	35.33333
MAHARASTHRA	2730.583	1498.833	1258.416667	3143.25	7520.333	317.1667
RAJASTHAN	1307.545	1249.909	2498.545455	751.2727	3227.273	61.54545



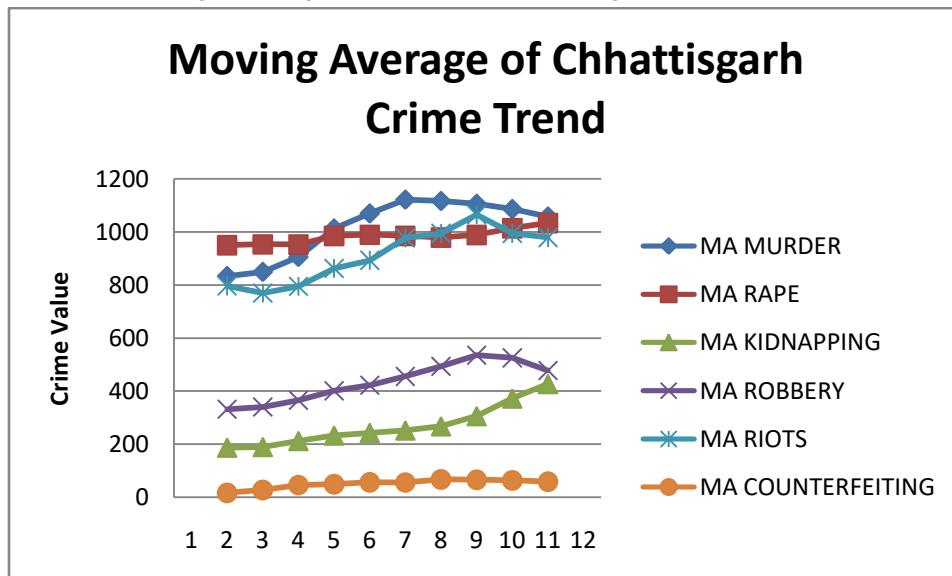
III. Moving Average Trends: How the levels of crimes are fluctuating over 12 years:

AIM: Our main purpose is to smooth out the different fluctuations the curves are having

- Original fluctuations and variations for Chhattisgarh Crime data

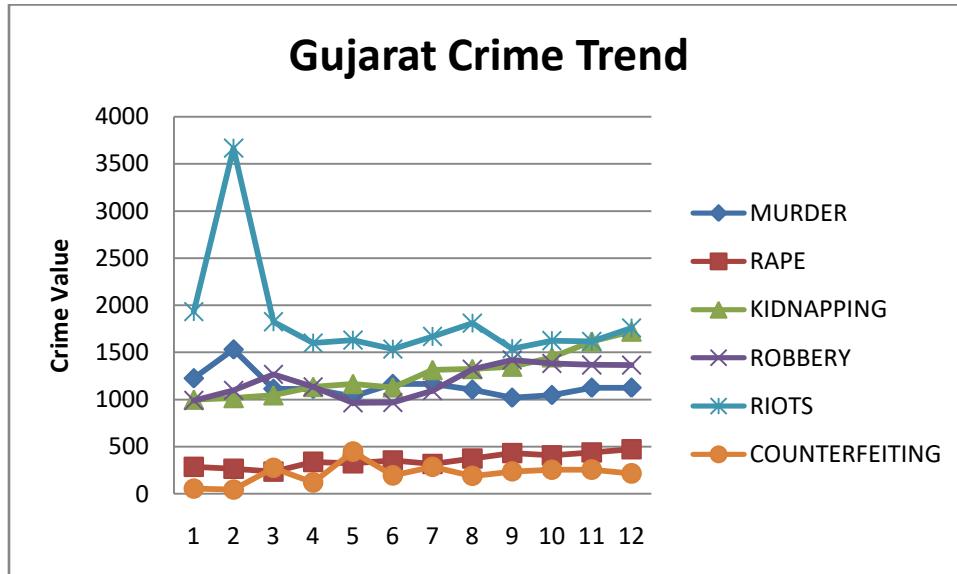


- 3 Year Moving average trends for Chhattisgarh Crime data

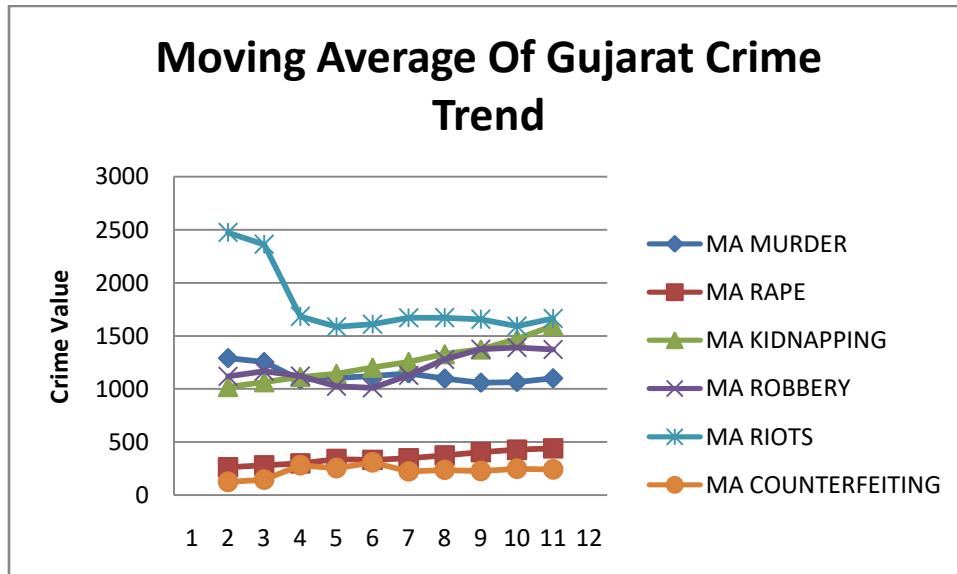


Observation: Here, we see that trend of Riots, Robbery and Murder is moving in the downward direction and trend of Kidnapping may go to upward direction and trend of Rape may go to slightly upward and Fraud is as constant in future at Chhattisgarh.

- Original fluctuations and variations for Gujarat Crime data

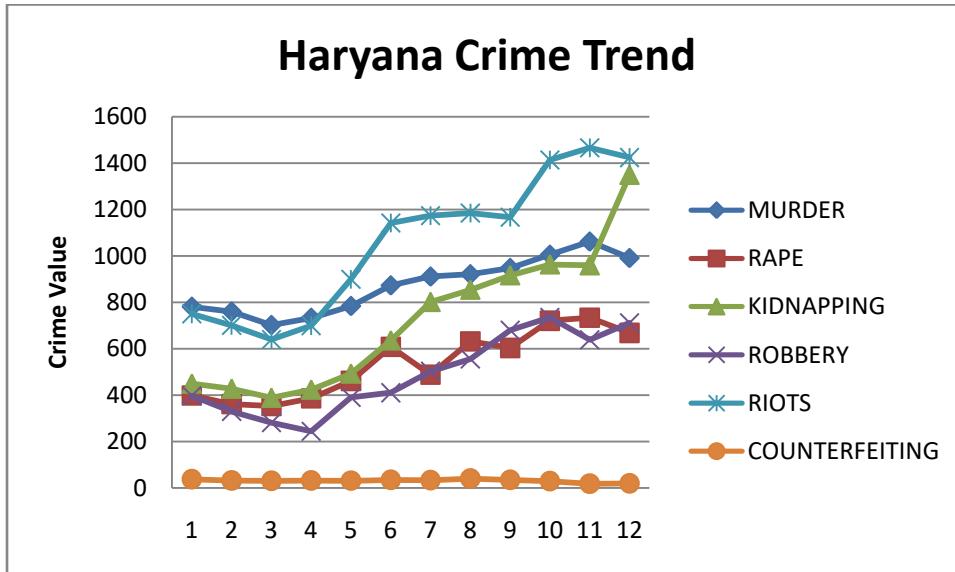


- 3 Year Moving average trends for Gujarat Crime data

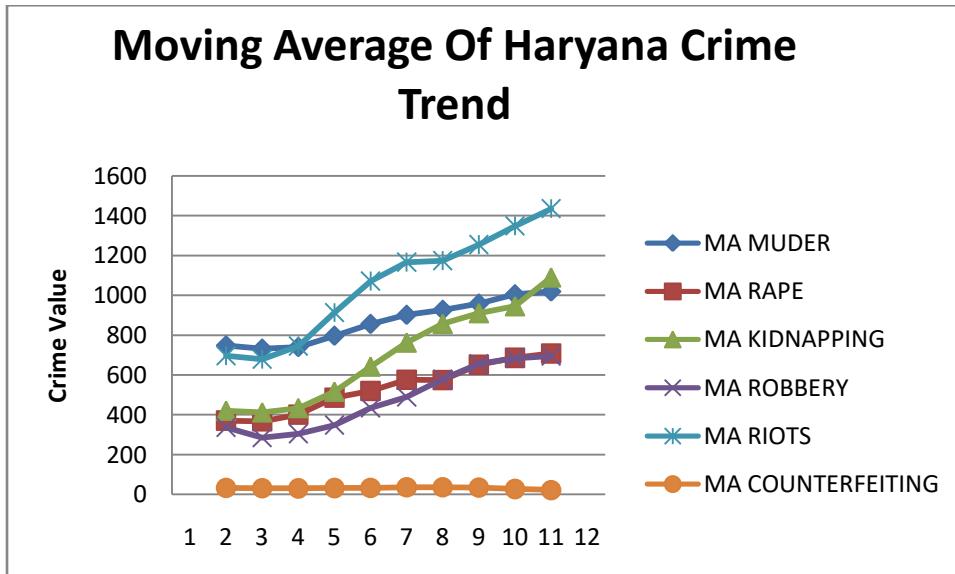


Observation: Here, we see that trend of Riots and Murder is moving in the downward direction and trend of Kidnapping and Robbery may go to upward direction and trend of Rape may go to slightly upward and Fraud is as constant in future at Gujarat.

- Original fluctuations and variations for Haryana Crime data

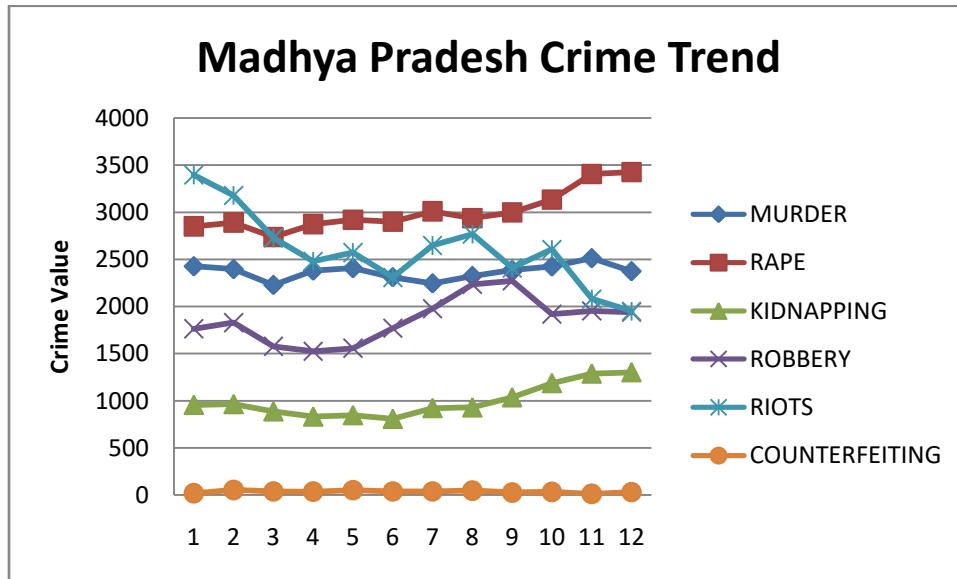


- 3 Year Moving average trends for Haryana Crime data

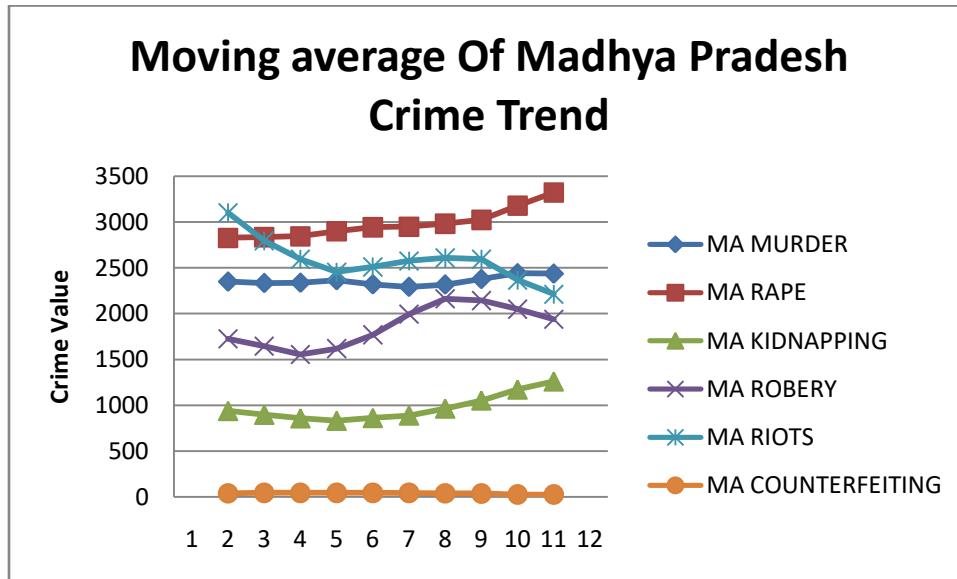


Observation: Here, we see that trend Murder, Rape, Kidnapping, Robbery, Riots may go to upward direction and Fraud is as constant in future at Haryana.

- Original fluctuations and variations for Madhya Pradesh Crime data

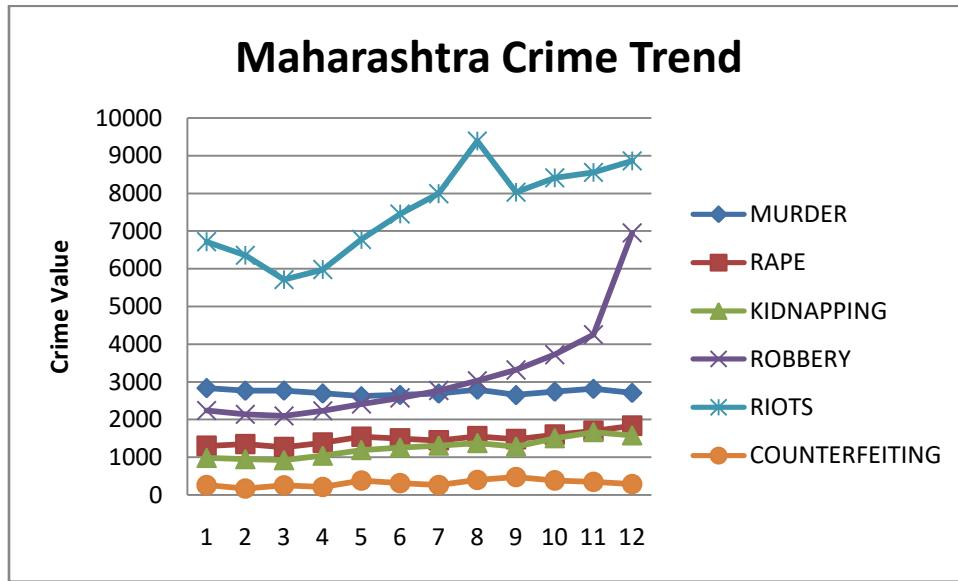


- 3 Year Moving average trends for Madhya Pradesh Crime data

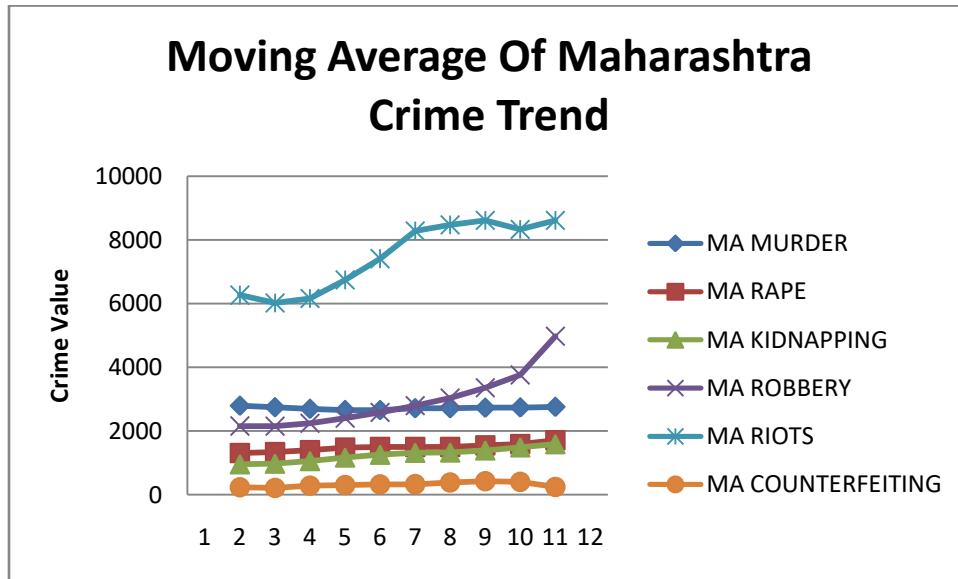


Observation: Here, we see that trend of Riots and Robbery is moving in the downward direction and trend of Kidnapping and Rape may go to upward direction and trend of Murder is approximately constant and Fraud is also constant in future at Madhya Pradesh.

- Original fluctuations and variations for Maharashtra Crime data

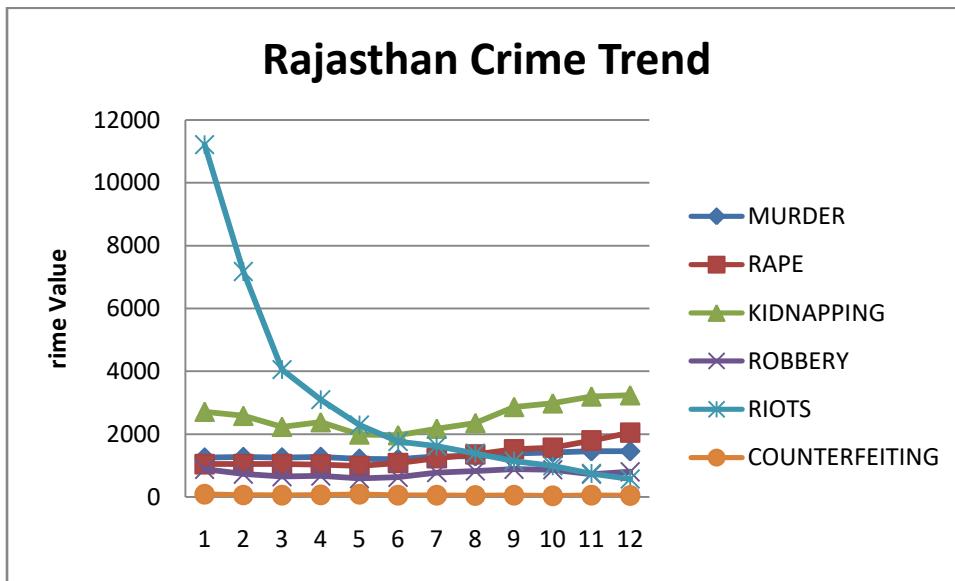


- 3 Year Moving average trends for Maharashtra Crime data

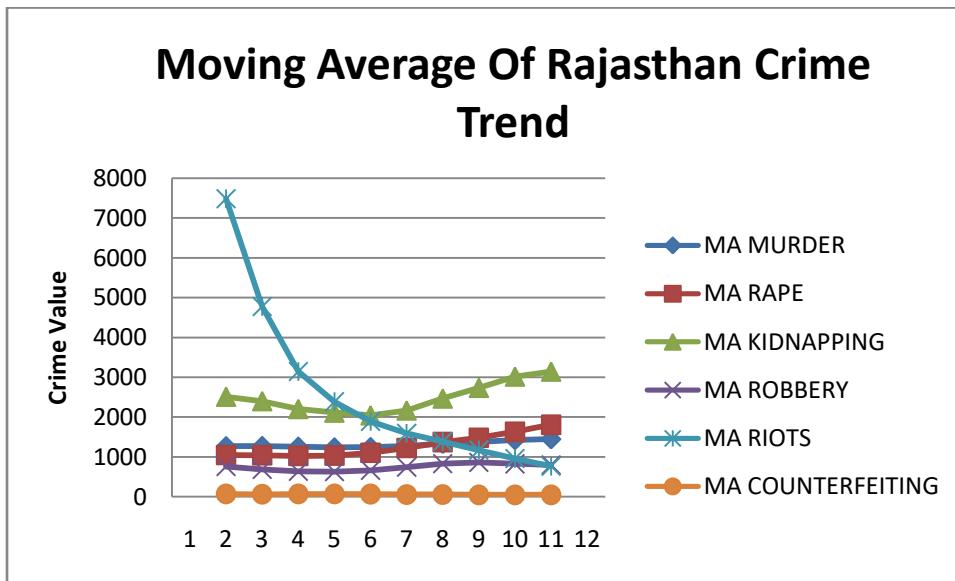


Observation: Here, we see that trend of Fraud is very slowly moving in the downward direction and trend of Riots and Robbery may go to upward direction and trend of Murder is approximately constant and trend of Rape and Kidnapping is moving slowly to upwards in the future at Maharashtra.

- Original fluctuations and variations for Rajasthan Crime data



- 3 Year Moving average trends for Rajasthan Crime data



Observation: Here, we see that trend of Riots is moving in the downward direction and trend of Kidnapping may go to upward direction and trend of Rape is moving slowly to upward direction and trend of Murder and Robbery is approximately constant and trend of Fraud is also constant in future at Rajasthan.

IV. Chi-Square Test:

Hypothesis:

H_0 : There is no significant relation between different states and the crimes

H_1 : There is significant relation between different states and the crimes

Table 4.1

Showing the observed values of our given data set

STATE	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	FRAUD	TOTAL
CHHATTISGARH	1005	986.5	278.75	426.083333	908.083333	48.1666667	3652
GUJARAT	1147.91667	354.083333	1270.5	1198.58333	1849.33333	215.166667	6036
HARYANA	872.583333	534.1666667	721.3333333	489.25	1054.66667	31	3703
MADHYA PRADESH	2366.58333	3007.25	996.5	1859.5	2593.66667	35.3333333	10859
MAHARASTHRA	2730.58333	1498.833333	1258.416667	3143.25	7520.33333	317.166667	16468
RAJASTHAN	1307.54545	1249.909091	2498.545455	751.272727	3227.27273	61.5454545	9096
TOTAL	9430	7631	7024	7868	17153	708	49814

Table 4.2

Showing the Expected values of our given data set

STATE	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	FRAUD	TOTAL
CHHATTISGARH	691.338981	559.4493917	514.9485687	576.824507	1257.53314	51.9054081	3652
GUJARAT	1142.64022	924.654033	851.1033846	953.371502	2078.44196	85.7888947	6036
HARYANA	700.993496	567.2620749	522.1398001	584.879833	1275.09453	52.6302646	3703
MADHYA PRADESH	2055.65443	1663.488758	1531.168266	1715.15261	3739.19836	154.337576	10859
MAHARASTHRA	3117.46176	2522.730718	2322.062713	2601.0805	5670.60674	234.057574	16468
RAJASTHAN	1721.91111	1393.415024	1282.577267	1436.69105	3132.12527	129.280283	9096
TOTAL	9430	7631	7024	7868	17153	708	49814

Conclusion:

At 5% level of significance,

Calculated value of $\chi^2_c = 6841.611006$

Table value of $\chi^2_{0.05} = 37.652$

Now Calculated value of $\chi^2_c >$ Tabular value of χ^2_c at 5% level of significance, reject H_0

Conclusion: Therefore, we come to the conclusion that **there is significant relation between different states and the crimes.**

V. Two-way ANOVA Test:

Here we have given the observations of 12 years and 6 states.

So, here $p=12$ and $q=6$

Therefore, the total no. of observation is: $n= 72$

We are using two way ANOVA model

Hypothesis to be tested:

Here we want to test whether there is any significant difference in the crime observations corresponding to the different years, i.e. the overall crimes over different years is similar or not similar.

$H_{01}: \alpha_i = 0, \forall i = 1(1)12$ Against $H_{11}: \text{Not } H_{01}$

We also want to test whether there is any significant difference in the crime observations corresponding to the different states, i.e. the overall crimes over different states is similar or not similar.

$H_{02}: \beta_j = 0, \forall j = 1(1)6$ Against $H_{12}: \text{Not } H_{02}$

Table 5.1

Showing the data for the two way ANOVA

YEAR/STATES	CHHATTISGARH	GUJARAT	HARYANA	MADHYA PRADESH	MAHARASHTRA	RAJASHTHAN
2001	3263	5486	2812	11410	14349	17218
2002	3175	7623	2610	11311	13745	12889
2003	2909	5760	2393	10197	13037	9315
2004	3305	5442	2517	10127	13556	8540
2005	3616	5562	3057	10354	14935	7175
2006	3709	5347	3703	10133	15761	6718
2007	3695	5843	3908	10837	16489	7178
2008	4135	6124	4185	11237	18556	7280
2009	3924	5998	4347	11126	17244	7874
2010	4139	6165	4865	11303	18373	7871
2011	4104	6418	4876	11249	19344	7999
2012	3857	6659	5163	11022	22234	8182

Since the values are too large so we divide each value by 100. Then our adjusted data is as follows:-

Table 5.2

Showing the adjusted data for the two way ANOVA

YEAR/STATE	CHHATTISGARH	GUJARAT	HARYANA	MADHYA PRADESH	MAHARASHTRA	RAJASHTHAN
2001	32.63	54.86	28.12	114.1	143.49	172.18
2002	31.75	76.23	26.1	113.11	137.45	128.89
2003	29.09	57.6	23.93	101.97	130.37	93.15
2004	33.05	54.42	25.17	101.27	135.56	85.4
2005	36.16	55.62	30.57	103.54	149.35	71.75
2006	37.09	53.47	37.03	101.33	157.61	67.18
2007	36.95	58.43	39.08	108.37	164.89	71.78
2008	41.35	61.24	41.85	112.37	185.56	72.8
2009	39.24	59.98	43.47	111.26	172.44	78.74
2010	41.39	61.65	48.65	113.03	183.73	78.71
2011	41.04	64.18	48.76	112.49	193.44	79.99
2012	38.57	66.59	51.63	110.22	222.34	81.82

Table 5.2**Showing two way ANOVA table**

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	Observed F	Tabulated F
Years	11	3918.093	356.1903	$F_1=1.164327$	$F_{0.05;11,55}=1.90$
States	5	145979.4	29195.88	$F_2=95.43649$	$F_{0.05;5,55}=2.38$
Error	55	16825.57098	305.9195		
Total	71	166723.0702	-----	-----	-----

Conclusion:

At 5% level of significance,

Now Observed value of $F_1 <$ Tabular value of $F_{0.05;11,55}$

So, we will accept H_{01} at 5% level of significance.

Therefore, we come to the conclusion that the crimes over the years are similar.

Now Observed value of $F_2 >$ Tabular value of $F_{0.05;5,55}$

So, we will reject H_{02} at 5% level of significance.

Therefore, we come to the conclusion that the overall crimes over different states are not similar.

VI .Crime index:

We have found out the different crime indexes for the western and middle part of India by the previous prescribed formula. And, now we are going to give a chart of crime indexes corresponding to 6 states due to 4 selected years (2003, 2006, 2009, and 2012). It is as follows:-

Table no.-6.1

Showing Crime Index for 6 states in 2003

STATE	CRIME INDEX	COMMENT
CHHATTISGARH	124.3281819	NOT SAFE
GUJARAT	166.7249549	NOT SAFE
HARYANA	83.67573698	SAFE
MADHYA PRADESH	108.4154425	MODERATE
MAHARASHTRA	94.36416708	SAFE
RAJASTHAN	76.27563388	SAFE

Table no.-6.2

Showing Crime Index for 6 states in 2006

STATE	CRIME INDEX	COMMENT
CHHATTISGARH	165.0451318	NOT SAFE
GUJARAT	144.2588626	NOT SAFE
HARYANA	126.0174665	NOT SAFE
MADHYA PRADESH	108.2960802	MODERATE
MAHARASHTRA	113.7151115	MODERATE
RAJASTHAN	70.26621064	SAFE

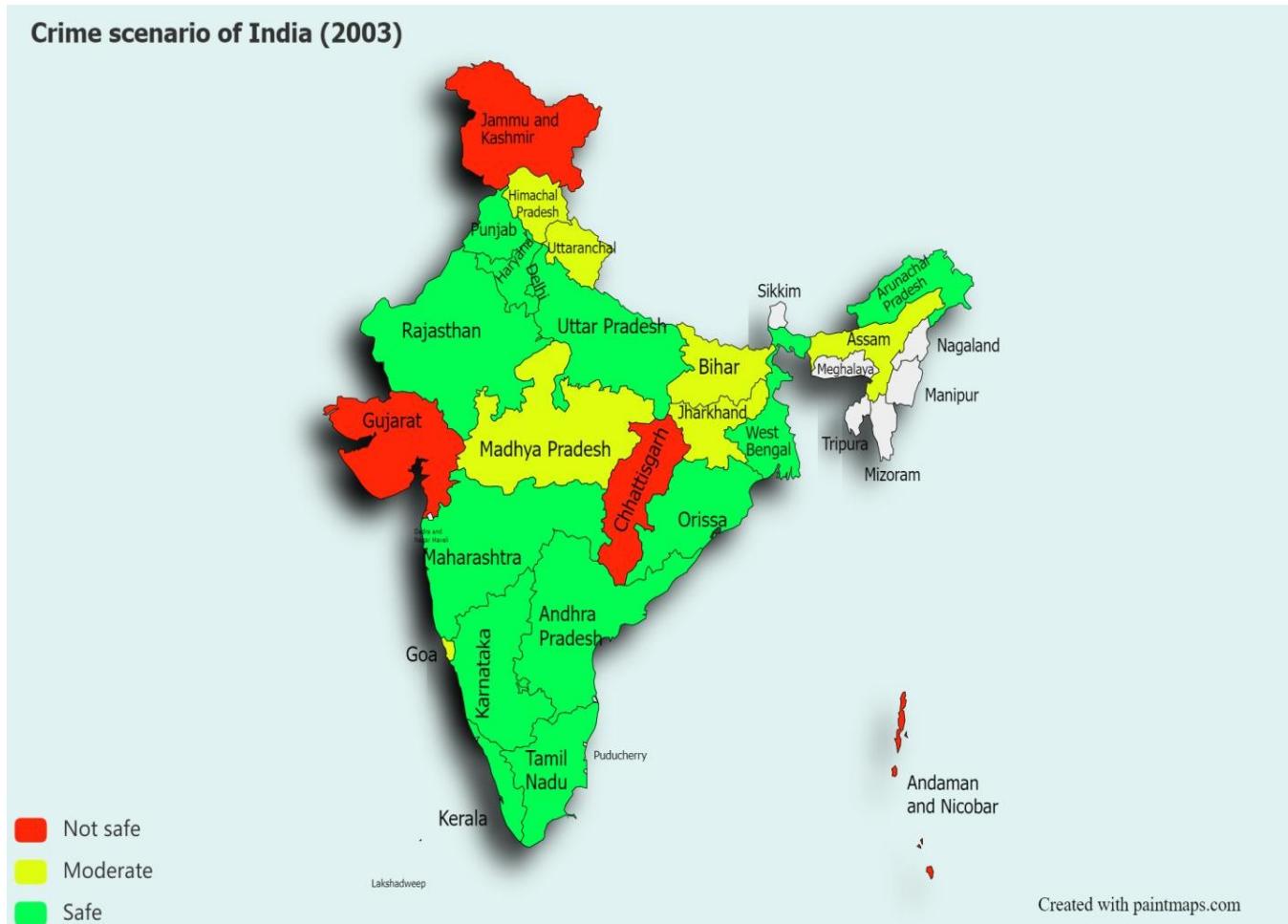
Table no.-6.3Showing Crime Index for 6 states in 2009

STATE	CRIME INDEX	COMMENT
CHHATTISGARH	247.7973578	NOT SAFE
GUJARAT	170.9039906	NOT SAFE
HARYANA	149.6655493	NOT SAFE
MADHYA PRADESH	108.9438999	MODERATE
MAHARASHTRA	130.9683598	NOT SAFE
RAJASTHAN	89.56071255	SAFE

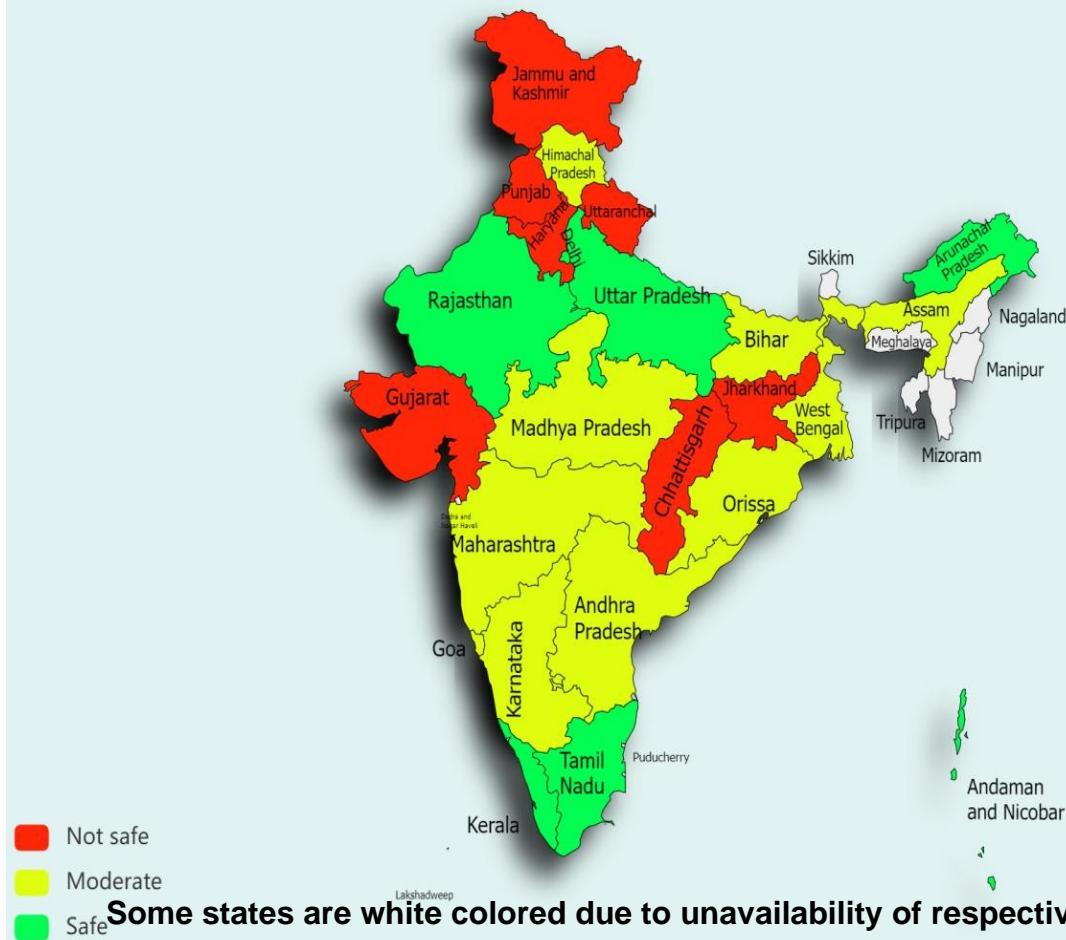
Table no.-6.4Showing Crime Index for 6 states in 2012

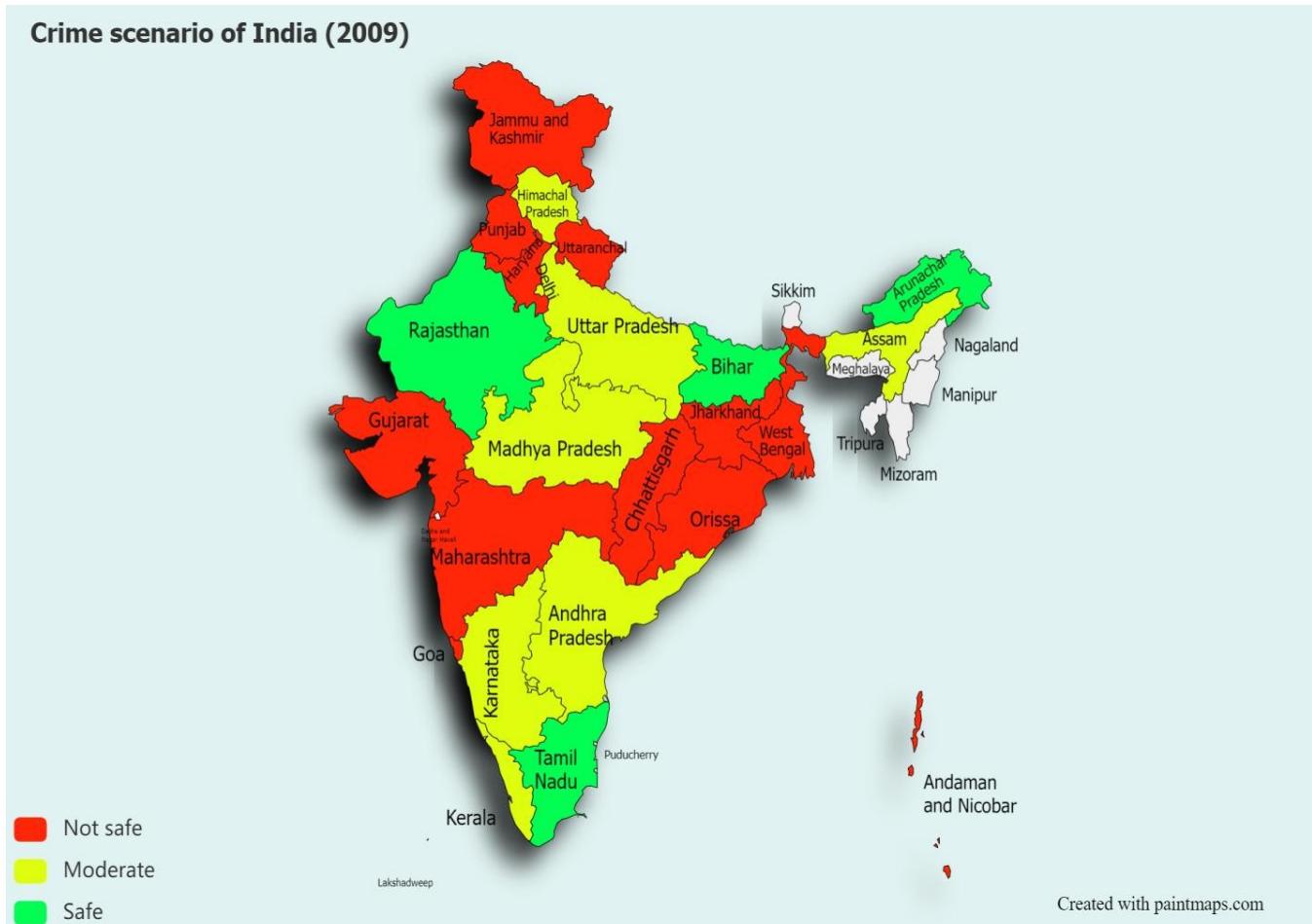
STATE	CRIME INDEX	COMMENT
CHHATTISGARH	221.2486908	NOT SAFE
GUJARAT	175.2047698	NOT SAFE
HARYANA	169.6978591	NOT SAFE
MADHYA PRADESH	114.1305846	MODERATE
MAHARASHTRA	158.2532406	NOT SAFE
RAJASTHAN	96.93850913	SAFE

VII. Overall crime scenario of India:

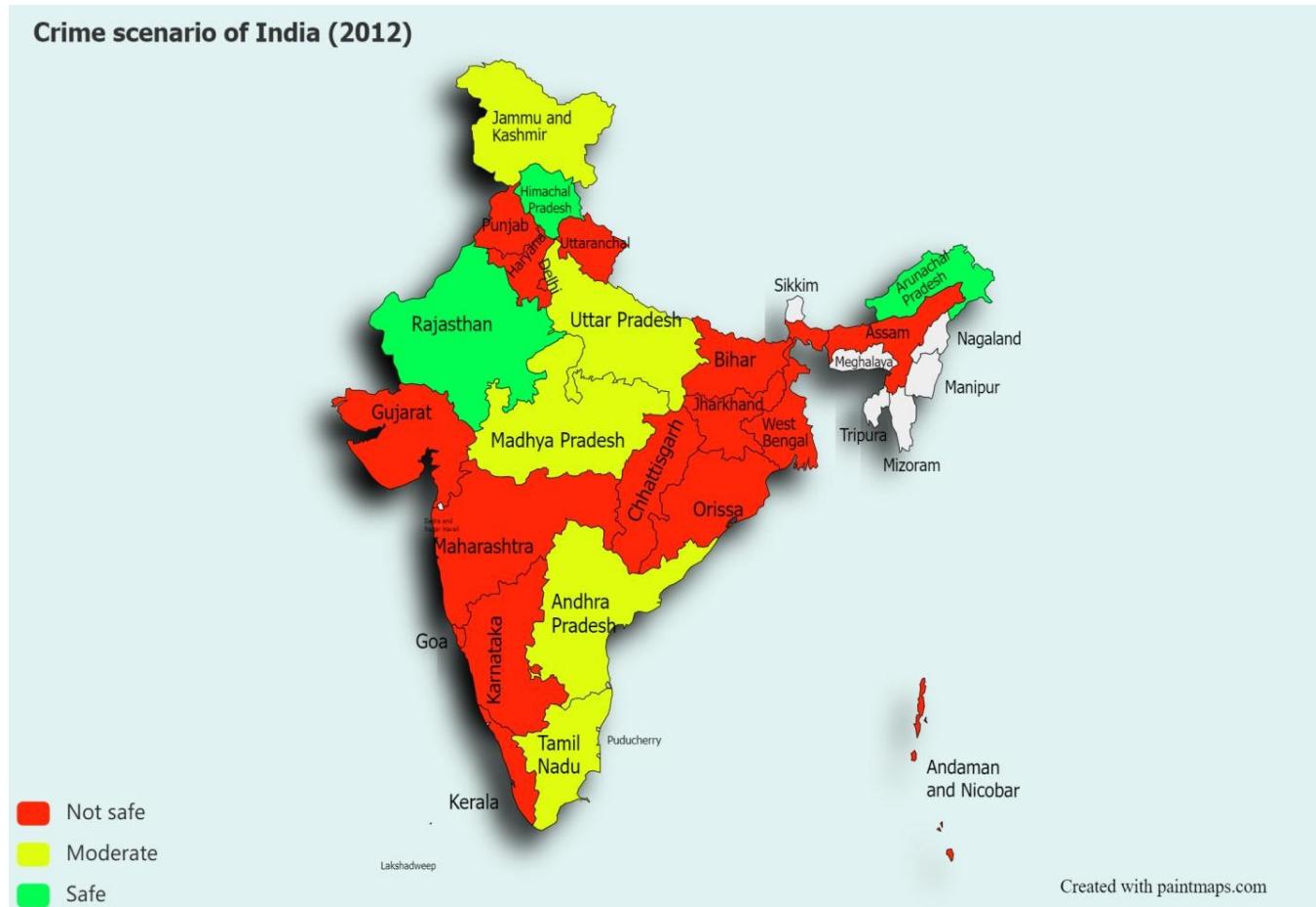


Some states are white colored due to unavailability of respective data.

Crime scenario of India (2006)



Some states are white colored due to unavailability of respective data.



Some states are white colored due to unavailability of respective data.

Conclusion

The above project on **Crime Data Analysis of the Western and Middle part of India** analyses the dependency of crime aspects (viz. Murder, Rape, Kidnapping, Robbery, Counterfeiting) in the six major states (viz. Gujarat, Rajasthan, Haryana, Madhya Pradesh, Chhattisgarh, and Maharashtra) considering the data of 12 years(2001-2012). Further, it helps to have an idea on crime observations corresponding to years and corresponding to states is similar or not. The crime index derived from the data analysis clarifies the nature of these selected states on the basis of safety and security. As the number of criminal activities in India is rapidly increasing day by day, it is the need of the hour to know the nature of the state, whether it is safe for public or not. Therefore, the indication of crime prone areas is required to be identified. This project analyses the crime data to indicate the crime prone areas. As a result, it can be used as a tool to study the law and order situation of those states during the period of 2001-2012. So, the above project is not only useful to understand the crime index of those states during the specified time period of 2001-2012, but also helpful in understanding of required steps to lower the criminal activities.

Bibliography

The sites which are used in the making of

this project- www.wikipedia.org

Link-<https://www.kaggle.com/datasets/rajanand/crime-in-india>

www.howtostats.com

The books that are used-

Fundamentals of Statistics Volume 1 & Volume 2 by Gun Gupta

and Dasgupta The source of this data-

Link-<https://www.kaggle.com/datasets/rajanand/crime-in-india>

Acknowledgement

I am indebted to number of person for helping me in the preparation of this project. Firstly, Dr. Manas Kabi, Principal, Asutosh College, University of Calcutta. Without whose help I couldn't have been a part of this prestigious college.

I owe a deep debt of gratitude to my supervisor prof. Dr. Shirsendu Mukherjee for necessary guidance, for this presentation of this dissertation, valuable comments and suggestions. I am extremely grateful to him for me the necessary stimulus, support and valuable time.

Special thanks to Prof. Dr. Dhiman Dutta, Head of the Department of Statistics, Asutosh College. I am greatly indebted to prof. Parthasarathi Bera and prof. Oindrila Bose(Faculty members) often took pains and stood by me in adverse circumstances. Without their encouragement and inspiration, it was not possible for me to complete this project.

Finally, my earnest thanks go to my friends who were always beside me when I needed them without any excuses and made these three years worthwhile.

This project is not only a mere project. It is the memories spend with the whole department which has created a mutual understanding among us. There are many emotions related to this piece of work, especially respect and duty towards teachers and vice versa; educational attachment with my friends; social attachment with my college.

Table 1.3

Haryana:

YEAR	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
2001	781	398	449	397	750	37
2002	760	361	427	330	700	32
2003	702	353	388	281	639	30
2004	733	386	423	244	699	32
2005	784	461	492	390	899	31
2006	873	608	635	410	1142	35
2007	911	488	801	502	1173	33
2008	921	631	854	555	1184	40
2009	948	603	916	679	1166	35
2010	1005	720	963	734	1414	29
2011	1062	733	959	638	1466	18
2012	991	668	1349	711	1424	20

Table 1.4

Madhya Pradesh:

YEAR	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
2001	2425	2851	956	1764	3395	19
2002	2395	2891	964	1829	3178	54
2003	2227	2738	887	1578	2729	38
2004	2379	2875	832	1526	2479	36
2005	2405	2921	847	1557	2573	51
2006	2309	2900	808	1770	2308	38
2007	2244	3010	922	1975	2648	38
2008	2322	2937	929	2234	2768	47
2009	2386	2998	1036	2270	2409	27
2010	2423	3135	1187	1919	2606	33
2011	2511	3406	1288	1952	2080	12
2012	2373	3425	1302	1940	1951	31

Table 1.5

Maharashtra:

YEAR	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
2001	2839	1302	985	2239	6719	265
2002	2768	1352	953	2136	6360	176
2003	2772	1268	926	2096	5714	261
2004	2696	1388	1045	2229	5980	218
2005	2621	1545	1194	2411	6779	385
2006	2656	1500	1261	2574	7453	317
2007	2693	1451	1312	2770	7993	270
2008	2795	1558	1379	3031	9388	405
2009	2653	1483	1286	3314	8030	478
2010	2744	1599	1508	3721	8412	389
2011	2818	1701	1669	4249	8556	351
2012	2712	1839	1583	6949	8860	291

Table 1.6

Rajasthan:

YEAR	MURDER	RAPE	KIDNAPPING	ROBBERY	RIOTS	COUNTERFEITING
2001	1259	1049	2718	889	11214	89
2002	1276	1051	2587	734	7178	63
2003	1262	1050	2238	655	4052	58
2004	1279	1038	2384	673	3101	65
2005	1221	993	1993	590	2290	88
2006	1209	1085	1970	631	1767	56
2007	1303	1238	2177	778	1626	56
2008	1297	1355	2358	829	1390	51
2009	1395	1519	2870	886	1145	59
2010	1421	1571	2985	872	986	36
2011	1461	1800	3204	727	751	56
2012	1461	2049	3243	807	573	49