A comparative study on Cyber Crime between two vast countries of the World – A Statistical Investigation

Report submitted to

SISTER NIVEDIA UNIVERSITY

For the award of the degree

Bachelor of Science in Statistics

By

Jeet Polley

***Enrollment Id**: 2111209001003*

***Registration No**: 210010032663*

Under the supervision of

Dr. Anindita Ghosal

Head of the Department

Department of Statistics

Sister Nivedita University

Newtown, Kolkata, West Bengal

CONTENTS

Sl. No.	Topic	Page No.
1	Abstract	1
2	Introduction	2
3	Objective	2
4	Importance of the study	3
5	Data Source	4
6	Methodology	5
7	All the Datasets	6-8
8	Analysis	9
8.1	Graphical Representation	9-16
8.2	Analysis Regarding ANOVA Model	16-24
8.3	Forecast Cybercrime cases in India	25-28
8.4	Forecast Cybercrime cases in USA	29-32
8.5	Comparison between India and USA	33-34
8.6	Predicting the rank of Top 10 states in India for future	34-69
9	Conclusion	70
10	Appendix	71
10.1	R Codes	71-85
11	References	86-89
12	Acknowledgement	90

SNU
SISTER NIVEDITA UNIVERSITY

B.Sc. FINAL YEAR PROJECT COMPLETION CERTIFICATE

This is to certify that **JEET POLLEY** (**Registration No: 210010032663**) has prepared the project work entitled **"A comparative study on Cyber Crime between two vast countries of the World – A Statistical Investigation** "under the supervision of **Dr. Anindita Ghosal** based on the survey of literatures in his area of interest for the partial fulfilment of the B.Sc. (Hons.) Statistics from Sister Nivedita University.

AGhosal 25/5/24

Dr Anindita Ghosal

Assistant Professor and HOD

25/5/24

Signature of External with date

AGhosal 25/5/24

Signature of Supervisor with date

ABSTRACT

"A comparative study on Cyber Crime between two vast countries of the World – A Statistical Investigation."

This project presents a comprehensive comparative study on cybercrime between two of the world's most populous and technologically advanced countries: India and the United States. The primary objective is to analyze and compare cybercrime trends across all states in both nations using an ANOVA (Analysis of Variance) model if it's can't happen then we will go for some Non-Parametric Tests and gets the proper results. In the digital age, cybercrime has become a significant challenge, affecting individuals, businesses, and governments. By examining historical cybercrime data from various **TOP 10 states** in **INDIA**, this study seeks to identify patterns and underlying factors contributing to cybercrime incidents over different years.

The research employs an ANOVA model to assess the variations in cybercrime rates among different states within each country, considering key factors such as population density, some particular facts such as **personal revenge, fraud, sexual exploitation, anger** and as well as year-to-year changes. This analysis helps in understanding the distribution and intensity of cybercrime across different regions and time periods.

Furthermore, the project utilizes time series analysis techniques to forecast cybercrime trends for **the next 10 years in both India and the USA on the basis of the past 20 years data (2003-2022)** . **And also predict the future cybercrime values for TOP 10 states for INDIA from top 15 states of India on the basis of the past 20 years data (2003-2022)**. This involves analyzing past data to forecast for upcoming future years.

By comparing the forecasted cybercrime trends for both countries, the study aims to provide a detailed understanding of potential future challenges and opportunities in combating cybercrime. The findings will highlight areas that require more stringent security measures and policies, and also identify potential opportunities for collaboration between the two countries in terms of cybersecurity practices and innovations.

In conclusion, this project seeks to contribute to the body of knowledge on cybercrime by providing a statistical investigation and comparative analysis of cybercrime trends in India and the USA. It aims to offer valuable insights for policymakers, law enforcement agencies, cybersecurity professionals, and researchers, helping them to develop effective strategies to mitigate the impact of cybercrime in the future.

INTRODUCTION

In the rapidly evolving digital landscape, cybercrime has emerged as one of the most pressing challenges facing nations around the globe. With the increasing reliance on technology for personal, professional, and governmental activities, the threat of cybercrime has expanded significantly. Cybercriminals exploit vulnerabilities in digital systems to commit a wide range of cases like anger, sexual exploitation, fraud, personal revenge and others.

This project undertakes a comparative study of cybercrime trends between **two of the world's largest and most technologically strong countries: India and the United States**. Both nations, despite their geographical and cultural differences, face some challenges in combating cybercrime. By examining and comparing the cybercrime trends within these countries, this study aims to provide a clear understanding of the dynamics of cybercrime.

OBJECTIVES

The primary objective of this research is to analyze historical cybercrime data from various states in India and the United States, using an Analysis of Variance (ANOVA) model or some non-parametric tests (if ANOVA not applicable) to identify patterns and underlying factors contributing to cybercrime incidents over different years. By doing so, the study seeks to:

A. **Identify Variations in Cybercrime Rates:** Assess the differences in cybercrime rates among different states within each country, considering key factors such as population density, internet penetration, economic conditions, and legislative measures.
B. **Forecast Future Values**: Utilize time series analysis techniques to predict cybercrime trends for the next 10 years in both India and the United States as well as for top 10 states for both the countries providing a long-term perspective on potential cybercrime challenges and opportunities.
C. **Compare Cybercrime Dynamics**: Draw comparisons between the cybercrime scenarios in India and the USA to identify commonalities and differences, which could inform better cybersecurity practices and policies.
D. **Predict TOP 10 states for INDIA**: Here we want to forecast the number of cybercrime cases in currently TOP 15 states in INDIA on the basis of the past 20 years values (2003-2022) and this forecasted results we are going to predict the TOP 10 states on the basis of cybercrime cases for INDIA.

IMPORTANCE OF THE STUDY

The significance of this study lies in its potential to contribute to the global understanding of cybercrime and its impact. In both India and the United States, cybercrime not only poses a direct threat to individuals and businesses but also has broader implications for national security and economic stability. By analyzing cybercrime data at a state level, this research can highlight specific regional vulnerabilities and strengths, offering targeted insights for policymakers and law enforcement agencies.

Furthermore, the comparative nature of this study allows for the identification of unique and shared challenges faced by India and the United States. These insights can foster international cooperation in the fight against cybercrime, encouraging the sharing of best practices and the development of coordinated strategies to enhance cybersecurity resilience.

DATA SOURCE

In the context of analyzing cybercrime data for India and the USA, the data collection procedure involves meticulous steps to ensure accuracy, reliability, and comprehensiveness. Below is an outline of the procedure, specifically for collecting the cybercrime data for India.

- The data containing number of cybercrimes for different factors for different states and union territories of India for the years 2017, 2018, 2019, 2020 is obtained from the official website of National Crime Records Bureau (NCRB), Ministry of Home Affairs, India.
- The data containing the total number of cybercrimes for different states and union territories of India for the years 2003 to 2022 is obtained from the official website of National Crime Records Bureau (NCRB), Ministry of Home Affairs, India.
- The data containing the total number of cybercrimes for different states of USA for the years 2003 to 2022 is obtained from the official website of FBI, USA.

METHODOLOGY

A. **Data Analysis via Graphical Representations**: Here we are going to represent our dataset graphically viz. dataset corresponding to the state wise cybercrime in India depends upon different factors for the years (2017-2020), dataset corresponding to the past 20 years cybercrime data for India, dataset corresponding to the past 20 years cybercrime data for USA, dataset corresponding to the past 20 years cybercrime data across all the states for India. And we are going to represent all this data set with some statistical tools like bar diagram, line diagram, pie chart &others.

B. **Test Regarding ANOVA**: In this step here, we are going to test the ANOVA procedure for the dataset corresponding to the state wise cybercrime in India depends upon different factors for the years (2017-2020) but ANOVA is applicable if and only if the test for normality (**Shapiro-wilk test**) gets accepted. If the test for normality gets rejected then we should go for some non-parametric tests like **Kruskal Walli's** and if it's also gets rejected then we should go for some pairwise comparison test.

C. **Forecasting**: Here we are going to for forecast our dataset corresponding to the past 20 years cybercrime data for India & dataset corresponding to the past 20 years cybercrime data for USA dataset for the years 2003-2022. And try to get the forecasted values for upcoming 10 years like 2023 to 2032.

D. **Comparison**: In this step here, we are going to compare the forecasted values for upcoming 10 years which we can get from the previous step and compare on the basis of that.

E. **Prediction the ranking of Top 10 states in India:** Here we are going to forecast the dataset corresponding to the past 20 years cybercrime data across all the states for India. Here we are basically select only top 15 states corresponding to the past 20 years data and forecast them. After gets the forecasted values we can rank them as TOP 10 state corresponding to forecasted values for upcoming 10 years (2023-2032).

DATASET

- **Dataset of the state wise (Top 10) cybercrime data for India corresponding to some different factors**.

States (TOP 15)	Personal Revenge	Anger	Fraud	Sexual Exploitation	Year
Karnataka	36	12	2764	55	2017
Karnataka	27	10	5441	85	2018
Karnataka	12	4	11381	90	2019
Karnataka	147	13	9680	191	2020
Uttar Pradesh	41	208	3450	117	2017
Uttar Pradesh	47	73	2351	343	2018
Uttar Pradesh	301	81	3549	430	2019
Uttar Pradesh	78	210	4674	560	2020
Maharashtra	47	80	2171	462	2017
Maharashtra	99	129	1998	724	2018
Maharashtra	48	45	3551	557	2019
Maharashtra	36	105	3413	612	2020
Telangana	14	201	529	58	2017
Telangana	19	3	732	77	2018
Telangana	11	4	2013	78	2019
Telangana	96	24	4436	85	2020
Andhra Pradesh	24	5	537	61	2017
Andhra Pradesh	34	26	733	92	2018
Andhra Pradesh	16	17	1211	84	2019
Andhra Pradesh	83	39	1149	169	2020
Assam	246	83	48	217	2017
Assam	239	46	389	113	2018
Assam	555	263	243	289	2019
Assam	654	164	242	483	2020
Jharkhand	31	11	460	14	2017
Jharkhand	16	6	783	16	2018
Jharkhand	8	0	964	15	2019
Jharkhand	4	4	1069	13	2020
Bihar	12	5	397	16	2017
Bihar	5	8	351	8	2018
Bihar	27	7	844	8	2019
Bihar	84	34	1218	32	2020
Rajasthan	2	3	331	29	2017
Rajasthan	9	11	499	60	2018
Rajasthan	17	45	938	103	2019

Rajasthan	22	10	641	67	2020
Gujarat	6	5	305	24	2017
Gujarat	17	32	401	23	2018
Gujarat	5	20	363	32	2019
Gujarat	6	31	875	37	2020

- **Dataset corresponding to the cybercrime cases for previous 20 years (2003-2022) for India**:

Year	INDIA (Under IT Act)
2003	201
2004	68
2005	179
2006	142
2007	217
2008	288
2009	420
2010	966
2011	1791
2012	601
2013	201
2014	7201
2015	8045
2016	12317
2017	13635
2018	18495
2019	30729
2020	29633
2021	27427
2022	31905

- **Dataset corresponding to the cybercrime cases for previous 20 years (2003-2022) for USA**:

Year	No. of cases in Usa Under IC3
2003	124509
2004	207449
2005	231493
2006	207492
2007	206884
2008	275,284
2009	336,655
2010	303,809
2011	314,246
2012	289874
2013	262813
2014	269422
2015	288012
2016	298728
2017	301,580
2018	351,937
2019	467,361
2020	791,790
2021	847,376
2022	800,944

- **Dataset corresponding to the past 20 years cybercrime data across top 15 states for India**:

Year	Karnatak	ar Prad	Assam	aharash	ajasth	elanga	arkhan	mil Na	hra Pra	Kerala	Haryana	Odisha	aya Pr	Gujara	est Ben
2003	0	2	0	17	0	0	0	10	107	1	1	2	0	29	0
2004	14	2	0	17	0	0	0	14	8	2	0	1	0	2	0
2005	38	4	1	26	18	0	0	22	14	3	8	6	0	2	0
2006	27	0	1	35	4	0	0	8	14	12	1	12	5	5	6
2007	40	5	0	49	16	0	0	10	16	38	0	0	6	1	2
2008	57	2	1	37	4	0	0	21	25	65	0	3	9	17	0
2009	97	14	2	53	27	0	0	18	30	64	0	2	16	20	13
2010	153	32	18	142	52	0	0	52	105	148	1	7	30	35	49
2011	151	101	31	306	122	0	8	37	349	227	42	7	90	52	43
2012	25	44	0	90	7	0	25	2	25	43	116	13	55	10	113
2013	0	2	0	17	0	0	0	10	107	1	1	2	0	29	0
2014	1010	1659	379	511	542	688	93	146	171	401	135	49	148	105	316
2015	1414	2161	483	348	639	472	180	126	393	248	208	43	162	103	295
2016	1101	2639	696	2380	941	593	259	144	616	283	401	317	258	362	478
2017	3152	4490	941	586	950	455	530	173	426	199	382	60	238	104	398
2018	5777	5513	1617	518	775	585	687	218	417	258	382	217	448	250	181
2019	12007	9353	1989	551	1074	###	1015	268	340	218	346	410	447	294	181
2020	10740	9131	2827	699	616	306	967	535	323	346	490	720	370	325	76
2021	8125	7586	3840	537	596	655	832	831	171	460	414	730	297	444	62
2022	12549	8952	1417	746	825	360	811	1484	418	375	429	755	407	437	91
Statew	56477	51692	14243	7665	7208	5743	5407	4129	4075	3392	3357	3356	2986	2626	2304

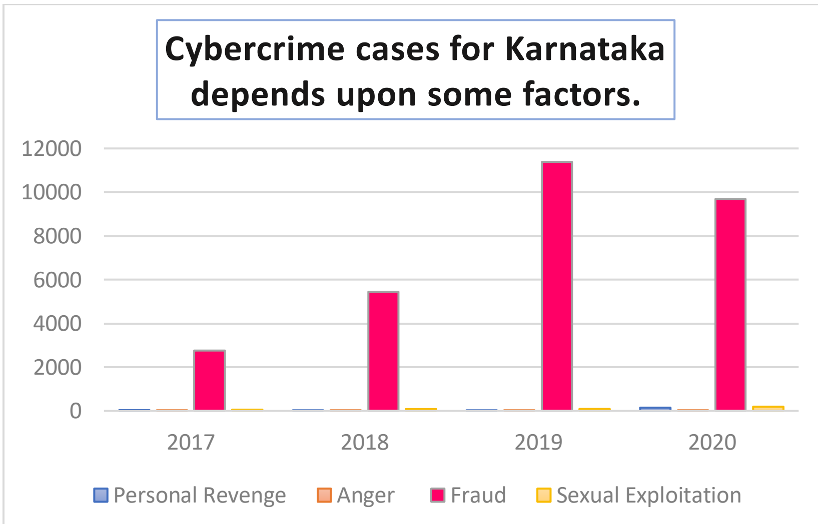
ANALYSIS

○ Data Analysis with Graphical Representation:

- **Diagram for the dataset of the state wise (Top 10) cybercrime data for India corresponding to some different factors.**

I. For the state ***KARNATAKA:***

Figure-1

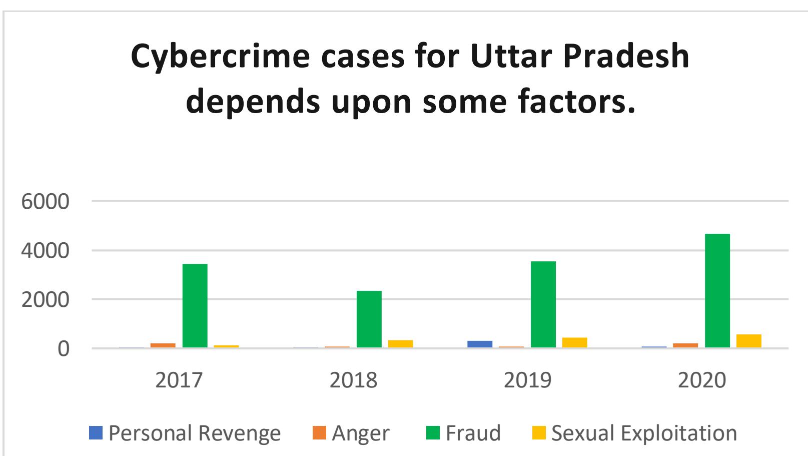


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2019.
b) Most of the cybercrimes happened due the factor fraud for every year.

II. For the state ***UTTAR PRADESH:***

Figure-2

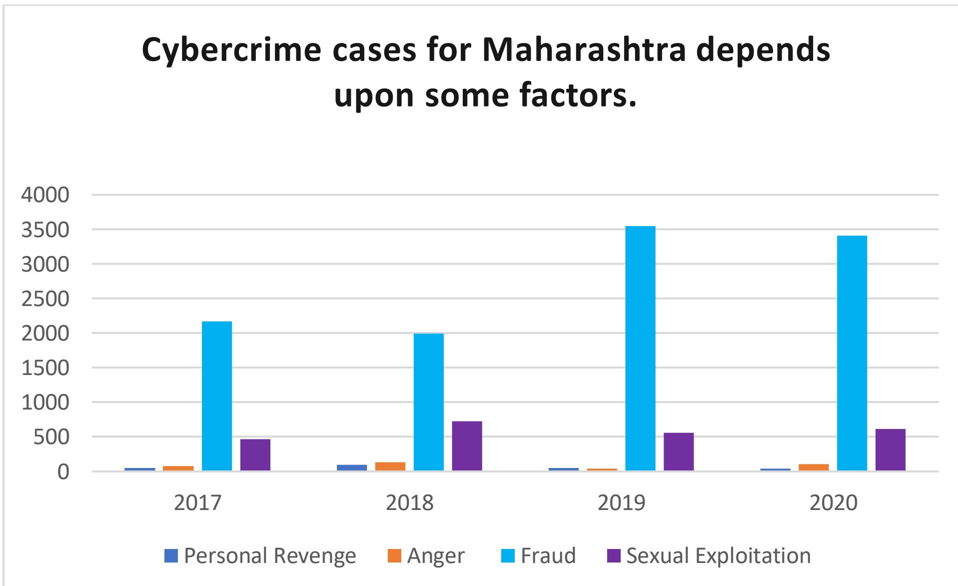


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2020.
b) Most of the cybercrimes happened due the factor fraud for every year.

III. For the state ***MAHARASHTRA***:

Figure-3

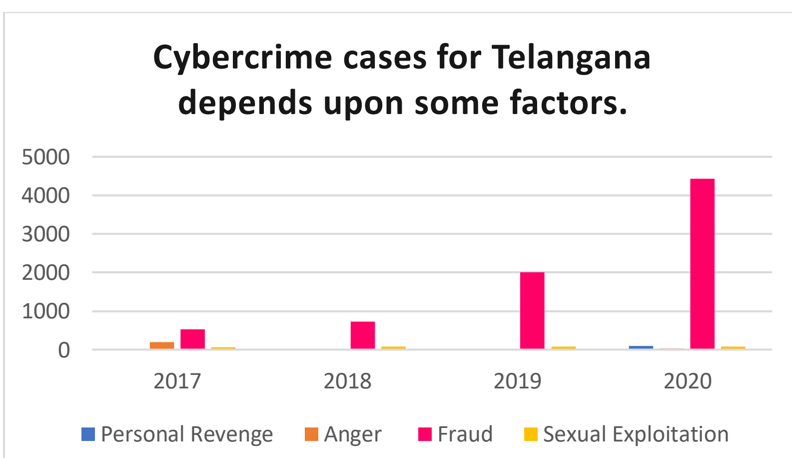


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2019.
b) Most of the cybercrimes happened due the factor fraud for every year.

IV. For the state ***TELANGANA:***

Figure-4

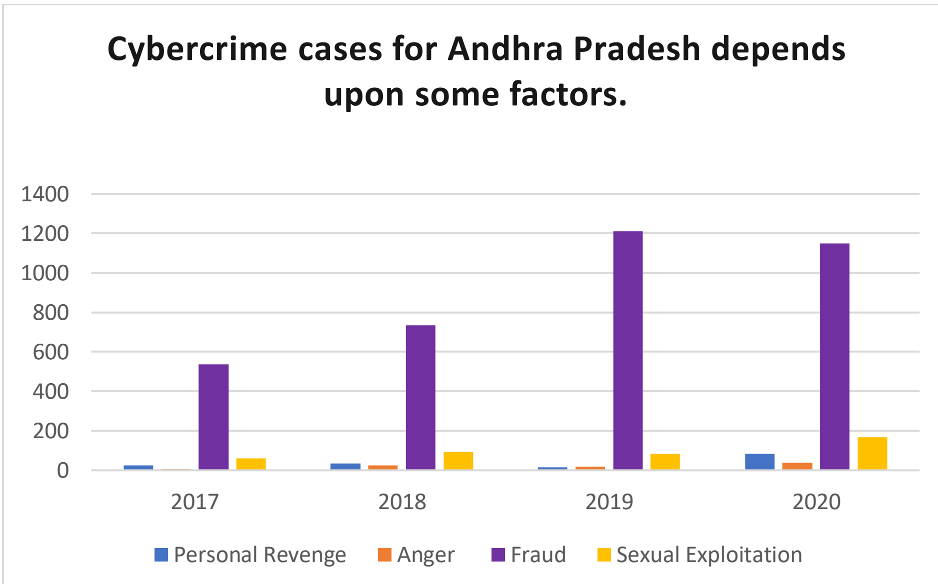


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2019.
b) Most of the cybercrimes happened due the factor fraud for every year.

V. For the state ***ANDHRA PRADESH:***

Figure-5

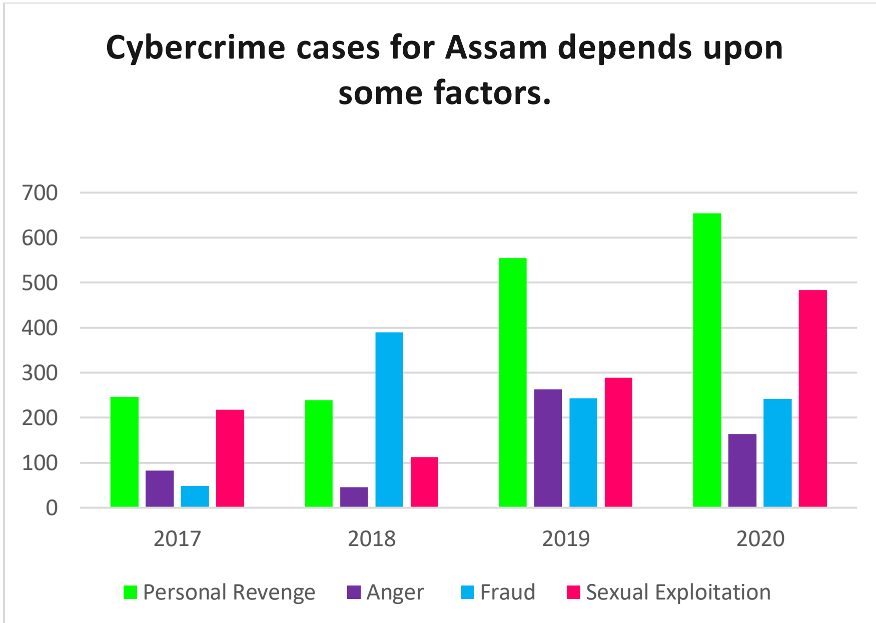


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2019.
b) Most of the cybercrimes happened due the factor fraud for every year.

VI. For the state ***ASSAM:***

Figure-6

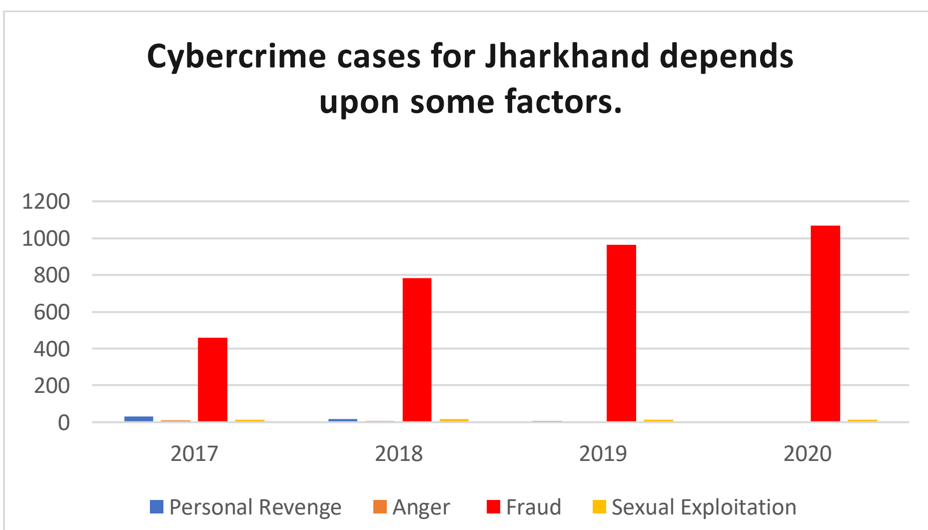


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2020.
b) Most of the cybercrimes happened due the factor personal revenge for the year 2017,2019, 2020 and due to the factor fraud in 2018.

VII. For the state ***JHARKHAND:***

Figure-7

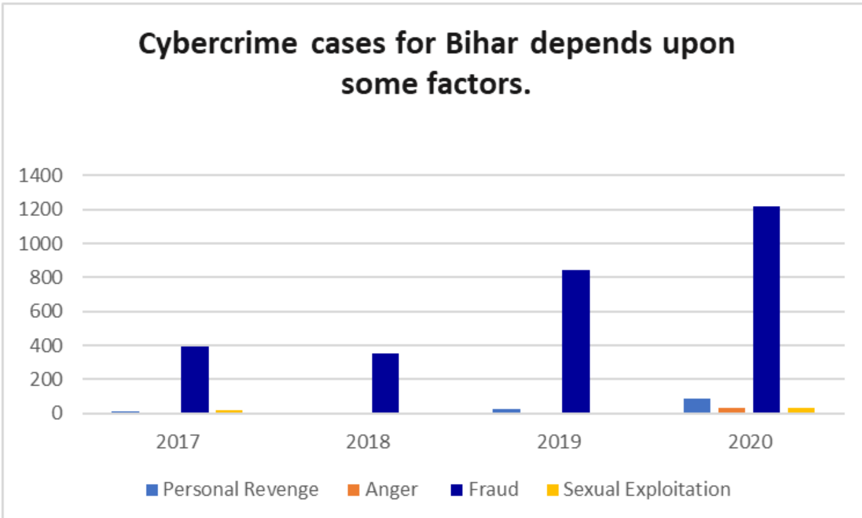


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2020.
b) Most of the cybercrimes happened due the factor fraud for every year.

VIII. For the state ***BIHAR:***

Figure-8

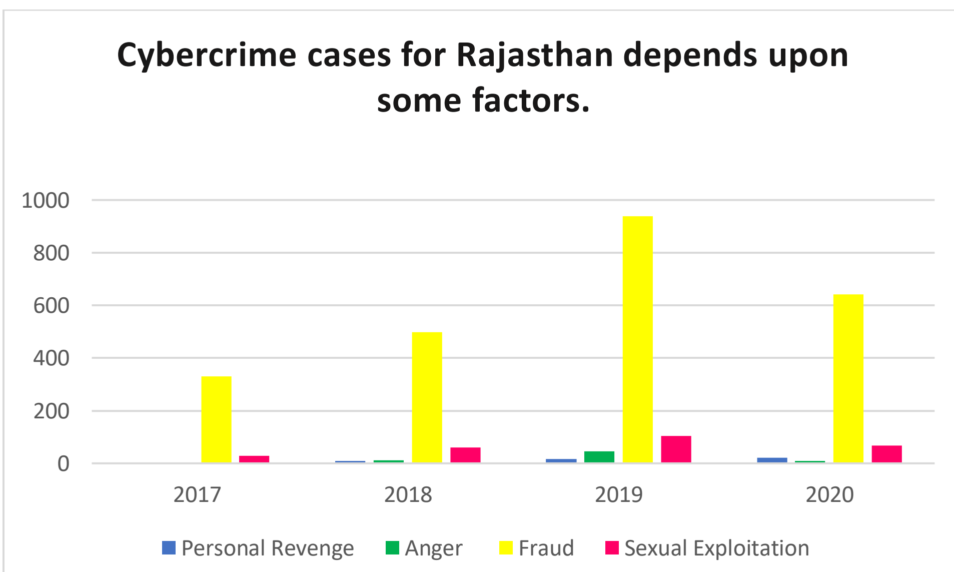


From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2020.
b) Most of the cybercrimes happened due the factor fraud for every year.

IX. For the state ***RAJASTHAN:***

Figure-9



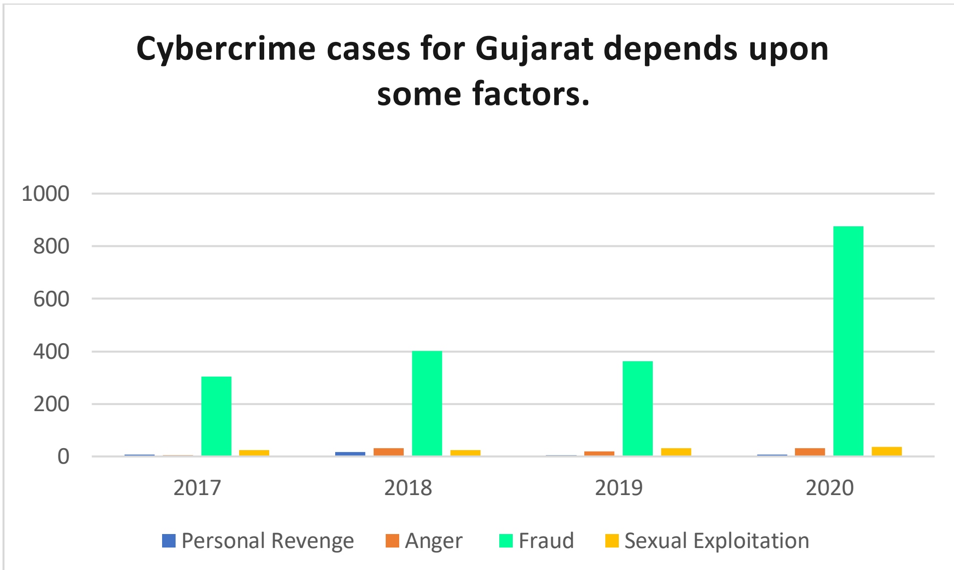
From this multiple bar diagram, we can observe that:

a) Maximum cybercrime cases are recorded in 2019.

b) Most of the cybercrimes happened due the factor fraud for every year.

X. For the state ***GUJARAT:***

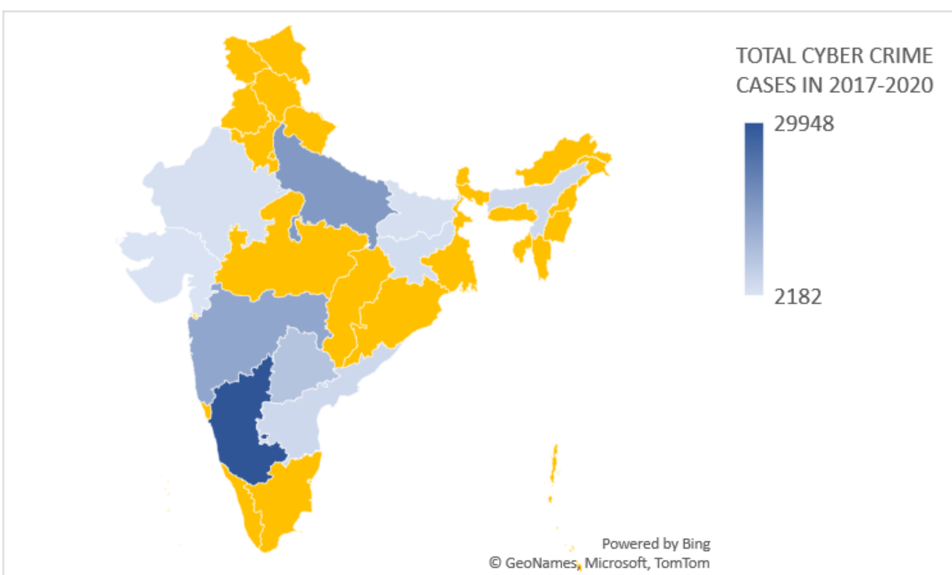
Figure-10



From this multiple bar diagram, we can observe that:

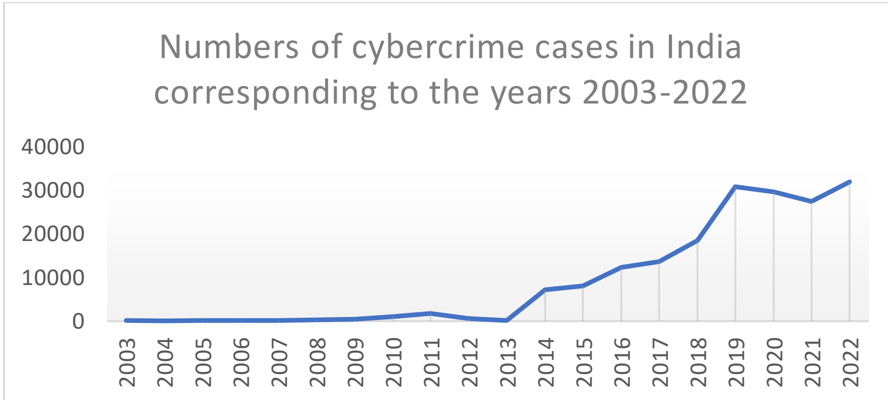
a) Maximum cybercrime cases are recorded in 2020.
b) Most of the cybercrimes happened due the factor fraud for every year.

Heatmap for cybercrime cases in India (Top 10 sates only) in 2017-2020



- Diagram for dataset corresponding to the cybercrime cases for previous 20 years (2003-2022) for **INDIA**:

Figure-11

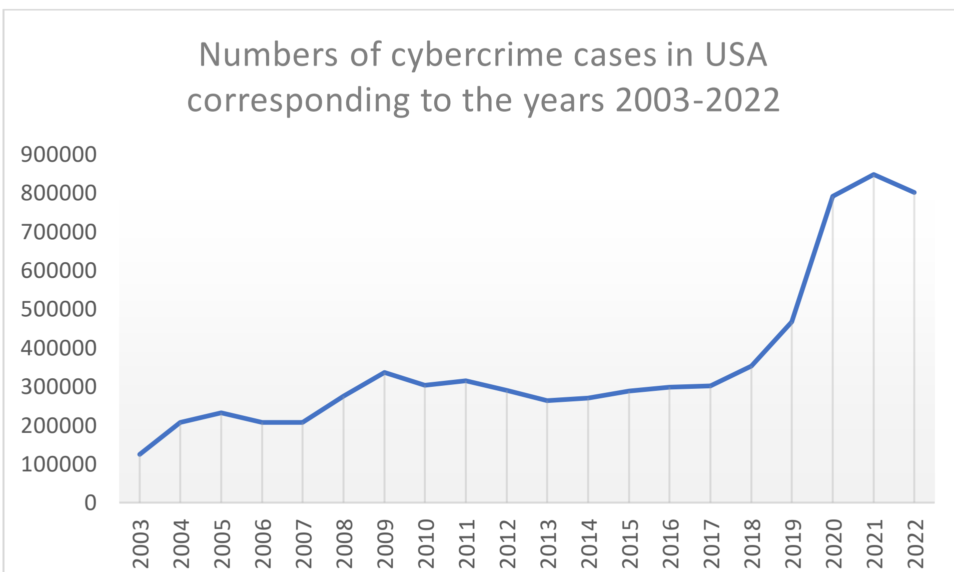


From this line diagram, we can observe that:

a) Maximum cybercrime cases are recorded in the year 2022 for India.
b) Moreover, this dataset follows an increasing trend for the corresponding past 20 years data.

- Diagram for dataset corresponding to the cybercrime cases for previous 20 years (2003-2022) for **USA**:

Figure-12



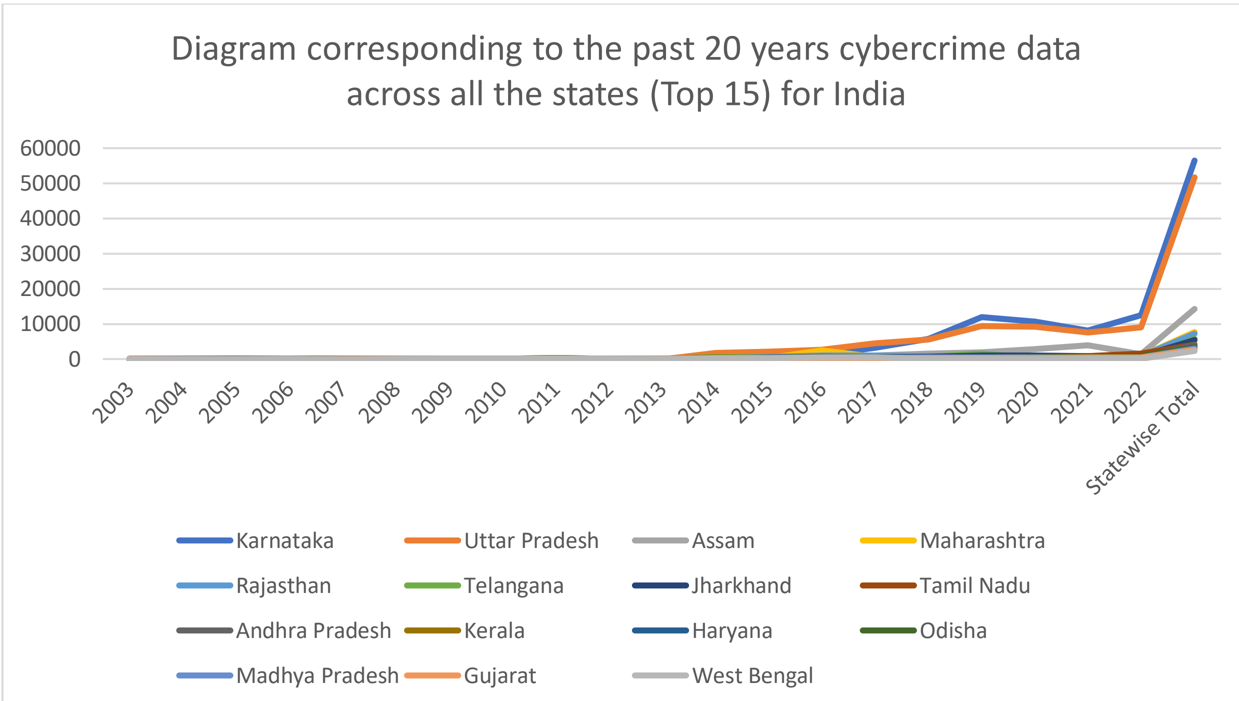
From this line diagram, we can observe that:

a) Maximum cybercrime cases are recorded in the year 2021 for India.

b) Moreover, this dataset follows an increasing trend for the corresponding past 20 years data.

- Diagram for dataset corresponding to the past 20 years cybercrime data across all the states (Top 15) for India:

Figure-13



From this line diagram, we can observe that:

a) Maximum cybercrime cases are recorded in the year 2021 for all the states of India.
b) Moreover, this dataset follows an increasing trend for the corresponding past 20 years data.
c) Here Karnataka and Uttar Pradesh are dominate the number of cybercrime cases for past 20 years cybercrime data.

Analysis regarding the ANOVA model:

Here, One-way Analysis of Variance (ANOVA) is a statistical method used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. In the context of our study, we are comparing the mean number of incidents for different types of cybercrimes (Personal Revenge, Anger, Fraud, Sexual Exploitation) across various states in India. The goal is to understand if the incidence rates of these cybercrimes differ significantly between states.

ANOVA Model

The one-way ANOVA model can be expressed mathematically as:

$$Y_{ij}=\mu+\tau_i+\epsilon_{ij}$$

where:

- Y_{ij} is the j^{th} observation in the i^{th} group, i=1(1)K, j=1(1)n_i
- μ is the overall mean,
- τ_i is the effect of the i^{th}group (state),
- $\epsilon_{ij}$is the random error associated with the j^{th} observation in the i^{th} group, assumed to be normally distributed with mean 0 and variance σ^2.

Assumptions

One-way ANOVA relies on the following assumptions to produce valid results:

Independence: The observations within each group and between groups are independent.

Normality: The data for each group should be approximately normally distributed.

Homogeneity of Variances: The variances among the groups should be approximately equal (homoscedasticity).

Hypotheses to be tested:

For each type of cybercrime, we establish the following hypotheses:

- **Null Hypothesis (H_0):** The means of the cybercrime incidents are equal across all states.

$$H_0: \mu_1=\mu_2=\cdots=\mu_k$$

where μ_i is the mean number of incidents for the i^{th}tate, and k is the total number of states.

- **Alternative Hypothesis (H_1):** At least one state has a different mean number of cybercrime incidents compared to others.

$$H_1\text{: At least one } \mu_i \text{ is different}$$

ANOVA TABLE:

Sources of variations	Degrees of freedom	SS	MS	F-statistic
Due to states	k-1	SSA= $\sum_{i=1}^{k} n_i(\bar{y}_{i0} - \bar{y}_{00})^2$	MSA=$\frac{SSA}{k-1}$	F=$\frac{MSA}{MSE}$
Due to error	n-k	SSE=$\sum_{i=1}^{k}\sum_{j=1}^{n_i} n_i(\bar{y}_{i0} - \bar{y}_{i0})^2$	MSE=$\frac{SSE}{n-k}$	
Total	n-1	TSS=$\sum_{i=1}^{k}\sum_{j=1}^{n_i} n_i(\bar{y}_{i0} - \bar{y}_{00})^2$		

where:

- k is the number of groups (states), k=10
- n_i is the number of observations in the i^{th}group, n_i=4
- N is the total number of observations, N=40
- $\bar{Y}_i$ is the mean of the i^{th}group,
- $\bar{Y}$ is the overall mean of all observations.

Apply ANOVA Test in R

Perform the one-way ANOVA test for each type of cybercrime using the aov() function in R.

Interpretation of Results

The ANOVA output provides a summary that includes the F-statistic and the p-value:

F-statistic: Indicates the ratio of the variance between the group means to the variance within the groups.

P-value: If the p-value is less than the significance level (as, 0.05), we reject the null hypothesis, indicating significant differences in the mean number of incidents of the cybercrime across all the states.

But all this procedure of testing ANOVA happens if and only if, All the assumptions of ANOVA testing are getting satisfied then only it happens.

- **Test for Normality:**

To check the assumption of normality in R package before applying ANOVA, we can use the **Shapiro-Wilk** test for each type of cybercrime data across different states. The Shapiro-Wilk test helps us determine if a data comes from a normally distributed population. We can perform the Shapiro-Wilk test in R for the **dataset of the state wise (Top 10) cybercrime data for India corresponding to some different factors**.

Interpretation of Results

p-value: The p-value from the Shapiro-Wilk test indicates whether the data significantly deviates from a normal distribution.

If the p-value is **greater than the significance level** (as, 0.05), we fail to reject the null hypothesis and conclude that the data is normally distributed.

If the p-value is **less than or equal to the significance level**, we reject the null hypothesis and conclude that the data is not normally distributed.

By examining the p-values for each group, we can determine if the normality assumption is met for each type of cybercrime across different states. This will help us to check the validity of the subsequent one-way ANOVA test.

Result for our dataset (dataset of the state wise (Top 10) cybercrime data for India corresponding to some different factors):

Factor	p-value
Personal Revenge	0.000169
Fraud	0.000000828
Anger	0.0156
Sexual Exploitation	0.00838

From the above result we can see that all the p-values which we get form the Shapiro-wilk test are less than 0.05. so, we reject the null hypothesis and conclude that the data is not normally distributed. So, here as our normality assumptions for ANOVA is violated. So, we can't apply ANOVA for this dataset.

- **Checking Normality with Graphical Representation:**

To check the normality of your data visually, you can use **box plots** and **Q-Q plots** (quantile-quantile plots). These plots help you assess the distribution of your data and identify any deviations from normality. Here we can generate all these plots via R-programing.

i. **Box Plot:** From here we can check the symmetry of the data and the presence of outliers.

Figure-14 (For Personal Revenge)

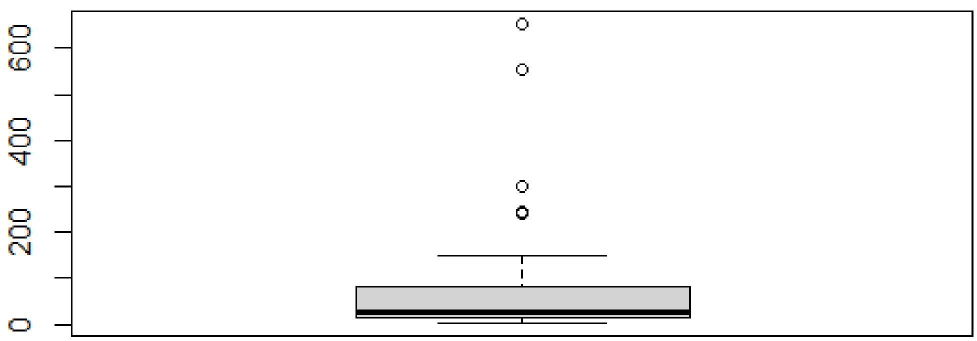


Figure-15 (For Fraud)

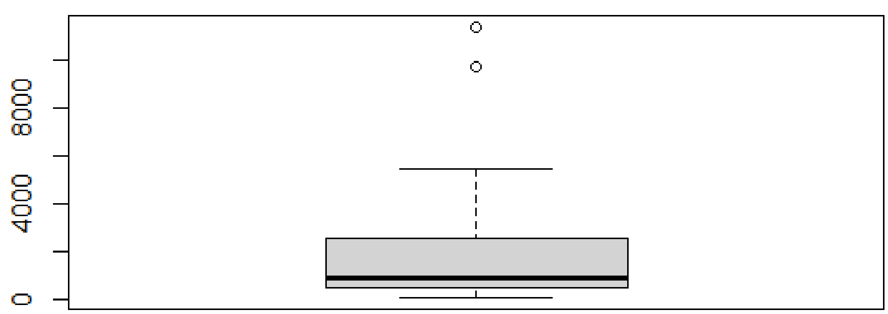


Figure-16 (For Anger)

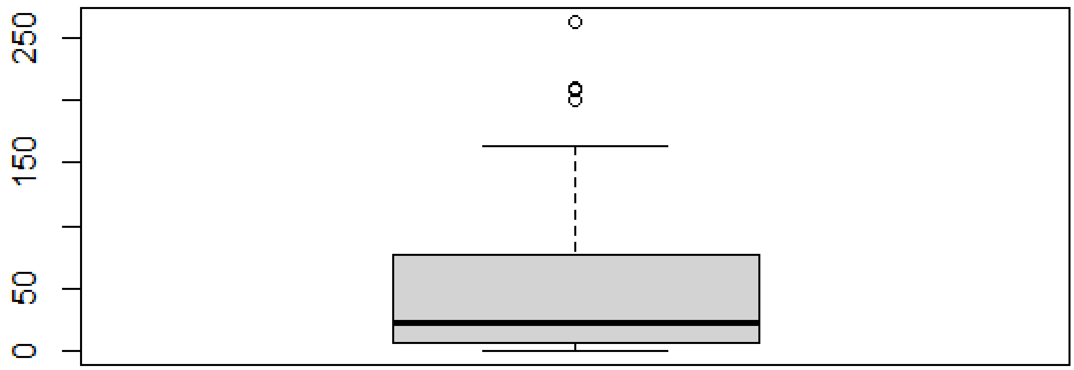
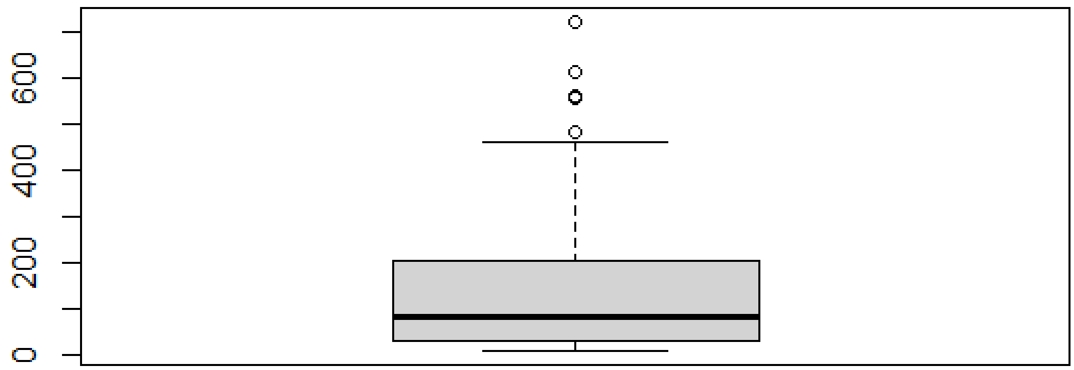


Figure-17 (For Sexual Exploitation)



From all the above plots we can conclude that,

here our data is not normally distributed, as the box plot aren't symmetric and also the median aren't near the center of the box.

ii. **Q-Q Plot:** Q-Q plots compare the quantiles of your data against the quantiles of a normal distribution.

Figure-18 (For Personal Revenge)

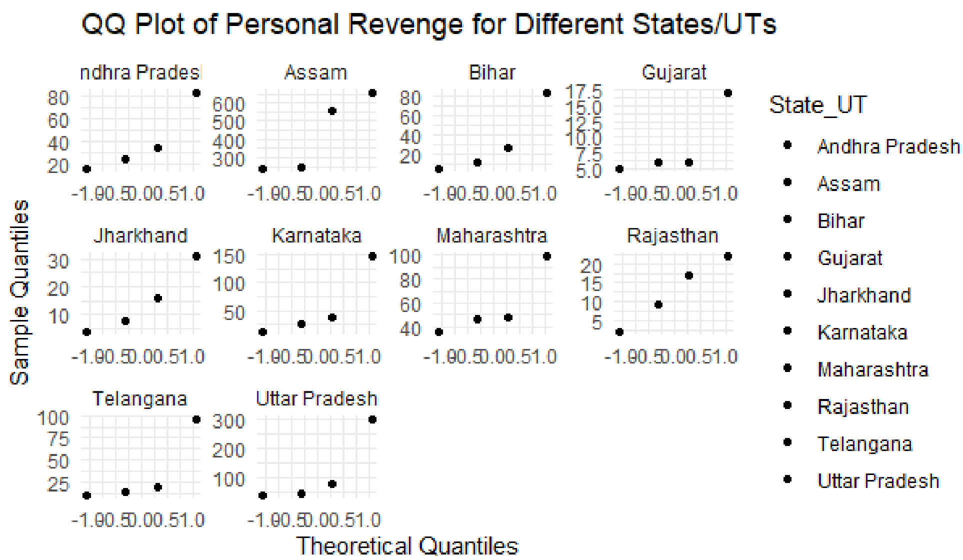


Figure-19 (For Anger)

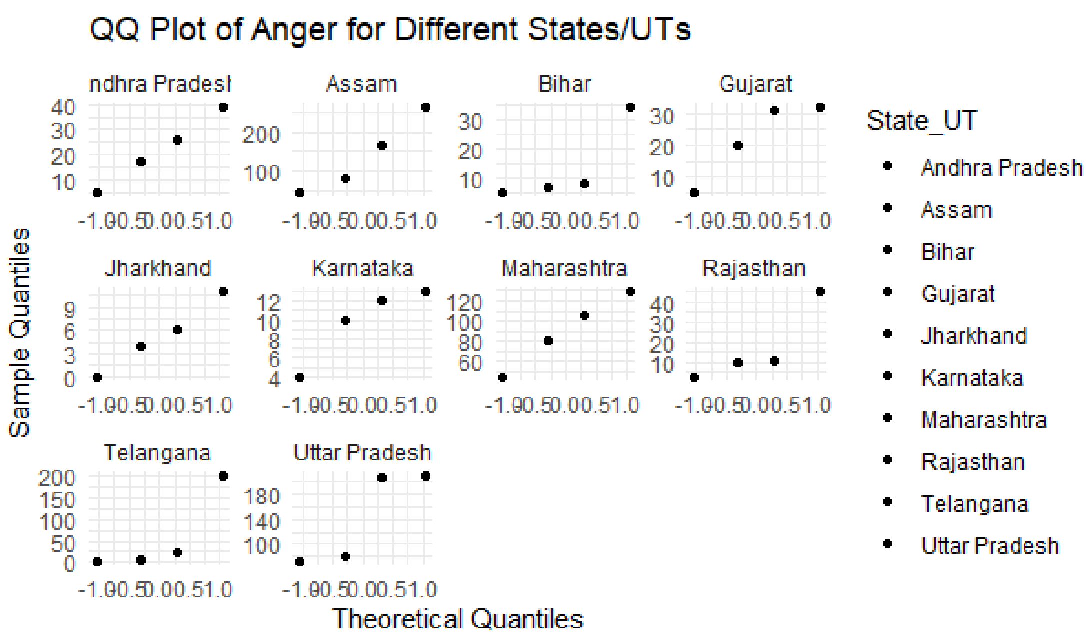


Figure-20 (For Fraud)

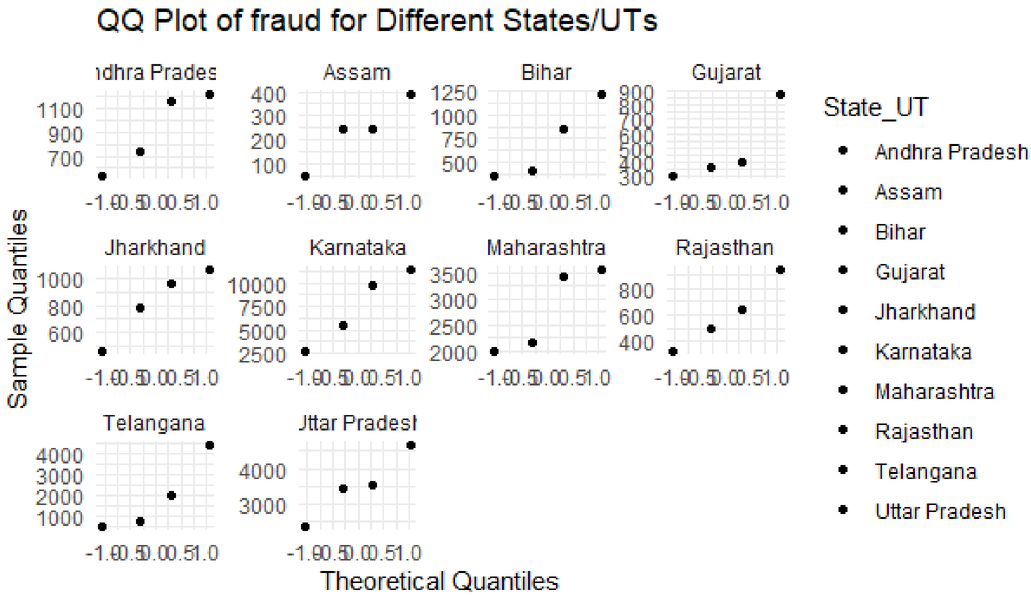
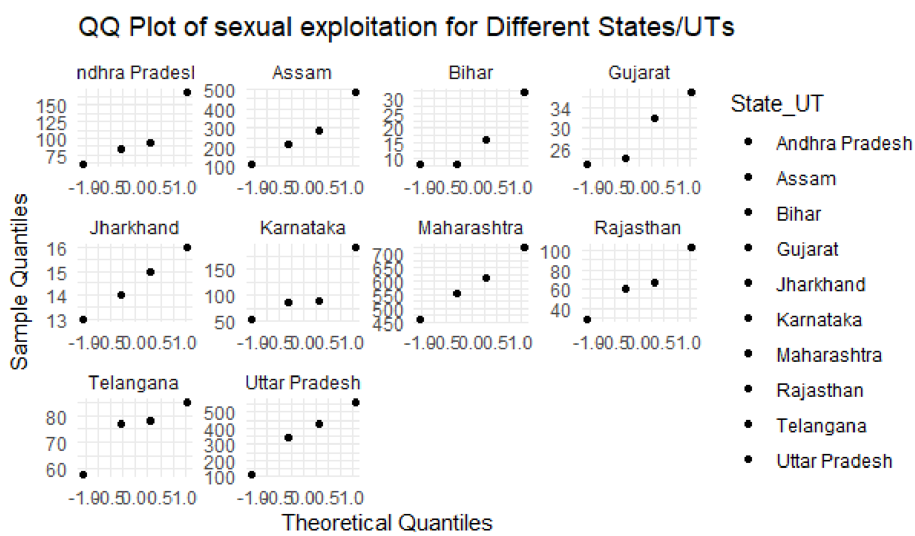


Figure-21 (For Sexual Exploitation)



From all the above plots we can conclude that, our data set it not normally distributed.

So, when the normality assumption for ANOVA is violated, it's generally acceptable to proceed directly to **non-parametric tests** like the **Kruskal-Wallis test** without checking for homoscedasticity (equal variances). Non-parametric tests do not require the assumption of normality and are less sensitive to differences in variances across groups.

Kruskal-Walli's test: We can use the Kruskal-Walli's test to compare the distributions of our data across different groups. We can perform the Kruskal-Wallis test for our dataset in R.

Interpretation of Results:

p-value: The p-value from the Kruskal-Walli's test indicates if there is a significant difference in the distributions across groups.

If the p-value is less than 0.05, it suggests significant differences among groups.

If the p-value is greater than 0.05, it suggests no significant differences among groups.

Result for our dataset (dataset of the state wise (Top 10) cybercrime data for India corresponding to some different factors):

Factor	p-value
Personal Revenge	0.00294
Anger	0.00354
Fraud	0.000391
Sexual Exploitation	0.0000636

From the above result we can conclude that,

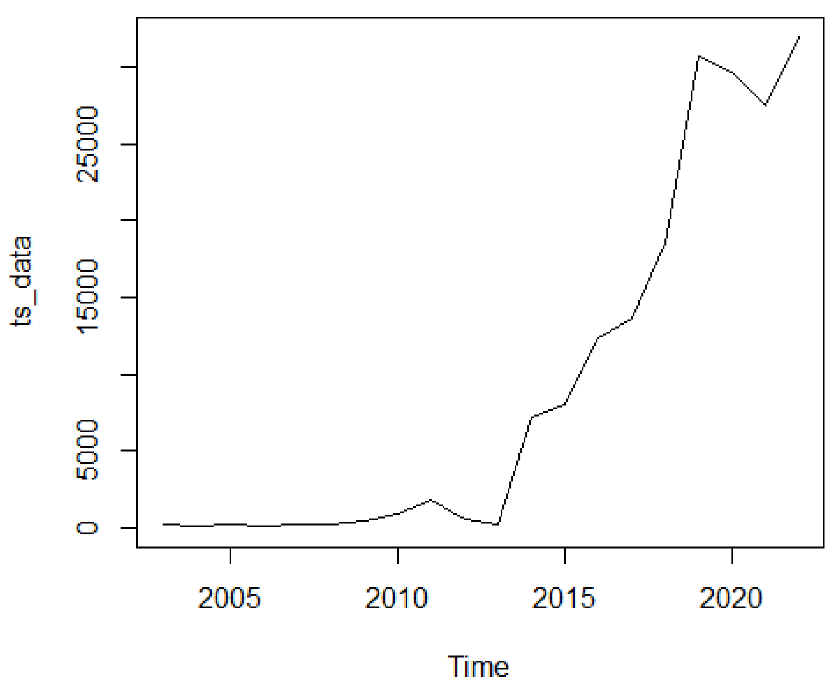
For our dataset, it suggests there is significant difference present among all the groups (States).

○ Forecasting cybercrime cases for India:

Now, we will try to forecast the data of future values of cybercrimes based on allover India using the future the data of past 20 years data. For forecasting future time points, we have considered the dataset corresponding to the cybercrime cases for previous 20 years (2003-2022) for India. And we are trying to predict the future values for the next 10 years (2023-2032).

Now, for the said purpose, we need to find the order of moving average (MA) process and the order of autoregression (AR) process. We have to plot the autocorrelation function (ACF) to find the order of moving average (MA) process. And also plot the partial autocorrelation function (PCF) to find the order of moving average (AR) process. We are trying all these things through R-programing.

Figure-22 (Time Series Plot)



From this time series plot we can observe that there is an increasing trend followed by this past 20 years data.

Figure-23(ACF Plot)

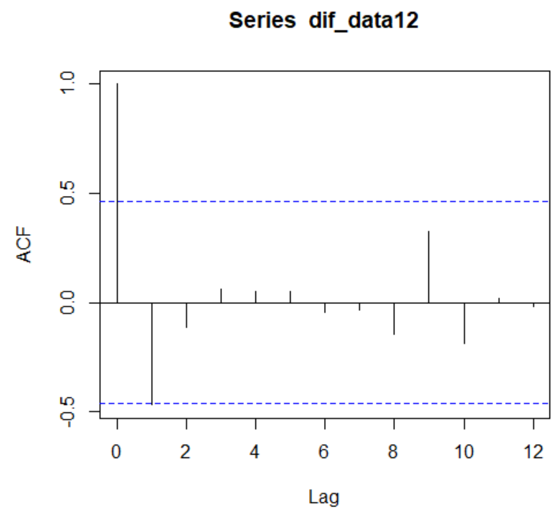
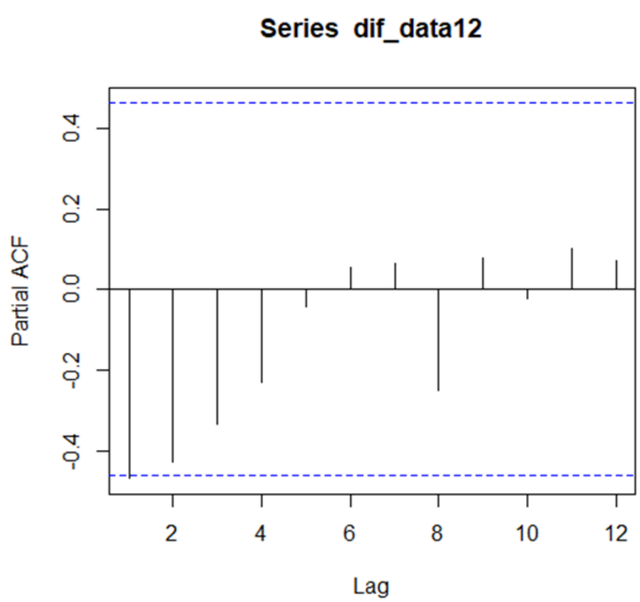
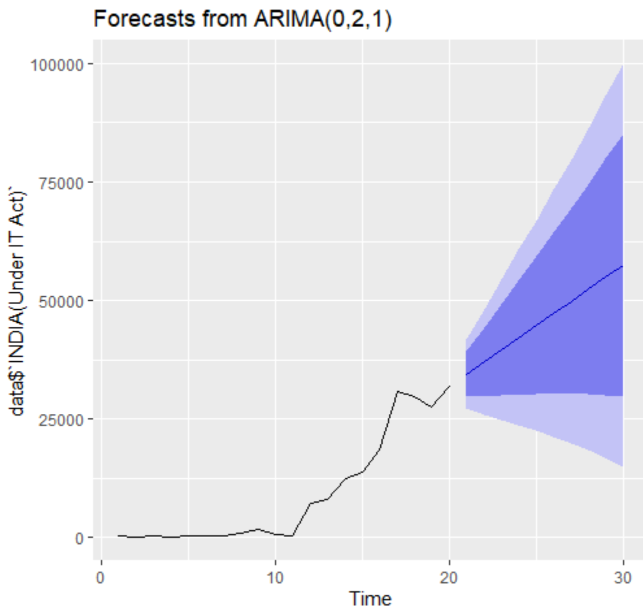


Figure-24(PACF Plot)



So, from the above plots we can understand that the fitted autoregression model is ARIMA (0,2,1).

Figure-25(Forecasting Graph)

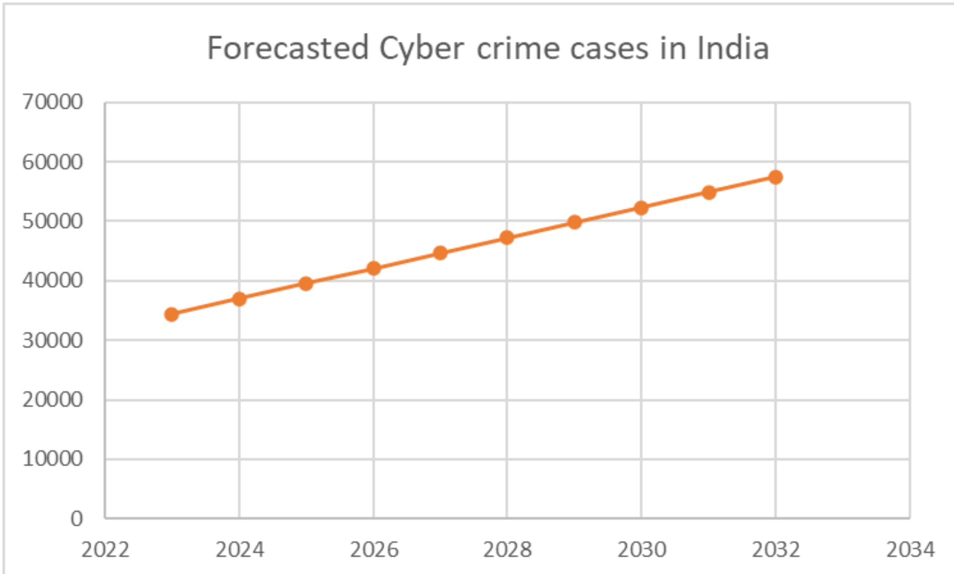


From the above forecasting plot, it can be observed the increasing trend regarding the number of future cybercrime cases in India.

Forecasted Values:

Year	Forecasted Cyber crime case:
2023	34459.37
2024	37013.74
2025	39568.12
2026	42122.49
2027	44676.86
2028	47231.23
2029	49785.61
2030	52339.98
2031	54894.35
2032	57448.72

Figure-26(Forecasted values for India)



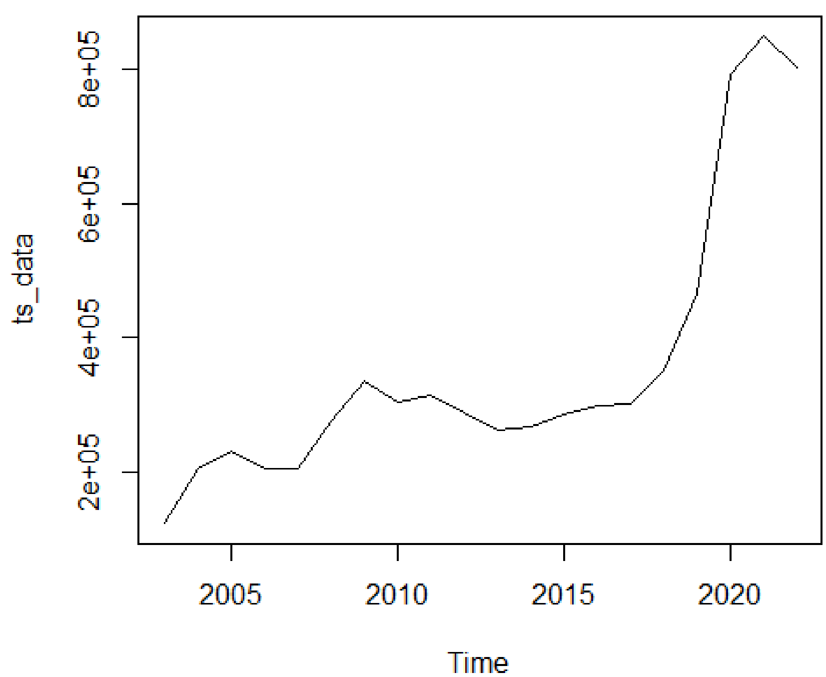
From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a strictly increasing trends.

○ Forecasting cybercrime cases for USA:

Now, we will try to forecast the data of future values of cybercrimes based on allover USA using the future the data of past 20 years data. For forecasting future time points, we have considered the dataset corresponding to the cybercrime cases for previous 20 years (2003-2022) for USA. And we are trying to predict the future values for the next 10 years (2023-2032).

Now, for the said purpose, we need to find the order of moving average (MA) process and the order of autoregression (AR) process. We have to plot the autocorrelation function (ACF) to find the order of moving average (MA) process. And also plot the partial autocorrelation function (PCF) to find the order of moving average (AR) process. We are trying all these things through R-programing.

Figure-27 (Time Series Plot)



From this time series plot we can observe that there is an increasing trend followed by this past 20 years data

Figure-28(ACF Plot)

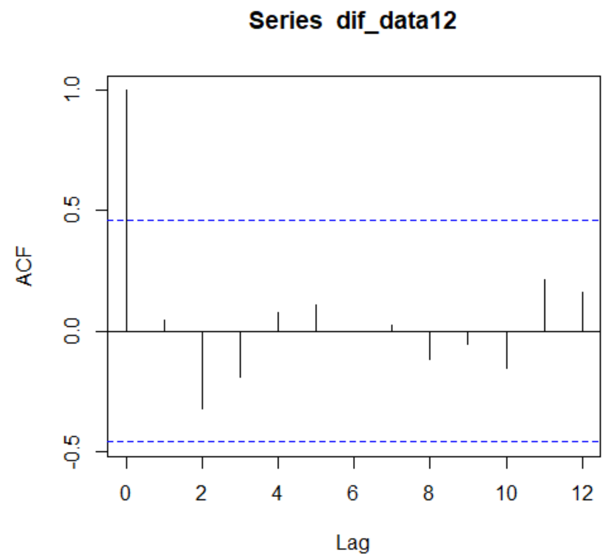
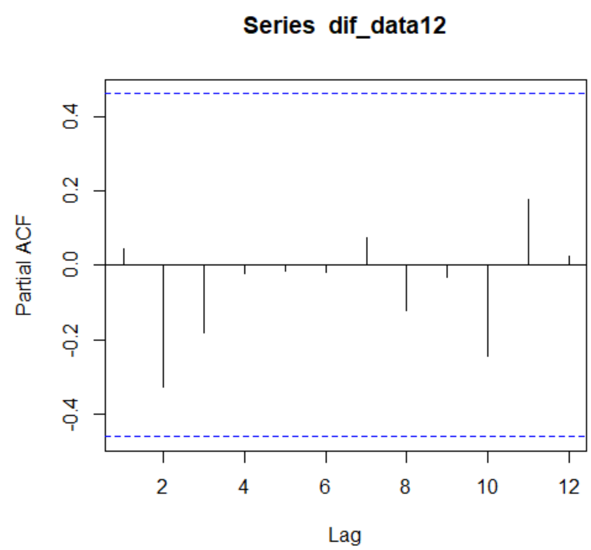
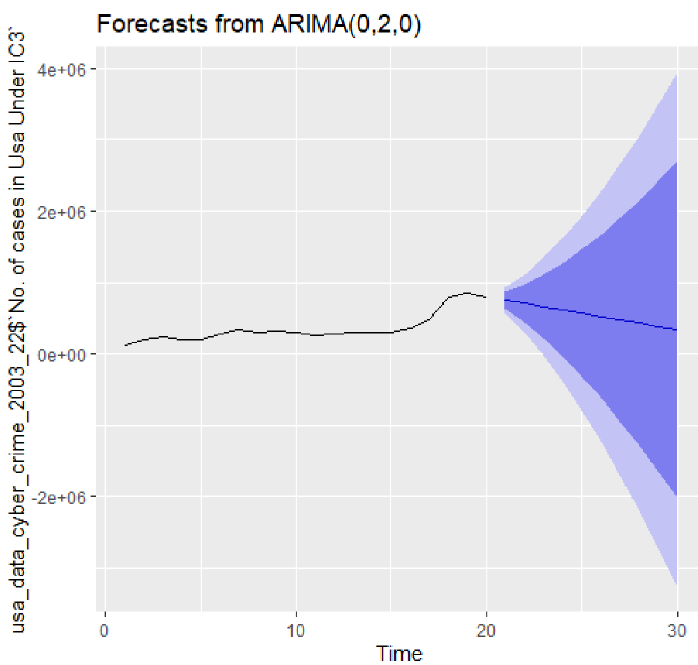


Figure-29(PACF Plot)



So, from the above plots we can understand that the fitted autoregression model is ARIMA (0,2,0).

Figure-30(Forecasting Graph)

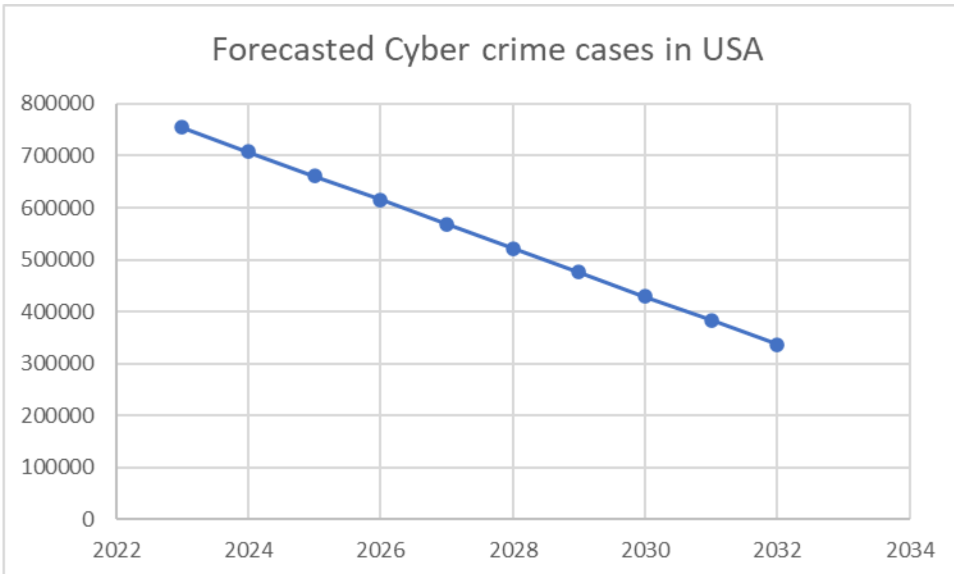


From the above forecasting plot, it can be observed the decreasing trend regarding the number of future cybercrime cases in USA.

Forecasted Values:

Year	Forecasted Cyber crime cases
2023	754512
2024	708080
2025	661648
2026	615216
2027	568784
2028	522352
2029	475920
2030	429488
2031	383056
2032	336624

Figure-31(Forecasted values for USA)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a strictly decreasing trends.

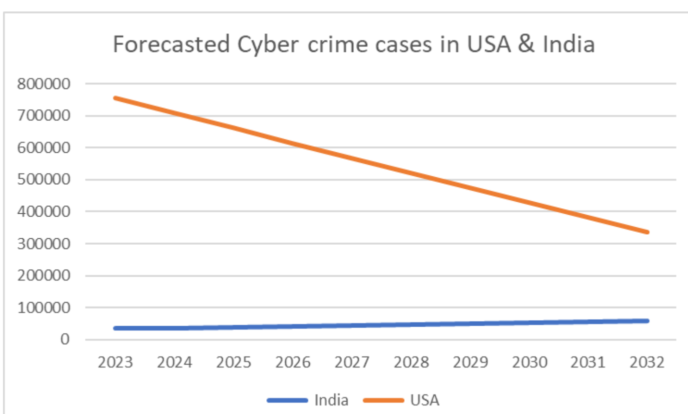
○ Comparison between India and USA:

We can compare between India and USA with respect to our forecasted future cybercrime values for upcoming 10 years (2023-2032).

Table for Comparison

INDIA				USA			
Original Values		Forecasted Values		Original Values		Forecasted Values	
Year	Cyber crime cases	Year	Forecasted Cyber crime cases	Year	Cyber crime cases	Year	Forecasted Cyber crime cases
2003	201	2023	34459.37	2003	124509	2023	754512
2004	68	2024	37013.74	2004	207449	2024	708080
2005	179	2025	39568.12	2005	231493	2025	661648
2006	142	2026	42122.49	2006	207492	2026	615216
2007	217	2027	44676.86	2007	206884	2027	568784
2008	288	2028	47231.23	2008	275,284	2028	522352
2009	420	2029	49785.61	2009	336,655	2029	475920
2010	966	2030	52339.98	2010	303,809	2030	429488
2011	1791	2031	54894.35	2011	314,246	2031	383056
2012	601	2032	57448.72	2012	289874	2032	336624
2013	201			2013	262813		
2014	7201			2014	269422		
2015	8045			2015	288012		
2016	12317			2016	298728		
2017	13635			2017	301,580		
2018	18495			2018	351,937		
2019	30729			2019	467,361		
2020	29633			2020	791,790		
2021	27427			2021	847,376		
2022	31905			2022	800,944		

Figure-32 (Line diagram for comparison)



From the above table and the line diagram we can conclude that, after 10 years from now the forecasted cybercrime cases in India shows an increasing tendency as well as the forecasted cybercrime cases in USA shows a decreasing tendency. But here after 10

years also after the decreasing trend then also the number of cybercrimes in USA is much more than the number of cybercrimes in India after 10 years after an increasing trend.

○ Prediction the ranking of Top 10 states in India:

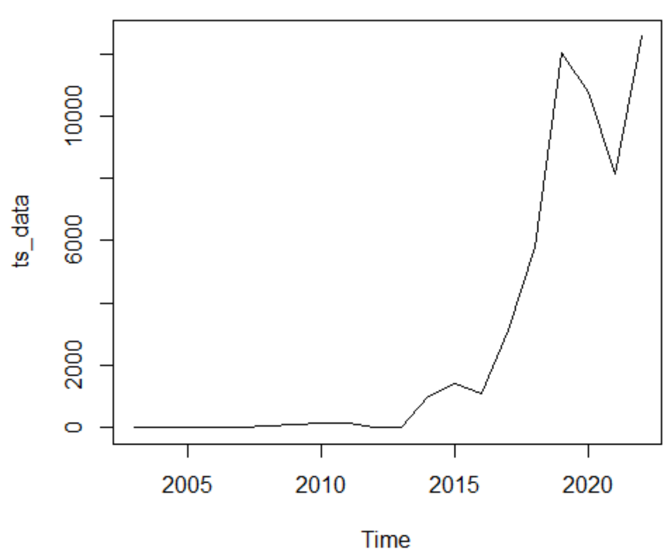
Now, we are trying to forecast the dataset corresponding to the past 20 years cybercrime data across top 15 states for India for upcoming 10 years (2023-2032). And after getting the forecasted values on the basis of the forecasted values of top 15 states we can rank them as top 10 states in India for the cybercrime cases for the years 2023 to 2032.

Now, we will try to forecast the data of future values of cybercrimes based on all 15 states separately using the future the data of past 20 years data. For forecasting future time points, we have considered the dataset corresponding to the cybercrime cases for previous 20 years (2003-2022). And we are trying to predict the future values for the next 10 years (2023-2032).

Now, for the said purpose, we need to find the order of moving average (MA) process and the order of autoregression (AR) process. We have to plot the autocorrelation function (ACF) to find the order of moving average (MA) process. And also plot the partial autocorrelation function (PCF) to find the order of moving average (AR) process. We are trying all these things through R-programing.

a) Karnataka:

Figure-33 (TS Plot)



From this time series plot we can observe that there is an increasing trend followed by this past 20 years data.

Figure-34 (ACF Plot)

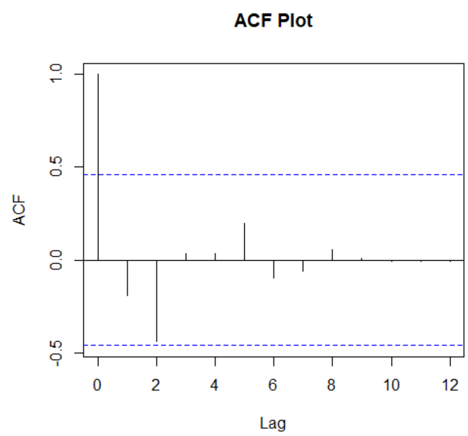
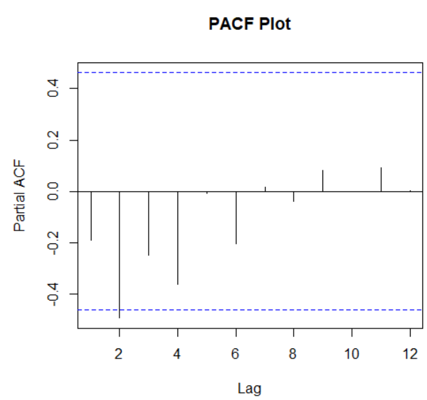
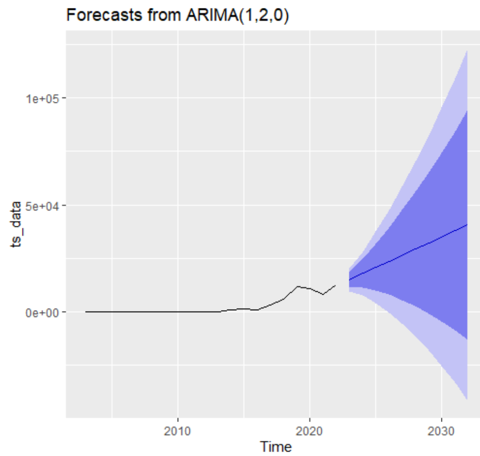


Figure-35 (ACF Plot)



So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,2,0).

Figure-36(Forecasting Graph)

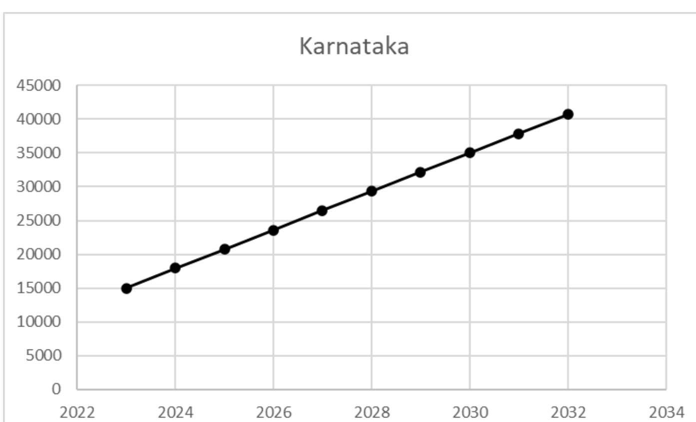


From the above forecasting plot, it can be observed the increasing trend regarding the number of future cybercrime cases in Karnataka.

Forecasted Values:

year	Karnataka
2023	14950.78
2024	17933.52
2025	20749.36
2026	23613.14
2027	26463.15
2028	29317.12
2029	32169.95
2030	35023.11
2031	37873.17
2032	40729.26

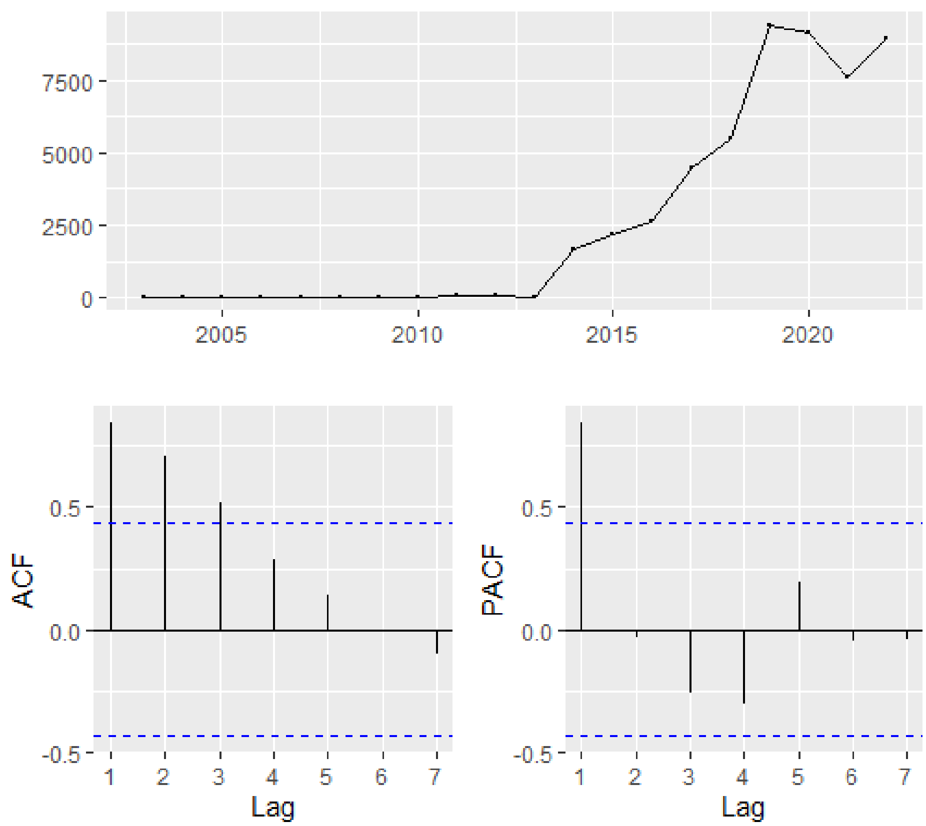
Figure-37 (line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a strictly increasing trends.

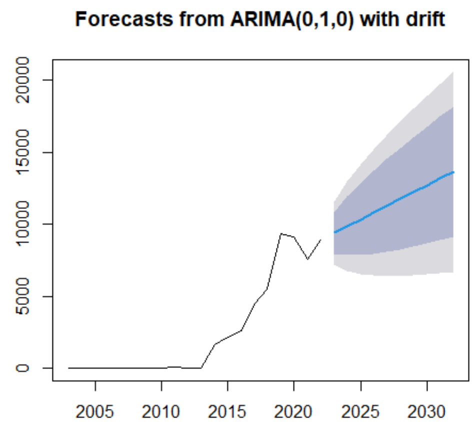
b) Uttar Pradesh:

Figure-38 (Time Series, ACF & PACF Plot)



From this time series plot we can observe that there is an increasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (0,1,0).

Figure-39(Forecasting Graph)

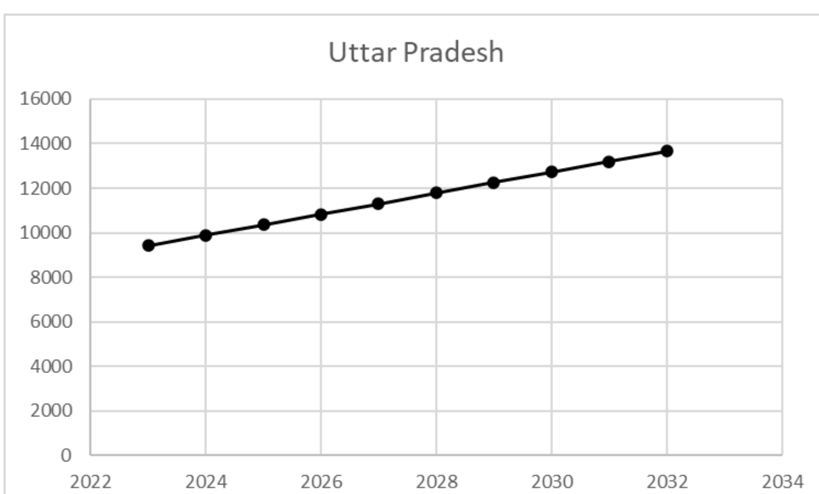


From the above forecasting plot, it can be observed the increasing trend regarding the number of future cybercrime cases in Uttar Pradesh.

Forecasted Values:

year	Uttar Pradesh
2023	9423.053
2024	9894.105
2025	10365.158
2026	10836.211
2027	11307.263
2028	11778.316
2029	12249.368
2030	12720.421
2031	13191.474
2032	13662.526

Figure-40 (line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a strictly increasing trends.

c) Assam:

Figure-41 (TS Plot)

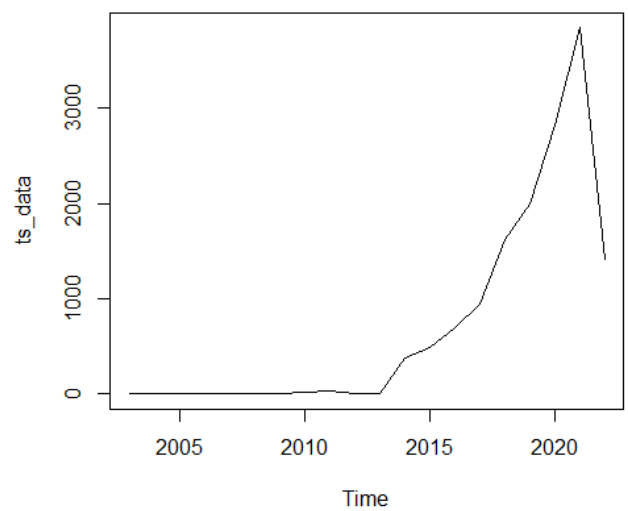


Figure-42 (ACF Plot)

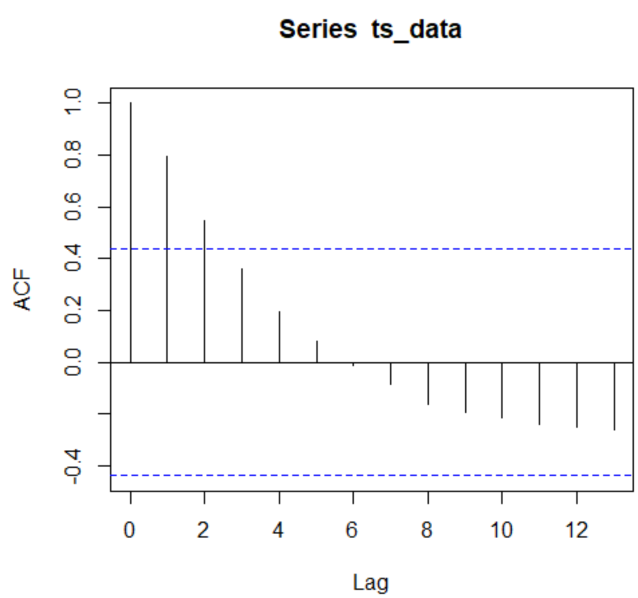
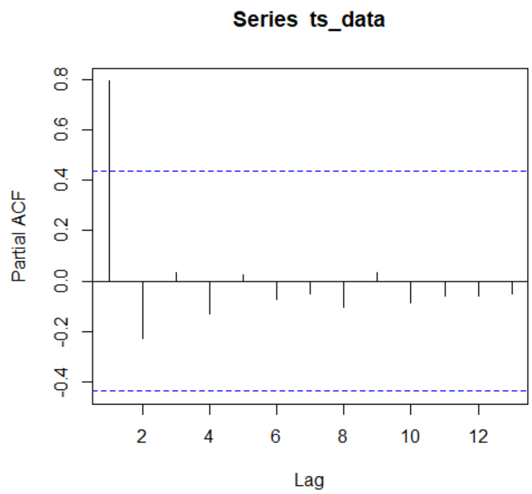
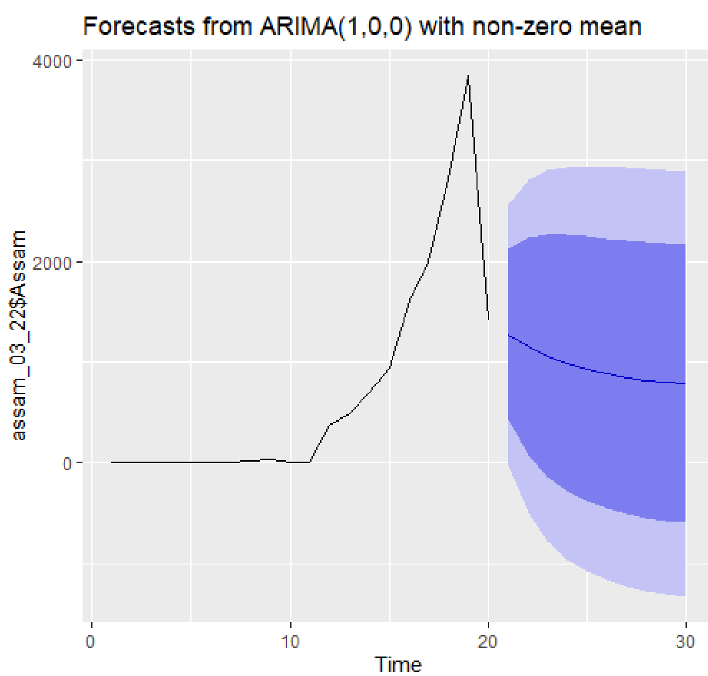


Figure-43 (PACF Plot)



From this time series plot we can observe that there is a decreasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,0,0).

Figure-44(Forecasting Graph)

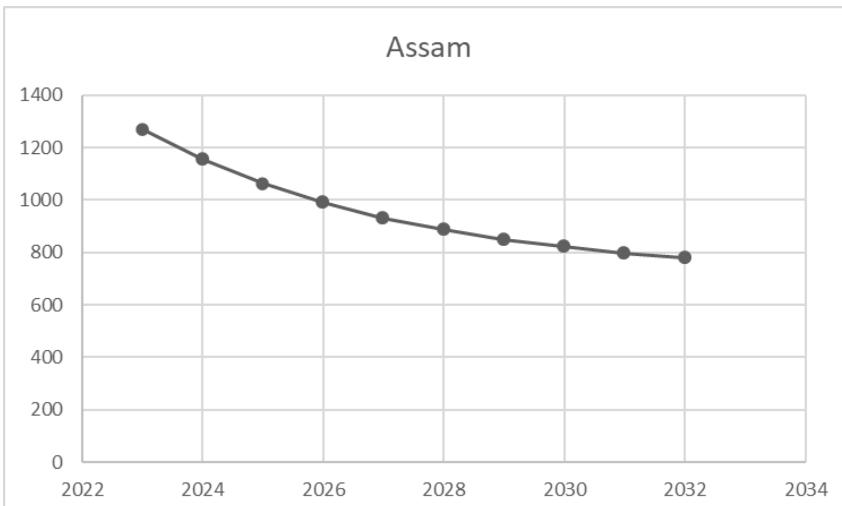


From the above forecasting plot, it can be observed the decreasing trend regarding the number of future cybercrime cases in Assam.

Forecasted Values:

year	Assam
2023	1270.1359
2024	1153.8286
2025	1061.7202
2026	988.7761
2027	931.0087
2028	885.2605
2029	849.0307
2030	820.3389
2031	797.6168
2032	779.6222

Figure-45 (line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a decreasing trend.

d) Maharashtra:

Figure-46 (TS plot)

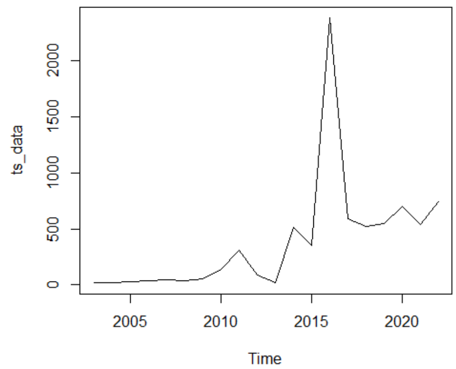


Figure-47 (ACF Plot)

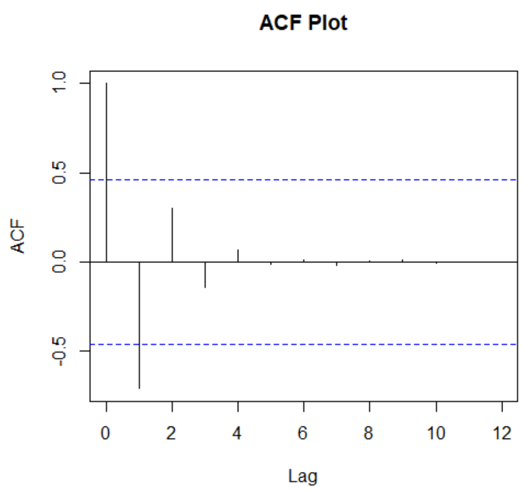
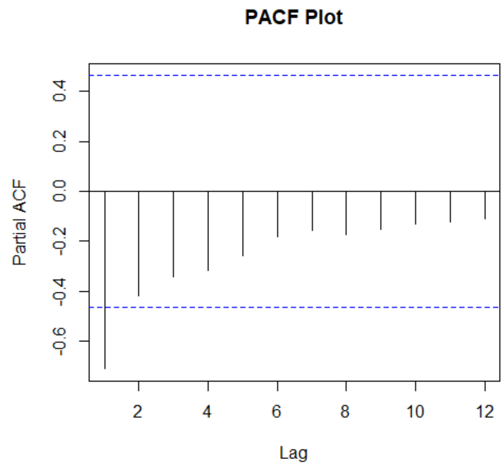
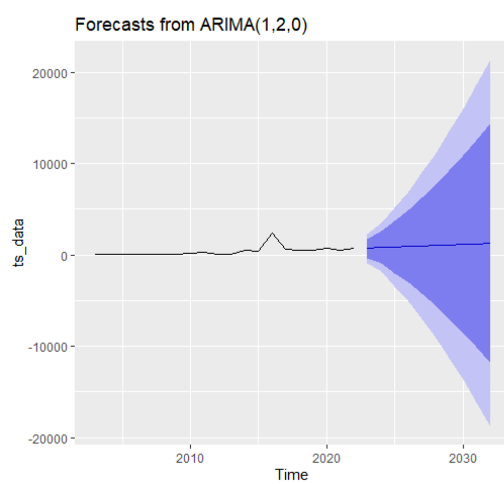


Figure-48 (PACF Plot)



From this time series plot we can observe that it is a dynamic data followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,2,0).

Figure-49(Forecasting Graph)

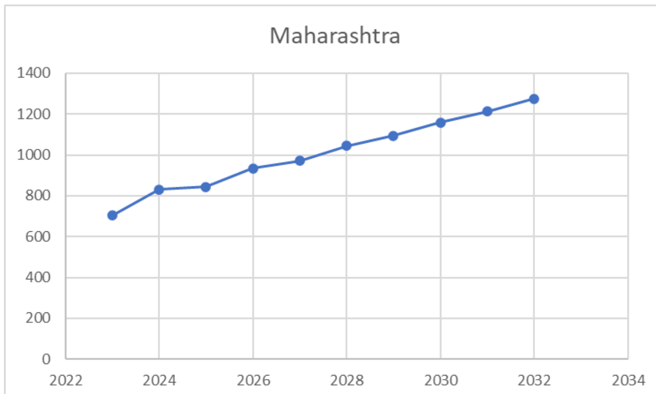


From the above forecasting plot, it can be observed the dynamic trend regarding the number of future cybercrime cases in Maharashtra.

Forecasted Values:

year	Maharashtra
2023	702.8972
2024	831.1039
2025	842.902
2026	933.8023
2027	970.9509
2028	1044.625
2029	1093.4791
2030	1159.199
2031	1213.4583
2032	1275.5053

Figure-50 (line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

e) Jharkhand:

Figure-51 (TS Plot)

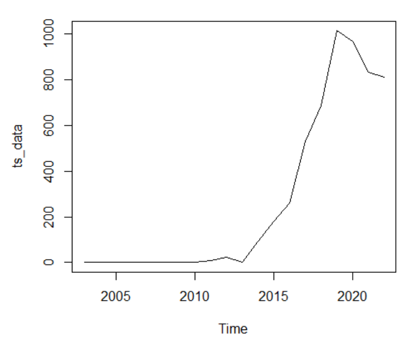


Figure-52 (ACF Plot)

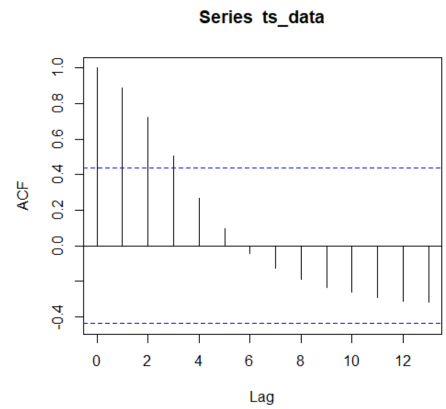
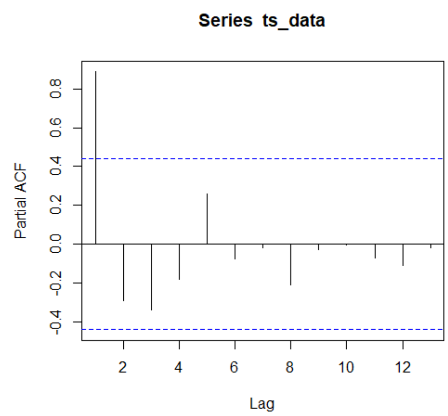
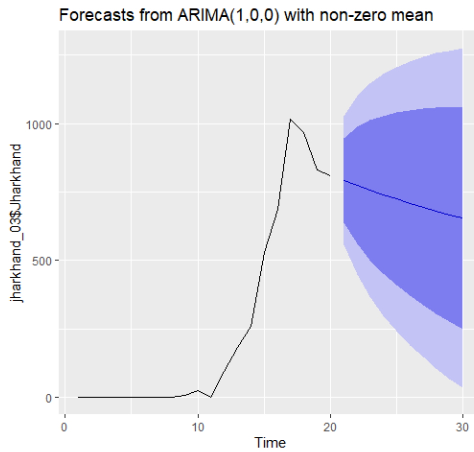


Figure-53 (PACF Plot)



From this time series plot we can observe that it is a decreasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,0,0).

Figure-54(Forecasting Graph)

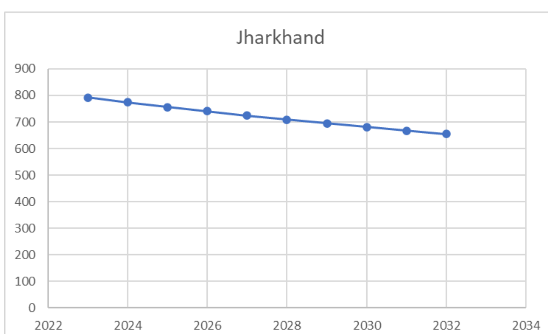


From the above forecasting plot, it can be observed the decreasing trend regarding the number of future cybercrime cases in Jharkhand.

Forecasted Values:

year	Jharkhand
2023	792.1144
2024	774.027
2025	756.7041
2026	740.1134
2027	724.224
2028	709.0062
2029	694.4316
2030	680.473
2031	667.1045
2032	654.301

Figure-55 (line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a decreasing trend.

f) Rajasthan:

Figure-56 (TS Plot)

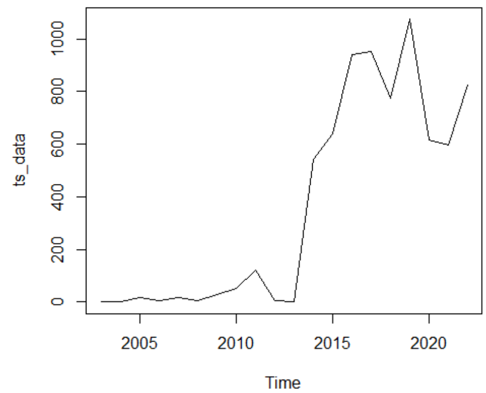


Figure-57 (ACF Plot)

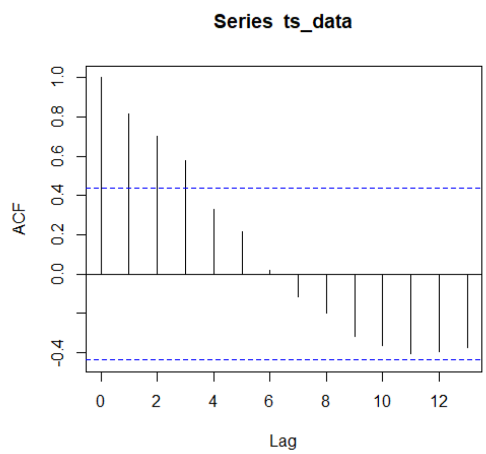
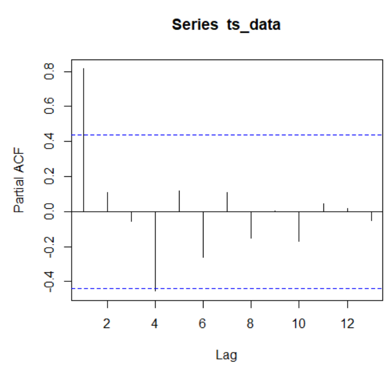
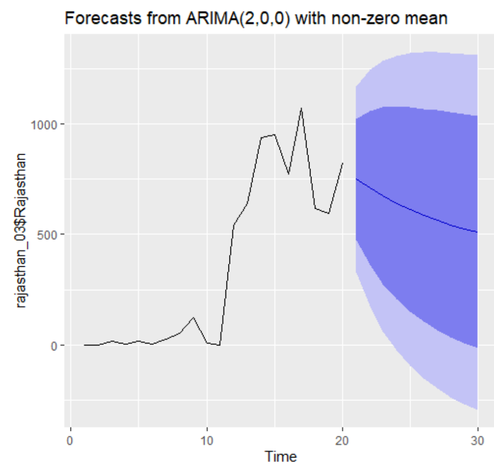


Figure-58 (PACF Plot)



From this time series plot we can observe that it is a dynamic data followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (2,0,0).

Figure-59 (Forecasting Graph)

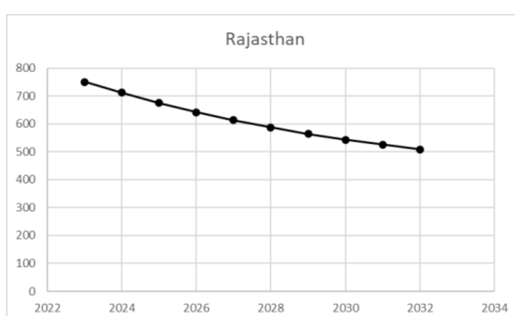


From the above forecasting plot, it can be observed the decreasing trend regarding the number of future cybercrime cases in Rajasthan.

Forecasted Values:

year	Rajasthan
2023	750.0979
2024	711.6961
2025	674.6748
2026	642.0313
2027	612.9632
2028	587.108
2029	564.1076
2030	543.6471
2031	525.4459
2032	509.2546

Figure-60 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a decreasing trend.

g) Telangana:

Figure-61 (Time Series Diagram)

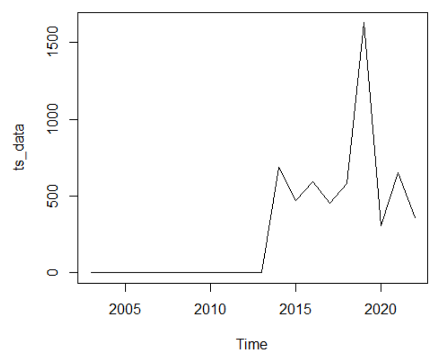


Figure-62 (ACF Plot)

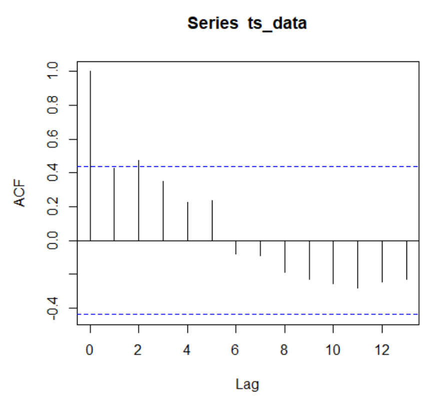
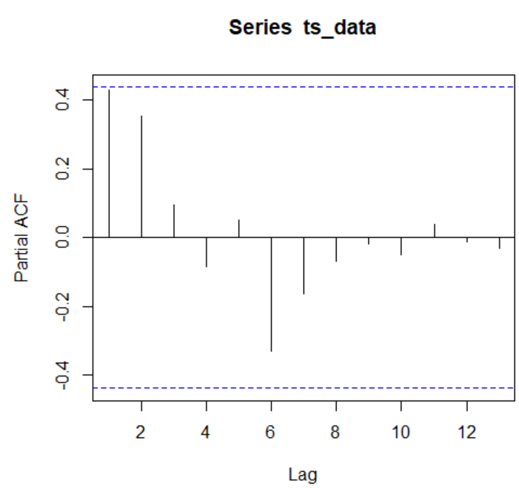
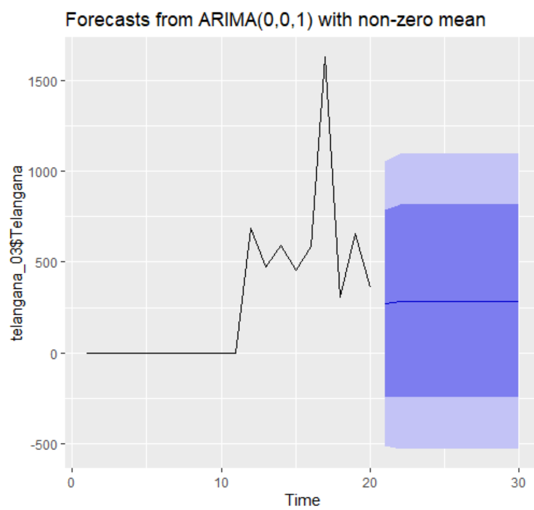


Figure-63 (PACF Plot)



From this time series plot we can observe that it is a dynamic data followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (0,0,1).

Figure-64 (Forecasting Graph)

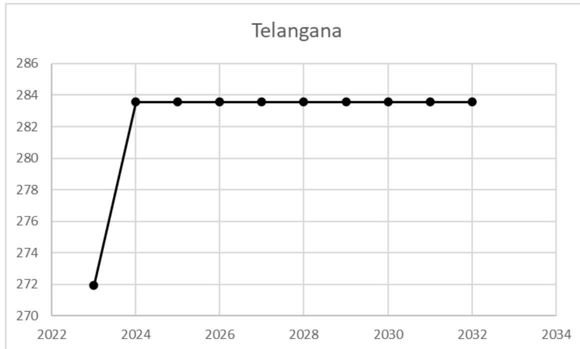


From the above forecasting plot, it can be observed the similar trend regarding the number of future cybercrime cases in Telangana.

Forecasted Values:

year	Telangana
2023	271.9122
2024	283.5758
2025	283.5758
2026	283.5758
2027	283.5758
2028	283.5758
2029	283.5758
2030	283.5758
2031	283.5758
2032	283.5758

Figure-65 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

h) Tamil Nadu:

Figure-66 (TS Plot)

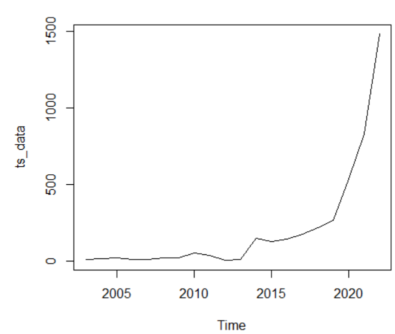


Figure-67 (ACF Plot)

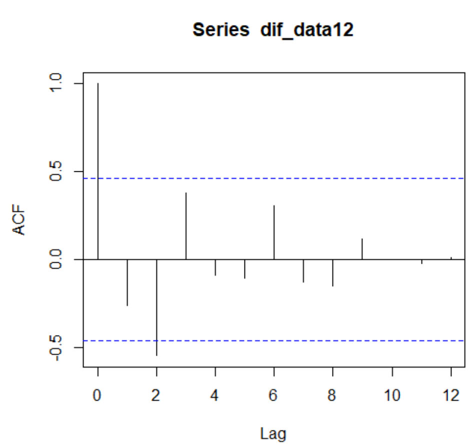
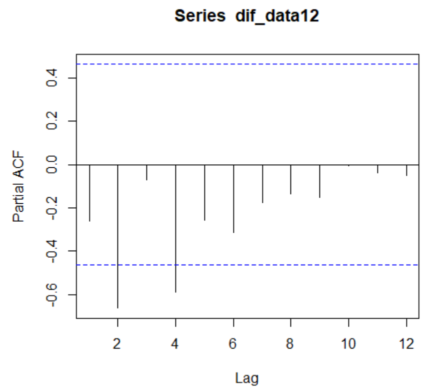
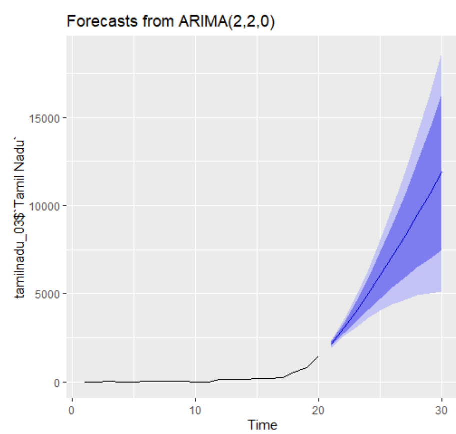


Figure-68 (PACF Plot)



From this time series plot we can observe that it is an increasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (2,2,0).

Figure-69 (Forecasting Graph)

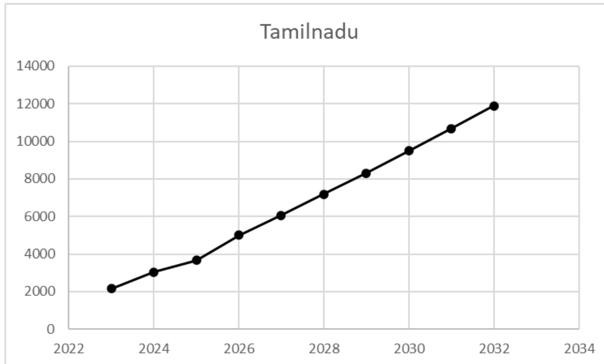


From the above forecasting plot, it can be observed an increasing trend regarding the number of future cybercrime cases in Tamil Nadu.

Forecasted Values:

year	Tamilnadu
2023	2163.603
2024	3053.879
2025	3665.434
2026	5001.461
2027	6053.332
2028	7178.832
2029	8315.619
2030	9496.01
2031	10684.207
2032	11898.257

Figure-70 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

i) Andhra Pradesh:

Figure-71 (TS Plot)

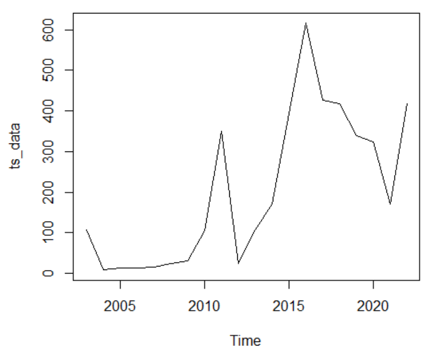


Figure-72 (ACF Plot)

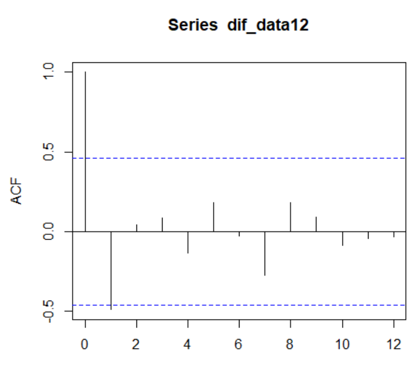
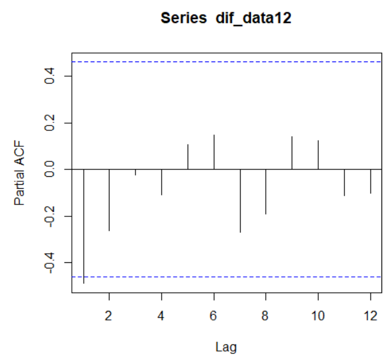
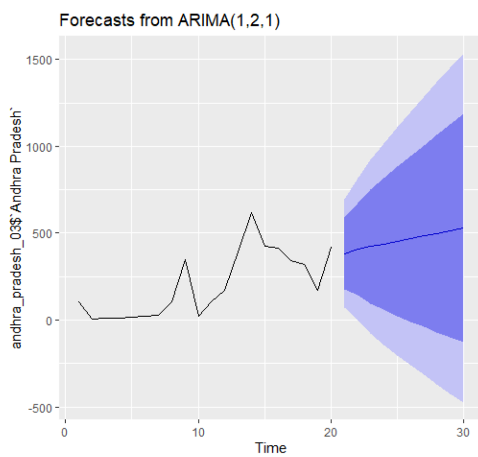


Figure-73 (PACF Plot)



From this time series plot we can observe that it is an increasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,2,1).

Figure-74 (Forecasting Graph)

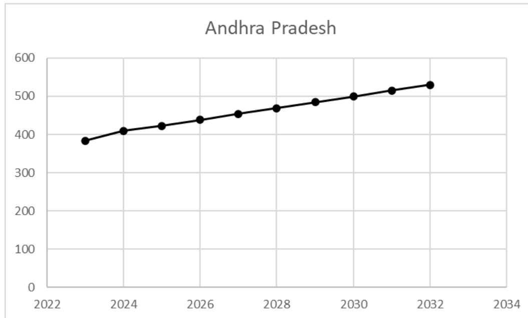


From the above forecasting plot, it can be observed an increasing trend regarding the number of future cybercrime cases in Andhra Pradesh.

Forecasted Values:

year	Andhra Pradesh
2023	383.8822
2024	409.6961
2025	422.7332
2026	438.4941
2027	453.6744
2028	468.9784
2029	484.2561
2030	499.5394
2031	514.8214
2032	530.1038

Figure-75 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

j) *Kerala:*

Figure-76 (TS Plot)

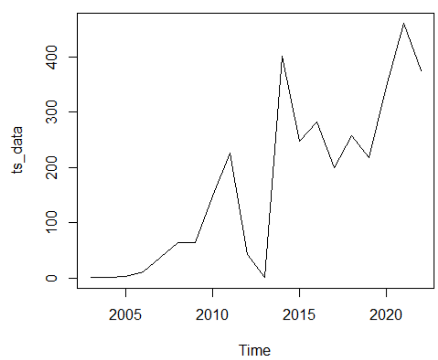


Figure-77 (ACF Plot)

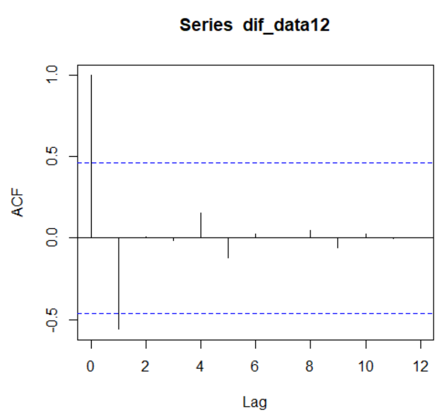
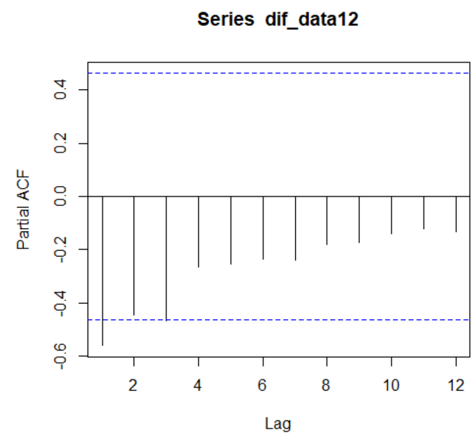
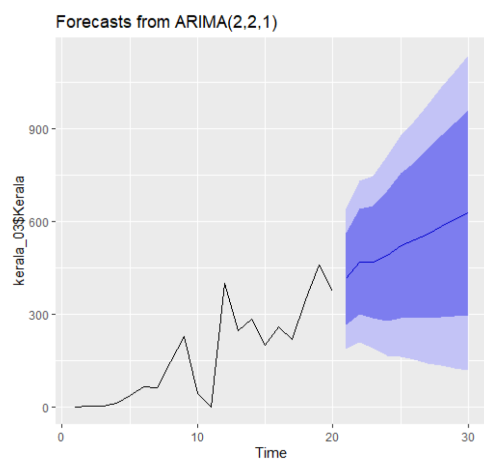


Figure-78 (PACF Plot)



From this time series plot we can observe that it is an increasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (2,2,1).

Figure-79 (Forecasting Graph)

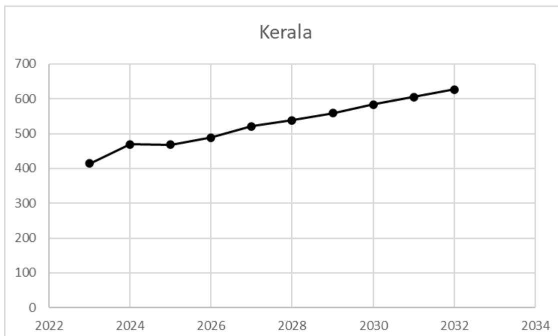


From the above forecasting plot, it can be observed an increasing trend regarding the number of future cybercrime cases in Kereala.

Forecasted Values:

year	Kerala
2023	414.2218
2024	469.2477
2025	468.4711
2026	489.0455
2027	520.6865
2028	538.6271
2029	559.0313
2030	583.5199
2031	605.046
2032	626.451

Figure-80 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

k) ***Haryana:***

Figure-81 (TS Plot)

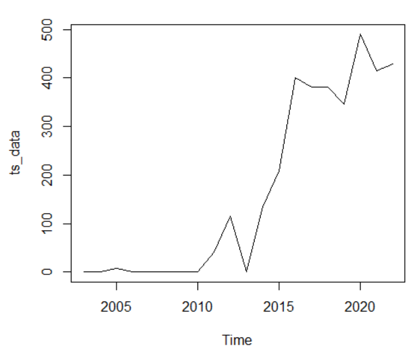


Figure-82 (ACF Plot)

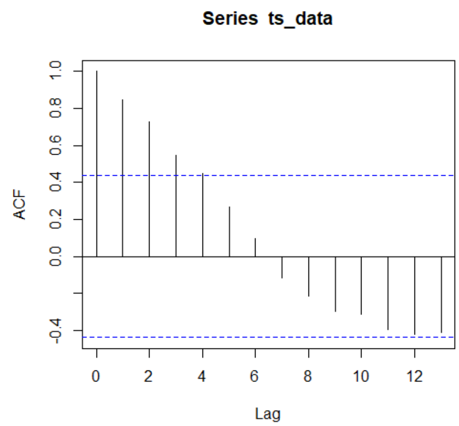
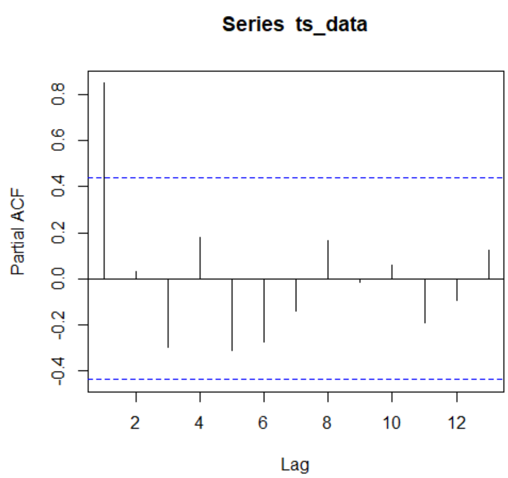
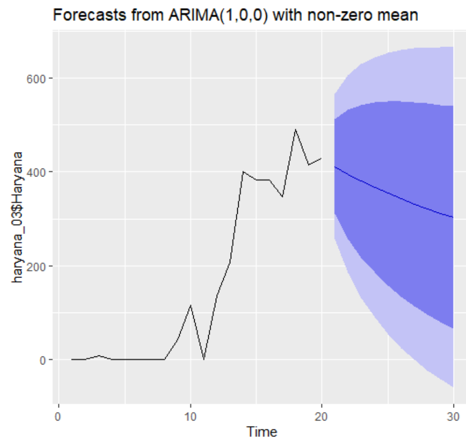


Figure-83 (PACF Plot)



From this time series plot we can observe that it is a decreasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,0,0).

Figure-84 (Forecasting Graph)

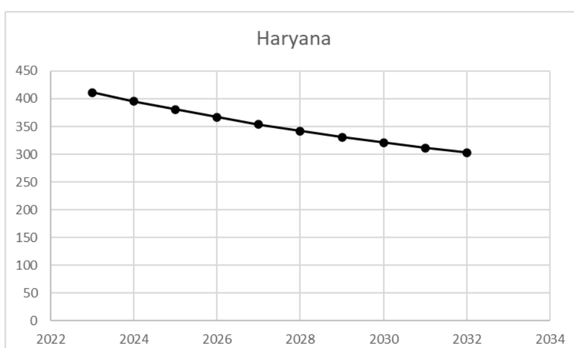


From the above forecasting plot, it can be observed a decreasing trend regarding the number of future cybercrime cases in Haryana.

Forecasted Values:

year	Haryana
2023	411.5815
2024	395.4542
2025	380.5224
2026	366.6975
2027	353.8974
2028	342.0462
2029	331.0735
2030	320.9142
2031	311.508
2032	302.799

Figure-85 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a decreasing trend.

1) ***Odisha:***

Figure-86 (TS Plot)

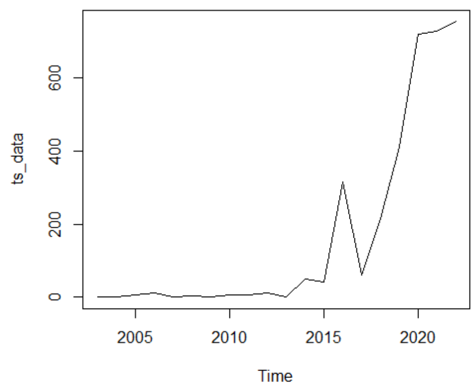


Figure-87 (ACF Plot)

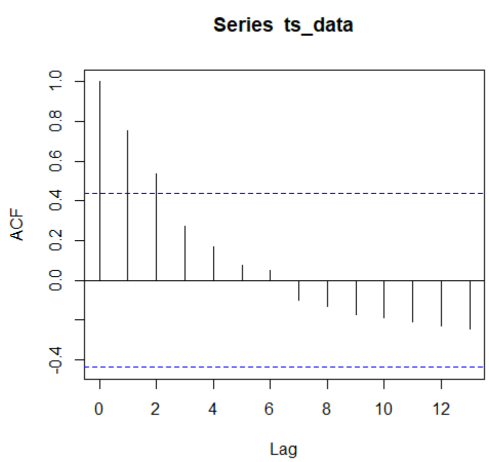
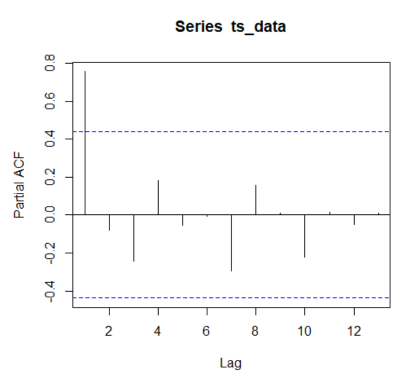
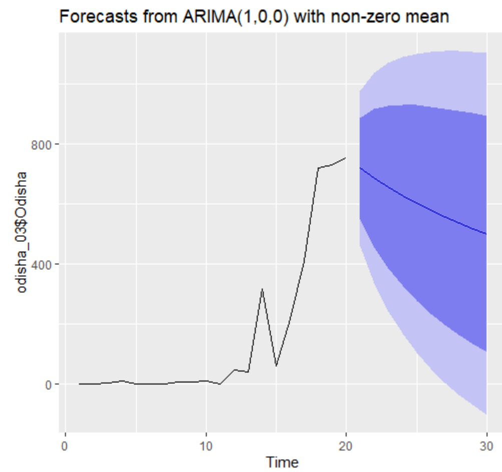


Figure-88 (PACF Plot)



From this time series plot we can observe that it is a decreasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,0,0).

Figure-89 (Forecasting Graph)

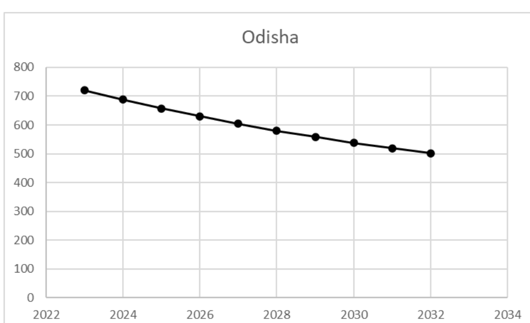


From the above forecasting plot, it can be observed a decreasing trend regarding the number of future cybercrime cases in Odisha.

Forecasted Values:

year	Odisha
2023	719.9453
2024	687.5006
2025	657.4715
2026	629.6782
2027	603.9542
2028	580.1454
2029	558.1093
2030	537.7139
2031	518.837
2032	501.3655

Figure-90 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a decreasing trend.

m) Madhya Pradesh:

Figure-91 (TS Plot)

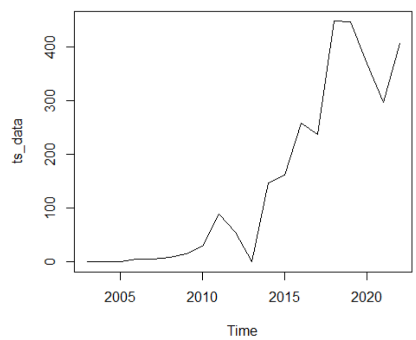


Figure-92(ACF Plot)

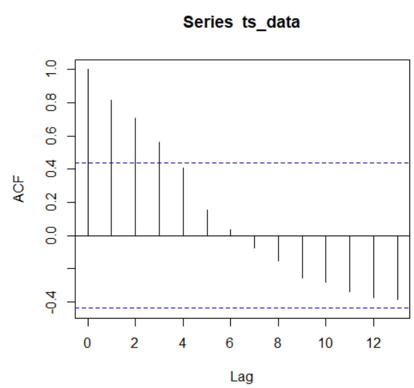
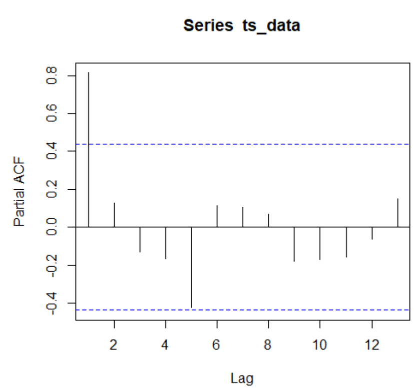
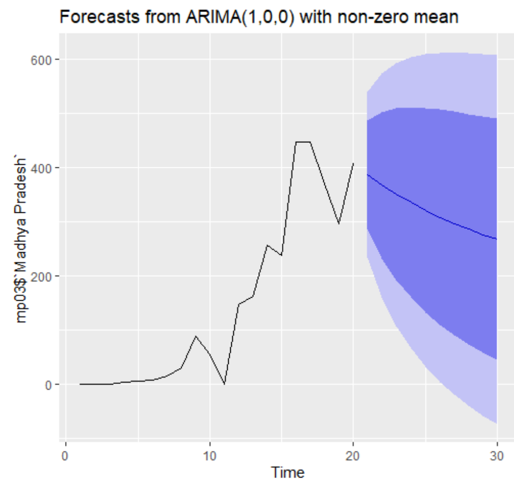


Figure-93(PACF Plot)



From this time series plot we can observe that it is a decreasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,0,0).

Figure-94 (Forecasting Graph)

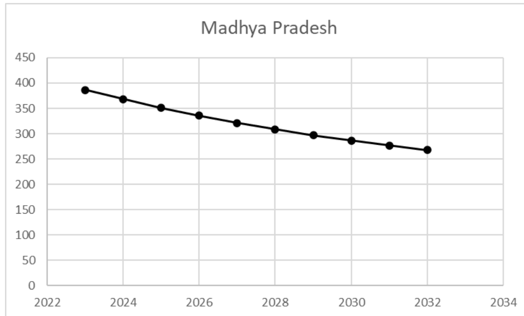


From the above forecasting plot, it can be observed a decreasing trend regarding the number of future cybercrime cases in Madhya Pradesh.

Forecasted Values:

year	Madhya Pradesh
2023	386.5844
2024	367.9791
2025	351.0237
2026	335.5718
2027	321.49
2028	308.657
2029	296.962
2030	286.304
2031	276.5911
2032	267.7395

Figure-95 (Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows a decreasing trend.

n) ***Gujarat:***

Figure-96 (TS Plot)

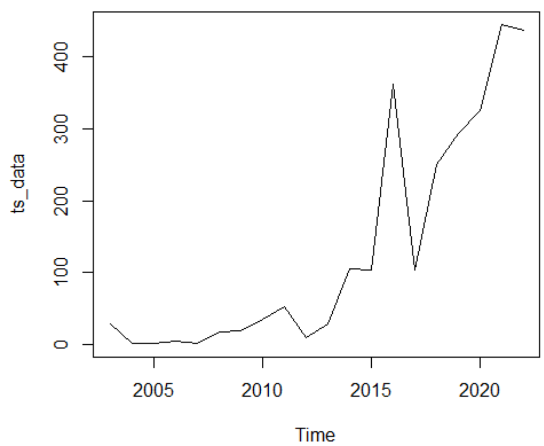


Figure-97 (ACF Plot)

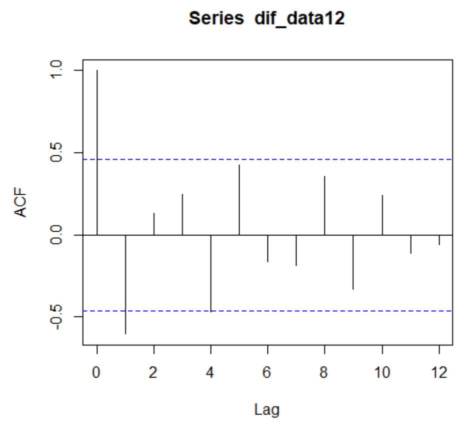
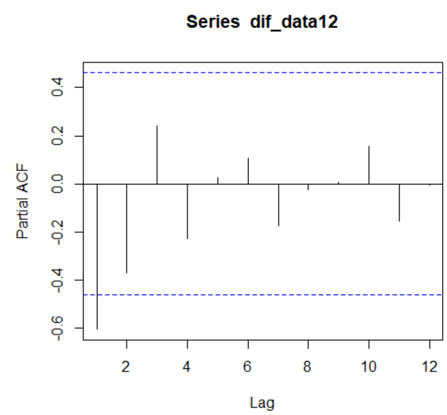
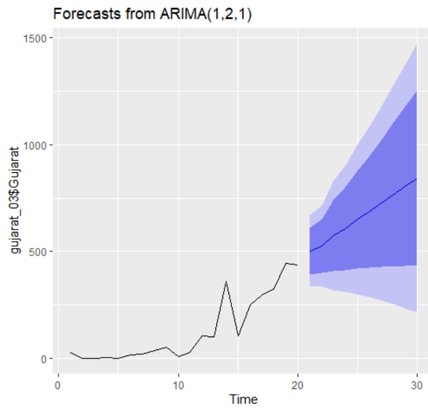


Figure-98 (PACF Plot)



From this time series plot we can observe that it is an increasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,2,1).

Figure-99 (Forecasting Graph)

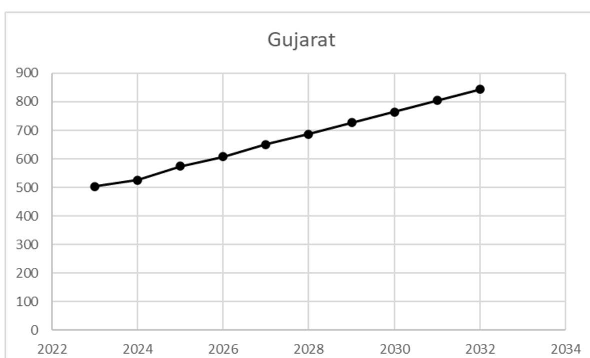


From the above forecasting plot, it can be observed an increasing trend regarding the number of future cybercrime cases in Gujarat.

Forecasted Values:

year	Gujarat
2023	503.7397
2024	525.6755
2025	574.8339
2026	607.452
2027	650.1199
2028	686.6816
2029	726.9534
2030	764.971
2031	804.3582
2032	842.9133

Figure-100(Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

o) ***West Bengal:***

Figure-101 (TS Plot)

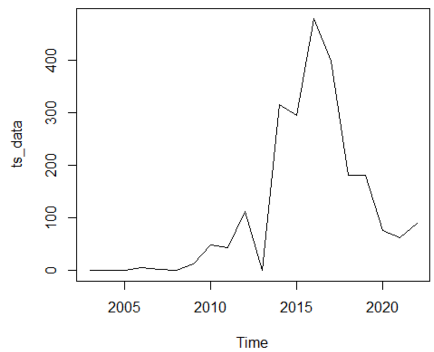


Figure-102 (ACF Plot)

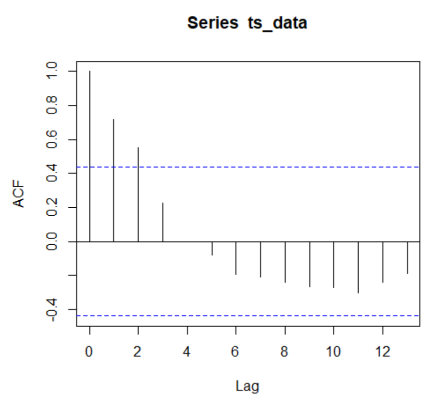
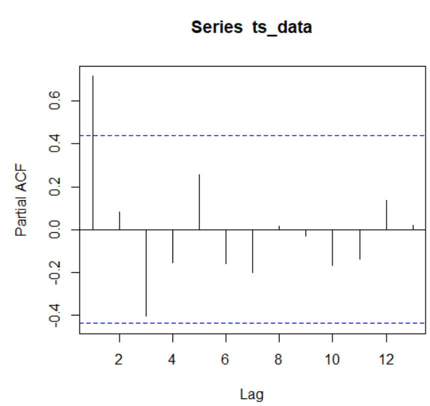
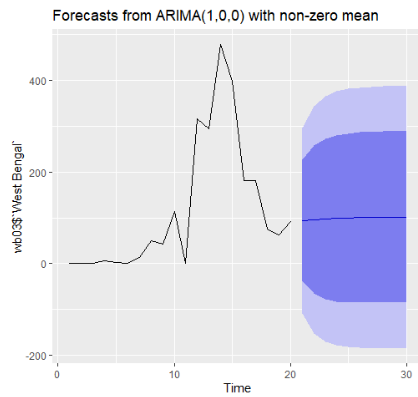


Figure-103(PACF Plot)



From this time series plot we can observe that it is an increasing trend followed by this past 20 years data. So, from the above plots we can understand that the fitted autoregression model is ARIMA (1,0,0).

Figure-104 (Forecasting Graph)

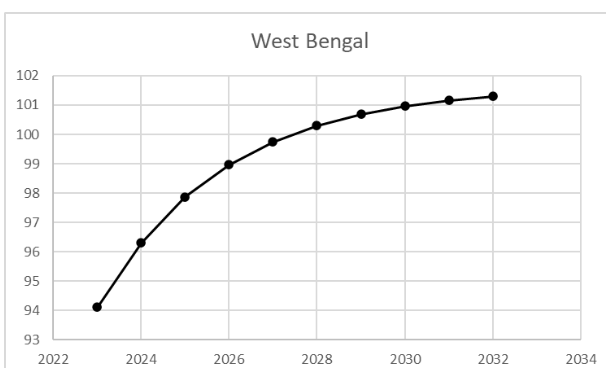


From the above forecasting plot, it can be observed an increasing trend regarding the number of future cybercrime cases in West Bengal.

Forecasted Values:

year	West Bengal
2023	94.10816
2024	96.30711
2025	97.86281
2026	98.96343
2027	99.74209
2028	100.29297
2029	100.6827
2030	100.95843
2031	101.1535
2032	101.29151

Figure-105(Line diagram)



From the above line diagram, we can observe that, the forecasted data for the years 2023 to 2032 shows an increasing trend.

- **Now we have the forecasted values of the number of cybercrimes for the years 2023 to 2032 for the top 15 states of India as per previous 20 years data. Now, we are going to rank them as per upcoming 10 years forecasted values and find the TOP 10 states in India corresponding to the forecasted cybercrime number for upcoming 10 years.**

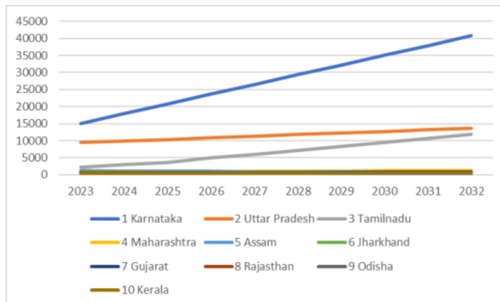


Table for Ranking

Rank	State	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032
1	Karnataka	14950.78	17933.5	20749	23613.1	26463.15	29317.1	32170	35023.1	37873.2	40729.26
2	Uttar Pradesh	9423.053	9894.11	10365	10836.2	11307.26	11778.3	12249.4	12720.4	13191.5	13662.53
3	Tamilnadu	2163.603	3053.88	3665.4	5001.46	6053.332	7178.83	8315.62	9496.01	10684.2	11898.26
4	Maharashtra	702.8972	831.104	842.9	933.802	970.9509	1044.63	1093.48	1159.2	1213.46	1275.505
5	Assam	1270.136	1153.83	1061.7	988.776	931.0087	885.261	849.031	820.339	797.617	779.6222
6	Jharkhand	792.1144	774.027	756.7	740.113	724.224	709.006	694.432	680.473	667.105	654.301
7	Gujarat	503.7397	525.676	574.83	607.452	650.1199	686.682	726.953	764.971	804.358	842.9133
8	Rajasthan	750.0979	711.696	674.67	642.031	612.9632	587.108	564.108	543.647	525.446	509.2546
9	Odisha	719.9453	687.501	657.47	629.678	603.9542	580.145	558.109	537.714	518.837	501.3655
10	Kerala	414.2218	469.248	468.47	489.046	520.6865	538.627	559.031	583.52	605.046	626.451

So, from the above table we can conclude that, our TOP 10 states of India for upcoming 10 years corresponding to the forecasted cybercrime cases are *Karnataka, Uttar Pradesh, Tamil Nadu, Maharashtra, Assam, Jharkhand, Gujarat, Rajasthan, Odisha, Kerala.*

CONCLUSION

This project gives a detailed comparison of cybercrime trends between India and the United States, two of the world's biggest and most populated countries. We used historical data from the top 10 states in India over 20 years (2003-2022) and applied statistical models to find patterns and reasons behind cybercrime. We started by using an ANOVA (Analysis of Variance) model to look at the differences in cybercrime rates among states, considering factors like personal revenge, fraud, sexual exploitation, and anger. Since ANOVA might not always work due to certain data requirements, we also used non-parametric tests to ensure our results were accurate. This approach gave us a clear picture of how cybercrime varies by region and over time. Additionally, we used time series analysis to predict cybercrime trends for the next 10 years in both India and the USA. By studying past data, we aimed to forecast future trends, which can help identify potential challenges and opportunities in fighting cybercrime. Comparing these forecasts for India and the USA provided valuable insights into the changing nature of cyber threats in both countries. We also specifically predicted future cybercrime rates for the top 10 states in India, chosen from the top 15 states based on past data. Our findings highlighted the importance of understanding regional differences and the need for tailored strategies to address specific cybercrime issues.

The results of this study have an important implication for policymakers, cyber security professionals and researchers. By pinpointing key areas that need stronger security measures and policies, the study offers guidance on developing effective strategies to reduce the impact of cybercrime. It also gives the benefits of international cooperation between India and the USA in cybersecurity practices and innovations.

In conclusion, this project adds to our understanding of cybercrime by providing a detailed analysis and comparison of trends in India and the USA. It highlights the need for ongoing monitoring, flexible strategies and international collaboration to effectively combat the ever-changing threat of cybercrime.

APPENDIX

R CODES:

Code for Shapiro-wilk test:

```
library(readxl)

top_10_india_17_20 <- read_excel("Python Scripts/top 10 india 17-20.xlsx")

View(top_10_india_17_20)

data= top_10_india_17_20

View(data)

# Load required package

library(dplyr)

# Sample data

data <- data.frame(

  State_UT = c(rep("Karnataka", 4), rep("Uttar Pradesh", 4), rep("Maharashtra", 4),

        rep("Telangana", 4), rep("Andhra Pradesh", 4), rep("Assam", 4),

        rep("Jharkhand", 4), rep("Bihar", 4), rep("Rajasthan", 4), rep("Gujarat", 4)),

  y1=data$Personal_Revenge,

  y2=data$Anger,

  y3=data$Fraud,

  y4=data$Sexual_Exploitation,

  Y=c(y1,y2,y3,y4),

   y1= c(36, 27, 12, 147, 41, 47, 301, 78, 47, 99, 48, 36, 14, 19, 11, 96, 24, 34, 16, 83, 246, 239, 555, 654, 31, 16, 8, 4, 12, 5,
27, 84, 2, 9, 17, 22, 6, 17, 5, 6),

  y2 = c(12, 10, 4, 13, 208, 73, 81, 210, 80, 129, 45, 105, 201, 3, 4, 24, 5, 26, 17, 39, 83, 46, 263, 164, 11, 6, 0, 4, 5, 8, 7, 34,
3, 11, 45, 10, 5, 32, 20, 31),

  y3 = c(2764, 5441, 11381, 9680, 3450, 2351, 3549, 4674, 2171, 1998, 3551, 3413, 529, 732, 2013, 4436, 537, 733, 1211,
1149, 48, 389, 243, 242, 460, 783, 964, 1069, 397, 351, 844, 1218, 331, 499, 938, 641, 305, 401, 363, 875),

  y4= c(55, 85, 90, 191, 117, 343, 430, 560, 462, 724, 557, 612, 58, 77, 78, 85, 61, 92, 84, 169, 217, 113, 289, 483, 14, 16,
15, 13, 16, 8, 8, 32, 29, 60, 103, 67, 24, 23, 32, 37)

)

# Apply Shapiro-Wilk test for each variable

result <- data %>%

  summarise(across(-State_UT, shapiro_test))

print(result)

# Install and load required packages
```

```
install.packages("ggplot2")
library(ggplot2)
# Assuming your data is stored in a data frame named 'data'
# You can replace 'data' with the name of your actual data frame
# Sample data
data <- data.frame(
  State_UT = c(rep("Karnataka", 4), rep("Uttar Pradesh", 4), rep("Maharashtra", 4),
              rep("Telangana", 4), rep("Andhra Pradesh", 4), rep("Assam", 4),
              rep("Jharkhand", 4), rep("Bihar", 4), rep("Rajasthan", 4), rep("Gujarat", 4)),
  Personal_Revenge = c(36, 27, 12, 147, 41, 47, 301, 78, 47, 99, 48, 36, 14, 19, 11, 96,
                       24, 34, 16, 83, 246, 239, 555, 654, 31, 16, 8, 4, 12, 5, 27, 84, 2, 9,
                       17, 22, 6, 17, 5, 6),
  Anger = c(12, 10, 4, 13, 208, 73, 81, 210, 80, 129, 45, 105, 201, 3, 4, 24, 5, 26, 17,
        39, 83, 46, 263, 164, 11, 6, 0, 4, 5, 8, 7, 34, 3, 11, 45, 10, 5, 32, 20, 31),
  Fraud = c(2764, 5441, 11381, 9680, 3450, 2351, 3549, 4674, 2171, 1998, 3551, 3413, 529,
        732, 2013, 4436, 537, 733, 1211, 1149, 48, 389, 243, 242, 460, 783, 964, 1069,
        397, 351, 844, 1218, 331, 499, 938, 641, 305, 401, 363, 875),
  Sexual_Exploitation = c(55, 85, 90, 191, 117, 343, 430, 560, 462, 724, 557, 612, 58, 77,
                          78, 85, 61, 92, 84, 169, 217, 113, 289, 483, 14, 16, 15, 13, 16, 8,
                          8, 32, 29, 60, 103, 67, 24, 23, 32, 37)
)
```

Code for Box & Q-Q Plot:

```
# Creating a QQ plot
ggplot(data, aes(sample = Personal_Revenge, fill = State_UT)) +
  geom_qq() +
  facet_wrap(~ State_UT, scales = "free") +
  labs(title = "QQ Plot of Personal Revenge for Different States/UTs",
      x = "Theoretical Quantiles",
      y = "Sample Quantiles") +
  theme_minimal()
ggplot(data, aes(sample = data$Anger, fill = State_UT)) +
  geom_qq() +
  facet_wrap(~ State_UT, scales = "free") +
  labs(title = "QQ Plot of Anger for Different States/UTs",
```

```
      x = "Theoretical Quantiles",
      y = "Sample Quantiles") +
  theme_minimal()
ggplot(data, aes(sample = data$Fraud, fill = State_UT)) +
  geom_qq() +
  facet_wrap(~ State_UT, scales = "free") +
  labs(title = "QQ Plot of fraud for Different States/UTs",
      x = "Theoretical Quantiles",
      y = "Sample Quantiles") +
  theme_minimal()
ggplot(data, aes(sample = data$Sexual_Exploitation, fill = State_UT)) +
  geom_qq() +
  facet_wrap(~ State_UT, scales = "free") +
  labs(title = "QQ Plot of sexual exploitation for Different States/UTs",
      x = "Theoretical Quantiles",
      y = "Sample Quantiles") +
  theme_minimal()
boxplot(data$Personal_Revenge)
boxplot(data$Anger)
boxplot(data$Fraud)
boxplot(data$Sexual_Exploitation)
```

Code for Kruskal Test:

```
# Sample data
data <- data.frame(
  State_UT = c(rep("Karnataka", 4), rep("Uttar Pradesh", 4), rep("Maharashtra", 4),
            rep("Telangana", 4), rep("Andhra Pradesh", 4), rep("Assam", 4),
            rep("Jharkhand", 4), rep("Bihar", 4), rep("Rajasthan", 4), rep("Gujarat", 4)),
  Personal_Revenge = c(36, 27, 12, 147, 41, 47, 301, 78, 47, 99, 48, 36, 14, 19, 11, 96, 24, 34, 16, 83, 246, 239, 555, 654, 31,
16, 8, 4, 12, 5, 27, 84, 2, 9, 17, 22, 6, 17, 5, 6),
  Anger = c(12, 10, 4, 13, 208, 73, 81, 210, 80, 129, 45, 105, 201, 3, 4, 24, 5, 26, 17, 39, 83, 46, 263, 164, 11, 6, 0, 4, 5, 8, 7,
34, 3, 11, 45, 10, 5, 32, 20, 31),
  Fraud = c(2764, 5441, 11381, 9680, 3450, 2351, 3549, 4674, 2171, 1998, 3551, 3413, 529, 732, 2013, 4436, 537, 733,
1211, 1149, 48, 389, 243, 242, 460, 783, 964, 1069, 397, 351, 844, 1218, 331, 499, 938, 641, 305, 401, 363, 875),
  Sexual_Exploitation = c(55, 85, 90, 191, 117, 343, 430, 560, 462, 724, 557, 612, 58, 77, 78, 85, 61, 92, 84, 169, 217, 113,
289, 483, 14, 16, 15, 13, 16, 8, 8, 32, 29, 60, 103, 67, 24, 23, 32, 37)
```

```
)

# Reshape the data to long format
data_long <- data %>%
  pivot_longer(cols = -State_UT, names_to = "Variable", values_to = "Value")
# Function to perform Kruskal-Wallis test
kruskal_test <- function(data) {
  kruskal.test(Value ~ State_UT, data = data)$p.value
}
# Apply Kruskal-Wallis test for each variable
kruskal_results <- data_long %>%
  group_by(Variable) %>%
  summarise(p_value = kruskal_test(cur_data()), .groups = 'drop')
# Print results
print(kruskal_results)
```

Code for forecast India:

```
library(readxl)
ind_fore_2003_22 <- read_excel("Python Scripts/ind fore 2003-22.xlsx")
View(ind_fore_2003_22)
data=ind_fore_2003_22
library(tseries)
summary(data)
ts_data = ts(data = data$`INDIA(Under IT Act)`, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(dif_data12)
pacf(dif_data12)
library(forecast)
model = Arima(data$`INDIA(Under IT Act)`, order = c(0,2,1))
autoplot(forecast(model))
```

```
forecast(model)
```

Code for forecast USA:

```
library(readxl)
usa_data_cyber_crime_2003_22 <- read_excel("Python Scripts/usa data cyber crime 2003-22.xlsx")
View(usa_data_cyber_crime_2003_22)
head(usa_data_cyber_crime_2003_22)
library(tseries)
summary(usa_data_cyber_crime_2003_22)
ts_data = ts(data = usa_data_cyber_crime_2003_22$`No. of cases in Usa Under IC3`, start = 2003, end = 2022, frequency =
1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(dif_data12)
pacf(dif_data12)
library(forecast)
model = Arima(usa_data_cyber_crime_2003_22$`No. of cases in Usa Under IC3`, order = c(0,2,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Karnataka:

```
library(tseries)
library(forecast)
library(readxl)
karnataka_2003_22 <- read_excel("Python Scripts/karnataka 2003-22.xlsx")
View(karnataka_2003_22)
head(karnataka_2003_22)
summary(karnataka_2003_22)
ts_data <- ts(karnataka_2003_22$Karnataka, start = min(karnataka_2003_22$Year), end = max(karnataka_2003_22$Year),
frequency = 1)
plot(ts_data)
diff_log <- diff(log(ts_data))
```

```
dif_data1 <- diff(diff_log)
plot(dif_data1)
dif_data12 <- diff(dif_data1)
plot(dif_data12)
dif_data123 <- diff(dif_data12)
plot(dif_data123)
dif_data1234 <- diff(dif_data123)
plot(dif_data1234)
adf_test_result <- tseries::adf.test(ts_data)
# Print the test result
print(adf_test_result)
# First-order differencing
diff_ts_data <- diff(ts_data)
# Perform ADF test on differenced data
adf_test_result_diff <- tseries::adf.test(diff_ts_data)
# Print the test result
print(adf_test_result_diff)
# 2nd-order differencing
diff_ts_data_2 <- diff(diff_ts_data)
# Perform ADF test on differenced data
adf_test_result_diff <- tseries::adf.test(diff_ts_data_2)
# Print the test result
print(adf_test_result_diff)
acf_result <- acf(diff_ts_data_2, main = "ACF Plot")
pacf_result <- pacf(diff_ts_data_2, main = "PACF Plot")
model <- Arima(ts_data, order = c(1, 2, 0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Uttar Pradesh:

```
# Load necessary libraries
library(forecast)
library(ggplot2)
# Provided data
data_up <- data.frame(
```

```
  Year = 2003:2022,
  Uttar_Pradesh = c(2, 2, 4, 0, 5, 2, 14, 32, 101, 44, 2, 1659, 2161, 2639, 4490, 5513, 9353, 9131, 7586, 8952)
)
# Convert the data to a time series object
ts_data_up <- ts(data_up$Uttar_Pradesh, start = min(data_up$Year), end = max(data_up$Year), frequency = 1)
# Plot ACF and PACF to identify model parameters
ggtsdisplay(ts_data_up)
# Fit SARIMA model
sarima_model <- auto.arima(ts_data_up, seasonal = TRUE)
# Summary of the model
summary(sarima_model)
# Generate forecasts
forecast_result <- forecast(sarima_model, h = 10)
# Plot forecasts
plot(forecast_result)
print(forecast_result)
```

Code to forecast Assam:

```
library(readxl)
assam_03_22 <- read_excel("Python Scripts/assam 03-22.xlsx")
View(assam_03_22)
head(assam_03_22)
library(tseries)
summary(assam_03_22)
ts_data = ts(data = assam_03_22$Assam, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(assam_03_22$Assam, order = c(1,0,0))
```

```
autoplot(forecast(model))

forecast(model)
```

Code to forecast Maharashtra:

```
library(readxl)

maharastha_03_22 <- read_excel("E:/raw data ind 17-20/maharastha 03-22.xlsx")

View(maharastha_03_22)

install.packages("forecast")

library(forecast)

ts_data <- ts(maharastha_03_22$Maharashtra, start = min(maharastha_03_22$Year), end = max(maharastha_03_22$Year),
frequency = 1)

plot(ts_data)

adf_test_result <- tseries::adf.test(ts_data)

# Print the test result

print(adf_test_result)

# First-order differencing

diff_ts_data <- diff(ts_data)

# Perform ADF test on differenced data

adf_test_result_diff <- tseries::adf.test(diff_ts_data)

# Print the test result

print(adf_test_result_diff)

# 2nd-order differencing

diff_ts_data_2 <- diff(diff_ts_data)

# Perform ADF test on differenced data

adf_test_result_diff <- tseries::adf.test(diff_ts_data_2)

# Print the test result

print(adf_test_result_diff)

acf_result <- acf(diff_ts_data_2, main = "ACF Plot")

pacf_result <- pacf(diff_ts_data_2, main = "PACF Plot")

model <- Arima(ts_data, order = c(1, 2, 0))

autoplot(forecast(model))

forecast(model)
```

Code to forecast Rajasthan:

```
library(readxl)

rajasthan_03 <- read_excel("Python Scripts/rajasthan 03.xlsx")
```

```
View(rajasthan_03)

head(rajasthan_03)
library(tseries)
summary(rajasthan_03)
ts_data = ts(data = rajasthan_03$Rajasthan, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
is.na(dif_data1)
dif_data10<- na.omit(dif_data1)
dif_data10
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(rajasthan_03$Rajasthan, order = c(2,0,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Telangana:

```
library(readxl)
telangana_03 <- read_excel("Python Scripts/telangana 03.xlsx")
View(telangana_03)
head(telangana_03)
library(tseries)
summary(telangana_03)
ts_data = ts(data = telangana_03$Telangana, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
is.na(dif_data1)
dif_data10 <- na.omit(dif_data1)
dif_data10
```

```
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(telangana_03$Telangana, order = c(0,0,1))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Jharkhand:

```
library(readxl)
jharkhand_03 <- read_excel("Python Scripts/jharkhand 03.xlsx")
View(jharkhand_03)
head(jharkhand_03)
library(tseries)
summary(jharkhand_03)
ts_data = ts(data = jharkhand_03$Jharkhand, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(jharkhand_03$Jharkhand, order = c(1,0,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Tamil Nadu:

```
library(readxl)
tamilnadu_03 <- read_excel("Python Scripts/tamilnadu 03.xlsx")
View(tamilnadu_03)
data_ind=read.csv(file.choose())
data_us = read.csv(file.choose())
head(tamilnadu_03)
library(tseries)
```

```
summary(tamilnadu_03)


ts_data = ts(tamilnadu_03$`Tamil Nadu`, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(dif_data12)
pacf(dif_data12)
library(forecast)
model = Arima(tamilnadu_03$`Tamil Nadu`, order = c(2,2,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Andhra Pradesh:

```
library(readxl)
andhra_pradesh_03 <- read_excel("Python Scripts/andhra pradesh 03.xlsx")
View(andhra_pradesh_03)
head(andhra_pradesh_03)
library(tseries)
summary(andhra_pradesh_03)
ts_data = ts(andhra_pradesh_03$`Andhra Pradesh`, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(dif_data12)
pacf(dif_data12)
library(forecast)
model = Arima(andhra_pradesh_03$`Andhra Pradesh`, order = c(1,2,1))
autoplot(forecast(model))
```

```
forecast(model)
```

Code to forecast Kerala:

```
library(readxl)
kerala_03 <- read_excel("Python Scripts/kerala 03.xlsx")
View(kerala_03)
head(kerala_03)
library(tseries)
summary(kerala_03)
ts_data = ts(data = kerala_03$Kerala, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(dif_data12)
pacf(dif_data12)
library(forecast)
model = Arima(kerala_03$Kerala, order = c(2,2,1))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Haryana:

```
library(readxl)
haryana_03 <- read_excel("Python Scripts/haryana 03.xlsx")
View(haryana_03)
head(haryana_03)
library(tseries)
summary(haryana_03)
ts_data = ts(data = haryana_03$Haryana, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
```

```
plot(dif_data12)


acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(haryana_03$Haryana, order = c(1,0,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Odisha:

```
library(readxl)
odisha_03 <- read_excel("Python Scripts/odisha 03.xlsx")
View(odisha_03)
head(odisha_03)
library(tseries)
summary(odisha_03)
ts_data = ts(data = odisha_03$Odisha, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(odisha_03$Odisha, order = c(1,0,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Madhya Pradesh:

```
library(readxl)
mp03 <- read_excel("Python Scripts/mp03.xlsx")
View(mp03)
data_ind=read.csv(file.choose())
```

```
head(mp03)


library(tseries)
summary(mp03)
ts_data = ts(data = mp03$`Madhya Pradesh`, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(mp03$`Madhya Pradesh`, order = c(1,0,0))
autoplot(forecast(model))
forecast(model)
```

Code to forecast Gujarat:

```
library(readxl)
gujarat_03 <- read_excel("Python Scripts/gujarat 03.xlsx")
View(gujarat_03)
head(gujarat_03)
library(tseries)
summary(gujarat_03)
ts_data = ts(data = gujarat_03$Gujarat, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(dif_data12)
pacf(dif_data12)
library(forecast)
```

```
model = Arima(gujarat_03$Gujarat, order = c(1,2,1))
autoplot(forecast(model))
forecast(model)
```

Code to forecast West Bengal:

```
library(readxl)
wb03 <- read_excel("Python Scripts/wb03.xlsx")
View(wb03)
head(wb03)
library(tseries)
summary(wb03)
ts_data = ts(data = wb03$`West Bengal`, start = 2003, end = 2022, frequency = 1)
plot(ts_data)
diff_log = log(ts_data)
dif_data1 = diff(diff_log)
plot(dif_data1)
dif_data12 = diff(dif_data1)
plot(dif_data12)
acf(ts_data)
pacf(ts_data)
library(forecast)
model = Arima(wb03$`West Bengal`, order = c(1,0,0))
autoplot(forecast(model))
forecast(model)
```

REFERENCES

The links of all the websites are used in the analysis of data collection are given below:

State/UT-wise Cyber Crime Motives during 2017:

https://data.gov.in/resource/stateut-wise-cyber-crime-motives-during-2017

State/UT-wise Cyber Crime Motives during 2018:

https://data.gov.in/resource/stateut-wise-cyber-crime-motives-during-2018

State/UT-wise Cyber Crime Motives during 2019:

https://data.gov.in/resource/stateut-wise-cyber-crime-motives-during-2019

State/UT-wise Cyber Crime Motives during 2020:

https://data.gov.in/resource/stateut-wise-cyber-crime-motives-during-2020

State/UT-wise Cybercrime in India in 2003:

https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2003.pdf

State/UT-wise Cybercrime in India in 2004:

https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2004.pdf

State/UT-wise Cybercrime in India in 2005:

https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2005.pdf

State/UT-wise Cybercrime in India in 2006:

https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2006.pdf

State/UT-wise Cybercrime in India in 2007:

https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2007.pdf

State/UT-wise Cybercrime in India in 2008:

https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2008.pdf

State/UT-wise Cybercrime in India in 2009:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2009.pdf

State/UT-wise Cybercrime in India in 2010:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2010.pdf

State/UT-wise Cybercrime in India in 2011:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2011.pdf

State/UT-wise Cybercrime in India in 2012:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2012.pdf

State/UT-wise Cybercrime in India in 2013:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2013.pdf

State/UT-wise Cybercrime in India in 2014:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2014.pdf

State/UT-wise Cybercrime in India in 2015:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2015.pdf

State/UT-wise Cybercrime in India in 2016:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2016.pdf

State/UT-wise Cybercrime in India in 2017:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2017.pdf

State/UT-wise Cybercrime in India in 2018:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2018.pdf

State/UT-wise Cybercrime in India in 2019:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2019.pdf

State/UT-wise Cybercrime in India in 2020:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2020.pdf

State/UT-wise Cybercrime in India in 2021:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2021.pdf

State/UT-wise Cybercrime in India in 2022:
https://ncrb.gov.in/uploads/2022/July/11/custom/crime-in-india/table-18.1-2022.pdf

Cybercrime data for USA in 2003:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2003-internet-crime-report

Cybercrime data for USA in 2004:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2004-internet-crime-report

Cybercrime data for USA in 2005:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2005-internet-crime-report

Cybercrime data for USA in 2006:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2006-internet-crime-report

Cybercrime data for USA in 2007:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2007-internet-crime-report

Cybercrime data for USA in 2008:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2008-internet-crime-report

Cybercrime data for USA in 2009:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2009-internet-crime-report

Cybercrime data for USA in 2010:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2010-internet-crime-report

Cybercrime data for USA in 2011:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2011-internet-crime-report

Cybercrime data for USA in 2012:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2012-internet-crime-report

Cybercrime data for USA in 2013:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2013-internet-crime-report

Cybercrime data for USA in 2014:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2014-internet-crime-report

Cybercrime data for USA in 2015:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2015-internet-crime-report

Cybercrime data for USA in 2016:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2016-internet-crime-report

Cybercrime data for USA in 2017:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2017-internet-crime-report

Cybercrime data for USA in 2018:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2018-internet-crime-report

Cybercrime data for USA in 2019:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2019-internet-crime-report

Cybercrime data for USA in 2020:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2020-internet-crime-report

Cybercrime data for USA in 2021:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2021-internet-crime-report

Cybercrime data for USA in 2022:

https://www.fbi.gov/news/press-releases/fbi-releases-the-internet-crime-complaint-center-2022-internet-crime-report

The references of some books:

- An outline of Statistical Theory (Vol. I) – A.K.Gun, M.K.Gupta, B.Dasgupta
- An outline of Statistical Theory (Vol. II) – A.K.Gun, M.K.Gupta, B.Dasgupta
- Fundamental of Statistical Theory (Vol. I) – A.K.Gun, M.K.Gupta, B.Dasgupta
- Fundamental of Statistical Theory (Vol. II) – A.K.Gun, M.K.Gupta, B.Dasgupta
- Fundamental of Mathematical Statistics – S.C.Gupta & V.K.Kapoor
- Fundamental of Applied Statistics – S.C.Gupta & V.K.Kapoor

ACKNOWLEDGEMENT

I am extremely grateful to my supervisor to this project Dr. Anindita Ghosal, Head of the department, Department of Statistics, Sister Nivedita University for guiding and supervising me throughout the project and she was continuously encouraged me to complete the project work sincerely. I have tried to follow her advice and guidance to make my work perfect and flawless.

Jeet Polley