Chapter 05

Court Disposal of Cyber Crime Cases

5.1 Introduction:

Cybercrime has arisen as a serious threat to people, organisations, and governments in our increasingly digital environment. Cybercrimes, which range from hacking and identity theft to online fraud and data breaches, have far-reaching repercussions that go beyond those of traditional criminal activities. As a result, courts all over the world have had to adjust to the distinct problems that these offences present and create specialised procedures for their handling

The legal procedure by which these offences are handled and adjudicated in a court of law is referred to as the disposition of cybercrimes. In addition to the eventual resolution of cases and the implementation of just punishments, it involves the investigation, prosecution, and sentencing of people engaged in cybercriminal acts. Given the complexity and ongoing evolution of cybercrimes, handling these cases offers special difficulties for legal systems around the world.

The technical sophistication of the offences is one of the main difficulties courts encounters when dealing with cybercrimes. It is crucial for judges, attorneys, and law enforcement agencies to have a thorough understanding of digital systems, networks, and internet security standards since cybercriminals frequently use advanced techniques and technology. In order to effectively investigate and prosecute cybercrimes, making sure that the evidence is correctly gathered, kept, and presented in court requires specialised training and skills.

The fact that cybercrime is a worldwide problem makes it difficult to resolve cases involving it. Because the digital world has no physical borders, cybercriminals can target victims in various countries from anywhere in the world. As a result, it might be difficult for authorities to identify and extradite cybercriminals because they must act within global legal frameworks and cooperative agreements. A partnership between numerous law enforcement agencies and international organisations is frequently necessary to effectively tackle cybercrimes due to their transnational character.

Additionally, there are continual difficulties in dealing with cybercrime cases due to the rapid advancement of technology and the appearance of new cyberthreats. Courts must stay up to date on the most recent trends and advancements in the digital world as criminals modify their methods to take advantage of vulnerabilities in developing technologies. To ensure they can accurately comprehend and assess the technical components of cybercrime cases, judges, prosecutors, and defence attorneys must get ongoing training and education.

To address these issues, numerous nations have set up specialised cybercrime departments inside their judicial systems in recent years. These teams frequently include specialists in computer science, cybersecurity, and digital forensics who collaborate closely with law enforcement officials and prosecutors to find and convict hackers. In order to provide direction on difficult technological issues when handling cybercrime cases, courts have started to depend on technical consultants and expert witnesses.

In conclusion, because of the technical complexity, international scope, and dynamic nature of cyber threats, the disposition of cybercrime cases offers particular difficulties for judges. In order to effectively manage cybercrimes and provide justice in the digital era, judicial institutions must adapt and outfit themselves with the required tools, knowledge, and collaboration frameworks as technology develops.

In this section, we analyse how court matters involving cybercrime are handled. From the provided data, we will attempt to extract some key information, identify patterns, and construct a model in accordance with those findings to see how well it can forecast the future.

5.2 Analysis of Court Disposal of Cyber Crime Cases (State/UT-wise)

In this section, we'll analyse the data from State/UT-level court disposal of cybercrime both conceptually and statistically. Additionally, we will also be dealing with the data of completed trials, conviction rate, pending cases and pendency rate across the states and UTs. We will attempt to gather some relevant information, establish certain trends, construct a model, and attempt to predict future occurrences.

5.2.1 Total No. Cases in The Court

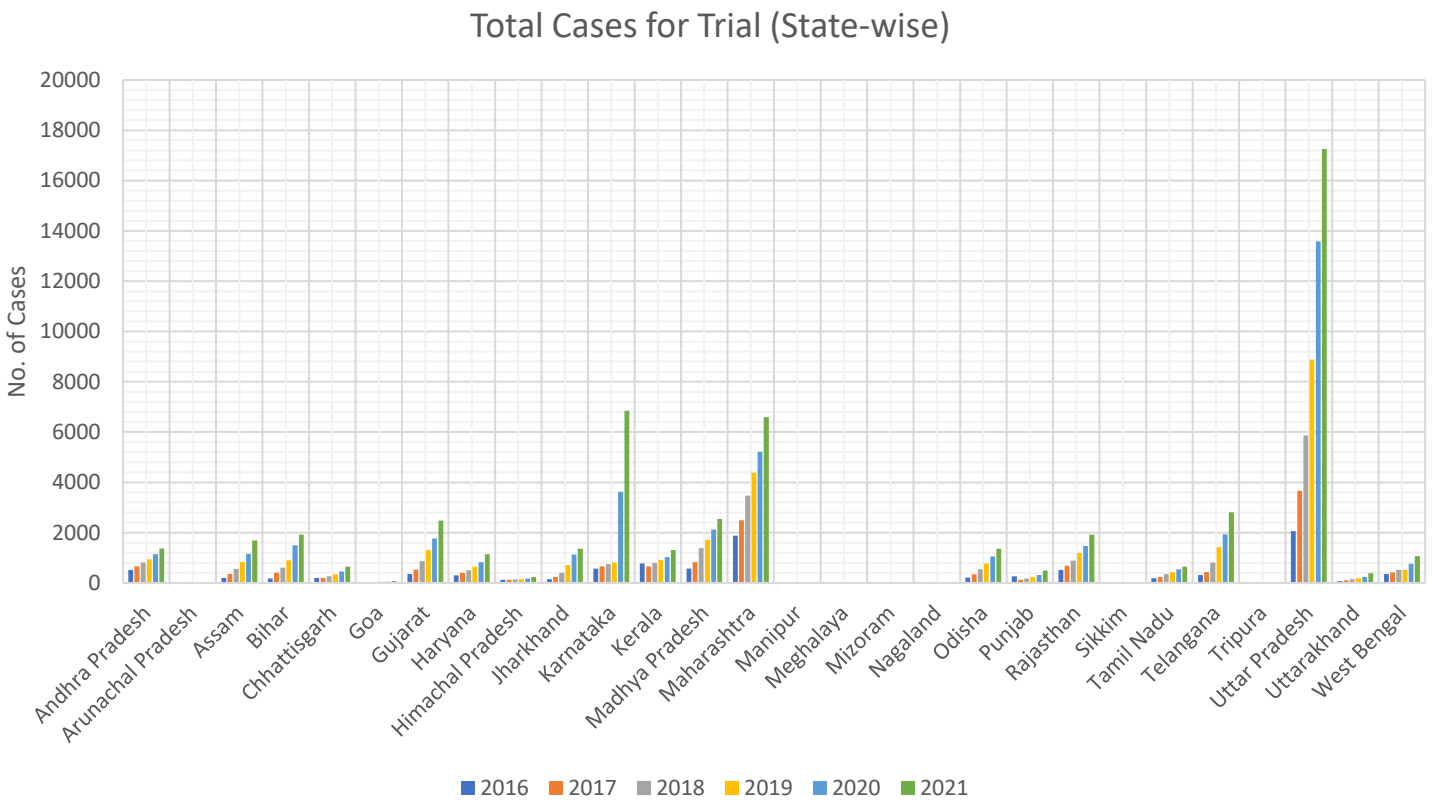
In this section we will deal with the total number of cases that were sent to trial in the court each year. It is determined by adding the total number of cases from the previous year's pending trial to the number of cases sent to trial during the current year.

Data Visualisation

In this part, we'll create graph for total no. of cases in the court for both State-wise and UT-wise with the provided data and derive some valuable conclusions from the graph as well from the data.

State-wise

Below is a graph showing how many instances there were in each state from 2016 to 2021.



We can draw the following inference from the above graph: -

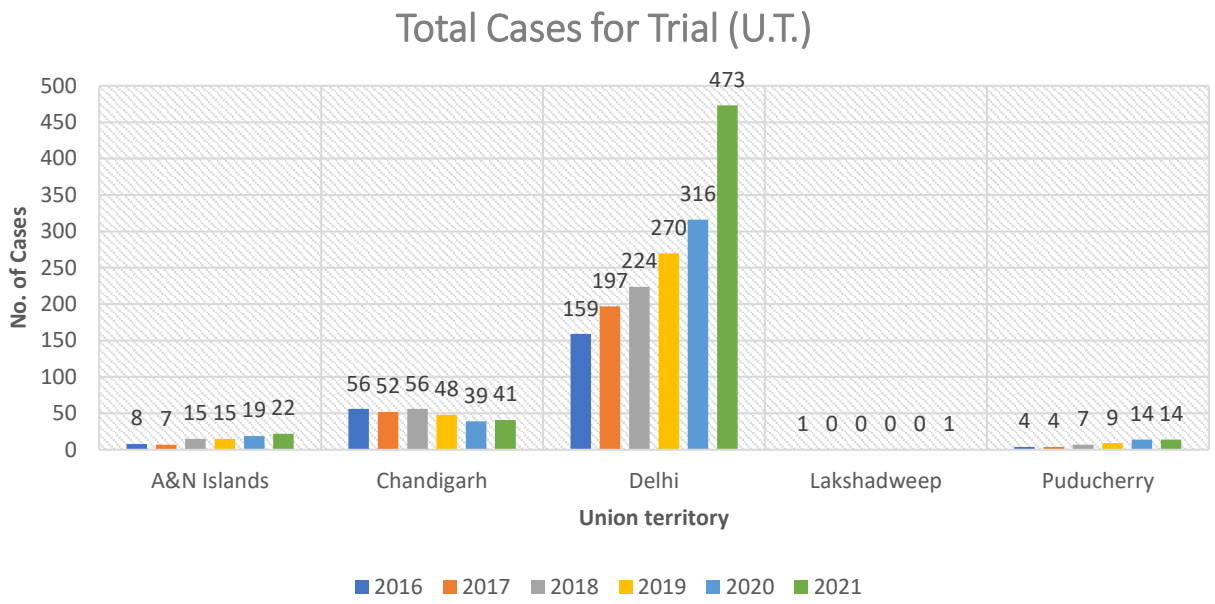
- **Overall Trend:** Over the years, cybercrime instances have increased overall across all states. 9,891 instances were recorded across all states in 2016, and 54,289 cases were reported across all states in 2021.
- **Variations in Cybercrime:** There are various degrees of cybercrime incidences in various states. When compared to other states, states like Uttar Pradesh, Maharashtra, and Karnataka routinely record more incidences of cybercrime.
- **Yearly Increases:** The number of reported cybercrime instances has generally been on the rise in the majority of states. This suggests an increase in cybercrime events as well as maybe more knowledge and reporting of such crimes.

- **Rapid Rise in Some States:** Over the years, the incidence of cybercrime has significantly increased in states like Uttar Pradesh, Maharashtra, and Karnataka. This shows that these areas require improved cybersecurity safeguards and awareness initiatives.
- **State-by-State Comparisons:** The data enables analysis of cybercrime instances in various states. As an illustration, Uttar Pradesh recorded the most instances (17,258) in 2021, followed by Maharashtra (6,596). Less occurrences have been reported in smaller states like Sikkim and Mizoram.

NOTE: - We have excluded the data of Jammu & Kashmir as it undergoes through a major change in its political status on August 5, 2019.

UT-wise

The graph for the no. of cases in the UT during the Year 2016-2021 are given below.



We can draw the following inference from the above graph: -

- **Variation in Cybercrime Cases:** Over time, there have been varying numbers of cybercrime cases reported in the Union Territories. Chandigarh and Delhi are the two cities with the most reported instances each year.
- **Rising Trend:** In most Union Territories, there has been an overall rise in the incidence of cybercrime. The reported cases in A&N Islands, Chandigarh, Delhi, and Puducherry gradually increased between 2016 and 2021.
- **Stable or Fluctuating Cases:** Some Union Territories show a pattern in the number of cybercrime cases that is either steady or variable. With the exception of one recorded case in 2021, Lakshadweep had few instances, whilst Puducherry's levels have remained largely stable over time.
- **High Incidence:** Compared to other Union Territories, Delhi routinely reports a greater number of cybercrime instances. There were 473 incidents in 2021, which highlights the region's need for more robust cybersecurity measures.
- **Low Incidence:** With just one case recorded in 2021, Lakshadweep has the lowest incidence of cybercrime incidents. When comparing the number of instances, it's crucial to take each Union Territory's population and technical setup into account.
- **Comparisons between Years:** In several Union Territories, the volume of cybercrime cases varies from one year to the next. For instance, Chandigarh had a drop in reported cases from 2017 to 2020, followed by a small rise in 2021.

NOTE: - We have excluded the data of Daman & Diu and D&N Haveli as it undergoes through a major change in its political status on November 26, 2019. Also we haven't included the data of Ladakh and Jammu & Kashmir as well as it became a U.T. on August 5 ,2019.

Statistical Analysis

In this portion, we'll employ some statistical techniques to extract a few necessary details from the data on the total number of Cyber Crime cases in the court.

Testing of Variability and Location (State-wise): -

Levene's test: - A statistical test called the Levene's test, sometimes referred to as the Levene's test for equality of variances, is used to determine if the variances of two or more groups or samples are statistically substantially different from one another. Before doing parametric tests like analysis of variance (ANOVA) or t-tests, which assume equal variances across groups, it is frequently used as a preliminary study.

The Levene's test compares the absolute deviations of each observation from the mean of the group. Based on these deviations, the test generates a test statistic and determines if it substantially deviates from what would be predicted if variances were equal.

Hypothesis:

- Null hypothesis (H_o): The variances are equal across all groups.
- Alternative hypothesis (H_1): At least one group has a different variance.

Test Statistics: The associated p-value shows the likelihood of observing the test statistic under the null hypothesis, and the test statistic follows an F-distribution.

A significant p-value indicates that there are significant variance differences between the groups, which is evidence against the null hypothesis of equal variances. The use of alternative statistical tests that do not assume equal variances or the consideration of data transformations to address the problem of unequal variances may be appropriate in such circumstances.

Level of Significance (α): The level of significance used for this analysis is 0.05 or 5%.

Output:

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value  Pr(>F)
## group   5  2.3633 0.04215 *
##       162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation: The test has been run on a dataset with six groups based on the results of Levene's Test for Homogeneity of Variance (with the centre set as the median). The degrees of freedom (Df) associated with the groups are 5. The test's F value was calculated to be 2.3633, and the associated p-value is 0.04215. The p-value is less than 0.05 (*), according to the provided significance codes, which denotes a statistically significant result.

There is evidence to suggest that the variances across the years being compared differ statistically significantly. This indicates that the assumption of equal variances may not hold, and care should be taken when using statistical tests, such as specific types of analysis of variance (ANOVA) or t-tests, that make this assumption.

Kruskal-Wallis test: - The Kruskal–Wallis test by ranks, Kruskal–Wallis H test (named after William Kruskal and W. Allen Wallis), or one-way ANOVA on ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of equal or different sample sizes. Since it is a nonparametric method, the Kruskal–Wallis test does not assume a normal distribution of the residuals, unlike the analogous one-way analysis of variance.

Hypotheses:

- Null hypothesis (H_o): The distributions of the groups are identical.
- Alternative hypothesis (H_1): At least one group has a different distribution.

Test Statistics: The test statistic in the Kruskal-Wallis test is based on the ranks and follows a chi-squared distribution with (k-1) degrees of freedom, where k is the number of groups being compared. The p-value associated with the test statistic is used to determine whether the differences between the groups are statistically significant.

Level of Significance (α): The level of significance used for this analysis is 0.05 or 5%.

Output:

```
##a Kruskal-Wallis rank sum test that was performed on a variable called "cases" grouped by the factor variable "year" .
```

```
##  Kruskal-Wallis rank sum test
## Kruskal-Wallis chi-squared = 13.265, df = 5, p-value = 0.02102
```

Interpretation: The test statistic (Kruskal-Wallis chi-squared) is calculated to be 13.265, with 5 degrees of freedom (df). The p-value associated with the test statistic is reported as 0.02102. since the p-value (0.02102) is less than the significance level of 0.05, there is sufficient evidence to reject the null hypothesis that the distributions of the variable "cases" are the same across all the levels of the factor variable "year." Therefore, we can conclude that there are significant differences in the "cases" variable among the different years.

Post-hoc test for Kruskal-Wallis: The Dunn's Test: - The results of the Kruskal-Wallis test indicate if there are differences between the groups, but they do not specify which groups are distinctive from one another. Post-hoc testing can be used to identify which groups vary from other groups. The Dunn test (1964) is the most frequently used post-hoc test for the Kruskal-Wallis test.

Hypotheses:

- Null hypothesis: The groups are sampled from populations with identical distributions. Typically, that the sampled populations exhibit stochastic equality.
- Alternative hypothesis (two-sided): The groups are sampled from populations with different distributions. Typically, that one sampled population exhibits stochastic dominance.

Test Statistics: The Dunn's (1964) z-test approximation is calculated as the difference in mean rank scores divided by the rank pooled variance estimate for two groups. The comparisons are presented in a matrix format, where the rows represent the reference groups (Row Mean) and the columns represent the groups being compared (Col Mean). The values in the matrix indicate the mean difference between the groups. The mean differences are shown in each cell. If the p-value for a comparison is less than or equal to $\alpha/2$, the null hypothesis of no difference between the groups can be rejected.

Level of Significance (α): The level of significance used for this analysis is 0.05 or 5%.

Output:

```
## Kruskal-Wallis chi-squared = 13.2652, df = 5, p-value = 0.02
##
##
```

```
##                            Comparison of x by group
##                                  (Bonferroni)
## Col Mean-|
## Row Mean |        y16        y17        y18        y19        y20
## ---------+-------------------------------------------------------
##      y17 |  -0.446468
##          |     1.0000
##          |
##      y18 |  -1.119605  -0.673136
##          |     1.0000     1.0000
##          |
##      y19 |  -1.807853  -1.361385  -0.688248
##          |     0.5297     1.0000     1.0000
##          |
##      y20 |  -2.427414  -1.980945  -1.307808  -0.619560
##          |     0.1141     0.3570     1.0000     1.0000
##          |
##      y21 |  -2.976913  -2.530445  -1.857308  -1.169060  -0.549499
##          |    0.0218*     0.0854     0.4745     1.0000     1.0000
##
## alpha = 0.05
## Reject Ho if p <= alpha/2
```

Interpretation: Looking at the comparison between "y16" and "y21," the difference in mean ranks is -2.976913, and the p-value is 0.0218 (marked with "*"). Since this p-value is less than the significance level alpha/2 of 0.025, we can conclude that there is a significant difference between the "y16" and "y21" groups.

For all other groups, the p-value is greater than the significance level alpha/2 of 0.025. Thus we conclude that there is no significant difference between any other group.

Model Fitting And Goodness of Fit:

By offering a structured method for comprehending the data, estimating relationships, and making predictions or inferences, model fitting plays a crucial role in statistical analysis. Combining statistical methods with mathematical optimisation and carefully weighing assumptions and constraints are all part of the process.

The following steps are typically included in the model fitting process:

Model Selection: Selecting a model or family of models that is suitable for the data and the research question is known as model selection. This decision may be supported by theoretical arguments, domain expertise, or prior research.

Parameter Estimation: Estimating the parameters of the model using different estimation methods, such as maximum likelihood estimation or least squares estimation. The estimated parameters are the values with the best fit, minimising the discrepancy between the model predictions and the observed data.

Goodness-of-Fit Assessment: Assessing the goodness-of-fit is the process of determining how well the fitted model fits the data. This is accomplished by evaluating the goodness-of-fit indicators using likelihood-based statistics, measures of variance explained (like R-squared), or residual analysis. If the fit is good, the model has done a good job of capturing the patterns and variability in the data.

Model Validation: Model validation is the process of determining whether the fitted model is valid by evaluating how well it performs on independent or holdout data. This procedure aids in determining whether the model's performance endures beyond the data used for fitting.

Any given data set can be fitted with a wide variety of models, but the best one is the one that passes the goodness of fit test. The exponential model, for which all the parameters and other helpful statistics are provided below, is the best fitted one out of all the models that we tried to fit in "R" for the given data set.

Output:

```
## Residuals:
##         1         2         3         4         5         6
##  0.002605 -0.018211  0.017887 -0.009696  0.025550 -0.018135
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.857e+02  1.005e+01  -68.25 2.76e-07 ***
##   year       3.447e-01  4.977e-03   69.25 2.60e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02082 on 4 degrees of freedom
## Multiple R-squared:  0.9992, Adjusted R-squared:  0.999
## F-statistic:  4796 on 1 and 4 DF,  p-value: 2.605e-07
```

Interpretation:

- **Residuals:** The residuals are the differences between the actual values observed and those predicted by the regression model. Each observation in the dataset has a corresponding six residuals in this instance.
- **Coefficients:** The regression model's estimated coefficients are shown in the coefficients section. The intercept and the coefficient for the variable "year" are the two coefficients in this instance.
    - The intercept, which represents the expected value of the response variable when the predictor variable "year" is zero, is estimated to be -6.857e+02.
    - The calculated coefficient for "year" is 3.447e-01. It shows how the response variable changes when "year" is raised by one unit.
    - The t-value, p-value, and estimated standard error for each coefficient are provided. The p-value represents the likelihood of finding such a value if the true coefficient were zero, while the t-value assesses the significance of the coefficient. Quick information on the level of significance is provided by the significance codes, where "***" denotes a highly significant coefficient.
- **R-squared multiple and Adjusted R-squared:** A measurement of how well the regression model fits the data is the multiple R-squared. It shows the proportion of the response variable's variance that the predictor variable or variables can account for. The multiple R-squared in this instance is reported as 0.9992, which denotes a very high level of explanation.
- The multiple R-squared is adjusted for the number of predictors in the model using the adjusted R-squared. It penalizes the addition of unnecessary predictors. The adjusted R-squared in this instance is 0.999, indicating a very good fit even when the number of predictors is taken into consideration.
- **F-statistic and p-value:** The F-statistic tests the overall significance of the regression model. It assesses whether the predictors, as a group, significantly contribute to explaining the variation in the response variable. The reported F-statistic is 4796, with 1 and 4 degrees of freedom. The associated p-value is 2.605e-07, indicating a highly significant model.

Overall, based on the provided output, the regression model appears to have highly significant coefficients, a high degree of explanation (R-squared), and a good overall fit to the data.

NOTE: - Less observations are there to fit this model (only 6 from 2016-2020). At least 10 observations are required for a good fit.

Prediction:

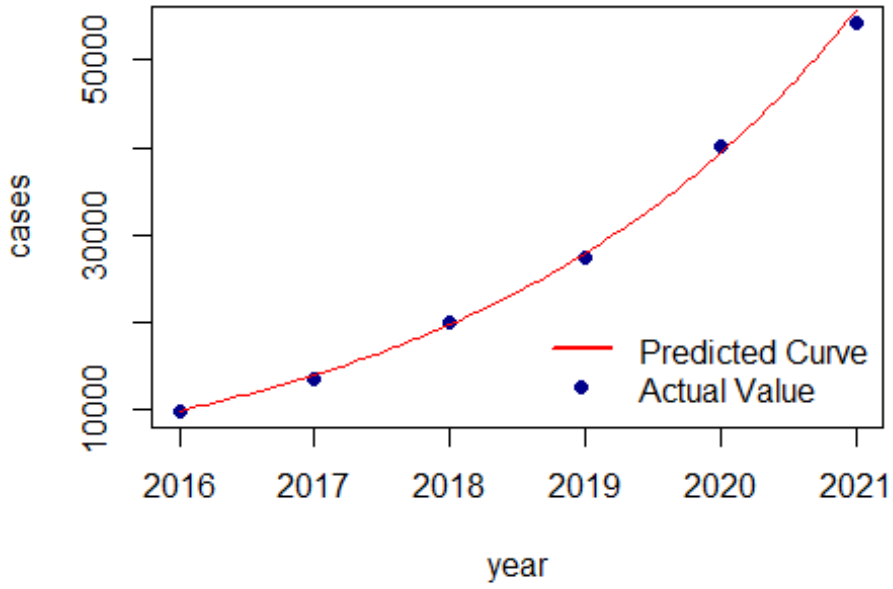
The model's prediction of the number of cybercrime cases from 2016 to 2021 is shown below, along with a comparison of how well it matched the actual data that was made available to us.

To check the dispersion of actual values to the predicted value, we also plotted a curve of the fitted model and plotted the actual values alongside it on the same curve.

With the aid of our fitted model, we have also predicted the number of cybercrime instances that may occur in 2022.

Output:

```
##      year actual predicted
## [1,] 2016   9891  9865.267
## [2,] 2017  13674 13925.302
## [3,] 2018  20011 19656.238
## [4,] 2019  27478 27745.733
## [5,] 2020  40178 39164.447
## [6,] 2021  54289 55282.515
## [7,] 2022     NA 78033.948
```



Interpretation: Although there may be some variation and differences between the predictions and the actual values, it seems that they are generally close to each other. However, since there is only a limited subset of the data provided, it is difficult to draw comprehensive conclusions about the overall performance of the model.

In the year 2022, there may be around 78034 cases of cybercrime that are held for the court's trial if the total number of cases continued the same pattern.

5.2.2 Cases in Which Trial Were Completed:

The data shown shows the number of trials that were successfully completed in different Indian states between 2017 and 2021. The given data deals with 3 types of categories namely

1. **Cases Convicted: -** It refers to legal cases in which the accused person(s) have been found guilty or proven to be accountable for the offence they were charged with It signifies that the court has determined the accused person's guilt based on the evidence presented during the trial.

2. **Cases Discharge: -** It refers to legal cases in which the accused person(s) are released or relieved from the charges filed against them; this implies the court has decided not to proceed with the trial or dropped the charges for a number of reasons and
3. **Cases Acquitted:** - It refers to court proceedings in which the accused person or persons are found not guilty of the charges levelled against them because the court finds that the evidence presented at the trial was insufficient to prove their guilt beyond a reasonable doubt.

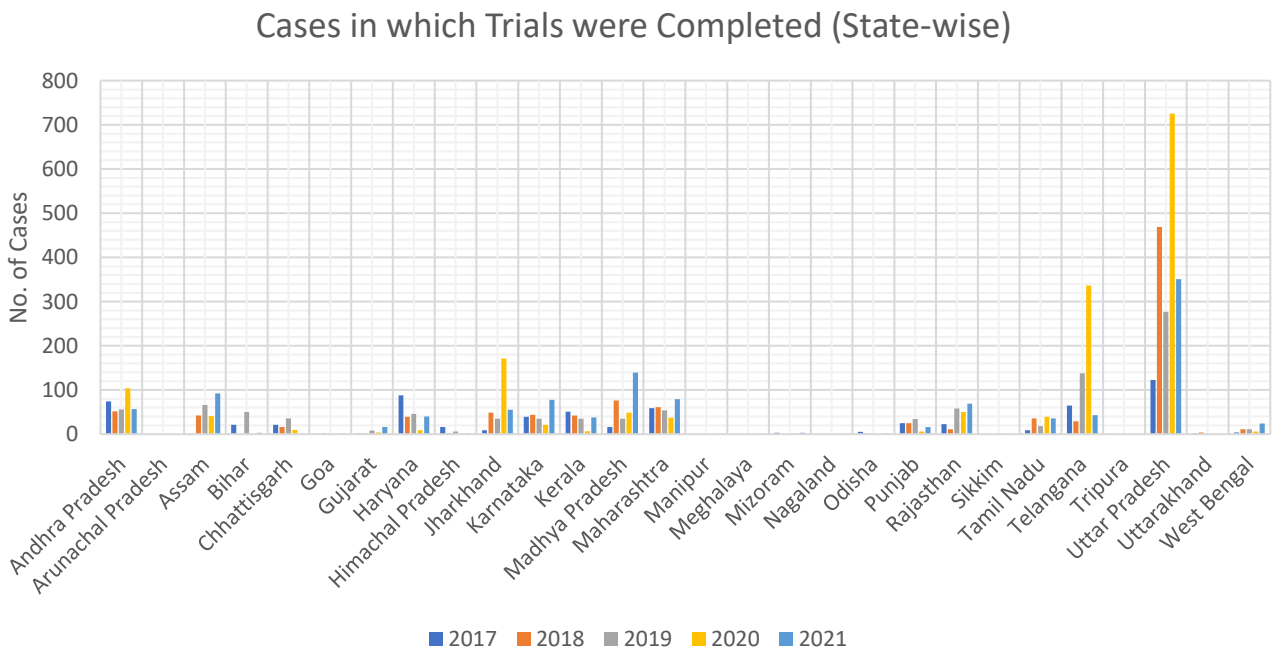
The results of these trials, such as convictions, discharges, or acquittals, aren't mentioned in detail, though. The data does not include information on the ultimate judgement or verdict in each case; it just shows the conclusion of trials.

Data Visualisation

In this section we will plot graph of completed trials yearly both state-wise and UT-wise and try to draw some conclusion from it.

State-wise

A clustered column chart comparing trial completion rates across Indian states over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

- **Overall Trend:** From 663 completed trials in 2017 to 1147 completed trials in all states by 2021. This shows a general increase trend in the number of trials finished over this time.
- **State-specific Variations:** The number of successfully completed trials varies significantly between states.

  - **High Number of Completed Trials:** During the five-year period, several states, like Uttar Pradesh, Maharashtra, Tamil Nadu, Telangana, and Karnataka, consistently had a high number of completed trials. Trials have routinely been conducted in these states.
  - **Patterns of Variation:** Over time, the number of successfully concluded trials varied in several states. For instance, the number of completed trials in Andhra Pradesh increased in 2020 but decreased in 2021. In Gujarat, completed trials increased significantly in 2021 compared to other years. from 2017 to 2020,

completed trials significantly decreased in Himachal Pradesh, although they slightly increased in 2021. With only a few trials performed each year, Goa and Mizoram displayed oscillations.

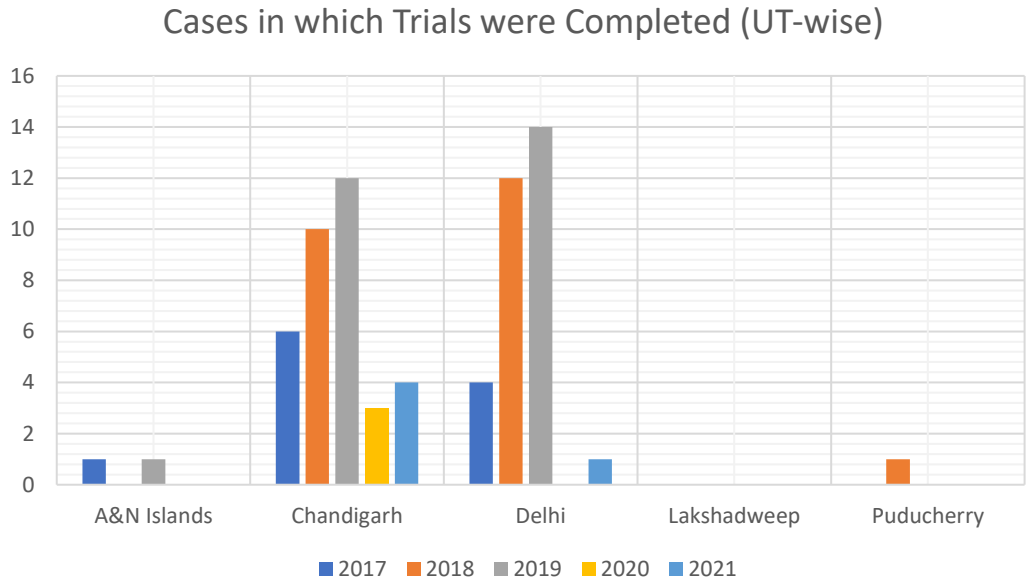
- **Low or No Trials:** On the other side, at this time, there were either very few or no trials conducted in states like Arunachal Pradesh, Manipur, Meghalaya, Nagaland, Sikkim, and Tripura.

- **Changes over Time:** The number of trials that were successfully completed changed over time. Although the overall number of trials completed rose, there were variations in some years. For instance, 1021 completed trials were finished in 2018, whereas 1622 were completed in 2020.
- **Regional Trends:** Some locations demonstrated greater rates of trial activity. For instance, compared to states in the north-eastern area, states in the northern region, such as Uttar Pradesh, Haryana, Punjab, and Himachal Pradesh, had a comparatively larger number of completed trials.

NOTE: We have excluded the data of Jammu & Kashmir as it undergoes through a major change in its political status on August 5, 2019.

UT-wise

A clustered column chart comparing trial completion rates across Indian Union Territories over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

- **A&N Islands:** One trial was finished in 2017 and another was finished in 2019. In the years 2018, 2020, and 2021, no trials were wrapped up.
- **Chandigarh:** Over the years, the number of trials that have been successfully concluded in Chandigarh has varied. Six trials were finished in 2017, and that number rose to ten in 2018 and then to twelve in 2019. In 2020, there were only 3 trials, while in 2021, there were somewhat more—4 trials.
- **Delhi:** The outcomes of the trials there have been uneven. Four trials were conducted in 2017; this number rose to 12 in 2018 and then to 14 in 2019. However, there were no trials finished in 2020 and only one trial was finished in 2021.
- **Lakshadweep:** Between the years of 2017 and 2021, no trials were finished in Lakshadweep.
- **Puducherry:** In Puducherry, a single trial was successfully concluded in 2018. In 2017, 2019, 2020, and 2021, no trials were finished.
- **Overall Trend: -** From 11 in 2017 to 27 in 2019, the overall number of Union Territories trials that were successfully concluded rose progressively. Only 4 trials were completed in 2020, a substantial decrease from the total number of completed trials. The overall number of successfully completed trials grew slightly to 5 in

2021, but it remained lower than in the years prior. Over the course of the time, there were ups and downs in the overall number of finished trials in the Union Territories, with some years seeing a rise and others a drop.

NOTE: - We have excluded the data of Daman & Diu and D&N Haveli as it undergoes through a major change in its political status on November 26, 2019. We haven't included the data of Ladakh and Jammu & Kashmir as well as it became a U.T. on August 5 ,2019.

5.2.3 Conviction Rate:

The percentage of cases with a conviction or guilty judgement relative to all cases with completed trials is referred to as the "conviction rate." It is a metric used to evaluate how well the criminal justice system performs in terms of obtaining convictions for criminal offences.

The number of convictions divided by the total number of cases that resulted in a verdict (convicted or acquitted) and multiplied by 100 to represent it as a percentage is the conviction rate.

The likelihood of a conviction can vary greatly based on a number of variables, including the nature of the offence, the relevant jurisdiction, the strength of the evidence, the calibre of the investigation, the skill of the prosecutors, the fairness of the judicial system, and other contextual considerations.

A high conviction rate implies an effective legal system and a strong prosecution by showing that a sizable number of cases that proceed to trial result in convictions. On the other side, a low conviction rate could make people wonder if the criminal justice system is indeed successful at obtaining convictions and holding people accountable for their claimed crimes.

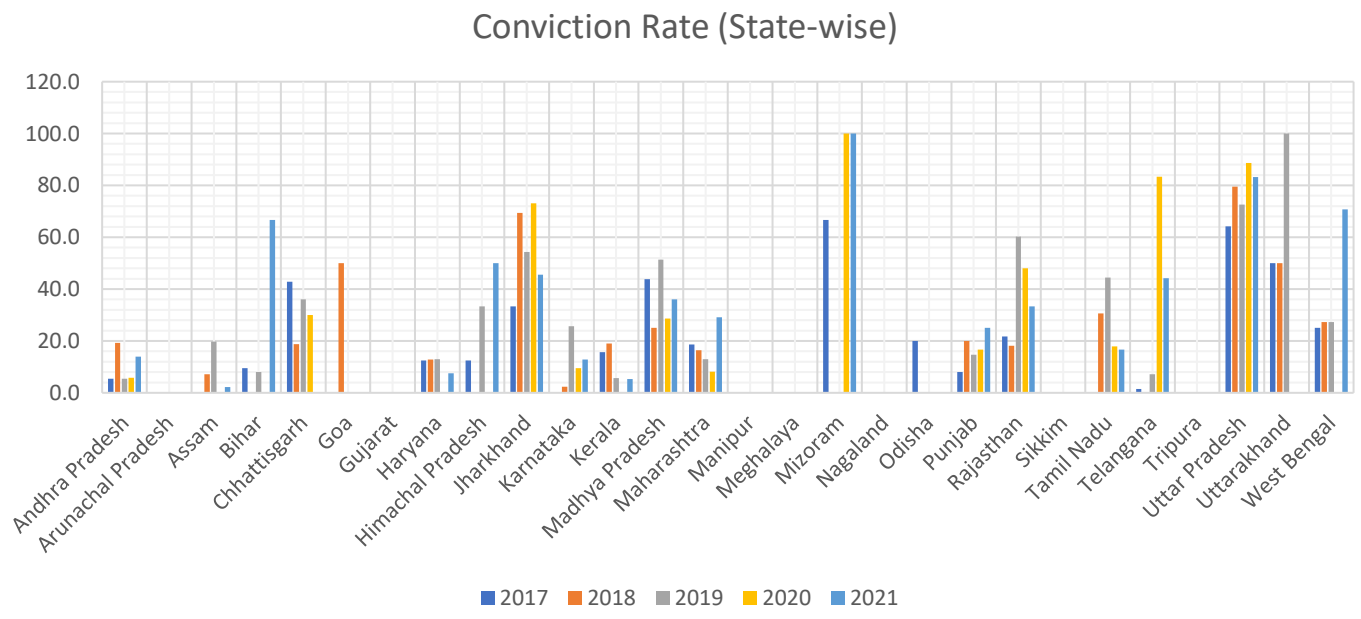
It's vital to remember that conviction rates shouldn't be used as the only indicator of a judicial system's efficiency or justice. When assessing the overall effectiveness of the criminal justice system, other elements including procedural protections, access to justice, and the preservation of human rights should also be taken into account.

Data Visualisation

In this section we will plot graph of conviction rate yearly both state-wise and UT-wise and try to draw some conclusion from it.

State-wise

A clustered column chart comparing conviction rates across Indian states over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

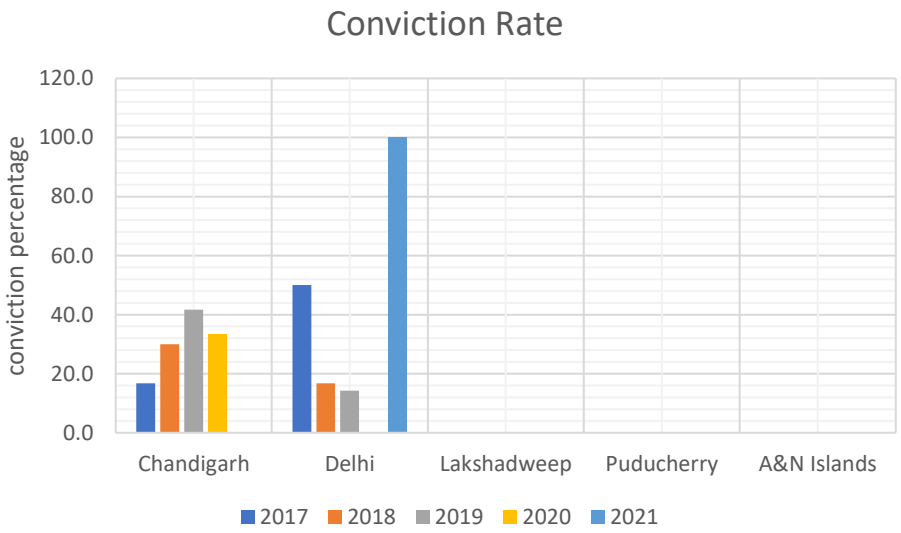
- **Varying Conviction Rates:** Conviction rates vary significantly between states and UTs, showing differences in the efficiency of the criminal justice system in various jurisdictions.
- **High Conviction Rates:** States like Uttar Pradesh, Uttarakhand, Rajasthan, Mizoram, and West Bengal have continuously had high conviction rates throughout the years. Conviction rates in these states have regularly outperformed the national average.
- **Conviction Rate Fluctuations:** Over the years, conviction rates have varied in a number of states. There have been fluctuating conviction rates in states including Tamil Nadu, Karnataka, Kerala, and Madhya Pradesh, with some years seeing greater rates and other years seeing lower rates.
- **Low Conviction Rates:** Over the specified time, several states have continuously had low conviction rates. Among these are Gujarat, Goa, and Manipur, where the conviction rate has often been zero or very near to zero. This suggests that it may be difficult to get convictions for cybercrimes in certain states.
- **Overall National Conviction Rate:** Over time, there have been changes in the overall conviction rate across all states. With the exception of 2017 and 2020, the total conviction rate is under 50% for the most of the years.

Note: - Some states' conviction rates for a given year cannot be calculated since there were zero cases with completed trials. Like in 2017, it is impossible to calculate the conviction rate for Arunachal Pradesh, Nagaland, Sikkim, and Manipur. Similar to how there were 7 states in 2018, 7 in 2019, 6 in 2020, and 6 in 2021 for which it is impossible to establish the conviction rate.

NOTE: - We have excluded the data of Jammu & Kashmir as it undergoes through a major change in its political status on August 5, 2019.

UT-wise

A clustered column chart comparing trial completion rates across Indian Union Territories over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

- **Chandigarh:** Over the years, the conviction rate in Chandigarh has changed. Starting at 16.7% in 2017 and rising to 41.7% in 2019, it fell to 0% in 2021. There was a small variation in between, but overall Chandigarh's conviction rates have been inconsistent.
- **Delhi:** The conviction rate there was 50% in 2017, dropped to 14.3% in 2019, then rose to 100% in 2021. Although the conviction rate is 100% but it is due to the reason that only 1 case was finished with results in the conviction of the accused. For 2020 conviction rate can’t be computed as no trial were completed in that year.

- **Lakshadweep: -** Conviction rate for the union territory cannot be computed for any given year as no trials were completed in any particular year.
- **Puducherry:** The information at hand indicates a conviction rate of 0% in 2018. nevertheless, it cannot be calculated for the other years because no trials were executed in those years.
- **A&N Islands:** In 2017 and 2019, there were no convictions in the A&N Islands as the conviction rate is 0% in these years. For other year conviction is impossible to calculate as no trial were executed.
- **Total Conviction Rates:** Over the specified time, there have been variations in the Union Territories' total conviction rates. The overall conviction rate was 27.3% in 2017, 21.7% in 2018, 25.9% in 2019, and 25.0% in 2020 before dropping to 25.0%. It further fell to 20% in 2021. These variations show differences in the prosecution and conviction success rates for criminal offences between Union Territories over the relevant years.

NOTE: - Some UT's conviction rates for a given year cannot be calculated since there were zero cases with completed trials. Like in 2017 it is impossible to calculate the conviction rate for Lakshadweep and Puducherry. Similarly for 2018 Lakshadweep and A&N Islands, for 2019 Lakshadweep and Puducherry, for 2020 Delhi Lakshadweep Puducherry A&N Islands and for 2021 Lakshadweep Puducherry A&N Islands conviction rate cannot be evaluated.

NOTE: - We have excluded the data of Daman & Diu and D&N Haveli as it undergoes through a major change in its political status on November 26, 2019. We haven't included the data of Ladakh and Jammu & Kashmir as well as it became a U.T. on August 5 ,2019.

5.2.4 Cases Pending Trial at End of the Year: -

The phrase "Cyber Crime Cases Pending Trial at End of the Year" refers to the quantity of Cyber Crime cases that are still pending trial as at the end of a specific year. This demonstrates the backlog of unresolved cybercrime cases and the workload and effectiveness of the legal system in processing such matters.

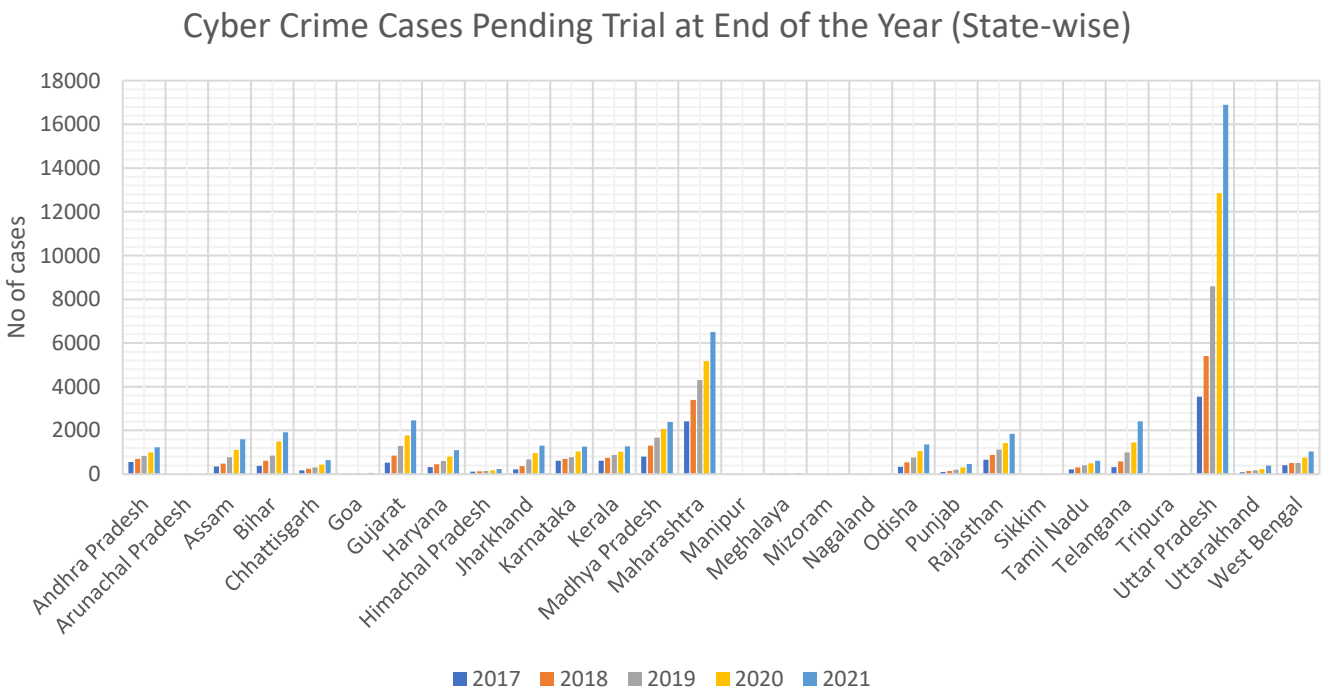
The amount of cybercrime cases still awaiting trial at the end of the year sheds light on the difficulties the judicial system has in handling this expanding problem. It's crucial to remember that a variety of factors, such as the volume of reported cybercrimes, the capability of the legal system, the complexity of the cases, and the effectiveness of case management, might affect the number of outstanding cases.

Data Visualisation

In this section we will plot graph of Pending Cyber Crime cases yearly both state-wise and UT-wise and try to draw some conclusion from it.

State-wise

A clustered column chart comparing trial completion rates across Indian states over a 5-year period (2017–2021) is shown below.



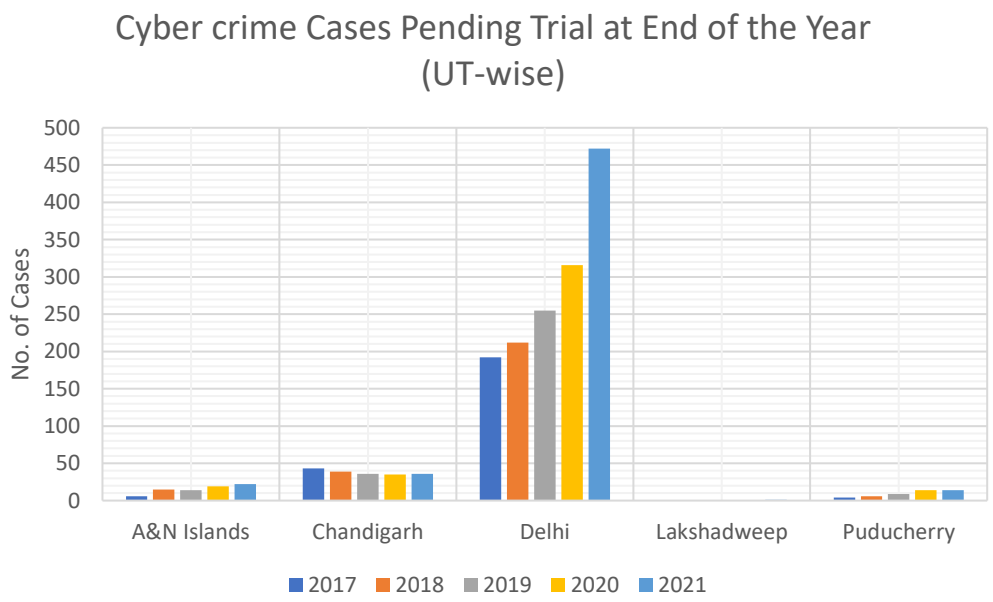
We can draw the following inference from the above graph: -

- **Rising Trend:** From 2017 to 2021, there have been an obvious rise in the overall number of cybercrime cases that are awaiting trial across all states, going from 12,930 to 47,160. This shows that the number of unsolved cybercrime cases in the legal system is expanding.
- **States with high Number of outstanding Cases:** Maharashtra regularly has the greatest number of outstanding cybercrime cases, with a considerable increase from 2,422 in 2017 to 6,498 in 2021. Uttar Pradesh, Gujarat, Madhya Pradesh, and Karnataka are among states that routinely have a large number of open cases.
- **Varying State-Wide Trends:** The amount of cybercrime cases awaiting trial varies significantly among states. Some states, including Assam, Bihar, Chhattisgarh, Haryana, Jharkhand, Kerala, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, Telangana, and West Bengal, exhibit an ongoing rise in the number of open cases over time.
- **States with Fluctuating Trends:** The number of active cybercrime cases varies in several states. For instance, the number of outstanding cases varies over time in Andhra Pradesh, Himachal Pradesh, Karnataka, Punjab, Tamil Nadu, and Uttarakhand, with both rises and falls.
- **Low Number of Pending Cases:** During the specified time, the number of pending cybercrime cases in Arunachal Pradesh, Manipur, Mizoram, Nagaland, Sikkim, and Tripura was comparatively low. However, the overall number of reported instances is also modest, as are the numbers of cases that were successfully resolved. It doesn't imply that these states have a high case disposition rate.

NOTE: - We have excluded the data of Jammu & Kashmir as it undergoes through a major change in its political status on August 5, 2019.

UT-wise

A clustered column chart comparing trial completion rates across Indian Union Territories over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

- **A&N Islands:** Over time, there have been more cybercrime cases in A&N Islands that are still undergoing trial. It began with 6 instances in 2017 and increased slowly to 22 cases in 2021, with peaks and valleys along the way.
- **Chandigarh:** Although there have been minor changes, the number of active cybercrime cases in Chandigarh has remained largely consistent. With slight fluctuations, the number of cases awaiting trial reduced from 43 in 2017 to 36 in 2021.
- **Delhi:** There have continuously been several cybercrime cases in Delhi that are awaiting trial. From 192 instances in 2017 to 472 cases in 2021, the number climbed significantly each year after that. This shows that there is a substantial backlog of cybercrime cases in Delhi.

- **Lakshadweep:** Until one case was listed as under trial in 2021, there were no pending cyber crime cases in Lakshadweep.
- **Puducherry:** Much to A&N Islands, Puducherry has seen a rise in the number of cybercrime cases that are now in the court system. Four instances were reported in 2017 and fourteen cases in 2021.
- **Overall Trend: -** Over the specified time, there was a steady rise in the overall number of cybercrime cases that were still for trial in Union Territories. From 245 instances in 2017 to 545 cases in 2021, it increased.
  The Union Territories' backlog of cybercrime cases has significantly increased, as shown by the fact that 2021 had the most active cases. The rise in the number of open cases in several Union Territories, including Chandigarh, Puducherry, and the A&N Islands, points to an increasing load on the legal system in dealing with cybercrime matters.
  Even tiny Union Territories are susceptible to similar situations, as at least one ongoing cybercrime case in Lakshadweep in 2021 shows.

NOTE: - We have excluded the data of Daman & Diu and D&N Haveli as it undergoes through a major change in its political status on November 26, 2019. We haven't included the data of Ladakh and Jammu & Kashmir as well as it became a U.T. on August 5 ,2019.

5.2.5 Pendency Rate:

The percentage of cases that are unresolved or still pending after a certain amount of time is referred to as the pendency rate. It serves as a gauge for the amount of cases accumulating in the court system's backlog.

The number of outstanding cases is divided by the total number of cases to determine the pendency rate, which is then multiplied by 100 to represent it as a percentage.

Different courts, regions, and case kinds may have varying pendency rates. Numerous variables have a role in it, including the volume of cases submitted, the judiciary's capacity, the effectiveness of case management systems, the accessibility of resources, and the complexity of legal processes.

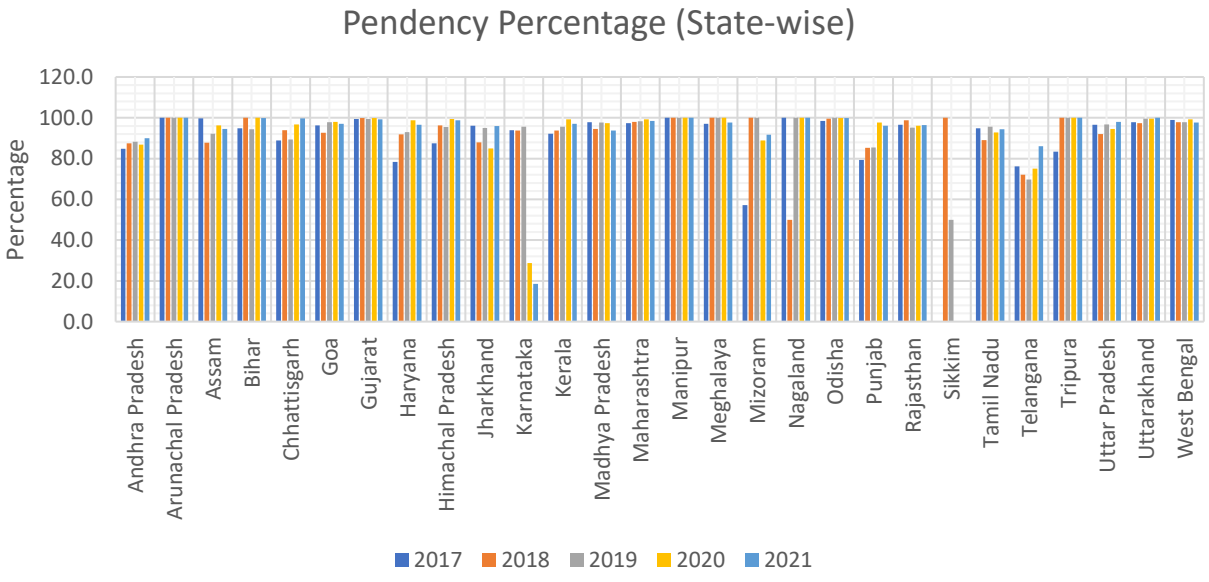
The pendency rate must be closely monitored in order to pinpoint areas that require improvement, such as process simplification, court capacity expansion, the adoption of technology-driven solutions, and the adoption of efficient case management techniques. The legal system can provide justice promptly and preserve public confidence by lowering the pendency rate.

Data Visualisation

In this section we will plot graph of Pendency percentage of Cyber Crime cases yearly both state-wise and UT-wise and try to draw some conclusion from it.

State-wise

A clustered column chart comparing Pendency percentage across Indian states over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

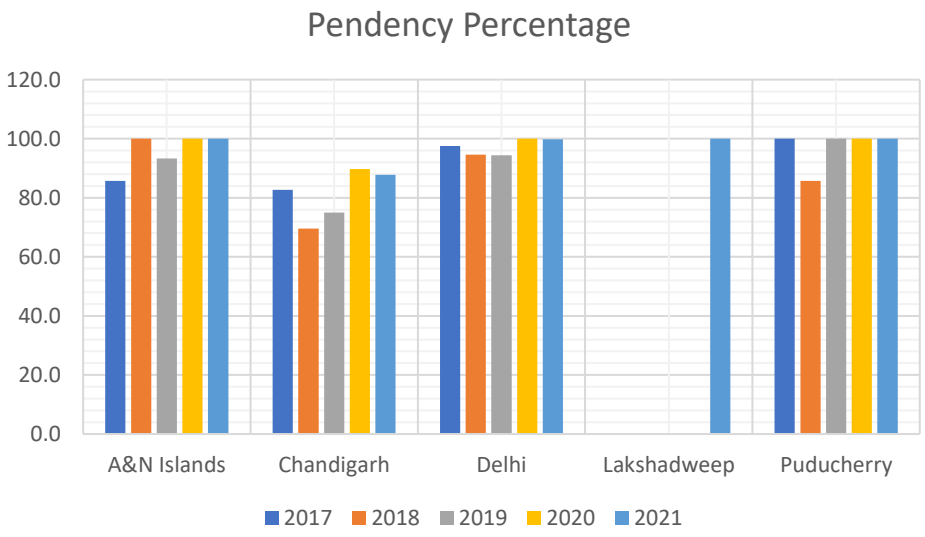
- **Varying Pendency Percentages:** The pendency percentages change over time and differ amongst states. While some states have lower percentages or inconsistent patterns, others exhibit substantially greater pendency percentages, suggesting a higher number of outstanding cases.
- **Fluctuating Pendency Percentages:** Over time, the pendency percentages in several states have changed. For instance, the pendency rates of West Bengal, Mizoram, and Karnataka all show different trends.
- **Higher Pendency Percentages: -** States with higher pendency percentages include Andhra Pradesh, Assam, Bihar, Chhattisgarh, Goa, Gujarat, Haryana, Himachal Pradesh, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Meghalaya, Odisha, Punjab, Rajasthan, Tamil Nadu, Telangana, Uttar Pradesh, Uttarakhand, and West Bengal. These states have pendency percentages that are higher than the national average for the time period.
- **States with Significant Improvements:** Over the years, certain states have made progress in lowering their pendency percentages. For instance, decreased pendency rates have been attained or regularly maintained in Uttarakhand, Mizoram, Meghalaya, and Manipur.
  Overall Average Pendency Percentage: The aggregate average pendency percentage for all states combined during the course of the specified time is 86.9%. This shows that, on average, 86.9% of cases were still open at the end of the year.
- **Pendency Rate Complexities: -** A high pendency rate does not always indicate that there are many cases still pending. For instance, the North-eastern states have a high pendency rate due to a low overall number of reported cases which may be due to potential court system constraints.

NOTE: - Pendency percentage of Sikkim for the year 2017 & 2021 as there were no cases of cyber crime in the court for trial in that particular year.

NOTE: - We have excluded the data of Jammu & Kashmir as it undergoes through a major change in its political status on August 5, 2019.

UT-wise

A clustered column chart comparing Pendency percentage across Indian Union Territories over a 5-year period (2017–2021) is shown below.



We can draw the following inference from the above graph: -

- **A&N Islands:** From 85.7% to 100%, the pendency percentage in the A&N Islands was continuously high over the years. This suggests that there is a substantial backlog of open cases in the area.
- **Chandigarh:** Values for the pendency percentage in Chandigarh have ranged from 69.6% to 89.7% over the years. Even though there were swings, Chandigarh had a sizable amount of cases that were still outstanding at the time.
- **Delhi:** Pendency rates in Delhi were also high, ranging from 94.4% to 100%. This shows that the Union Territory has a sizable backlog of open cases.

- **Lakshadweep:** The data shows a pendency proportion of 100% for the year 2021, indicating that all cases in Lakshadweep were unresolved and pending at the time. For all other years (2017-2020) pendency rate cannot be computed as no case was being reported during that time period
- **Puducherry:** Throughout all of the years, Puducherry likewise had high pendency percentages that remained at 100%. This suggests that there was a total backlog of open cases in Puducherry since all cases were left unresolved.
- **Overall:** Over the course of the time, the pendency % for Union Territories varied. In 2017, it was 94.2%; in 2018, it was 90.0%; in 2019, it was 91.8%; and in 2020 and 2021, it was 98.9%. These figures show that there is a substantial backlog of cases that need to be handled across Union Territories.

NOTE: - Pendency percentage of Lakshadweep for the year 2017 & 2021 as there were no cases of cyber crime in the court for trial in that particular year.

NOTE: - We have excluded the data of Daman & Diu and D&N Haveli as it undergoes through a major change in its political status on November 26, 2019. Also we haven't included the data of Ladakh and Jammu & Kashmir as well as it became a U.T. on August 5 ,2019.

5.3 Analysis of Court Disposal of Cyber Crime Cases (Crime Head-wise)

The crime head has been classified into three groups, including:

- Offences under Information Technology (I.T.) Act
- Offences under Indian Penal Code (IPC)
- Offences under Special and Local Law (SLL)

Offences under Information Technology (I.T.) Act includes: -

1. Tampering computer source documents
2. Computer Related Offences
3. Cyber Terrorism
4. Publication/transmission of obscene / sexually explicit act in electronic form
5. Interception or monitoring or decryption of Information
6. Un-authorized access/attempt to access to protected computer system
7. Abetment to Commit Offences
8. Attempt to Commit Offences
9. Other Sections of IT Act

Offences under Indian Penal Code (IPC) includes: -

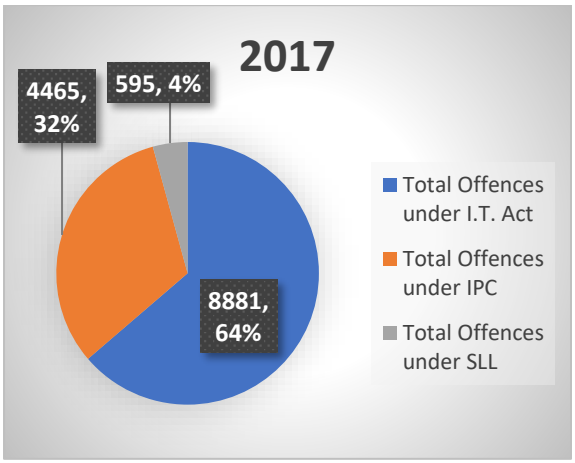
1. Abetment of Suicide (Online)
2. Cyber Stalking/Bullying of Women/Children
3. Data theft
4. Fraud
5. Cheating
6. Forgery
7. Defamation/Morphing
8. Fake Profile
9. Counterfeiting
10. Cyber Blackmailing/Threatening
11. Fake News on social media
12. Other Offences

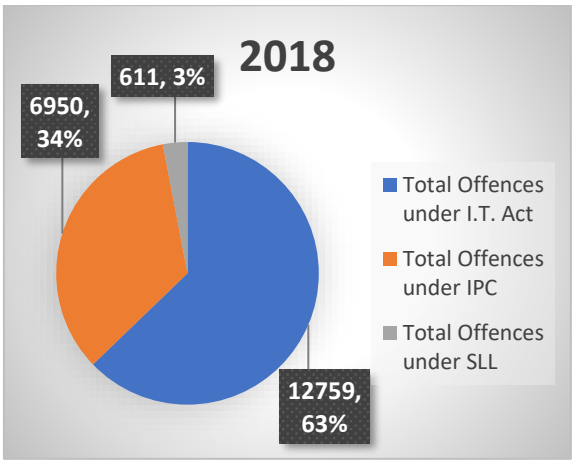
Offences under Special and Local Law (SLL) includes: -

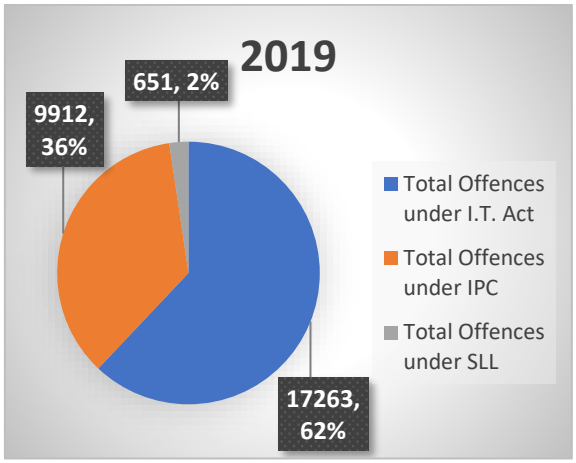
1. Gambling Act (Online Gambling)
2. Lotteries Act (Online Lotteries)
3. Copy Right Act
4. Trade Marks Act
5. Other SLL Crimes

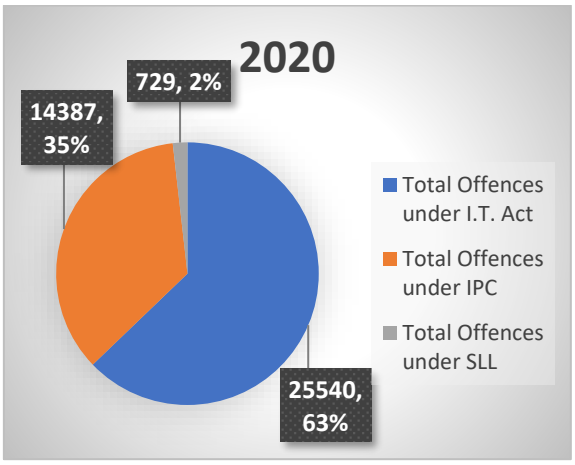
5.3.1 Pie-charts

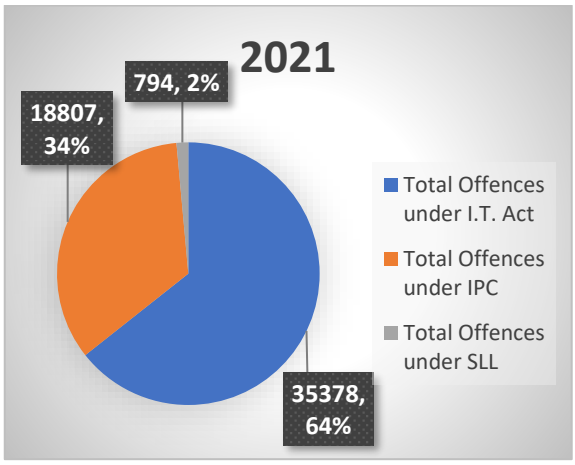
The head-to-head crime data from 2017 through 2021 are displayed in the pie charts below. We'll conclude some inferences from it.











We can draw the following inference from the above Pie charts: -

- **Average Number of Cases:** As can be observed on average, 2.6% of cases fall under the SLL Act, 34.2% of cases fall beneath the IPC, and 63.2% of cases fall underneath the I.T. Act over the duration of 5 years (2017-2021).
- **Overall Trends:** Over time, more offences have been committed in total under the I.T. Act, IPC, and SLL. The figures show that the recorded infractions are generally on the rise.
- **Magnitude of crimes:** Compared to the IPC and SLL, the overall number of crimes under the IT Act is constantly larger. This shows that crimes covered by the Indian Penal Code and Special and Local Laws are less common than those linked to information technology and cybercrimes.

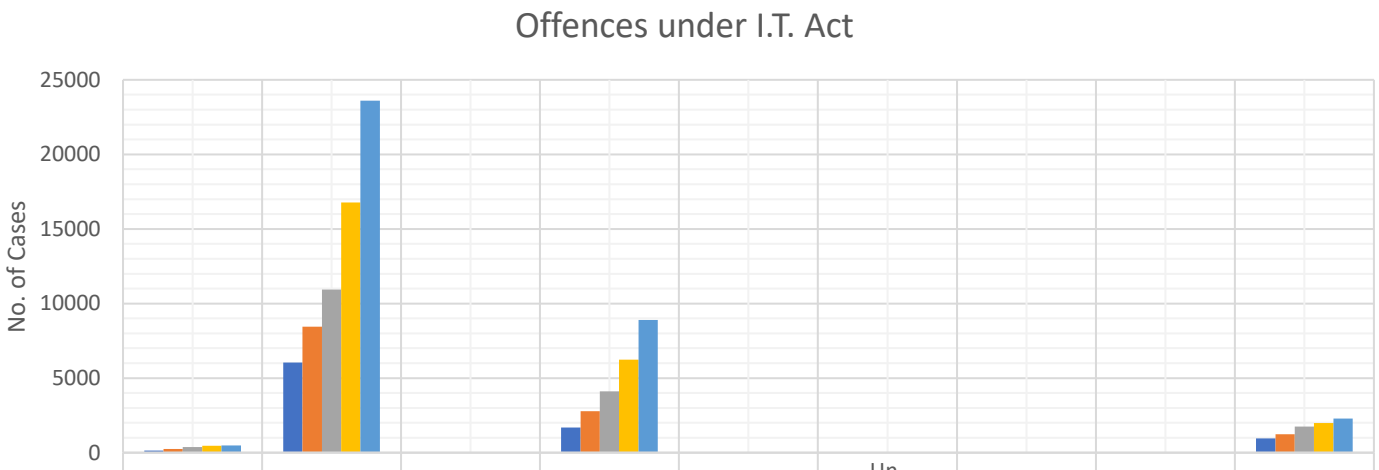
5.3.2 Analysis for Offences Under I.T. Act

In this section, we shall analyse the data that is provided to us under the crime head-wise of the I.T. Act category statistically and quantitatively. We'll make an effort to gather some intriguing data, spot some trends, develop a model, and predict likely outcomes.

Data Visualisation

In this section we will plot graph of offences under the Information Technology (I.T.) Act subcategory wise yearly and try to draw some conclusion from it.

A clustered column chart comparing the sub-categories of the offences under the I.T. Act over a 5-year period (2017–2021) is shown below.



	Tampering computer source documents	Computer Related Offences	Cyber Terrorism	Publication/ transmission of obscene / sexually explicit act in electronic form	Interception or Monitoring or decryption of Information	Un-authorized access/attempt to access to protected computer system	Abetment to Commit Offences	Attempt to Commit Offences	Other Sections of IT Act
2017	141	6048	16	1696	2	10	0	2	966
2018	253	8438	21	2790	2	9	1	10	1235
2019	383	10946	26	4118	10	11	0	20	1749
2020	456	16766	30	6239	13	11	1	32	1992
2021	483	23609	36	8903	14	11	5	35	2282

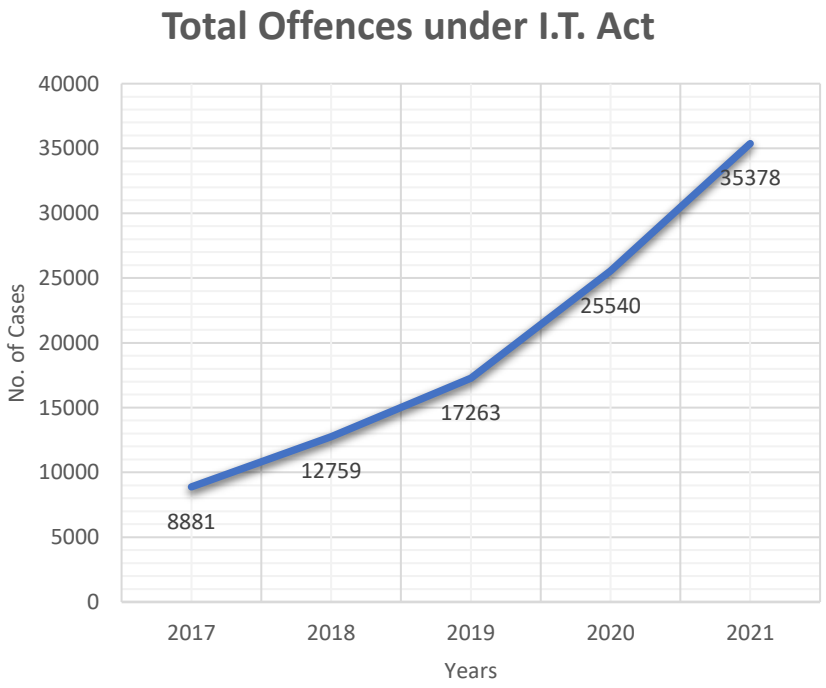
We can draw the following inference from the above graph: -

- **Major Contribution:** Only 4 sub-categories, out of a total of 9, contribute to more than 99.6% of crimes in this category over the course of 5 years (2017-2021), including "Tampering with Computer Source Documents," "Computer Related Offences," "Publication/Transmission of Obscene / Sexually Explicit Act in Electronic Form," and "Other Sections of IT Act." (99.66% in 2017, 99.66% in 2018, 99.61 in 2019, 99.65 in 2020, and 99.7 in 2021)
- **Computer source document tampering:** Over the years, there has been a steady rise in the reported occurrences of computer source document tampering, with a clear uptick starting in 2019.
- **Computer-Related Offences:** From 2017 to 2021, the number of computer-related offences climbed steadily, pointing to an upsurge in cybercrimes.
- **Cyberterrorism:** Despite being relatively rare, occurrences of cyberterrorism have steadily increased over time, suggesting a potential problem in this field.
- **Publishing/transmission of obscene/sexually explicit actions:** From 2017 to 2021, there was a considerable increase in the number of instances involving the publishing or transmission of obscene or sexually explicit acts in electronic form, underscoring the necessity to address such offences.
- **Information interception, monitoring, or decryption:** Although there has been a minor rise over time, the recorded instances of information interception, monitoring, or decryption have remained very low.
- **Unauthorised access to protected computer systems:** Over the years, there have been only minor variations in the number of incidents involving unauthorised access to or efforts to gain access to protected computer systems.
- **Abetment to Commit Offences:** According to the data, there have been fewer recorded instances of crimes using the I.T. Act in recent years.
- **attempts to Commit Crimes:** Over the years, there has been a steady rise in the reported incidents of efforts to commit crimes, indicating a rising propensity for cybercrimes.
- **Other Sections of IT Act:** Over time, there have been increasingly more instances that fall under other sections of the IT Act, showing that a wider variety of cybercrimes are being recorded.

Statistical Analysis

This phase will cover all the total no. of offences under the IT Act that occurred between 2017 and 2021.

Graphical representation of total offences under I.T. Act is given below: -



We can draw the following inference from the above graph: -

- **Increasing trend:** The frequency of violations of the Information Technology Act has been rising over time. The overall number of offences has an increase tendency from 2017 to 2021.

- **Huge surge:** During this time, there has been a large increase in the overall number of offences. Between 2017 and 2021, the number of offences more than quadrupled, demonstrating a significant uptick in the conduct of crimes covered by the I.T. Act.
- **Accelerated growth:** It seems as though the growth rate is speeding up with time. The rise in offences is not linear; rather, it has been progressively steeper in recent years. This shows that cybercrimes are becoming more prevalent and that tougher measures are required to combat them.
- **Rising awareness and enforcement:** The increase in offences may be a sign that authorities are doing a better job of enforcing their laws and raising public awareness of digital crimes. This may be the outcome of improved digital connection, broader technological use, and a stronger emphasis on cybersecurity.

- **Need for cybersecurity measures:** Strong cybersecurity measures are required for people, corporations, and governmental organisations, as evidenced by the rising number of offences under the IT Act. It highlights the need of putting into place solid security procedures and spending money on cybersecurity infrastructure to defend against online attacks.

Model Fitting & Goodness of Fit:

A broad range of models can be used to fit the given data set of total no. offences under the I.T. Act, but the model that passes the goodness of fit test is always the best option. Out of all the models we attempted to fit in "R" for the given data set, the exponential model, for which all the parameters and other useful information are supplied below, is the one that fits the data set the best.

Output: -

```
## Residuals:
##         1         2         3         4         5
## -0.001424  0.015062 -0.028445  0.017399 -0.002593
##
## Coefficients: Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.885e+02  1.353e+01  -50.88 1.67e-05 ***
## x            3.458e-01  6.702e-03   51.60 1.60e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02119 on 3 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9985
## F-statistic:  2663 on 1 and 3 DF,  p-value: 1.603e-05
```

Interpretation: The result from the previous step allows us to derive a total of 7 areas of interest. They are interpreted as follows:

- **Residuals:** These are the variations between the values that were seen and those that the regression model predicted. We got five residuals in this instance: -0.001424, 0.015062, -0.028445, 0.017399, and -0.002593. The residuals give an indication of how well the model captured the data's variability.
- **Coefficients:** The computed regression parameters are represented by the coefficients. The intercept, sometimes referred to as the constant term, is represented by the first coefficient, which is projected to have a value of -6.885e+02. The slope of the regression line is represented by the second coefficient, denoted by the letter "x," and it has an estimated value of 3.458e-01. These correlation coefficients show how the predictor variable (x) and the responder variable are related.

- **Standard Error:** The calculated coefficients' variability is measured by the standard error. The standard error is 1.353e+01 for the intercept and 6.702e-03 for the coefficient of x. Smaller standard errors indicate greater precision in the coefficient estimates.
- **t-value and p-values:** These numbers are used to evaluate the coefficients' statistical significance. By dividing the coefficient estimate by its standard error, the t-value is determined. High t-values (50.88 and 51.60, respectively) in this instance show that the intercept and coefficient of x are statistically significant. The p-value indicates the likelihood that such severe t-values would be seen if the null hypothesis—that there is no association between the variables—were true. Strong evidence is suggested against the null hypothesis in favour of a meaningful association by the extremely modest p-values (1.67e-05 and 1.60e-05).
- **R-squared:** The multiple R-squared (0.9989) is the proportion of the response variable's variation that the model's predictor variable or variables can account for. In this instance, the predictor variable(s) account for about 99.89% of the variation in the response variable. The adjusted R-squared (0.9985) provides a more trustworthy indicator of model fit by adjusting the R-squared value for the number of predictors and the sample size.
- **F-statistic and p-value:** The F-statistic evaluates the regression model's overall significance. It determines if there is a substantial relationship between the predictor variable(s) and the response variable overall. The F-statistic in this instance is 2663, and the p-value is 1.603e-05, which provides compelling evidence that the model as a whole is significant.

Overall, the study indicates that the predictor variable (x) and the response variable have a strong linear connection. High statistical significance and a high percentage of variation explained suggest that the model offers a good fit to the data.

NOTE: - Less observations are there to fit this model (only 5 from 2017-2021). At least 10 observations are required for a good fit.

Prediction:

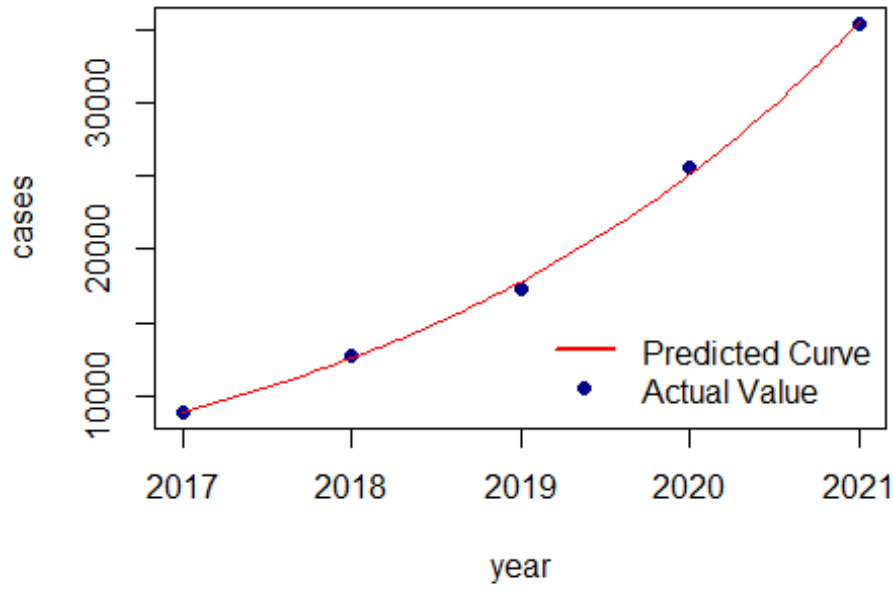
The model's prediction of the total number of offences under the I.T. Act from 2017 to 2021 is shown below in the tabular format using “R”, along with a comparison of how well it matched the actual data that was available to us.

To check the dispersion of actual values to the predicted value, we also plotted a curve of the fitted model and plotted the actual values alongside it on the same curve.

With the help of our fitted model, we have also predicted the number of cybercrime instances that may occur in 2022.

Output: -

```
##      year actual predicted
## [1,] 2017   8881  8893.657
## [2,] 2018  12759 12568.258
## [3,] 2019  17263 17761.097
## [4,] 2020  25540 25099.467
## [5,] 2021  35378 35469.838
## [6,] 2022     NA 50124.945
```



Inference: Although not always accurate, the projected values are generally near to the observed values, suggesting that the predictions had a reasonable degree of accuracy. Generally speaking, the projections mirror the upward trend of the actual values, though year to year fluctuations in accuracy may occur.

If the violations under the I.T. Act continued on their current trajectory, the expected value from the model for the year 2022 is 50125, suggesting that nearly 50125 cases could occur in that year.

Note: - Although the model fits quite well but as there are less observation for model fitting so the predicted value might not be that accurate.

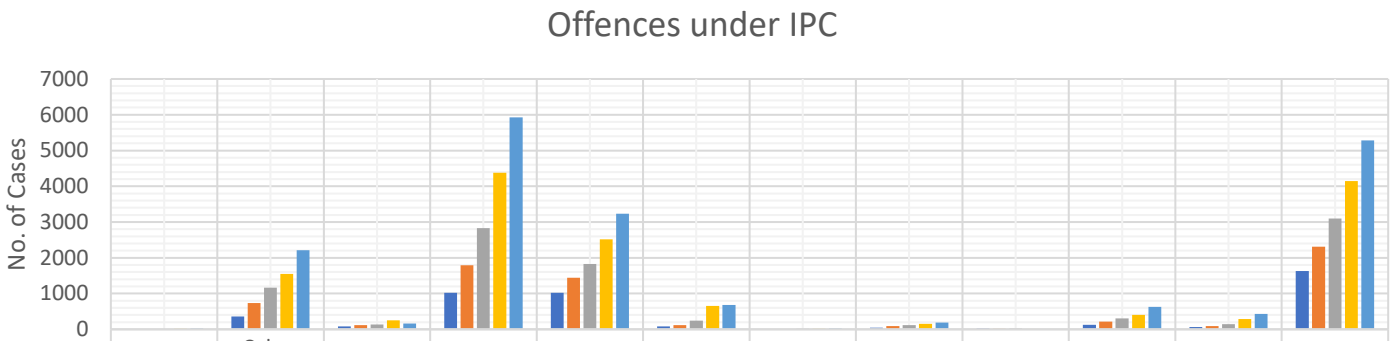
5.3.3 Analysis for Offences Under IPC

In this section, we shall analyse the data that is provided to us under the crime head-wise of the IPC category statistically and quantitatively. We'll make an effort to gather some intriguing data, spot some trends, develop a model, and predict likely outcomes.

Data Visualisation

In this section we will plot graph of offences under the Information Technology (I.T.) Act subcategory wise yearly and try to draw some conclusion from it.

A clustered column chart comparing the sub-categories of the offences under the IPC over a 5-year period (2017–2021) is shown below.



	Abetment of Suicide (Online)	Cyber Stalking/ Bullying of Women/ Children	Data theft	Fraud	Cheating	Forgery	Defamation/Morphing	Fake Profile	Counterfeiting	Cyber Blackmailing/Threatening	Fake News on Social Media	Other Offences
2017	1	361	85	1023	1025	80	6	46	25	123	61	1629
2018	6	733	119	1787	1444	121	11	92	23	214	91	2309
2019	10	1162	136	2828	1826	242	15	118	28	301	145	3101
2020	17	1548	248	4377	2516	656	19	153	13	406	287	4147
2021	27	2210	163	5929	3230	681	30	187	14	631	427	5278

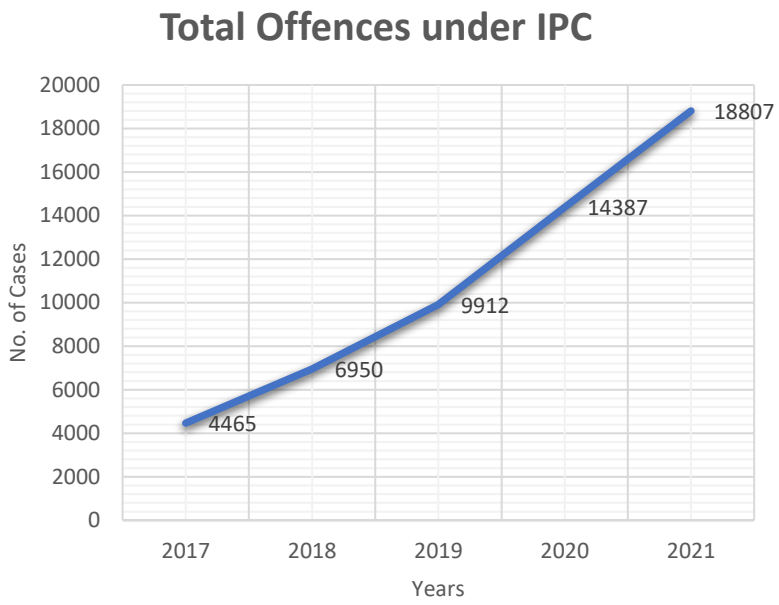
We can draw the following inference from the above graph: -

- **Major Contribution:** Out of the 12 sub-categories, only 7 of them—"Cyber Stalking/Bullying of Women/Children," "Fraud," "Cheating," "Forgery," "Cyber Blackmailing/Threatening," "Fake News on Social Media," and "Other Offenses"—contribute to more than 96.3% of crimes in this category over the course of five years (2017–2021) (**96.34%** in 2017, **96.38%** in 2018, **96.90%** in 2019, **96.87%** in 2020 & **97.76%** in 2021).
- **Abetment of Suicide (Online):** Over the years, there have been more examples of online abetment of suicide documented, highlighting a troubling trend in online harassment and its possible effects on people.
- **Cyberstalking/Bullying of Women/Children:** From 2017 to 2021, there were gradually more instances of cyberstalking and bullying of women and children, underscoring the need of addressing online safety and defending vulnerable populations.
- **Data Theft:** According to recorded incidents, there has been a modest decline in data theft in 2021 compared to prior years.
- **Fraud:** From 2017 to 2021, there were more recorded occurrences of fraud than ever before, which suggests that internet fraud is becoming more common.
- **Cheating:** From 2017 to 2021, there was a significant increase in the number of occurrences of cheating recorded under the IT Act.
- **Forgery:** Over the years, there has been a noticeable increase in the reported occurrences of forgery, showing a growing worry about online forgery and its possible repercussions.
- **Defamation/Morphing:** From 2017 to 2021, there was a gradual increase in the reported incidents of defamation and morphing, underscoring the influence of online content manipulation and reputational harm.
- **false Profile:** Over the years, there have been more instances of false profiles being created, which suggests that impersonation and other fraudulent actions are becoming more prevalent.
- **Counterfeiting:** Despite minor swings throughout the years, the recorded instances of counterfeiting have remained quite low when compared to other types of criminal activity.
- **Cyber-Blackmailing and -Threatening**: Over the years, there have been an increasing number of recorded occurrences of cyber-blackmailing and -threats, which highlights the need to combat online harassment and ensure people's safety online.
- **Fake News on social media:** From 2017 to 2021, there was an upward trend in the number of instances of fake news being circulated on social media, underscoring the importance of false information and its possible societal repercussions.
- **Other Offences**: The reported instances falling under other offences have steadily increased over the years, showing a wider spectrum of cybercrimes being reported.

Statistical Analysis

This phase will cover all of the total no. of offences under the IPC that occurred between 2017 and 2021.

Graphical representation of total offences under IPC is given below: -



We can draw the following inference from the above graph: -

- **Increasing trend:** The number of offences covered by the IPC has been on the rise throughout time. The overall number of offences has an increase tendency from 2017 to 2021.
- **Massive increase:** During this time, there has been a large increase in the overall number of offences. Between 2017 and 2021, the number of offences more than quadrupled, demonstrating a significant rise in the number of crimes covered by the IPC.
- **Accelerated growth:** It seems as though the growth rate is speeding up with time. The rise in offences is not linear; rather, it has been progressively steeper in recent years. This shows that criminal activity is becoming more prevalent and that more effective law enforcement measures are required.
- **Societal aspects**: A range of societal variables, including shifting economic conditions, social discontent, developing criminal tactics, and other underlying causes, may be to blame for the increase in offences.
- **Increased reporting and awareness:** A growth in the number of crimes covered by the IPC may also be a sign that more people are aware of their rights and the need of reporting criminal activity.
- **Need for crime prevention and law enforcement:** In order to address the growing number of offences, there is a need for efficient crime prevention tactics, law enforcement measures, and a strong criminal justice system. This data emphasises the need of these measures. It emphasises the necessity of preventative actions to stop criminal activity, provide public safety, and safeguard individual rights.

Model Fitting & Goodness of Fit:

Based on the observed data set and applying different techniques to it, the polynomial model with two degrees was found to be the model that fit the data set the best.

The outcome of fitting the polynomial model with two degrees in "R" is shown below.

Output:

```
## Residuals:
##      1      2      3      4      5
##   16.0   42.4 -223.2  255.2  -90.4
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      1.560e+09  2.725e+08   5.724   0.0292 *
## poly(x, degree = 2, raw = TRUE)1 -1.549e+06  2.700e+05  -5.737   0.0291 *
## poly(x, degree = 2, raw = TRUE)2  3.845e+02  6.686e+01   5.751   0.0289 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 250.2 on 2 degrees of freedom
## Multiple R-squared:  0.9991, Adjusted R-squared:  0.9981
## F-statistic:  1059 on 2 and 2 DF,  p-value: 0.0009435
```

Interpretation: - The "R" output shown above provides us with a variety of helpful information, including the following:

- **Residuals:** These are the differences between the actual values observed and those predicted by the regression model. There are five residuals in this instance: 16.0, 42.4, -223.2, 255.2, and -90.4. The residuals give an indicator of how effectively the model represents the variability in the data.
- **Coefficients:** The coefficients are estimates of the variables in a regression. Two polynomial terms are present in the model in this instance: a linear term (poly(x, degree = 2, raw = TRUE)1) and a quadratic term (poly(x, degree = 2, raw = TRUE)2). The quadratic term's coefficient estimates are 3.845e+02 and the linear term's are -1.549e+06, respectively. There is also an intercept term with an estimated value of 1.560e+09.
- **Standard Error:** The standard error monitors the variation in the calculated coefficients. The standard errors are 2.725e+08, 2.700e+05, and 6.686e+01 for the intercept, linear term, and quadratic term, respectively. Greater precision in the coefficient estimations is shown by smaller standard errors.
- **t-value and p-value:** both of these values evaluate the coefficients' statistical significance. By dividing the coefficient estimate by its standard error, the t-value is determined. The fact that all of the coefficients in this instance have t-values that are higher than the critical value (in absolute value) indicates that they are all statistically significant. The corresponding p-values represent the likelihood that such extreme t-values would be seen if the null hypothesis—that there is no association between the variables—were true. All coefficients' p-values are less than 0.05, which is a strong indicator of a meaningful association.
- **Residual standard error:** Standard deviation of the residuals is estimated to be 250.2. It indicates how far apart on average the measured data and regression curve are. A lower residual standard error shows that the model fits the data more well.
- **R-squared:** The multiple R-squared (0.9991) is a measure of the proportion of the response variable's variation that the model's predictor variable or variables is responsible for. The quadratic regression model in this instance accounts for around 99.91% of the variation in the response variable. The adjusted R-squared (0.9981) provides a more reliable measure of model fit by adjusting the R-squared value for the number of predictors and the sample size.
- **F-statistic and p-value:** The F-statistic evaluates the regression model's overall significance. It determines if there is a significant connection between the predictor variable(s) and the response variable overall. The F-statistic in this instance is 1059, and the p-value is 0.0009435, which provides compelling evidence that the model as a whole is significant.

Overall, the results of the study indicate that there is a significant quadratic connection between the predictor variable (x) and the response variable. High statistical significance and a high percentage of variation explained suggest that the model serves as an excellent fit to the data.

NOTE: - Less observations are there to fit this model (only 5 from 2017-2021). At least 10 observations are required for a good fit.

Prediction:

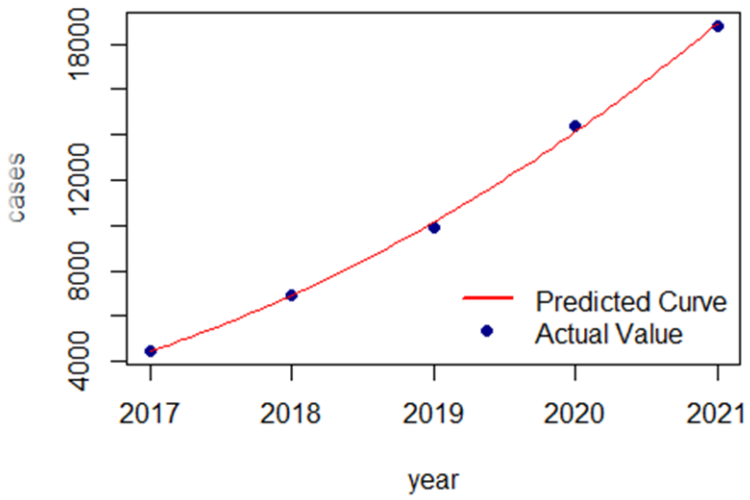
The total number of IPC offences predicted by the model from 2017 to 2021 is displayed below in tabular form using "R," along with a comparison of how well it matched the authentic information we had access to.

We additionally generated a curve of the fitted model and displayed the actual values next to it on the same curve to examine the dispersion of real values to the predicted value.

We have also estimated the potential number of cybercrime incidents for the year 2022 using the fitted model.

Output:

```
##      year actual predicted
## [1,] 2017   4465    4449.0
## [2,] 2018   6950    6907.6
## [3,] 2019   9912   10135.2
## [4,] 2020  14387   14131.8
## [5,] 2021  18807   18897.4
## [6,] 2022     NA   24432.0
```



Inference: In general, the predicted values show a respectable level of accuracy, roughly matching the actual values. Despite a few minor exceptions where the forecasts slightly over- or under-estimated the actual values, overall, the projections accurately reflect the variable's increasing trend.

The predicted figure from the model for the year 2022 is 24432, indicating that almost 24432 instances could occur in that year if violations under the IPC continued on their current track.

Note: - Although the model fits pretty well, the projected value might not be as accurate because there are less observations available for model fitting.

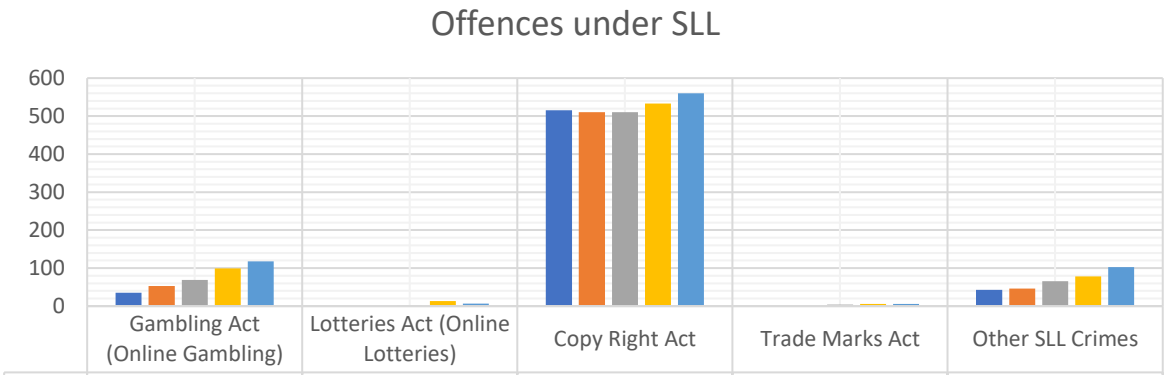
5.3.4 Analysis for Offences Under SLL

In this section, we shall analyse the data that is provided to us under the crime head-wise of the offences under the SLL category statistically and quantitatively. We'll make an effort to gather some intriguing data, spot some trends, develop a model, and predict likely outcomes.

Data Visualisation

In this section we will plot graph of offences under the SLL subcategory wise yearly and try to draw some conclusion from it.

A clustered column chart comparing the sub-categories of the offences under the SLL over a 5-year period (2017–2021) is shown below.



	Gambling Act (Online Gambling)	Lotteries Act (Online Lotteries)	Copy Right Act	Trade Marks Act	Other SLL Crimes
2017	35	1	515	1	43
2018	53	1	510	1	46
2019	69	2	510	4	66
2020	99	13	533	6	78
2021	118	7	560	6	103

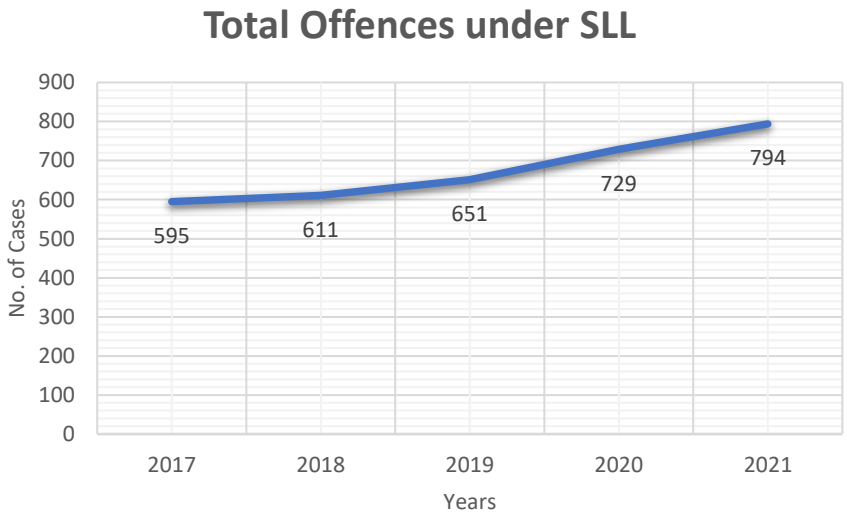
We can draw the following inference from the above graph: -

- **Major Contribution:** The three subcategories "Gambling Act (Online Gambling)," "Copy Right Act," and "Other SLL Crimes" account for more than 97.4% of the total offences committed in this category during the course of five years (2017-2021). (99.56% in 2017, 99.57% in 2018, 99.57% in 2019, 97.39% in 2020, and 98.36% in 2021)
- **Gambling Act (Online Gambling):** Over the years, there has been a progressive rise in the recorded incidents of offences under the Gambling Act relating to online gambling, showing the increasing popularity of these activities.
- **Lotteries Act (Online Lotteries):** Over the years, there have been notable swings in the reported incidents of offences under the Lotteries Act, notably relating to online lotteries, with a minor decline in 2021 when compared to earlier years.
- **Copyright Act:** Over the years, there have been only very little variations in the recorded incidents of violations of the Copyright Act.
- **Trademarks Act:** The documented instances of violations of the Trademarks Act have been modest and mostly steady throughout the years, demonstrating continuous enforcement of trademark laws.
- **Other SLL Crimes:** Over the years, there has been a steady rise in the recorded cases of other Special and Local Laws crimes, indicating that a wider range of offences are being reported and dealt with.

Statistical Analysis

This phase will cover all of the total no. of offences under the SLL that occurred between 2017 and 2021.

Graphical representation of total offences under SLL is given below: -



We can draw the following inference from the above graph: -

- **Steady Increase:** Over the years, there has been a consistent rising trend in the overall number of offences. The number of offences reported has consistently increased between 2017 and 2021.
- **Accelerating Rate:** It seems as though the rate of growth is picking up speed. The number of offences reported increased by just a minor amount from 2017 to 2018 (16 offences), but higher increases were seen in the following years (40 in 2019, 78 in 2020, and 65 in 2021).
- **Potential Causes:** A number of elements may be involved in the rise in offences. Laws may have changed, reporting and enforcement have risen, the population has grown, socioeconomic circumstances have changed, or criminal behaviour patterns have changed.
- **Need for Additional Analysis:** Although the data shows an increasing trend, more research is needed to understand the underlying causes and gauge the severity of these offences. Investigations on the exact offence categories, geographic distribution, demographics of offenders, and any potential relationships with societal norms or other external variables would be useful.

Model Fitting & Goodness of Fit:

The polynomial model with two degrees was discovered to be the model that fit the data set the best based on the observed data set and the application of various strategies to it.

The results of fitting a polynomial model with two degrees of "R" are displayed below.

Output:

```
## Residuals:
##      1      2      3      4      5
##  2.771 -3.686 -5.571 11.114 -4.629
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     3.950e+07  1.082e+07   3.650   0.0675 .
## poly(x, degree = 2, raw = TRUE)1 -3.917e+04  1.072e+04  -3.655   0.0674 .
## poly(x, degree = 2, raw = TRUE)2  9.714e+00  2.654e+00   3.660   0.0672 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.931 on 2 degrees of freedom
## Multiple R-squared:  0.993,  Adjusted R-squared:  0.986
## F-statistic: 141.7 on 2 and 2 DF,  p-value: 0.007009
```

Interpretation: - We can infer the following results from the output that "R" provided above.

- **Residuals:** These are the discrepancies between the values that were seen and those that the regression model predicted. Five residuals are present in this instance: 2.771, -3.686, -5.571, 11.114, and -4.629. The residuals give an idea of how effectively the model portrayed the data's variability.
- **Coefficients:** The computed regression parameters are represented by the coefficients. Two polynomial terms are present in the model, one of which is linear (poly(x, degree = 2, raw = TRUE)1) and the other quadratic (poly(x, degree = 2, raw = TRUE)2). The quadratic and linear terms' coefficient estimates are -3.917e+04 and 9.714, respectively. A further factor with an estimated value of 3.950e+07 is the intercept term.
- **Standard Error:** The standard error quantifies the variability of the calculated coefficients. The standard errors for the intercept, linear term, and quadratic term are 1.082e+07, 1.072e+04, and 2.654, respectively. The coefficient estimations are more precise when the standard errors are less.
- **t-value and p-value:** The coefficients' statistical significance is evaluated using the t-value and p-value. The t-value is obtained by subtracting the standard error from the coefficient estimate. This indicates that none of the coefficients are statistically significant since all of their t-values are below the crucial value (in absolute value). The linked p-values are all larger than 0.05, showing marginal evidence that there is no association, which is the null hypothesis.
- **Residual standard error:** This figure, 9.931, is an estimate of the residuals' standard deviation. It indicates how far apart on average the measured data and regression curve are. A lower residual standard error shows that the model fits the data more well.
- **R-squared:** The multiple R-squared (0.993) is the proportion of the response variable's variation that the model's predictor variable or variables can account for. The quadratic regression model in this particular case accounts for about 99.3% of the variation in the response variable. A more reliable indicator of model fit is provided by the adjusted R-squared (0.986), which corrects the R-squared value for the number of predictors and the sample size.
- **F-statistic and p-value:** The F-statistic assesses the regression model's overall significance. It determines if there is a substantial relationship between the predictor variable(s) and the response variable overall. The F-statistic in this instance is 141.7, and the p-value is 0.007009, which indicates only limited evidence that the model is generally significant.

Overall, the study reveals that the predictor variable (x) and the response variable have a quadratic connection. It is crucial to remember that the coefficients do not meet the criteria for statistical significance at standard levels ($p > 0.05$). As a result, care should be taken when interpreting the findings, and more research may be required.

NOTE: - Less observations are there to fit this model (only 5 from 2017-2021). At least 10 observations are required for a good fit.

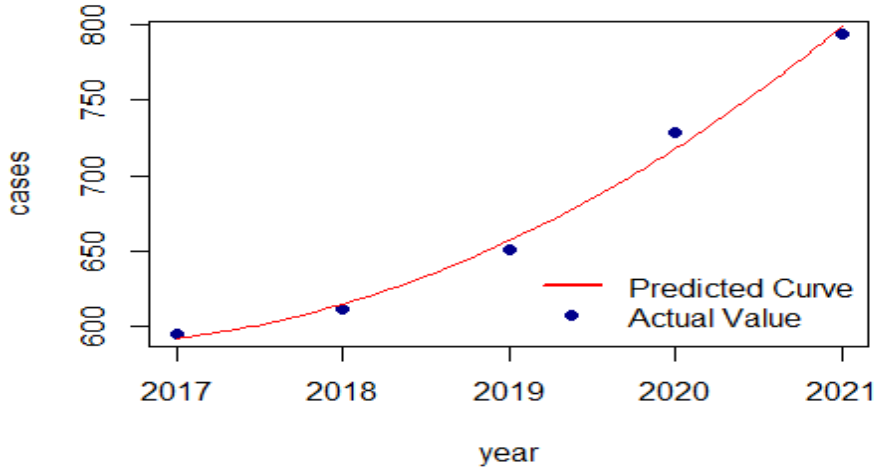
Prediction:

Below, using "R," is a tabular representation of the total number of offences under special and local law (SLL) projected by the model from 2017 to 2021, along with a comparison of how well it matched the real data we were granted access to.

In order to investigate the dispersion of real values from the predicted value, we also constructed a curve of the fitted model and plotted the actual values onto it on the same curve.

Using the fitted model, we also forecast the probable number of cybercrime instances for the year 2022.

```
##      year actual predicted
## [1,] 2017    595  592.2286
## [2,] 2018    611  614.6857
## [3,] 2019    651  656.5714
## [4,] 2020    729  717.8857
## [5,] 2021    794  798.6286
## [6,] 2022     NA  898.8000
```



Inference: The anticipated values match the actual numbers admirably well and with notable accuracy. Although these inconsistencies are modest, there are some deviations where the forecasts diverge somewhat, either underestimating or overestimating the actual values. Overall, the forecasts well represent the general trend and patterns that the variable has shown over time.

If violations of the SLL act remained on their current course, the projected figure from the model for 2022 is almost 900, meaning that around 900 occurrences might occur in that year.

Note: - The projected value may not be as accurate even though the model fits pretty well because there are fewer observations available for model fitting.

5.3.5 Predicted Offences for 2022: Percentage Contribution and Comparison

Following the model's prediction of the predicted number of cases for the I.T. Act, IPC, and SLL offences in 2022, which comes out to be roughly 50125, 24432, and 900, respectively. If we calculate the percentage contribution of each, it comes out to 66.42% for violations under the Information Technology (I.T.) Act, 32.37% for offences under the IPC, and 1.19 for offences under the SLL, which is broadly in line with the conclusion we reached from the pie charts (Section 5.3.1). Although there are a few variations since there are fewer observations available to fit the model, the numbers are still fairly close to those values.

5.4 Gist of The Analysis

In this section, we give a brief overview of the analysis that was done, emphasising the key discoveries, revelations, and conclusions drawn from the thorough investigation. This summary attempts to quickly summarise the analysis and highlight the main conclusions for simple understanding.

Emerging Trends in Incidents: Several significant tendencies are shown by an investigation of cybercrime incidents among states and union territories. Overall, there has been a sizable rise in cybercrime instances over time, showing the crimes' rising frequency. Compared to other states, Uttar Pradesh, Maharashtra, and Karnataka routinely report higher rates of cybercrime. The majority of states report an annual increase in the number of cybercrime incidents, which suggests both an increase in cybercrime incidents and maybe more awareness and reporting. Chandigarh and Delhi routinely record the largest number of cybercrime cases in the union territories, which overall show an increasing trend. Lakshadweep, which has a low incidence, and Puducherry, which has reasonably consistent levels, are two union territories that show stable or fluctuating incidences. Particularly notable is Delhi's high rate of cybercrime. When comparing the number of incidents in other union territories, it's crucial to take the demographic and technological environment into account as well. In various union areas, variations in cybercrime instances between years are seen, reflecting the dynamic character of these crimes. The overall implications of these findings highlight the rising threat of cybercrime and the need for improved cybersecurity measures, awareness campaigns, and efficient law enforcement activities throughout states and union territories.

Completed Trials: - Several important conclusions are drawn from the examination of successfully concluded cybercrime trials across states and union territories. The total number of successfully completed trials varies significantly between states. High numbers of successfully concluded trials were routinely found in states like Uttar Pradesh, Maharashtra, Tamil Nadu, Telangana, and Karnataka, showing a steady flow of judicial processes. However, certain states showed patterns of change throughout time. The number of successfully completed trials varied in Andhra Pradesh, Gujarat, and Himachal Pradesh, rising or falling in various years. On the other side, throughout the time period, very few or no trials were undertaken in states like Arunachal Pradesh, Manipur, Meghalaya, Nagaland, Sikkim, and Tripura.

Despite variances in certain years, the overall trend indicates an increase in the number of completed trials. According to the geographic study, states in the north had a greater proportion of successfully completed trials than those in the northeast. The number of successfully concluded trials in the union territory fluctuated over time in the A&N Islands, Chandigarh, Delhi, Lakshadweep, and Puducherry. While the number of trials in Chandigarh and Delhi fluctuated, the number of trials in the A&N Islands was very low. Puducherry had only done a single trial, whereas Lakshadweep had no completed trials.

From 2017 to 2019, there was a steady rise in the total number of finished trials in the union territory. However, there was a significant drop in 2020, which was followed by a little rise in 2021. The number of successfully ended trials in the union territories fluctuated from year to year, with an overall trend of ups and downs.

The findings emphasise the necessity for consistent and effective legal procedures to properly manage cybercrime matters by highlighting the disparities in trial activity among states and union territories.

Conviction Rate: - In respect to cybercrime cases, it was reveals that different states and union territories have different conviction rates. While some states continually have high conviction rates, others have peaks and valleys or consistently have low rates. The findings point to inequities in the criminal justice system's effectiveness and emphasise the difficulties in obtaining convictions for cybercrimes. For the bulk of the years considered, the national conviction rate remained below 50% overall.

Pending cases: - A number of important behaviours are shown by the study of cybercrime cases that are now for trial across states and union territories. The overall number of ongoing cybercrime cases increased significantly between 2017 and 2021, pointing to an increasing backlog in the legal system. Maharashtra typically has the most open cases, although other states with a significant number of open cases include Uttar Pradesh, Gujarat, Madhya Pradesh, and Karnataka. States display different tendencies; some show a steady increase in the number of pending cases, while others see variations over time. Arunachal Pradesh, Manipur, Mizoram, Nagaland, Sikkim, and Tripura are a few states with relatively few open cases; however, this does not necessarily mean that they have a high case disposition rate. While Chandigarh has maintained a reasonably constant level, the number of cybercrime cases awaiting trial has significantly climbed in A&N Islands and Puducherry. On the other hand, pending cases in Delhi have significantly increased, indicating a large backlog. Until a case was registered in 2021, there were no open cases in Lakshadweep. Overall, the number of active cybercrime cases in the union territory is on the rise, placing a significant pressure on the justice system to handle these cases. To promote effective justice delivery, the findings emphasise the necessity for excellent case management and prompt settlement of cybercrime cases.

5.5 Remedies

Cybercrime poses a significant threat in today's digital age, and it requires effective remedies to ensure justice and deterrence. In the above section we have identified the problem using the statistical analysis.

Some Suggestions for Strengthening Cybercrime Case Trials in Court:

- **Specialized Cybercrime Courts:** Establishing specialised cybercrime courts—or divisions within existing courts—can guarantee that jurists and judges have the requisite training and experience to handle matters involving cybercrime.
- **Training and capacity building:** Providing thorough training programmes on cybercrime laws, digital evidence collection, and the most recent technological advancements for judges, prosecutors, defence

attorneys, and court personnel can improve their comprehension and effectiveness in handling cybercrime cases.

- **Fast-Track Protocols:** Implementing accelerated protocols and schedules expressly for cybercrime cases helps hasten the trial process and lessen the backlog of ongoing cases.
- **Technical Expertise and Forensic Support:** Ensuring access to knowledgeable technical experts and forensic laboratories that can examine digital evidence, track cybercriminal activities, and offer expert testimony in court can strengthen the prosecution's case and improve the court's comprehension of complicated cyber issues.
- **Cybersecurity of Court Systems:** To preserve the integrity and secrecy of the processes, it is imperative that court systems have strong cybersecurity measures to protect digital evidence, case files, and the personal information of parties engaged in cybercrime cases.
- **International Coordination:** Improving co-operation with international law enforcement agencies, cybercrime squads, and judicial bodies can help with the investigation of transnational cybercrimes and the execution of successful trials by facilitating the exchange of information, mutual legal assistance, and extradition processes.
- **Standardized Procedures:** Creating standardised processes and rules can help ensure consistency and clarity in court proceedings when it comes to issues like the admission of digital evidence, chain of custody, and authentication in cybercrime cases.
- **Collaboration with Digital Service Providers:** Creating mechanisms for the judicial system's cooperation with digital service providers, such as social media platforms and technology firms, can hasten investigations and improve the trial process by securing evidence, preserving data, and obtaining pertinent information.
- **Public Education:** Raising public awareness of the value of cybercrime trials, the court's function, and the public's need to support and cooperate with the legal system can help build trust and promote active involvement from victims and witnesses in court processes.
- **Legislative Reforms:** The court trial process may be improved by routinely evaluating and amending cybercrime legislation to accommodate new threats, speed legal processes, and ensure appropriate and efficient consequences for cybercriminals.

Courts can successfully handle the particular difficulties presented by cybercrime cases by putting these remedies into place, ensuring fair and effective trials, and helping to discourage and lessen cybercrime generally.

5.6 Conclusion

The study of cybercrime occurrences throughout states and union territories has revealed a significant increase in cases, underscoring the urgent need for better cybersecurity measures, awareness programmes, and effective law enforcement efforts across the nation. The examination of concluded trials reveals differences in trial activity between jurisdictions, highlighting the need for uniform and efficient legal procedures to handle cybercrime-related issues. The disparate conviction rates and growing backlog of open cases reflect the challenges that the criminal justice system has in successfully combating cybercrimes.

There are numerous solutions that may be used to overcome these problems. This includes the creation of specialised courts for cybercrime, training and capacity building for judges and court staff, fast-track protocols for cybercrime cases, access to technical assistance and forensic support, enhancing the cybersecurity of court systems, international coordination, standardised procedures, collaboration with digital service providers, public awareness campaigns, and legislative reforms. By putting these solutions into practise, cybercrime cases in the northeastern states and across the nation can have a better court experience, resulting in fair and efficient trials, a decrease in the backlog of open cases, and a reduction in cybercrime activity.

It is critical to give priority to building the infrastructure, knowledge base, and regulatory framework needed to successfully combat cybercrimes. By doing this, we may promote a more secure online environment, defend people and organisations against online threats, and respect the ideals of justice in the face of developing technical difficulties.