(Read readme.md file first)

Udit Aggarwal

MBA(BA)

IIFT Delhi

<div align="center">

Report

Analytical Dashboard on State wise cases for Child Pornography

</div>

Introduction:

Dashboard provides a detailed overview of the State/UT-wise occurrences of Cyber Crimes Against Children from 2019 to 2021, specifically focusing on the distribution of cases related to Cyber Pornography, Hosting or Publishing Obscene Sexual Materials Depicting Children. The dashboard offers a nuanced perspective on the prevalence of various criminal activities, categorized into (CR) Child Rape, (CCS) Child Sexual Assault, (CON) Child Online Grooming, (PAR) Possession and Distribution of Child Abuse Material, (PCS) Child Sexual Exploitation, and (PCV) Child Victimization. Abbreviations used are (CR) Child Rape, (CCS) Child Sexual Assault, (CON) Child Online Grooming, (PAR) Possession and Distribution of Child Abuse Material, (PCS) Child Sexual Exploitation, and (PCV) Child Victimization

Summary of Dashboard: First tab i.e. Data tab covers summary of data. Second Tab i.e. Visualization tab covers interactive graphs and third tab i.e. Analysis tab covers ANOVA tests to check if cases have risen over the years.

<div align="center">

Detailed Analysis

</div>

Tab 1: Data

Part 1: Introduction

This part provides a brief introduction of dashboard and abbreviations used as provided earlier.

Part 2: Table

This part provides the data that is used in this in this dashboard. It is downloaded from data.gov.in, which is open-source website by government of India

Part 3: Structure

This part of the data provides detailed overview of different variables of the data. Variables are State/UT (Names of states) and number of cases registered under different heads during 2019, 2020 and 2021.

Part 4: Summary

Objective: To identify different statistical parameters for cases registered under different heads over years 2019, 2020 and 2021.

Analysis: Here summary function of R is used on all variables of data frame.

Result:

- Child Rape cases have mean 2.94 in 2019, 20.50 in 2020 and 26.92 in 2021. Also, there are states with 0 cases also and maximum case for a state is 27 in 2019, 161 in 2020 and 160 in 2021.
- Child Sexual Assault cases have mean 1.65 in 2019, 7.55 in 2020 and 12.53 in 2021. Also, there are states with 0 cases also and maximum case for a state is 23 in 2019, 91 in 2020 and 71 in 2021.
- Child Online Grooming cases have mean 0.05 in 2019, 0 in 2020 and 0.19 in 2021. Also, there are states with 0 cases also and maximum case for a state is 1 in 2019 and 3 in 2021.
- Possession and Distribution of Child Abuse Material cases have mean 2.97 in 2019, 10.30 in 2020 and 19.78 in 2021. Also, there are states with 0 cases also and maximum case for a state is 37 in 2019, 103 in 2020 and 130 in 2021.
- Child Sexual Exploitation cases have mean 2.45 in 2019, 10.6 in 2020 and 14.44 in 2021. Also, there are states with 0 cases also and maximum case for a state is 37 in 2019, 147 in 2020 and 79 in 2021.
- Child Victimization cases have mean 0.11 in 2019, 0 in 2020 and 0.19 in 2021. Also, there are states with 0 cases also and maximum case for a state is 3 in 2019 and 3 in 2021.

Implication: It shows cases per state for Child Rape, Child Sexual Assault, Possession and Distribution of Child Abuse Material and Child Sexual Exploitation have taken a significant jump from 2019 to 2021. However, cases per state for Child Online Grooming and Child Victimization have reduced or remained constant over years 2019 to 2021.

Tab 2: Visualization

Part 1: Crime Trends by States

Objective: To show state wise crime relating to children.

Analysis: In this part of dashboard, I have used an interactive histogram which has states on X axis and different cases in different years on y axis.

Result:

- Child Rape cases are highest in Kerala in 2019, Uttar Pradesh in 2020 and Karnataka in 2021.
- Child Sexual Assault cases are highest in Uttar Pradesh in 2019, Uttar Pradesh in 2020 and Kerala in 2021.
- Child Online Grooming cases are highest in Uttar Pradesh in 2019 and Karnataka in 2021.

- Possession and Distribution of Child Abuse Material cases are highest in Uttar Pradesh in 2019, Kerala in 2020 and Delhi in 2021.
- Child Sexual Exploitation cases are highest in Uttar Pradesh in 2019, Uttar Pradesh in 2020 and Kerala in 2021.
- Child Victimization cases are highest in Rajasthan in 2019 and Delhi in 2021.

Implication: This section portrays that similar states have emerged as states with the greatest number of cases in different years which shows these children related crimes are prominent in some states and government should take focused measures for these states.

Part 2: Distribution

Objective: To show trend among number of crimes per state.

Analysis: In this part of dashboard, I have used an interactive histogram and box plot which has name of crime on X axis and number of cases in different states on y axis.

Result:

- Most states have Child Rape cases between 0 - 5 in 2019, 0-20 in 2020 and 0-50 in 2021.

- Most states have Child Sexual Assault cases between 0 - 5 in 2019, 0-10 in 2020 and 0-10 in 2021.
- Most states have Child Online Grooming cases between 0 – 0.5 in 2019, 0-0.5 in 2020 and 0-1 in 2021.
- Most states have Possession and Distribution of Child Abuse Material cases between 0 - 5 in 2019, 0-20 in 2020 and 0-20 in 2021.
- Most states have Child Sexual Exploitation cases between 0 - 5 in 2019, 0-20 in 2020 and 0-20 in 2021.
- Most states have Child Victimization cases between 0 - 1 in 2019, 0-0.5 in 2020 and 0-1 in 2021.

Implication: This section portrays that most states have cases near to 0, which implies some states have more cases which is actually increasing average cases per state.

Tab 3: Analysis

Part 1: Introduction

This part of Dashboard statistically compares different crimes per state over years using test of Anova.

Part 2: Test for Child Rape Cases

Objective: To check the impact of different states and years on child rape cases.

Analysis: I have used Analysis of Variance (ANOVA) using the formula Cases ~ Year + Error (State / Year), which is suitable for examining the influence of the variable 'Year' on the

response variable 'Cases', while considering potential differences among States and across Years within each State. Here, null hypothesis will be no significant relation between variables and alternate hypothesis shows significant relation between variables.

Result:

- Error: State

  Df (Degrees of freedom): 1 for 'Year' and 34 for 'Residuals.'

  Sum Sq (Sum of Squares): 583 for 'Year' and 76116 for 'Residuals.'

  Mean Sq (Mean Squares): 583.2 for 'Year' and 2238.7 for 'Residuals.'

  F value: 0.261

  Pr(>F) (p-value): 0.613

  This part of the output tests the overall effect of 'Year' on 'Cases' across all States. The F value is a ratio of variance between the groups (due to 'Year') to variance within the groups (Residuals). The p-value (Pr(>F)) indicates whether the observed F value is statistically significant. In this case, since the p-value is 0.613 (greater than the commonly used threshold of 0.05), we fail to reject the null hypothesis, suggesting no significant effect of 'Year' on 'Cases' across States.

- Error: State:Year

  Df: 2 for 'Year' and 69 for 'Residuals.'

  Sum Sq: 11470 for 'Year' and 53096 for 'Residuals.'

  Mean Sq: 5735 for 'Year' and 770 for 'Residuals.'

  F value: 7.453

  Pr(>F): 0.00117 (**)

  This part of the output tests the interaction effect between 'State' and 'Year.' The significant p-value (0.00117) indicates that there is evidence to reject the null hypothesis, suggesting that the effect of 'Year' on 'Cases' varies significantly across different States.

Implication: In conclusion, while there isn't a significant overall effect of 'Year' on child rape cases across all States, there is a significant interaction effect between 'State' and 'Year,' implying that the influence of 'Year' on child rape cases differs among different States.

Part 3: Test for Child Sexual Assault cases

Objective: To check the impact of different states and years on Child Sexual Assault cases.

Analysis: I have used Analysis of Variance (ANOVA) using the formula Cases ~ Year + Error (State / Year), which is suitable for examining the influence of the variable 'Year' on the response variable 'Cases', while considering potential differences among States and across Years within each State. Here, null hypothesis will be no significant relation between variables and alternate hypothesis shows significant relation between variables.

Result:

- Error: State

  Df (Degrees of freedom): 1 for 'Year' and 34 for 'Residuals.'

  Sum Sq (Sum of Squares): 109 for 'Year' and 17344 for 'Residuals.'

  Mean Sq (Mean Squares): 108.6 for 'Year' and 510.1 for 'Residuals.'

  F value: 0.213

  Pr(>F) (p-value): 0.647

  This part of the output tests the overall effect of 'Year' on 'Cases' across all States. The F value is a ratio of variance between the groups (due to 'Year') to variance within the groups (Residuals). The p-value (Pr(>F)) indicates whether the observed F value is statistically significant. In this case, since the p-value is 0.647 (greater than the commonly used threshold of 0.05), we fail to reject the null hypothesis, suggesting no significant effect of 'Year' on 'Cases' across States.

- Error: State:Year

  Df: 2 for 'Year' and 69 for 'Residuals.'

  Sum Sq: 2200 for 'Year' and 9174 for 'Residuals.'

  Mean Sq: 1100 for 'Year' and 133 for 'Residuals.'

  F value: 8.271

  Pr(>F): 0.000603 (***)

  This part of the output tests the interaction effect between 'State' and 'Year.' The significant p-value (0.000603) indicates that there is evidence to reject the null hypothesis, suggesting that the effect of 'Year' on 'Cases' varies significantly across different States.

Implication: In conclusion, while there isn't a significant overall effect of 'Year' on Child Sexual Assault cases across all States, there is a significant interaction effect between 'State'

and 'Year,' implying that the influence of 'Year' on Child Sexual Assault cases differs among different States.

Part 4: Test for Child Online Grooming Cases

Objective: To check the impact of different states and years on Child Online Grooming cases.

Analysis: I have used Analysis of Variance (ANOVA) using the formula Cases ~ Year + Error (State / Year), which is suitable for examining the influence of the variable 'Year' on the response variable 'Cases', while considering potential differences among States and across Years within each State. Here, null hypothesis will be no significant relation between variables and alternate hypothesis shows significant relation between variables.

Result:

- Error: State

  Df (Degrees of freedom): 1 for 'Year' and 34 for 'Residuals.'

  Sum Sq (Sum of Squares): 0.014 for 'Year' and 6.229 for 'Residuals.'

  Mean Sq (Mean Squares): 0.01442 for 'Year' and 0.18319 for 'Residuals.'

  F value: 0.079

  Pr(>F) (p-value): 0.781

  This part of the output tests the overall effect of 'Year' on 'Cases' across all States. The F value is a ratio of variance between the groups (due to 'Year') to variance within the groups (Residuals). The p-value (Pr(>F)) indicates whether the observed F value is statistically significant. In this case, since the p-value is 0.781 (greater than the commonly used threshold of 0.05), we fail to reject the null hypothesis, suggesting no significant effect of 'Year' on 'Cases' across States.

- Error: State:Year

  Df: 2 for 'Year' and 69 for 'Residuals.'

  Sum Sq: 0.723 for 'Year' and 9.277 for 'Residuals.'

  Mean Sq: 0.3617 for 'Year' and 0.1344 for 'Residuals.'

  F value: 2.69

  Pr(>F): 0.075

  This part of the output tests the interaction effect between 'State' and 'Year.' The p-value (0.075) indicates that there is no evidence to reject the null hypothesis, suggesting that the effect of 'Year' on 'Cases' doesn't varies significantly across different States.

Implication: In conclusion, there isn't a significant overall effect of 'Year' on Child Online Grooming cases across all States, also there is no significant interaction effect between 'State' and 'Year,' implying no influence of 'Year' on Child Online Grooming cases differing among different States.

Part 5: Test for Possession and Distribution of Child Abuse Material Cases

Objective: To check the impact of different states and years on Possession and Distribution of Child Abuse Material cases.

Analysis: I have used Analysis of Variance (ANOVA) using the formula Cases ~ Year + Error (State / Year), which is suitable for examining the influence of the variable 'Year' on the response variable 'Cases', while considering potential differences among States and across Years within each State. Here, null hypothesis will be no significant relation between variables and alternate hypothesis shows significant relation between variables.

Result:

- Error: State

  Df (Degrees of freedom): 1 for 'Year' and 34 for 'Residuals.'

  Sum Sq (Sum of Squares): 251 for 'Year' and 32595 for 'Residuals.'

  Mean Sq (Mean Squares): 251.2 for 'Year' and 958.7 for 'Residuals.'

  F value: 0.262

  Pr(>F) (p-value): 0.612

  This part of the output tests the overall effect of 'Year' on 'Cases' across all States. The F value is a ratio of variance between the groups (due to 'Year') to variance within the groups (Residuals). The p-value (Pr(>F)) indicates whether the observed F value is statistically significant. In this case, since the p-value is 0.612 (greater than the commonly used threshold of 0.05), we fail to reject the null hypothesis, suggesting no significant effect of 'Year' on 'Cases' across States.

- Error: State:Year

  Df: 2 for 'Year' and 69 for 'Residuals.'

  Sum Sq: 5260 for 'Year' and 23896 for 'Residuals.'

  Mean Sq: 2629.9 for 'Year' and 346.3 for 'Residuals.'

  F value: 7.594

  Pr(>F): 0.00105 (**)

  This part of the output tests the interaction effect between 'State' and 'Year.' The significant p-value (0.00105) indicates that there is evidence to reject the null

hypothesis, suggesting that the effect of 'Year' on 'Cases' varies significantly across different States.

Implication: In conclusion, while there isn't a significant overall effect of 'Year' on Possession and Distribution of Child Abuse Material cases across all States, there is a significant interaction effect between 'State' and 'Year,' implying that the influence of 'Year' on Possession and Distribution of Child Abuse Material cases differs among different States.

Part 6: Test for Child Sexual Exploitation Cases

Objective: To check the impact of different states and years on Child Sexual Exploitation cases.

Analysis: I have used Analysis of Variance (ANOVA) using the formula Cases ~ Year + Error (State / Year), which is suitable for examining the influence of the variable 'Year' on the response variable 'Cases', while considering potential differences among States and across Years within each State. Here, null hypothesis will be no significant relation between variables and alternate hypothesis shows significant relation between variables.

Result:

- Error: State

  Df (Degrees of freedom): 1 for 'Year' and 34 for 'Residuals.'

  Sum Sq (Sum of Squares): 174 for 'Year' and 31483 for 'Residuals.'

  Mean Sq (Mean Squares): 174.1 for 'Year' and 924.6 for 'Residuals.'

  F value: 0.188

  Pr(>F) (p-value): 0.667

  This part of the output tests the overall effect of 'Year' on 'Cases' across all States. The F value is a ratio of variance between the groups (due to 'Year') to variance within the groups (Residuals). The p-value (Pr(>F)) indicates whether the observed F value is statistically significant. In this case, since the p-value is 0.667 (greater than the commonly used threshold of 0.05), we fail to reject the null hypothesis, suggesting no significant effect of 'Year' on 'Cases' across States.

- Error: State:Year

  Df: 2 for 'Year' and 69 for 'Residuals.'

  Sum Sq: 2805 for 'Year' and 14953 for 'Residuals.'

  Mean Sq: 1402.6 for 'Year' and 216.7 for 'Residuals.'

  F value: 6.472

Pr(>F): 0.00265 (**)

This part of the output tests the interaction effect between 'State' and 'Year.' The significant p-value (0.00265) indicates that there is evidence to reject the null hypothesis, suggesting that the effect of 'Year' on 'Cases' varies significantly across different States.

Implication: In conclusion, while there isn't a significant overall effect of 'Year' on Child Sexual Exploitation cases across all States, there is a significant interaction effect between 'State' and 'Year,' implying that the influence of 'Year' on Child Sexual Exploitation cases differs among different States.

Part 7: Test for Child Victimization Cases

Objective: To check the impact of different states and years on Child Victimization cases.

Analysis: I have used Analysis of Variance (ANOVA) using the formula Cases ~ Year + Error (State / Year), which is suitable for examining the influence of the variable 'Year' on the response variable 'Cases', while considering potential differences among States and across Years within each State. Here, null hypothesis will be no significant relation between variables and alternate hypothesis shows significant relation between variables.

Result:

- Error: State

    Df (Degrees of freedom): 1 for 'Year' and 34 for 'Residuals.'

    Sum Sq (Sum of Squares): 0.022 for 'Year' and 11.181 for 'Residuals.'

    Mean Sq (Mean Squares): 0.0215 for 'Year' and 0.3289 for 'Residuals.'

    F value: 0.066

    Pr(>F) (p-value): 0.8

    This part of the output tests the overall effect of 'Year' on 'Cases' across all States. The F value is a ratio of variance between the groups (due to 'Year') to variance within the groups (Residuals). The p-value (Pr(>F)) indicates whether the observed F value is statistically significant. In this case, since the p-value is 0.8 (greater than the commonly used threshold of 0.05), we fail to reject the null hypothesis, suggesting no significant effect of 'Year' on 'Cases' across States.

- Error: State:Year

    Df: 2 for 'Year' and 69 for 'Residuals.'

    Sum Sq: 0.685 for 'Year' and 11.981 for 'Residuals.'

Mean Sq: 0.3427 for 'Year' and 0.1736 for 'Residuals.'

F value: 1.973

Pr(>F): 0.147 (**)

This part of the output tests the interaction effect between 'State' and 'Year.' The significant p-value (0.00147) indicates that there no is evidence to reject the null hypothesis, suggesting that the effect of 'Year' on 'Cases' does not varies significantly across different States.

Implication: In conclusion, there isn't a significant overall effect of 'Year' on Child Victimization cases across all States, also there is no significant interaction effect between 'State' and 'Year,' implying no influence of 'Year' on Child Victimization cases differing among different States.