

**Title:**

In 2018, the Number of Women raped in India per each district & broken down by age group

**Abstract:**

Women being raped is an issue that is known, but it is important to see the numbers and analyze the data to understand the number of women that are becoming victims of this. Once we understand the mass impact, and learn where more cases are occurring then the actions can be taken to reduce the amount of rapes.

This dataset provides the number of women that were raped in 2018 in India. The dataset provides each district's name and if that district is ruled by State or the Union Territories. There are a total of 29 states and 7 Union Territories. For each district, the number of cases of women raped is provided. Following that, the number of women who were raped is broken down into age categories. The age category is 0-5, 6-11, 12-15, 16-17, 18-29, 30-44, 45-59, and 60 plus.

When looking at the average number of cases and the number of victims, there are fewer cases reported than the number of victims. This shows that not all women report cases, yet given the number of victims, a total number of 262 girls (below age of 18) were raped, and an average of 681 women (above age 18) were raped. In total, the average number of victims in 2018 in India was 943. This is a high number, especially since this data is only from a single year and in one country.

**Introduction:**

The dataset discussed in this report is about the number of women raped in India in each district in the year of 2018. The number of women raped have been broken down by age group, particularly to emphasize the number of girls under 18 and women 18 and above that are raped. The purpose of this analysis is to highlight the big issue of rape by showing how many little girls, middle age and elderly women are negatively impacted. Bringing light to the numbers, will give encouragement to help prevent the rape cases and also provide support for women who have been raped.

The report will first discuss the source of the data and describe the variables of the dataset. After that, describe the methods and the steps taken to produce the results and findings. After discussing this, the results of the tests and the code ran will be shown through tables and graphs. In the conclusion section, the findings, limitations, and future recommendations on further analysis will be discussed.

**Data and Data Preparation:**

I received this dataset from Kaggle, and was uploaded two years ago by Neha Prabhavalkar. She collected this data from the Open Government Data (OGD) Platform India. The attribute "State/UT" is a categorical- nominal data type that lists the districts in India. The Category attribute, is a categorical - symmetric data type, with a range of two values which are State and Union Territory. This tells the viewer if the district is ruled by the state or union. In total, there are a total of 27 states and 9 Union Territories, which sums up the 36 districts in India which are listed under the "State/UT" column. Following this, there is a column called "Cases Reported", this consists of numeric - ratio data type, where it tells the number of cases of rape that were reported. After this the number of rapes are broken down by age group, so they all are numeric. The first age group is "Child.Victims.of.Rape..Below.18.Yrs....Below.6.Years", which consists of the girls ages 0-5 years old that were raped. Secondly,

"Child.Victims.of.Rape..Below.18.Yrs....6.Years...Above", this consists of girls ages 6-11 that were raped. Third, "Child.Victims.of.Rape.. Below.18.Yrs... 12.Years...Above.", which is of ages 12-15. Fourth,"Child.Victims.of.Rape..Below.18.Yrs...16.

Years..Above", which is of ages (16-17). Lastly, "Child.Victims.of.Rape..Below.18.Yrs...total.Girl Child.Victim", is the total number of victims that were children, so below the ages of 18 that were raped. In the same way, cases of women that were raped were broken down by age. So, 18-29, 30-44, 45-59, 60+, and then the total number of women (18+) that were raped. The last column, "Total Victims" constis of the total number of women that were raped, so ages 0-60+. Below is a table, summarizing the above information of variables, description of the data in each variable, and the data type.

Variables	Description	Data Type
Category	District is ruled by the State or Union Territory	Categorial - Symmetry
State.UT	Name of district	Categorial - Nominal
Cases Reported	Number of Cases Reported	Numeric - Ratio
Child Victims of Rape: 0-5	Number of victims ages 0-5	Numeric - Ratio
Child Victims of Rape: 6-11	Number of victims ages 6-11	Numeric - Ratio
Child Victims of Rape: 12-15	Number of victims ages 12-15	Numeric - Ratio
Child Victims of Rape: 16-18	Number of victims ages 16-18	Numeric - Ratio
Total Number of Child Victims	Total number of child vctims	Numeric - Ratio
Adult Victims of Rape: 19-29	Number of victims ages 19-29	Numeric - Ratio
Adult Victims of Rape: 30-44	Number of victims ages 30-44	Numeric - Ratio
Adult Victims of Rape: 45-59	Number of victims ages 45-59	Numeric - Ratio
Adult Victims of Rape: 60+	Number of victims ages 60+	Numeric - Ratio
Total Number of Adult Victims:	Total number of adult victims	Numeric - Ratio
Total Number of Victims:	Total number of victims	Numeric - Ratio

No preprocessing steps were done to the data prior to the analysis as all the data was filled, with no missing information.

### Methods:

I found the outliers for each age group that was raped. To find the outliers I used the standard deviation, z score, and the IQR method. For the standard deviation method, I found the standard deviation of the number of cases for that age group, then established the upper and lower bound. If anything was higher than the upper limit or below the lower limit it was considered an outlier. For the z-score, I took the column of rape cases for the age group and subtract it by the average and divide that by standard deviation of that column. Anything greater than 3 or below -3 was considered an outlier. For the IQR method, the IQR was times by 1.5 and that value subtracted from Quartile 1, if anything was less than that value it was considered an outlier. Also, 1.5 times IQR was added to the Quartile 3, and if anything was more than that value it was considered an outlier. In this way three different methods were used to identify the outliers for each group and this can be seen in Table 1 through Table 11. To show the outliers visually, a histogram and box plot were made for the total number of child victims, adult victims, and the total number of victims age. Then I subsetted the data to show districts that had more than a 1000 total number of victims, child, and adult victims. A density plot was made to visually show the distribution of the number of victims over the subset. Also, I ran descriptive statistics over the subsetted data which included the minimum, maximum, mean, median, variance, and

standard deviation. After this I converted the categorical variable, Category, which tells whether the district is ruled by the State or Union Territory to a flag variable. So the code to set State to equal 1, and Union Territory looked like this

```
"Victims_Rape$Category_1_flag[Victims_Rape$Category=="State"] <- 1" and  
"Victims_Rape$Category_1_flag[Victims_Rape$Category!="State"] <- 0". The result can be  
seen in Figure 10. Lastly, I converted the numeric variable, Total Victims, to a categorical  
variable using binning using the cut function. Then I visualize the result through a bar plot, seen  
in Figure 11.
```

### **Results:**

Table 1 through Table 11 shows the outliers for each age group, obtained from each of the three methods (standard deviation, z score, IQR). The district and the number of victims in that age group are shown in the table. For the total number of child victims, the outliers are shown in Table 5, and visually the data can be seen in Figure 1 and Figure 2. The same visualization, histogram and box plot, are used to show the data for the total number of adult victims (Figure 3 and Figure 4) and the overall total number of victims(Figure 5 and Figure 6), alongside the tables.

Table 12 shows the data of those districts whose total number of adult victims were greater than 1000. Figure 7 helps visualize this through a density plot. Based on this visualization, the number of victims are around 1,500 but also there seems to be a few districts whose cases are around 2200. Table 13, provides descriptive statistics of this data showing that the average number of victims for districts greater than a 1000 victims are 1918. Similarly, Table 14, shows the districts whose total number of child victims were greater than a 1000. The density plot in Figure 8 and the descriptive statistics provided in Table 15, support that the average number of victims is 1500 as there seems to be a spike in the graph at 1200 and 2500 (Average of 1200 and 2500 is 1850). Table 16, shows the districts whose total number of victims raped were greater than 1000. Based on the table, the density plot in Figure 9, and the descriptive statistics provided in Table 17, support that the average number of cases is 2434. Table 16, also shows that even though the data is being looked at districts whose number of victims is greater than a 1000, it is still more than the number of cases reported for all of the districts. That means more women are getting raped, and the rapists are not bearing with the consequences they deserve because of not reporting it. Additionally, in Figure 10, there is a column called "Categorial\_Num, State=1, and "Categorial\_Num, Union Territory=1". These columns represent the conversion of the categorical variable from the State/UT column to a numeric value. In the "Categorial\_Num, State=1, the State is represented by 1 and Union Territory is 0. For the "Categorial\_Num, Union Territory=1", Union Territory is represented by 1 and State by 0. Lastly, the total number of victims is a numeric data type but converted into a categorical data type with equal width binning as seen in Figure 11.

Using all these results, the districts with a high number of victims raped compared to other districts can be identified for each age group. For ages 0-5, the districts are: Uttar Pradesh, Andhra Pradesh, Chhattisgarh, Kerala, Madhya Pradesh and Rajasthan. For ages 6-11, the districts are: Uttar Pradesh, Andhra Pradesh, Chhattisgarh, Kerala, Madhya Pradesh, and Rajasthan. For ages 12-15, the districts are: Madhya Pradesh, Andhra Pradesh, Chhattisgarh, Kerala, Punjab, Rajasthan, and Uttar Pradesh. For ages 16-18, the districts are Madhya Pradesh, Andhra Pradesh, Chattisgarh, Kerala, Rajasthan, Uttar Pradesh, and Uttrakhand. For ages 19-29, the districts are Rajasthan and Uttar Pradesh. Ages 30-44, the districts are

Rajasthan, Madhya Pradesh, and Uttar Pradesh. For ages 45-59, the districts are Rajasthan, Assam, and Uttar Pradesh. Lastly, for ages 60+, the districts are Kerala, Rajasthan, Chhattisgarh, Madhya Pradesh and Maharashtra. From this, the common districts with the large number of victims for children (ages 0-18) are: Uttar Pradesh, Andhra Pradesh, Chhattisgarh, Kerala, Madhya Pradesh, and Rajasthan. For adult victims (18 & above), the common districts are Rajasthan and Uttar Pradesh.

Table 1: Outliers for the Ages 0-5, that were Raped

Test	District	Total Cases Reported
Standard	Uttar Pradesh	67
Z-Score	Uttar Pradesh	67
IQR	Andhra Pradesh	16
IQR	Chhattisgarh	41
IQR	Kerala	48
IQR	Madhya Pradesh	54
IQR	Rajasthan	17
IQR	Uttar Pradesh	67

Table 2: Outliers for the Ages 6-11, that were Raped

Test	District	Total Cases Reported
Standard	Uttar Pradesh	174
Z-Score	Uttar Pradesh	174
IQR	Andhra Pradesh	57
IQR	Chhattisgarh	80
IQR	Kerala	129
IQR	Madhya Pradesh	142
IQR	Rajasthan	48
IQR	Uttar Pradesh	174

Table 3: Outliers for the Ages 12-15, that were Raped

Test	District	Total Cases Reported
Standard	Madhya Pradesh	1143
Z-Score	Madhya Pradesh	1143
IQR	Andhra Pradesh	181
IQR	Chhattisgarh	557
IQR	Kerala	334
IQR	Madhya Pradesh	1143
IQR	Punjab	151
IQR	Rajasthan	392
IQR	Uttar Pradesh	570

Table 4: Outliers for the Ages 16-18, that were Raped

Test	District	Total Cases Reported
Standard	Madhya Pradesh	1502
Z-Score	Madhya Pradesh	1502
IQR	Andhra Pradesh	251
IQR	Chhattisgarh	541
IQR	Kerala	645
IQR	Madhya Pradesh	1502
IQR	Rajasthan	575
IQR	Uttar Pradesh	600
IQR	Uttarkand	186

Table 5: Outliers for the Total Number of Child Victims of Rape

Test	District	Total Cases Reported
Standard	Madhya Pradesh	2841
Z-Score	Madhya Pradesh	2841
IQR	Madhya Pradesh	2841

Figure 1: Histogram: Data of Total Number of Child Victims of Rape

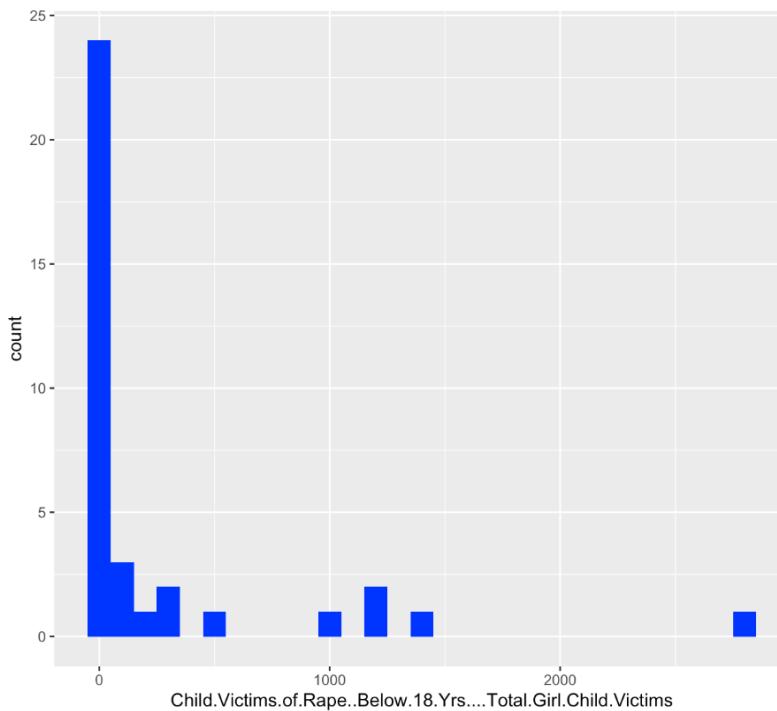


Figure 2: Boxplot: Data of Total Number of Child Victims of Rape

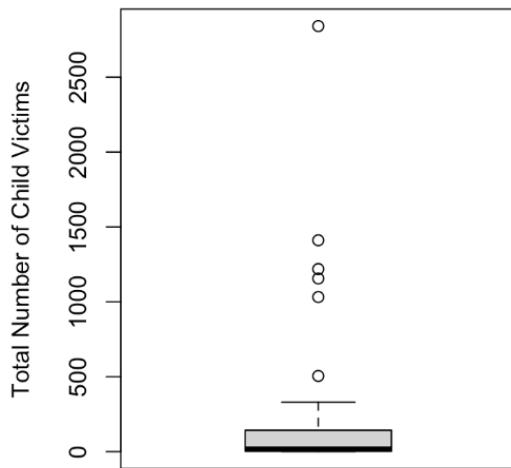


Table 6: Outliers for the Ages 19-29, that were Raped

Test	District	Total Cases Reported
Standard	NA	NA
Z-Score	NA	NA
IQR	Rajasthan	2263
IQR	Uttar Pradesh	603

Table 7: Outliers for the Ages 30-44, that were Raped

Test	District	Total Cases Reported
Standard	Rajasthan	912
Z-Score	Rajasthan	912
IQR	Madhya Pradesh	725
IQR	Rajasthan	912
IQR	Uttar Pradesh	603

Table 8: Outliers for the Ages 45-59, that were Raped

Test	District	Total Cases Reported
Standard	Rajasthan	129
Z-Score	Rajasthan	129
IQR	Assam	103
IQR	Rajasthan	129
IQR	Uttar Pradesh	70

Table 9: Outliers for the Ages 60+, that were Raped

Test	District	Total Cases Reported
Standard	Kerala	14
Z-Score	Rajasthan	129
IQR	Chhattisgarh	6
IQR	Kerala	14
IQR	Madhya Pradesh	11
IQR	Maharashtra	11

Table 10: Outliers for the Total Number of Adult Victims of Rape

Test	District	Total Cases Reported
Standard	Rajasthan	3305
Z-Score	Rajasthan	3305
IQR	Madhya Pradesh	2609
IQR	Rajasthan	3305
IQR	Uttar Pradesh	2911

Figure 3: Histogram: Data of Total Number of Adult Victims of Rape

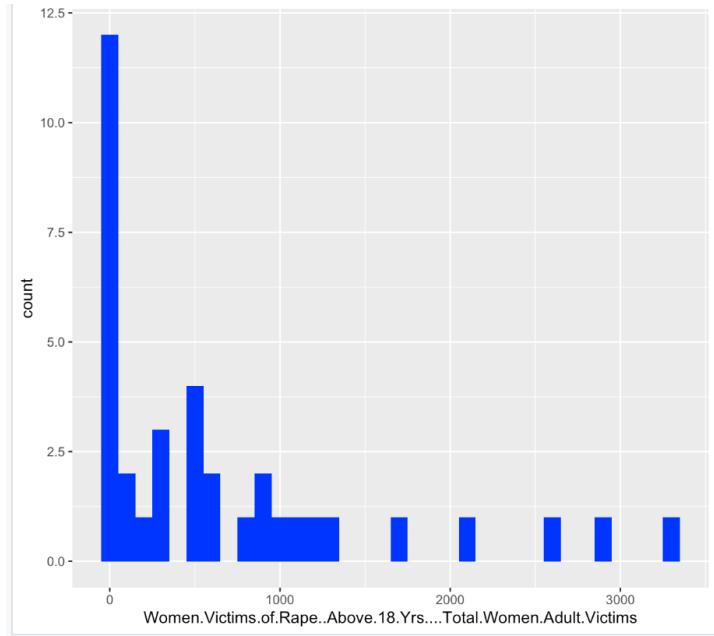


Figure 4: Boxplot: Data of Total Number of Adult Victims of Rape

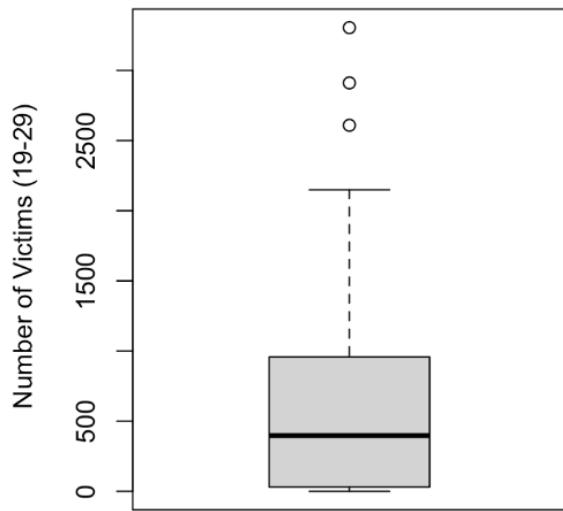


Table 11: Outliers for the Total Number of Victims of Rape

Test	District	Total Cases Reported
Standard	Madhya Pradesh	5450
Z-Score	Madhya Pradesh	5450
IQR	Rajasthan	4337
IQR	Madhya Pradesh	5450
IQR	Uttar Pradesh	4322

Figure 5: Histogram: Data of Total Number of Victims of Rape

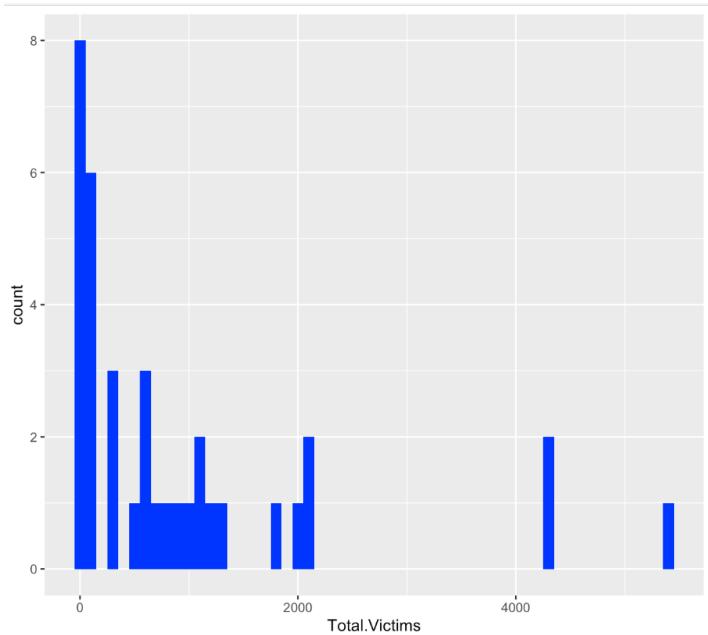


Figure 6: Boxplot: Data of Total Number of Victims of Rape

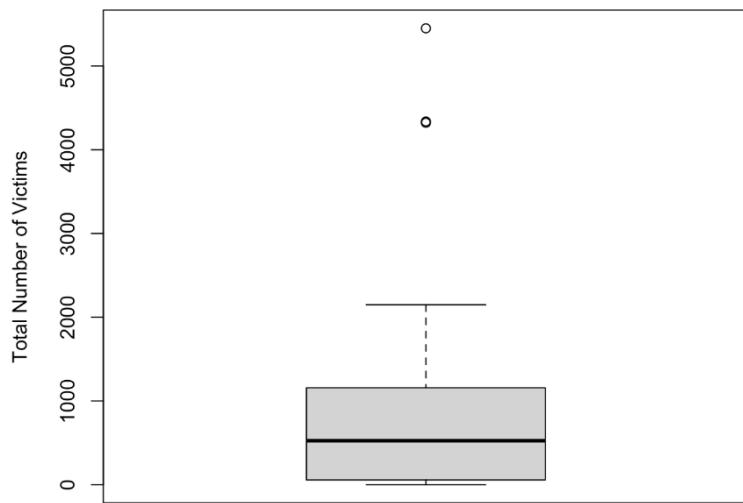


Table 12: Subsetted Data: Total Number of Adult Victims of Rape &gt;1000

State/UT	District	Total Number of Adult Women Victims
State	Assam	1678
State	Haryana	1294
State	Jharkhand	1029
State	Madhya Pradesh	2609
State	Maharashtra	2149
State	Rajasthan	3305
State	Uttar Pradesh	2911
State	West Bengal	1065
Union Territory	Delhi UT	1217

Figure 7: Density Plot: Total Number of Adult Victims of Rape &gt;1000

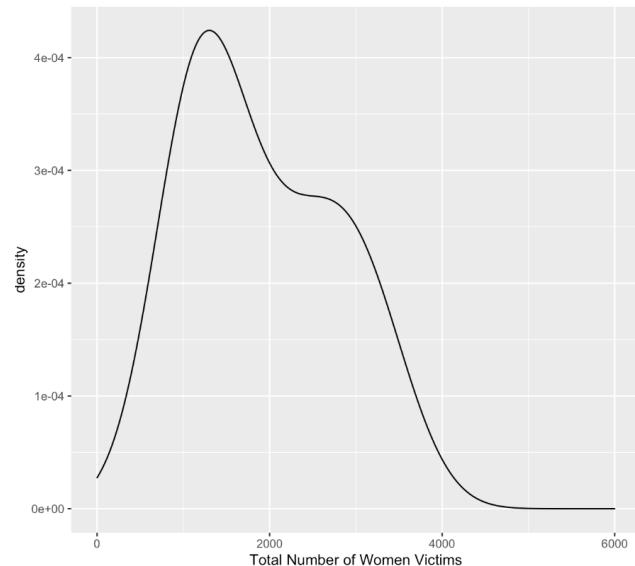


Table 13: Descriptive Statistics of the Total Number of Child Victims &gt;1000

	Number of Victims
Min	1029
Max	3305
Mean	1917.444
Median	1678
Variance	737120.5
Standard Deviation	858.5572

Table 14: Subsetted Data: Total Number of Child Victims of Rape &gt;1000

State/UT	District	Total Number of Child Victims
State	Chhattisgarh	1219
State	Kerala	1156
State	Madhya Pradesh	2841
State	Rajasthan	1032
State	Uttar Pradesh	1411

Figure 8: Density Plot: Total Number of Child Victims of Rape &gt;1000

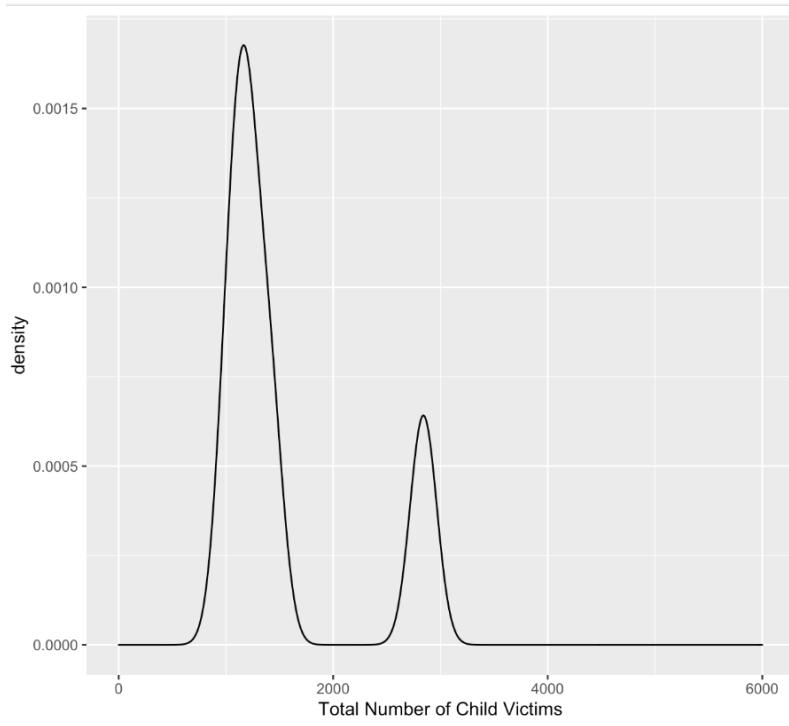


Table 15: Descriptive Statistics of the Total Number of Child Victims &gt;1000

	Number of Victims
Min	1032
Max	2841
Mean	1531.8
Median	1219
Variance	554366.7
Standard Deviation	744.5581

Table 16: Subsetted Data: Total Number of Victims of Rape &gt;1000

State/UT	District	Cases Reported	Total Number of Women Victims
State	Assam	1648	1767
State	Chhattisgarh	2091	2101
State	Haryana	1296	1296
State	Jharkhand	1090	1098
State	Kerala	1945	1972
State	Madhya Pradesh	5433	5450
State	Maharashtra	2142	2149
State	Rajasthan	4335	4337
State	Uttar Pradesh	3946	4322
State	West Bengal	1069	1069
Union Territory	Delhi UT	1215	1217

Figure 9: Density Plot: Total Number of Victims of Rape &gt;1000

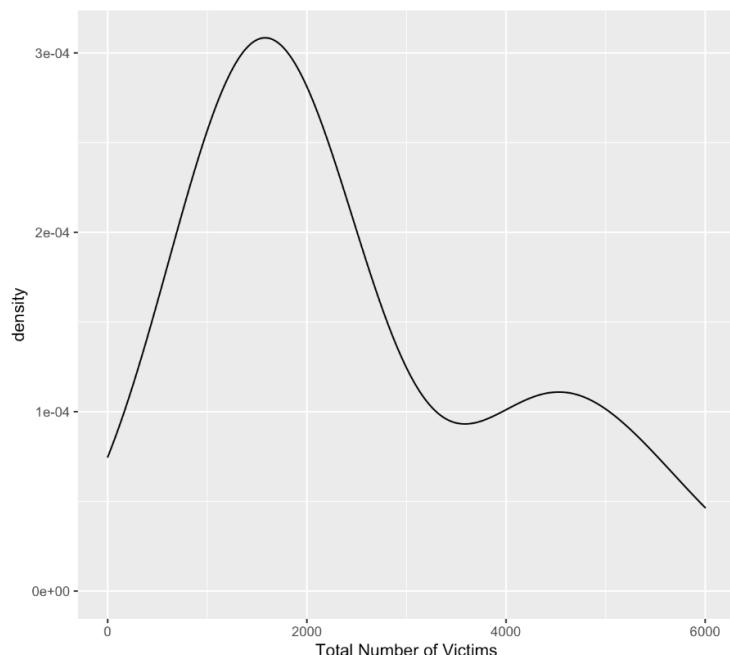


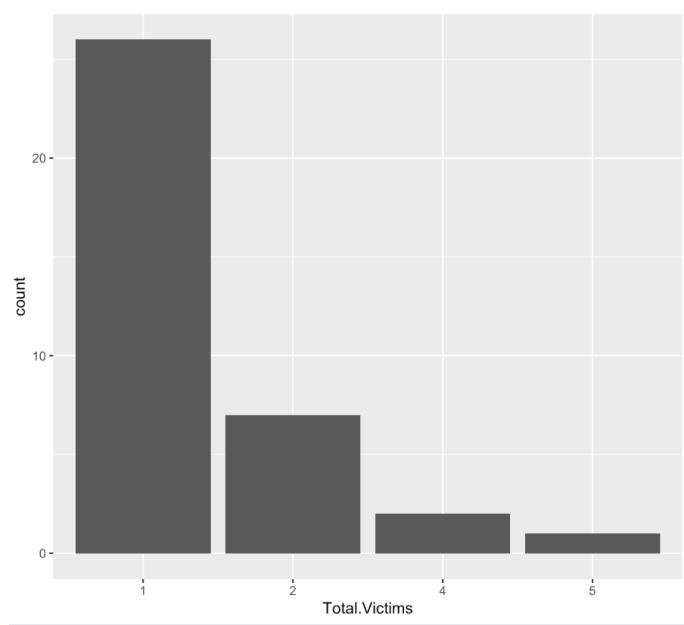
Table 17: Descriptive Statistics of the Total Number of Victims &gt;1000

	Number of Victims
Min	1069
Max	5450
Mean	2434.364
Median	1972
Variance	2355691
Standard Deviation	1534.826

Figure 10: Categorical Variable (State/Union Territory) to Numeric,

State/UnionTerritory	District	Categorial_Num, State=1	Categorial_Num, Union Territory=1
State	Arunachal Pradesh	1	0
State	Assam	1	0
State	Bihar	1	0
State	Chhattisgarh	1	0
State	Goa	1	0
State	Gujarat	1	0
State	Haryana	1	0
State	Himachal Pradesh	1	0
State	Jammu & Kashmir	1	0
State	Jharkhand	1	0
State	Karnataka	1	0
State	Kerala	1	0
State	Madhya Pradesh	1	0
State	Maharashtra	1	0
State	Manipur	1	0
State	Meghalaya	1	0
State	Mizoram	1	0
State	Nagaland	1	0
State	Odisha	1	0
State	Punjab	1	0
State	Rajasthan	1	0
State	Sikkim	1	0
State	Tamil Nadu	1	0
State	Telangana	1	0
State	Tripura	1	0
State	Uttar Pradesh	1	0
State	Uttarakhand	1	0
State	West Bengal	1	0
Union Territory	A & N Islands	0	1
Union Territory	Chandigarh	0	1
Union Territory	D&N Haveli	0	1
Union Territory	Daman & Diu	0	1
Union Territory	Delhi UT	0	1
Union Territory	Lakshadweep	0	1
Union Territory	Puducherry	0	1

Figure 11: Bar Chart: Total Number of Victim, Equal Width Bins



### Conclusion and Discussion:

Based on the results, the districts with higher numbers of victims compared to other districts for each age group are identified. As well as the common districts that seem to have a high number of victims for the ages 0-18 and 18 & above. With this information, further investigation can be done regarding why the cases are so high. Why is that children are more raped in Uttar Pradesh, Andhra Pradesh, Chattigrah, Kerala, Madhya Pradesh and Rajasthan? Is it the school, the transportation, the after school activities. Maybe more police or safety guards can be put in place from the routes kids walk to school. Similarly, why is that women are more likely to get raped in Rajasthan and Uttar Pradesh? Is it the work environment, way to work, lack of awareness, lack of self-defense? After finding the reason, actions can be taken which in turn will reduce the number of children and women being raped. Also, from the results it can be observed that the districts ruled by the state have more rape victims than the districts ruled by the Union Territory. More investigation can be done regarding this, and actions can be taken accordingly. A limitation to this dataset, is that this is just the number of women raped, different genders are not included. Additionally, this data is just from one county, India. Rape is a worldwide problem, analysis and actions need to be done for every part of the world.

### References:

- Bhardwaj, A. & Muraleedharan,A. (2016). *Political Map of India*. Maps of India.
- National Portal of India. *Open Government Data (OGD) Platform India*, 2022, <https://data.gov.in/>.
- NCRB. *National Crime Records Bureau*, 2018. <https://ncrb.gov.in/en>
- Prabhavalkar, N. (2021). *Crimes in India-2018 Victims of Rape*. Kaggle.