Two Decades of Women Safety in India (2001–2021)

An Exploratory Data Analysis of Crime Trends

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly
remount, call drive.mount("/content/drive", force_remount=True).
```

Importing the essential libraries for our project

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Reading our CSV data to analyse

```
url = 'https://drive.google.com/uc?
export=download&id=1j8tOcMfhRZPL1ijwYX3AMc6iuy_Udnve'
df = pd.read_csv(url)
df
```

```
{"type":"dataframe","variable_name":"df"}
```

Read first 5 rows of the data

```
df.head(30)
```

```
{"type":"dataframe","variable_name":"df"}
```

Now analyse the size and no of content we have in our data and then the statistic information

```
print(df.info())
print("*"*70)
print("Size of the dataset is:",df.size)
print("*"*70)
print("Shape of the dataset is:",df.shape)
print("*"*70)
print(df.describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 44 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   year                                    10000 non-null  int64
 1   state                                   10000 non-null  object
 2   district                                10000 non-null  object
 3   rape                                    10000 non-null  int64
 4   attempt_to_rape                         10000 non-null  int64
 5   gang_rape                               10000 non-null  int64
 6   murder_with_rape                        10000 non-null  int64
 7   kidnapping_and_abduction                10000 non-null  int64
 8   dowry_deaths                            10000 non-null  int64
 9   dowry_prohibition_act                   10000 non-null  int64
 10  cruelty_by_husband_or_relatives_498A    10000 non-null  int64
 11  acid_attack                             10000 non-null  int64
 12  attempt_to_acid_attack                  10000 non-null  int64
 13  assault_on_women_modesty_354            10000 non-null  int64
 14  assault_intent_disrobe_354B             10000 non-null  int64
 15  voyeurism_354C                          10000 non-null  int64
 16  stalking_354D                           10000 non-null  int64
 17  insult_to_modesty_509                   10000 non-null  int64
 18  trafficking                             10000 non-null  int64
 19  procuration_of_minor_girls              10000 non-null  int64
 20  importation_of_girls                    10000 non-null  int64
 21  immoral_traffic_act                     10000 non-null  int64
 22  indecent_representation_of_women        10000 non-null  int64
 23  women_killed_in_honour_killing          10000 non-null  int64
 24  witch_hunting                           10000 non-null  int64
 25  cyber_crime_obscenity_against_women     10000 non-null  int64
 26  cyber_stalking_bullying_against_women   10000 non-null  int64
 27  child_marriage_prohibition_act          10000 non-null  int64
 28  pocso_rape                              10000 non-null  int64
 29  pocso_assault                           10000 non-null  int64
 30  pocso_harassment                        10000 non-null  int64
 31  pocso_unnatural_offences                10000 non-null  int64
 32  pocso_other                             10000 non-null  int64
 33  abduction_for_marriage                  10000 non-null  int64
 34  abduction_for_illicit_intercourse       10000 non-null  int64
 35  attempt_to_kidnap                       10000 non-null  int64
 36  domestic_violence_act_cases             10000 non-null  int64
 37  abetment_to_suicide_women               10000 non-null  int64
 38  attempt_to_murder_women                 10000 non-null  int64
 39  insult_outraging_modesty_other          10000 non-null  int64
 40  sexual_harassment_workplace             10000 non-null  int64
 41  dowry_harassment                        10000 non-null  int64
 42  marital_rape_reports                    10000 non-null  int64
 43  total_cases                             10000 non-null  int64
dtypes: int64(42), object(2)
```

```
memory usage: 3.4+ MB
None
**********************************************************************
Size of the dataset is: 440000
**********************************************************************
Shape of the dataset is: (10000, 44)
**********************************************************************
                year          rape  attempt_to_rape     gang_rape  \
count  10000.000000  10000.000000     10000.000000  10000.000000
mean    2010.253000     17.861100        17.816000     17.810400
std        5.631178     10.291045        10.264944     10.236138
min     2001.000000      0.000000         0.000000      0.000000
25%     2005.000000      9.000000         9.000000      9.000000
50%     2010.000000     17.000000        17.000000     17.000000
75%     2015.000000     25.000000        25.000000     25.000000
max     2020.000000     55.000000        60.000000     56.000000

       murder_with_rape  kidnapping_and_abduction  dowry_deaths  \
count      10000.000000                10000.00000  10000.000000
mean          17.777600                   27.87320     17.844900
std           10.239728                   10.70823     10.238825
min            0.000000                    4.00000      0.000000
25%            9.000000                   19.00000      9.000000
50%           16.000000                   27.00000     17.000000
75%           25.000000                   35.00000     25.000000
max           55.000000                   70.00000     55.000000

       dowry_prohibition_act  cruelty_by_husband_or_relatives_498A  \
count           10000.000000                          10000.000000
mean               17.876300                             27.823600
std                10.198068                             10.773954
min                 0.000000                              3.000000
25%                 9.000000                             19.000000
50%                17.000000                             27.000000
75%                25.000000                             35.000000
max                61.000000                             76.000000

        acid_attack  ...  abduction_for_illicit_intercourse  \
count  10000.000000  ...                       10000.000000
mean      17.834200  ...                          17.858400
std       10.299676  ...                          10.270204
min        0.000000  ...                           0.000000
25%        9.000000  ...                           9.000000
50%       17.000000  ...                          17.000000
75%       25.000000  ...                          25.000000
max       58.000000  ...                          60.000000

       attempt_to_kidnap  domestic_violence_act_cases  \
count        10000.00000                 10000.000000
mean            17.83650                    17.803100
```

```
std              10.24629                    10.221026
min               0.00000                     0.000000
25%               9.00000                     9.000000
50%              17.00000                    17.000000
75%              25.00000                    25.000000
max              56.00000                    60.000000

       abetment_to_suicide_women  attempt_to_murder_women  \
count               10000.000000              10000.00000
mean                   17.884100                 17.78990
std                    10.281986                 10.20092
min                     0.000000                  0.00000
25%                     9.000000                  9.00000
50%                    17.000000                 17.00000
75%                    25.000000                 25.00000
max                    58.000000                 60.00000

       insult_outraging_modesty_other  sexual_harassment_workplace  \
count                    10000.000000                 10000.000000
mean                        17.815500                    17.824800
std                         10.284942                    10.221847
min                          0.000000                     0.000000
25%                          9.000000                     9.000000
50%                         17.000000                    17.000000
75%                         25.000000                    25.000000
max                         59.000000                    57.000000

       dowry_harassment  marital_rape_reports   total_cases
count      10000.000000          10000.000000  10000.000000
mean          17.872300             17.804700   1634.724800
std           10.331547             10.245968    842.604171
min            0.000000              0.000000    443.000000
25%            9.000000              9.000000    793.000000
50%           17.000000             17.000000   1581.000000
75%           25.000000             25.000000   2261.250000
max           59.000000             57.000000   3676.000000

[8 rows x 42 columns]
```

Dataset Analysis Summary 😁

- The dataset contains **10,000 rows** and **44 columns**.
- Using the `info()` function, we confirmed that there are **no null values** in the dataset.
- The `describe()` function provides the **statistical analysis** of the data, including measures such as mean, standard deviation, minimum, maximum, and quartiles.

```
df.columns
```

```
Index(['year', 'state', 'district', 'rape', 'attempt_to_rape',
'gang_rape',
       'murder_with_rape', 'kidnapping_and_abduction', 'dowry_deaths',
       'dowry_prohibition_act',
'cruelty_by_husband_or_relatives_498A',
       'acid_attack', 'attempt_to_acid_attack',
'assault_on_women_modesty_354',
       'assault_intent_disrobe_354B', 'voyeurism_354C',
'stalking_354D',
       'insult_to_modesty_509', 'trafficking',
'procuration_of_minor_girls',
       'importation_of_girls', 'immoral_traffic_act',
       'indecent_representation_of_women',
'women_killed_in_honour_killing',
       'witch_hunting', 'cyber_crime_obscenity_against_women',
       'cyber_stalking_bullying_against_women',
       'child_marriage_prohibition_act', 'pocso_rape',
'pocso_assault',
       'pocso_harassment', 'pocso_unnatural_offences', 'pocso_other',
       'abduction_for_marriage', 'abduction_for_illicit_intercourse',
       'attempt_to_kidnap', 'domestic_violence_act_cases',
       'abetment_to_suicide_women', 'attempt_to_murder_women',
       'insult_outraging_modesty_other',
'sexual_harassment_workplace',
       'dowry_harassment', 'marital_rape_reports', 'total_cases'],
      dtype='object')
```

We have to analyse the datatype of each column

df.dtype

```
df.dtypes
```

```
year                                    int64
state                                  object
district                               object
rape                                    int64
attempt_to_rape                         int64
gang_rape                               int64
murder_with_rape                        int64
kidnapping_and_abduction                int64
dowry_deaths                            int64
dowry_prohibition_act                   int64
cruelty_by_husband_or_relatives_498A    int64
acid_attack                             int64
attempt_to_acid_attack                  int64
assault_on_women_modesty_354            int64
assault_intent_disrobe_354B             int64
voyeurism_354C                          int64
```

```
stalking_354D                              int64
insult_to_modesty_509                      int64
trafficking                                int64
procuration_of_minor_girls                 int64
importation_of_girls                       int64
immoral_traffic_act                        int64
indecent_representation_of_women           int64
women_killed_in_honour_killing             int64
witch_hunting                              int64
cyber_crime_obscenity_against_women        int64
cyber_stalking_bullying_against_women      int64
child_marriage_prohibition_act             int64
pocso_rape                                 int64
pocso_assault                              int64
pocso_harassment                           int64
pocso_unnatural_offences                   int64
pocso_other                                int64
abduction_for_marriage                     int64
abduction_for_illicit_intercourse          int64
attempt_to_kidnap                          int64
domestic_violence_act_cases                int64
abetment_to_suicide_women                  int64
attempt_to_murder_women                    int64
insult_outraging_modesty_other             int64
sexual_harassment_workplace                int64
dowry_harassment                           int64
marital_rape_reports                       int64
total_cases                                int64
dtype: object
```

##Lets have a look on the skewness of our data

Firstly lets have a look on skewness --Skewness helps data analysts understand the asymmetry of a data distribution, indicating whether extreme values are concentrated on one side of the mean

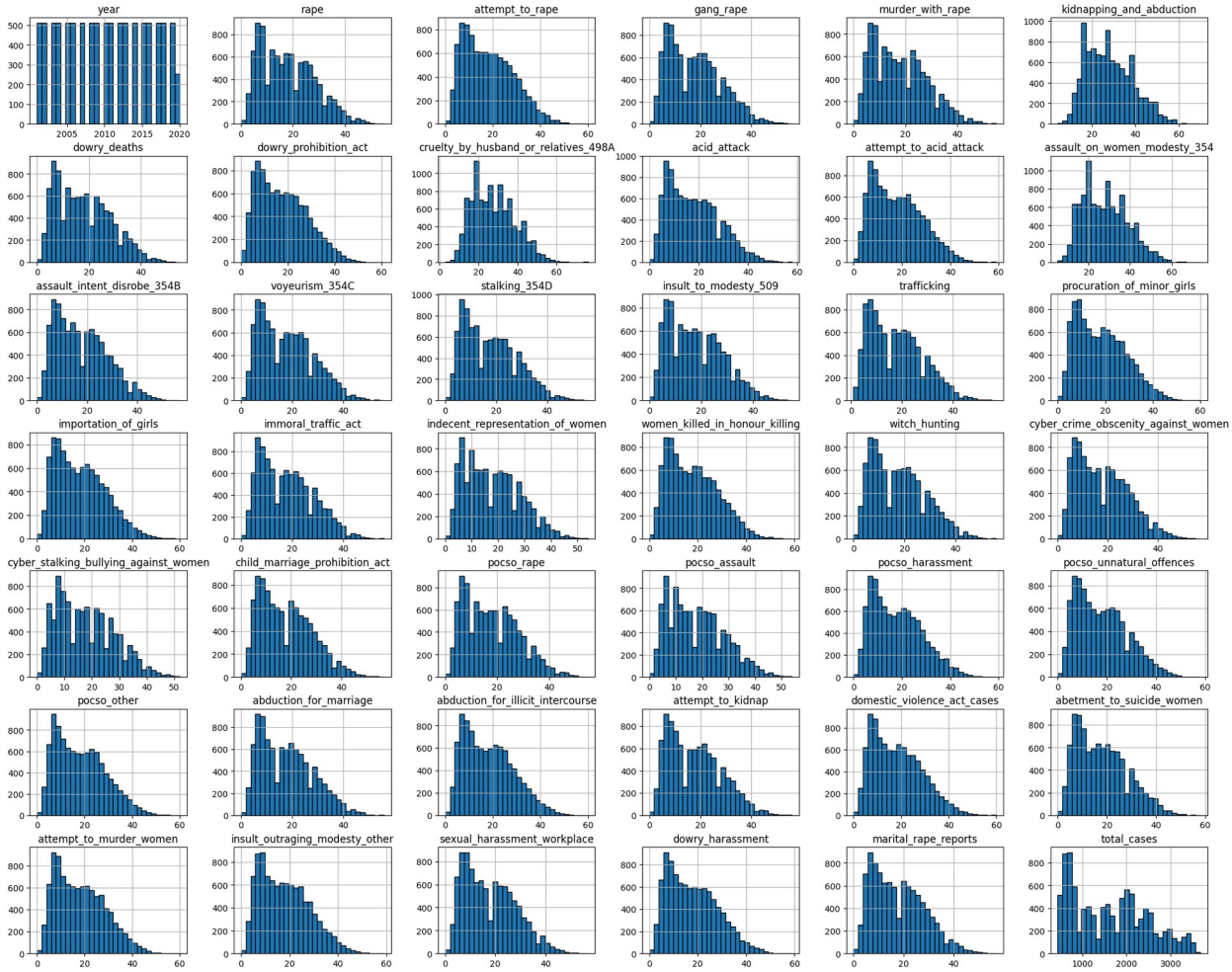
```
skewness = df.skew(numeric_only=True)
skewness
```

```
year                                    0.006824
rape                                    0.520890
attempt_to_rape                         0.533621
gang_rape                               0.559385
murder_with_rape                        0.544564
kidnapping_and_abduction                0.479961
dowry_deaths                            0.518284
dowry_prohibition_act                   0.538136
cruelty_by_husband_or_relatives_498A    0.469812
acid_attack                             0.545617
```

```
attempt_to_acid_attack                    0.533860
assault_on_women_modesty_354              0.441971
assault_intent_disrobe_354B               0.552664
voyeurism_354C                            0.532775
stalking_354D                             0.563124
insult_to_modesty_509                     0.514559
trafficking                               0.547969
procuration_of_minor_girls                0.532007
importation_of_girls                      0.553537
immoral_traffic_act                       0.525546
indecent_representation_of_women          0.513228
women_killed_in_honour_killing            0.527994
witch_hunting                             0.527033
cyber_crime_obscenity_against_women       0.545949
cyber_stalking_bullying_against_women     0.515016
child_marriage_prohibition_act            0.556025
pocso_rape                                0.526582
pocso_assault                             0.542816
pocso_harassment                          0.566727
pocso_unnatural_offences                  0.546533
pocso_other                               0.546226
abduction_for_marriage                    0.545291
abduction_for_illicit_intercourse         0.549432
attempt_to_kidnap                         0.526885
domestic_violence_act_cases               0.559063
abetment_to_suicide_women                 0.558741
attempt_to_murder_women                   0.533533
insult_outraging_modesty_other            0.544977
sexual_harassment_workplace               0.528809
dowry_harassment                          0.545642
marital_rape_reports                      0.541792
total_cases                               0.372521
dtype: float64
```

Lets Visualise the skewness

```
df.hist(figsize=(25, 20), bins=30, edgecolor='black')
plt.suptitle("Distribution of Features", fontsize=16)
plt.show()
```



Univariate and Bivariate Analysis

Now we are going to Analyse the total cases in each state of India

```
r=df[['state','district','total_cases']]
r.describe()
```


```
{"summary":"{\n  \"name\": \"r\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"total_cases\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 3139.283365096296,\n
\"min\": 443.0,\n        \"max\": 10000.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
1634.7248,\n          1581.0,\n          10000.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```

##For Analysing the total cases in each state we use the group by function -grouping total cases by their state -Finding the total cases of the state

```
r=df.groupby('state')['total_cases'].sum()
r
```

```
state
Andaman and Nicobar Islands                   274535
Andhra Pradesh                                187089
Arunachal Pradesh                             445161
Assam                                         241283
Bihar                                         876635
Chandigarh                                    888178
Chhattisgarh                                  484416
Dadra and Nagar Haveli and Daman and Diu      611507
Delhi                                         236732
Goa                                           186520
Gujarat                                       227607
Haryana                                       224636
Himachal Pradesh                              187142
Jammu and Kashmir                             240948
Jharkhand                                     728108
Karnataka                                     335808
Kerala                                        591354
Ladakh                                        687954
Lakshadweep                                   225392
Madhya Pradesh                                444497
Maharashtra                                   470311
Manipur                                       200492
Meghalaya                                     404514
Mizoram                                       407988
Nagaland                                      667720
Odisha                                        498677
Puducherry                                    795917
Punjab                                        956294
Rajasthan                                     401083
Sikkim                                        829631
Tamil Nadu                                    266936
Telangana                                     889374
Tripura                                       252904
Uttar Pradesh                                 237442
Uttarakhand                                   238349
West Bengal                                   504114
Name: total_cases, dtype: int64
```

```
r.describe()
```

```
count        36.000000
mean     454090.222222
std      244171.316812
```

```
min       186520.000000
25%       238122.250000
50%       406251.000000
75%       625560.250000
max       956294.000000
Name: total_cases, dtype: float64
```

Now from above analysis we saw that the state having minimum and maximum total case is as follows::

```
# State with minimum cases
min_state = r.idxmin()
min_cases = r.min()

# State with maximum cases
max_state = r.idxmax()
max_cases = r.max()

print("State with Minimum Crime Cases:", min_state, "->", min_cases)
print("State with Maximum Crime Cases:", max_state, "->", max_cases)
```

```
State with Minimum Crime Cases: Goa -> 186520
State with Maximum Crime Cases: Punjab -> 956294
```

Analysing the top10 states whole total crimes is higher then other

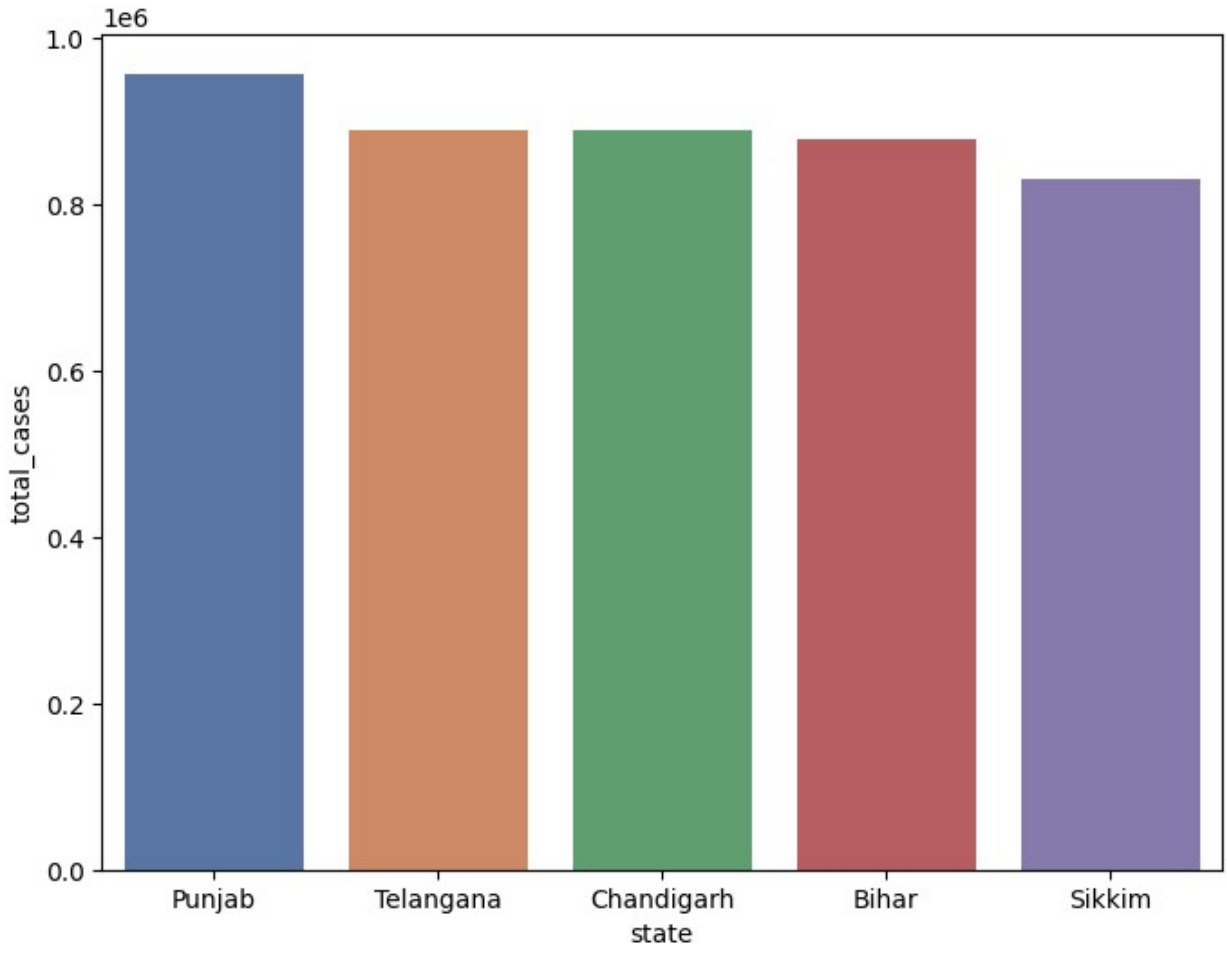
-- Ploting the bar graph after grouping data -- Sort_value help us to see the data in descending order

```
r=df.groupby('state')
['total_cases'].sum().sort_values(ascending=False).head(5)
plt.figure(figsize=(8, 6))
sns.barplot(r,palette="deep")
plt.show()
```

```
/tmp/ipython-input-2692999933.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.barplot(r,palette="deep")
```



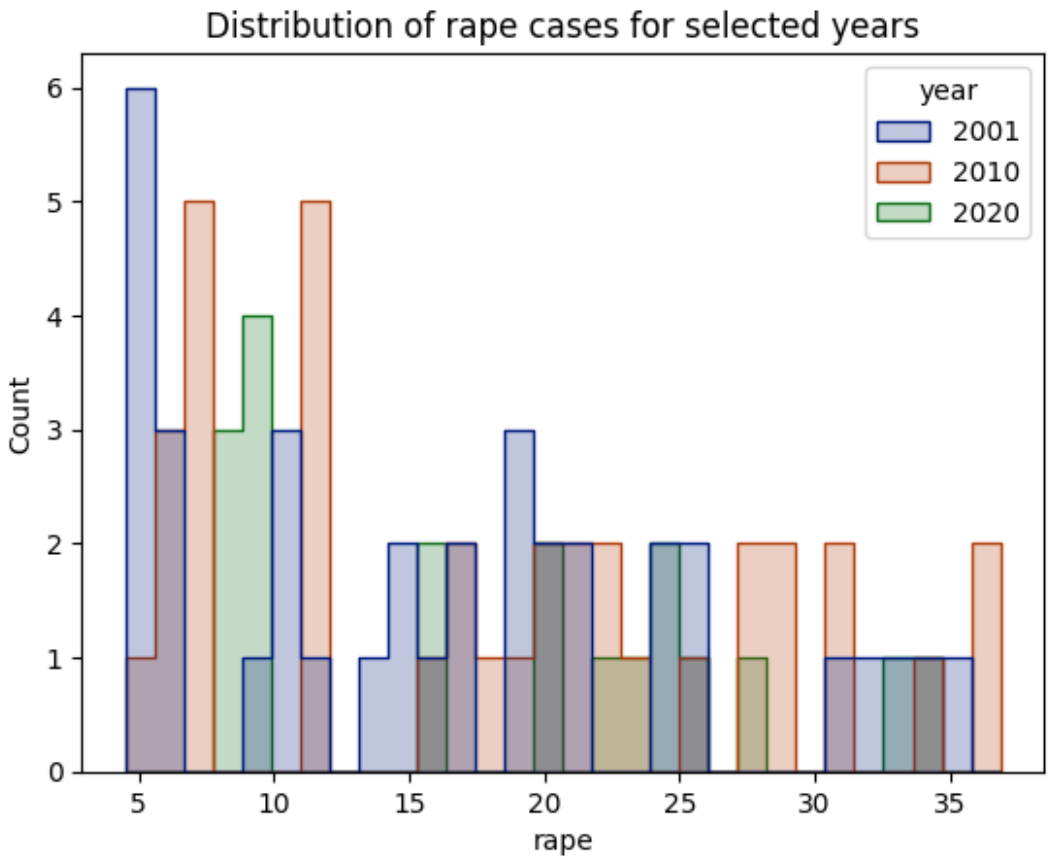
Distribution of Rape Cases Across States (2001, 2010, 2020)

This histogram shows the **average number of rape cases per state** for the years 2001, 2010, and 2020.

Key Takeaways:

- Average rape cases per state have **increased over the last 20 years**.
- Some states consistently report higher numbers — these may need **focused attention**.
- The trend may reflect **better reporting, increased awareness, or actual rise in incidents**.

```
d1 = df.groupby(['year','state'], as_index=False)['rape'].mean()
subset = d1[d1['year'].isin([2001, 2010, 2020])]
sns.histplot(data=subset, x='rape', hue='year', bins=30,
element="step",palette='dark')
plt.title("Distribution of rape cases for selected years")
plt.show()
```



Trend of Average Rape Cases per State (2001–2021)

This line plot shows the **average number of rape cases per state** for each year from 2001 to 2021.

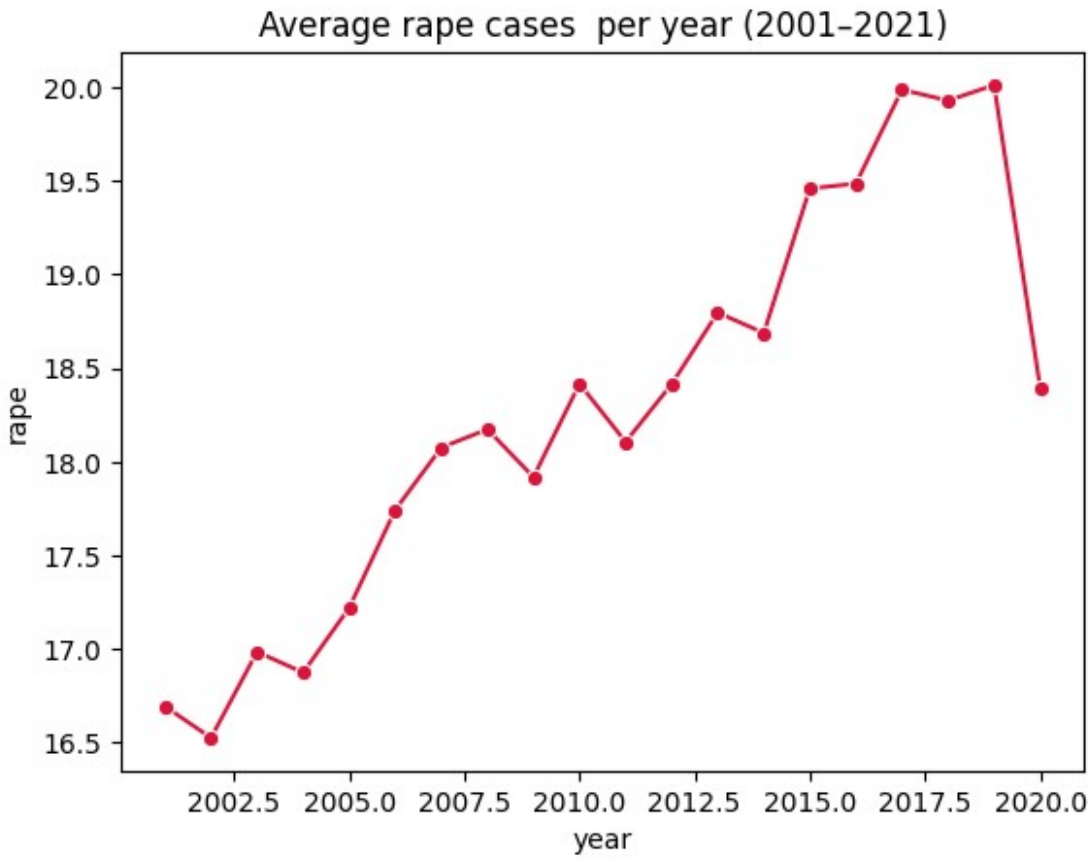
Observations:

- There is a **gradual increase** in average cases over the 21-year period.
- **Early 2000s (2001–2005):** Lower average cases, indicating fewer reported incidents or lower reporting.
- **2010–2015:** A noticeable rise, possibly due to improved reporting mechanisms or increased awareness.
- **2016–2021:** The trend continues upward, suggesting either a real increase in cases or better data collection.
- The plot highlights **year-to-year variations**, but the overall trend is clearly **increasing**.

Key Takeaways:

- Average rape cases per state have **risen consistently over two decades**.
- The upward trend may reflect **better reporting, awareness campaigns, or actual increase in incidents**.
- This analysis helps identify **long-term trends** rather than individual state variations.

```
trend = d1.groupby('year')['rape'].mean().reset_index()
sns.lineplot(data=trend, x='year', y='rape', marker='o',
color="crimson")
plt.title("Average rape cases  per year (2001–2021)")
plt.show()
```



Share of Selected Crimes in 2020

This pie chart shows the **distribution of four major crime categories** against women in 2020:

- **Rape**
- **Dowry Deaths**
- **Domestic Violence Act Cases**

- **Cruelty by Husband or Relatives (498A)**

Observations:

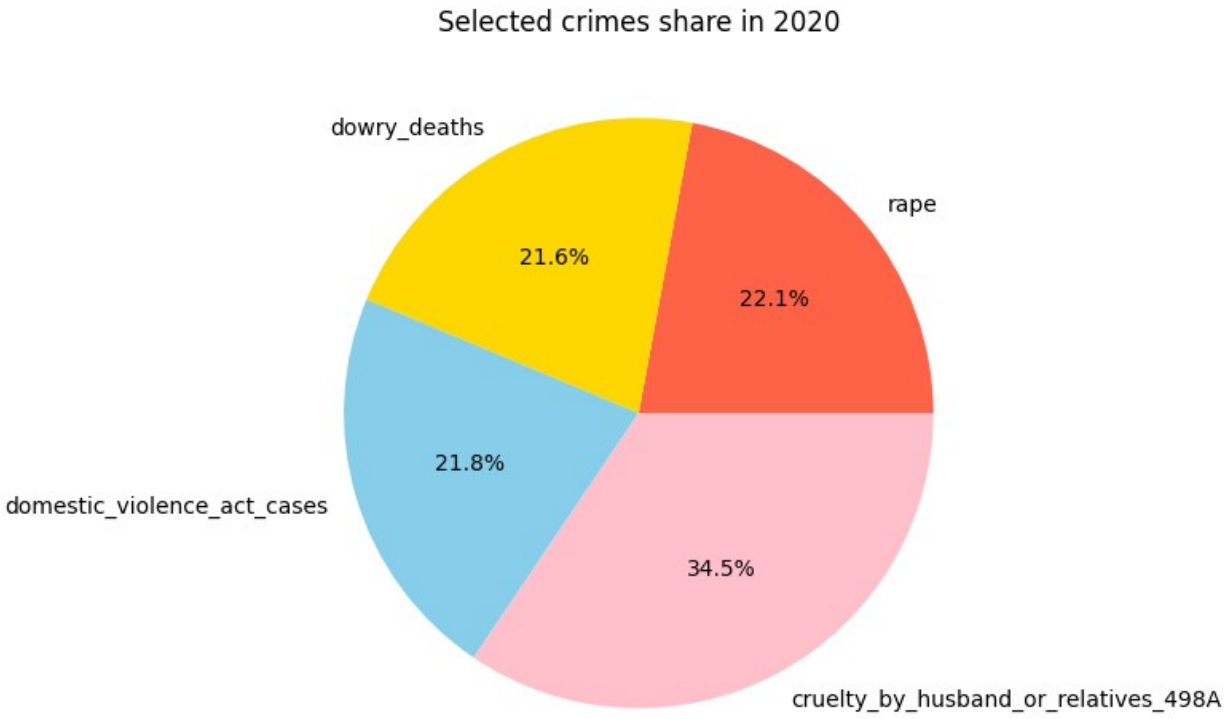
- The largest slice represents the crime with the **highest number of reported cases** among the selected categories.
- Smaller slices indicate **less frequent but still significant crimes**.
- The pie chart helps understand **which types of crimes contribute most to the overall burden** in 2020.

Key Takeaways:

- Rape or Domestic Violence-related cases may dominate the share, highlighting **critical areas for policy and awareness**.
- Even smaller slices like Dowry Deaths and Cruelty under 498A are important and **require attention**.
- Such proportion-based visualization complements line and histogram plots by **showing composition rather than trend**.

```
year_data = df[df['year'] == 2020]
selected =
year_data[['rape','dowry_deaths','domestic_violence_act_cases','cruelt
y_by_husband_or_relatives_498A']].sum()

plt.figure(figsize=(6,6))
plt.pie(selected, labels=selected.index, autopct='%1.1f%%',
colors=['tomato','gold','skyblue','Pink'])
plt.title("Selected crimes share in 2020")
plt.show()
```



Distribution of Total Women-related Crime Cases Across States (2018–2020)

This boxplot shows the **distribution of total women-related crime cases** per state for the years **2018, 2019, and 2020**.

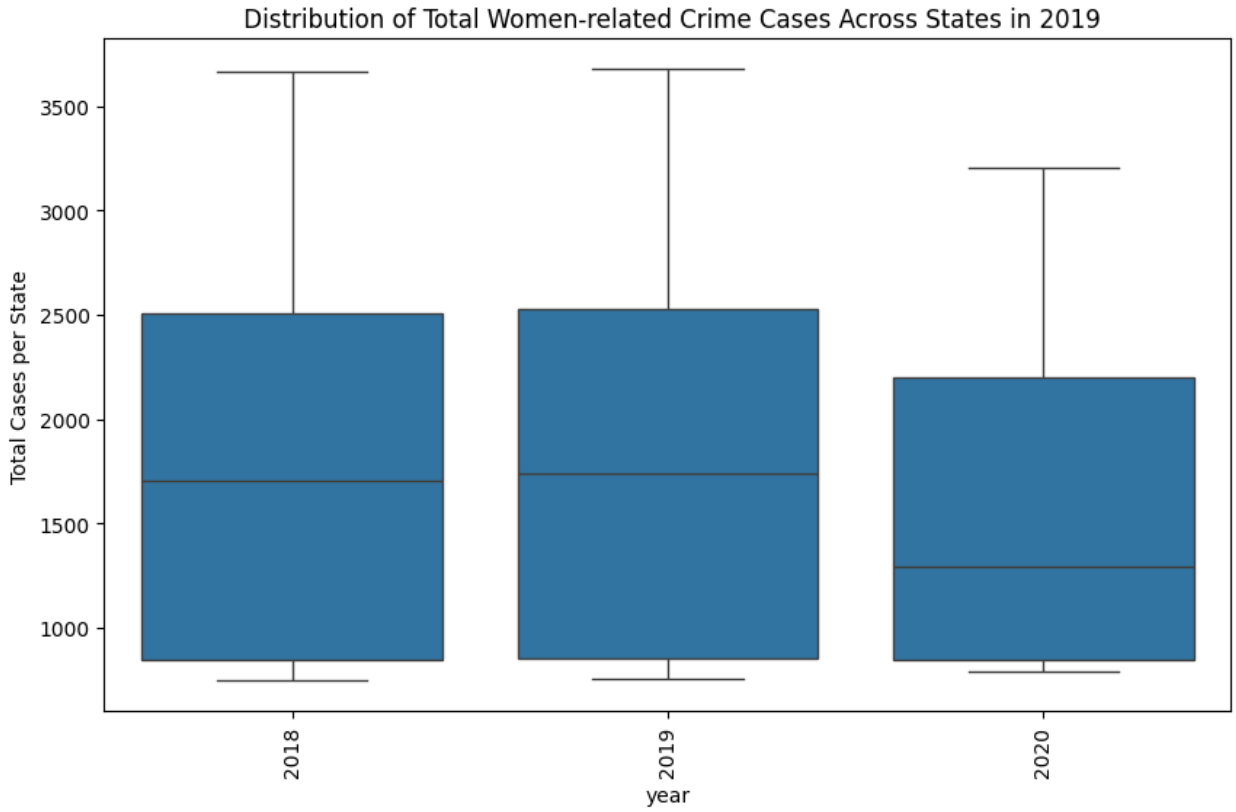
Observations:

- Each box represents the **spread of total cases** across all states for that year.
- The **central line in each box** shows the **median number of cases**.
  - A rising median over the years indicates an **overall increase in reported crimes**.
- The **height of the box** (interquartile range, Q1–Q3) shows the **variation among states**.
  - Wider boxes indicate more disparity; narrower boxes indicate more uniform reporting.
- **Whiskers** represent the **minimum and maximum within 1.5×IQR**.
- **Dots outside the whiskers** are **outlier states** with unusually high or low total crime cases.

```
plt.figure(figsize=(10,6))
d1 = df[df['year'].isin([2018, 2019, 2020])]

sns.boxplot(data=d1, x='year', y='total_cases')
plt.xticks(rotation=90)
plt.title(" Distribution of Total Women-related Crime Cases Across
```

```
States in 2019")
plt.ylabel("Total Cases per State")
plt.show()
```



##Dowry Death and Domestic Violance cases

Heatmap Analysis of Dowry Deaths (State vs Year)

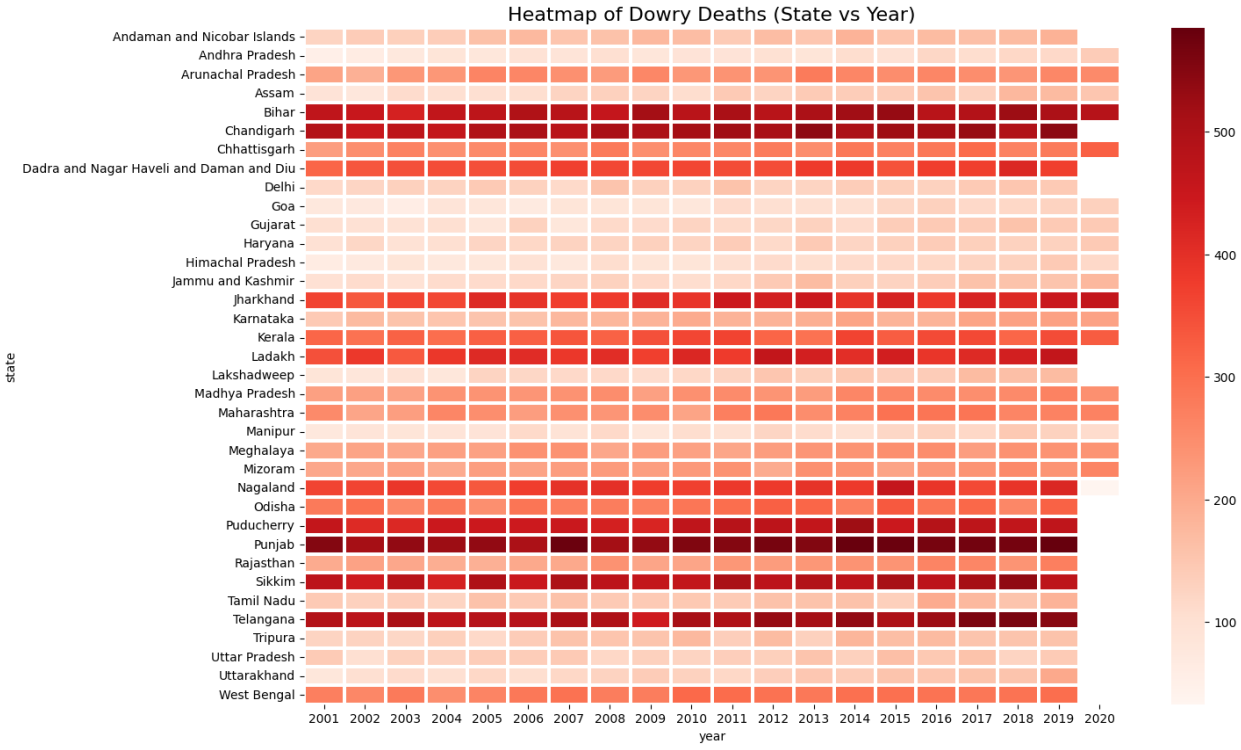
The heatmap above shows the **distribution of dowry deaths across different states and years**.

- Darker shades represent **higher numbers of dowry deaths**, while lighter shades represent **fewer cases**.

- This visualization makes it easier to identify **regional and temporal patterns**:
  - Some states consistently show darker shades across years, indicating **persistently high dowry death cases**.
  - A few states show **gradual decline or increase** in cases over the years.
  - States with very light colors have **lower incidence** of dowry-related deaths.

```
# 4. Heatmap (State vs Year)
heatmap_data = df.pivot_table(index="state", columns="year",
values="dowry_deaths", aggfunc="sum")
plt.figure(figsize=(15,10))
```

```
sns.heatmap(heatmap_data, cmap="Reds", linewidths=1.5)
plt.title("Heatmap of Dowry Deaths (State vs Year)", fontsize=16)
plt.show()
```



Now this will help us to analyse:

- That in past few year do our security and system helps to reduce the domestic violance in womens even
- Even having wast education have these cases get reduced

```
data=df[df['year'].isin([2018,2019,2020])]
data=data.groupby(['state', 'year'])
['domestic_violence_act_cases'].mean().reset_index()
data
```

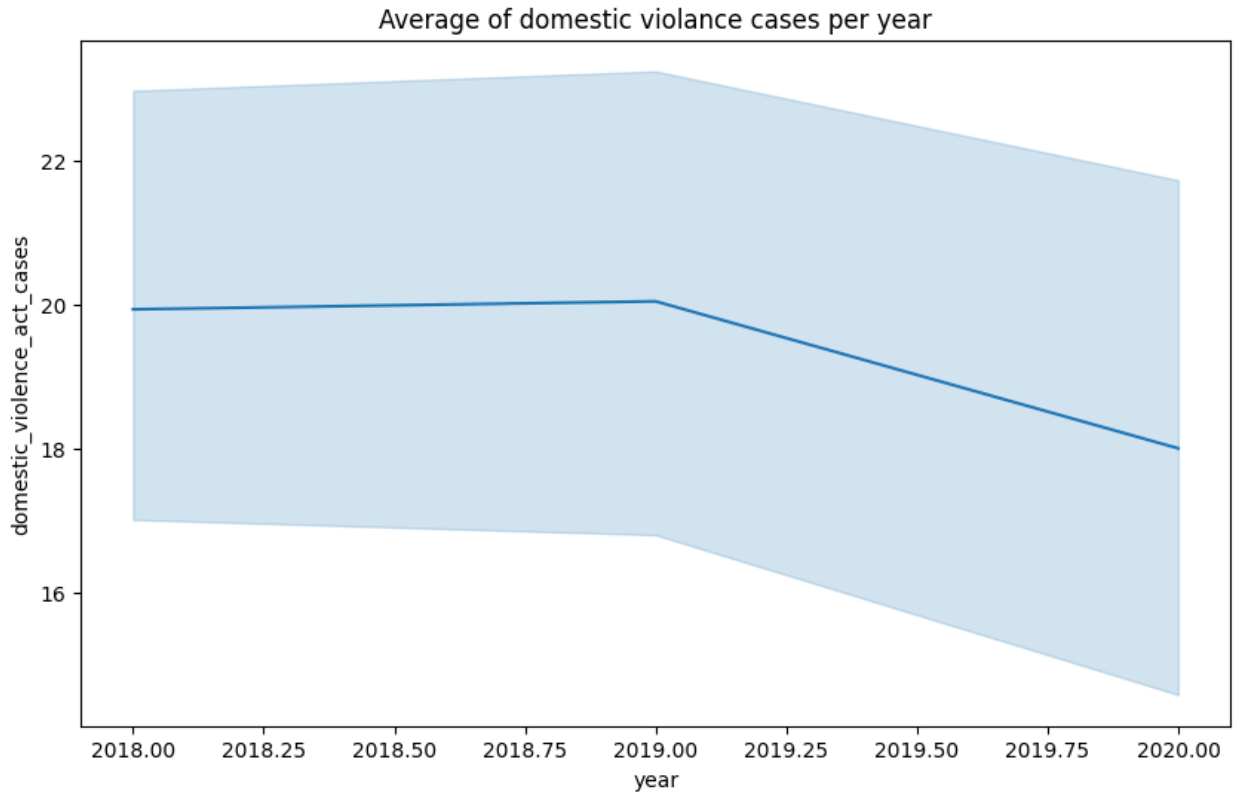

```
{"summary":"{\n  \"name\": \"data\",\n  \"rows\": 91,\n  \"fields\":
[\n    {\n      \"column\": \"state\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 36,\n
\"samples\": [\n          \"West Bengal\",\n          \"Jammu and
Kashmir\",\n          \"Puducherry\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"year\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 2018,\n
\"max\": 2020,\n        \"num_unique_values\": 3,\n
\"samples\": [\n          2018,\n          2019,\n          2020\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"domestic_violence_act_cases\",\n
```

```
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
9.131850444785,\n        \"min\": 6.642857142857143,\n        \"max\":
41.46153846153846,\n        \"num_unique_values\": 86,\n
\"samples\": [\n          16.22222222222222,\n
14.461538461538462,\n          13.789473684210526\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"data"}
```


Line bar will help us to analyse the data

-We had grouped the data with respect to the state and year -After having the mean lets analyse the data

```
plt.figure(figsize=(10,6))
sns.lineplot(x='year',y='domestic_violence_act_cases',data=data)
plt.title("Average of domestic violance cases per year")
plt.show()
```



Heatmap of Domestic Violence Act Cases (2018–2020)

In this visualization, we created a **pivot table** of states (rows) versus years (columns: 2018, 2019, 2020) and plotted a **heatmap**.

Why are we doing this?

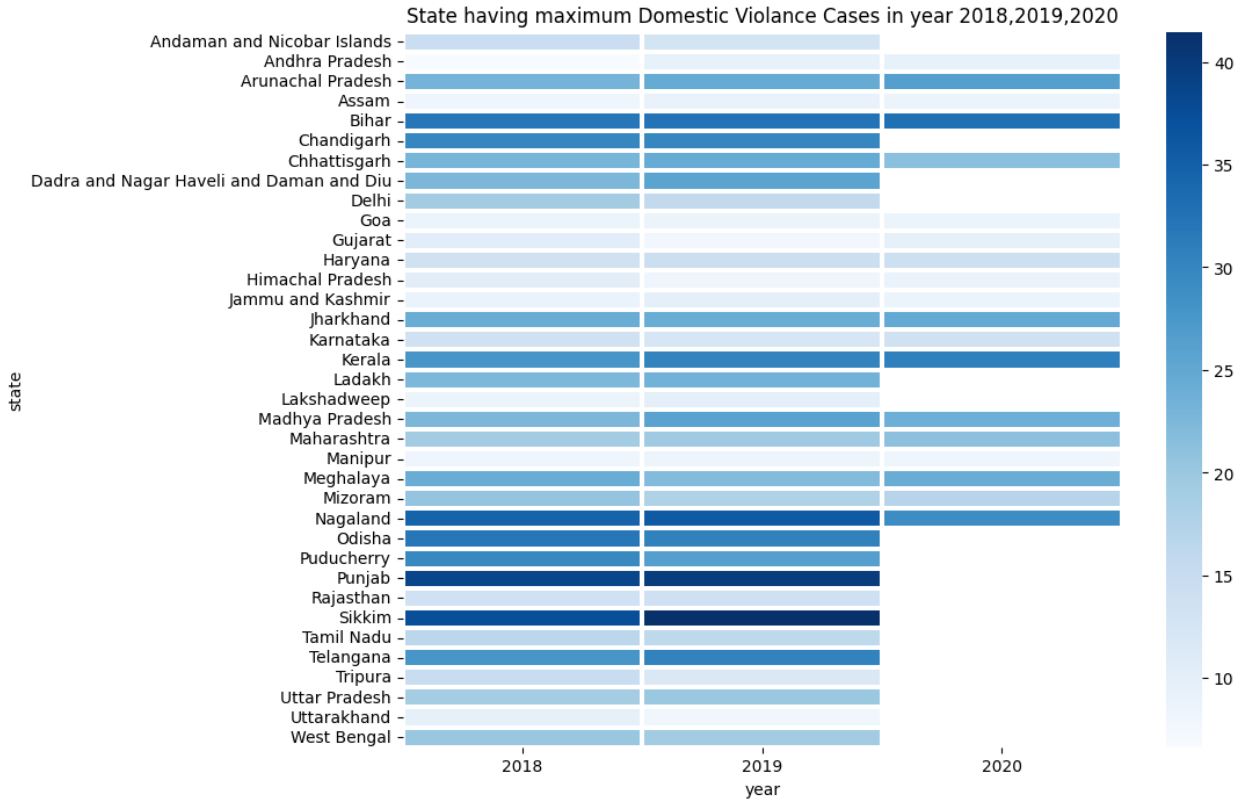
- To compare **state-wise domestic violence cases across multiple years** in one single chart.
- Heatmap helps us quickly identify where the values are **high (darker shades)** and **low (lighter shades)** without scrolling through large tables.
- Instead of checking each state separately, we can **visually detect patterns and outliers**.

What insights do we get?

- **States with consistently darker shades** across 2018–2020 → these states have **persistently high domestic violence cases**.
- **Lighter shades** indicate states with relatively fewer cases.
- By comparing columns (years), we can observe if cases are **increasing, decreasing, or stable** across time.
- It gives both a **spatial (state-wise)** and **temporal (year-wise)** perspective.

Overall, this heatmap allows us to **pinpoint which states need urgent attention** in terms of domestic violence cases and also whether the situation is improving or worsening over time.

```
plt.figure(figsize=(10,8))
pivot=df.pivot_table(index='state',columns='year',values='domestic_vio
lence_act_cases',aggfunc='mean')
pivot = pivot.loc[:, [2018, 2019, 2020]]
sns.heatmap(data=pivot,cmap='Blues',linewidths=1.5)
plt.title("State having maximum Domestic Violance Cases in year
2018,2019,2020")
plt.show()
```



Analysis of Acid Attack Cases (2018–2020)

Acid attacks are among the most brutal crimes against women. They not only cause **severe physical injuries and disfigurement** but also leave victims with **lifelong trauma**, affecting their ability to move freely, feel safe, and live independently.

Why are we analyzing this?

- Acid attacks directly **restrict women's freedom** in public spaces, as fear of such crimes can discourage women from education, employment, and social participation.
- By studying the **state-wise and year-wise trend**, we can identify **regions with higher risk** and track whether the situation is improving or worsening over time.
- It highlights **where stricter law enforcement, awareness programs, and victim support** are most urgently needed.

What does the grouped data show?

- We have grouped the dataset by **state and year (2018–2020)** to calculate the **average number of acid attack cases**.
- This grouping helps us compare how different states perform over the years and locate the **hotspots of the crime**.

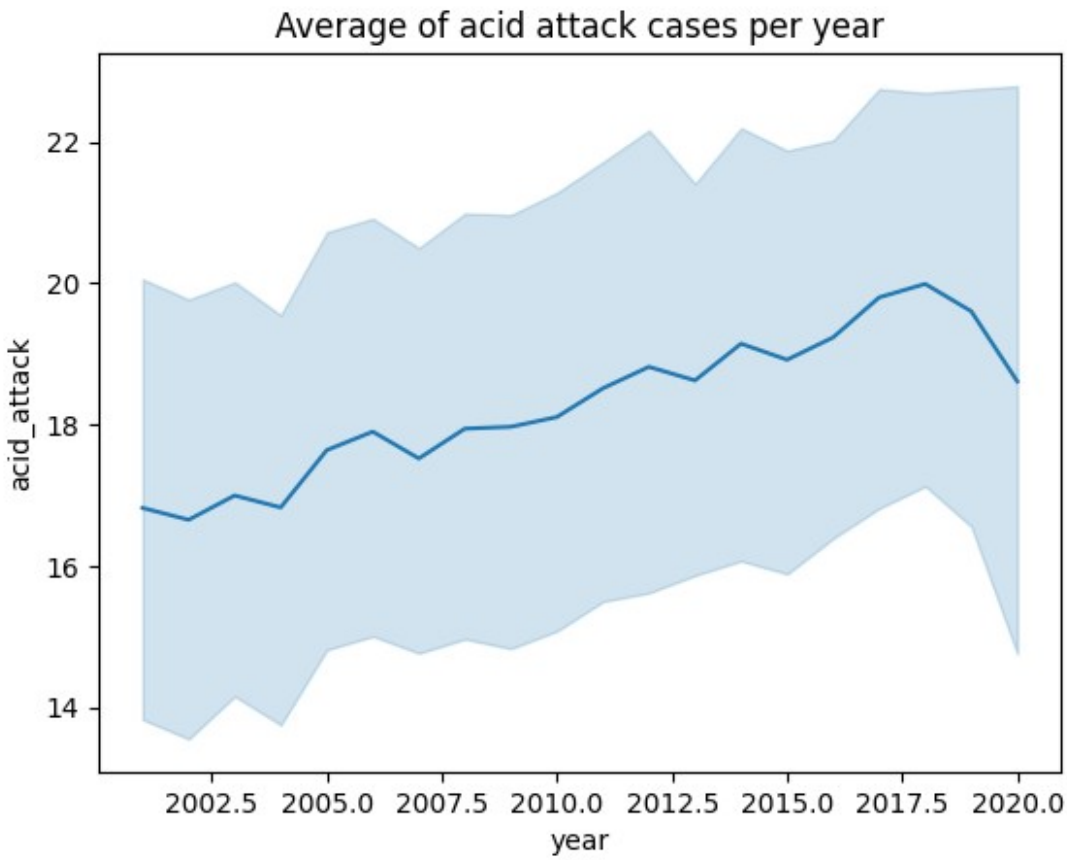
- A further visualization (heatmap or barplot) will make these differences even clearer.

 Overall, analyzing acid attack data is crucial for understanding the **threat landscape faced by women**, and it guides **policy makers, NGOs,**

```
# data=df[df['year'].isin([2018,2019,2020])]
data=df.groupby(['state', 'year'])['acid_attack'].mean().reset_index()
data
```


```
{"summary":"{\n  \"name\": \"data\",\n  \"rows\": 703,\n  \"fields\":
[\n    {\n      \"column\": \"state\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 36,\n
\"samples\": [\n          \"West Bengal\",\n          \"Jammu and
Kashmir\",\n          \"Puducherry\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"year\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5,\n        \"min\": 2001,\n
\"max\": 2020,\n        \"num_unique_values\": 20,\n
\"samples\": [\n          2001,\n          2018,\n          2016\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"acid_attack\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
9.331381458747197,\n        \"min\": 4.214285714285714,\n
\"max\": 41.0,\n        \"num_unique_values\": 542,\n
\"samples\": [\n          17.76923076923077,\n
32.266666666666666,\n          15.384615384615385\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"data"}
```


```
sns.lineplot(x='year',y='acid_attack',data=data)
plt.title("Average of acid attack cases per year")
plt.show()
```



The line chart above represents the **average number of acid attack cases reported each year**.

Why this visualization?

- Using a **line chart** helps us observe the **trend of acid attacks over time**.
- It gives a clear view of whether the crime rate is **increasing, decreasing, or staying stable** year by year.

Key Insights

- Each point on the line corresponds to the **mean value of acid attack cases** for that particular year.
- A **rising line** suggests that the problem is worsening and requires urgent intervention.
- A **falling line** indicates that stricter laws, awareness, or enforcement may be having a positive effect.

```
df.columns
```

```
Index(['year', 'state', 'district', 'rape', 'attempt_to_rape',
'gang_rape',
       'murder_with_rape', 'kidnapping_and_abduction', 'dowry_deaths',
```

```
       'dowry_prohibition_act',
'cruelty_by_husband_or_relatives_498A',
       'acid_attack', 'attempt_to_acid_attack',
'assault_on_women_modesty_354',
       'assault_intent_disrobe_354B', 'voyeurism_354C',
'stalking_354D',
       'insult_to_modesty_509', 'trafficking',
'procuration_of_minor_girls',
       'importation_of_girls', 'immoral_traffic_act',
       'indecent_representation_of_women',
'women_killed_in_honour_killing',
       'witch_hunting', 'cyber_crime_obscenity_against_women',
       'cyber_stalking_bullying_against_women',
       'child_marriage_prohibition_act', 'pocso_rape',
'pocso_assault',
       'pocso_harassment', 'pocso_unnatural_offences', 'pocso_other',
       'abduction_for_marriage', 'abduction_for_illicit_intercourse',
       'attempt_to_kidnap', 'domestic_violence_act_cases',
       'abetment_to_suicide_women', 'attempt_to_murder_women',
       'insult_outraging_modesty_other',
'sexual_harassment_workplace',
       'dowry_harassment', 'marital_rape_reports', 'total_cases'],
      dtype='object')


r=df.groupby('state')
['sexual_harassment_workplace'].mean().reset_index().sort_values(by='s
exual_harassment_workplace',ascending=False)
r=r.head(10)
r
```


```
{"summary":"{\n  \"name\": \"r\",\n  \"rows\": 10,\n  \"fields\": [\n
{\n      \"column\": \"state\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 10,\n
\"samples\": [\n          \"Chandigarh\",\n          \"Punjab\",\n
\"Puducherry\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"sexual_harassment_workplace\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 4.791170242584993,\n
\"min\": 22.44090909090909,\n        \"max\": 37.28744939271255,\n
\"num_unique_values\": 10,\n        \"samples\": [\n
26.64819944598338,\n          36.4,\n          27.176470588235293\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\n}","type":"dataframe","variable_name":"r"}
```


Workplace Harassment Cases Across States (2018–2020)

This bar plot visualizes the **total reported cases of sexual harassment in the workplace** for each state in the years **2018, 2019, and 2020**.

What does the code do?

- It calculates the **sum of 'sexual_harassment_workplace' cases** for each state and year (2018-2020).
- It then creates a **bar plot** where:
  - The **x-axis** represents each **state**.
  - The **y-axis** shows the **total number of workplace harassment cases**.
  - Different colored bars within each state represent the data for the years **2018, 2019, and 2020**, making it easy to compare yearly trends within a state.

Why this visualization?

- **State-wise comparison:** Quickly compare which states report higher or lower numbers of workplace harassment cases.
- **Yearly trend within states:** Observe if cases are increasing, decreasing, or staying stable within specific states over the three years.
- **Identification of hotspots:** Pinpoint states with consistently high numbers, indicating areas that may require more targeted interventions.

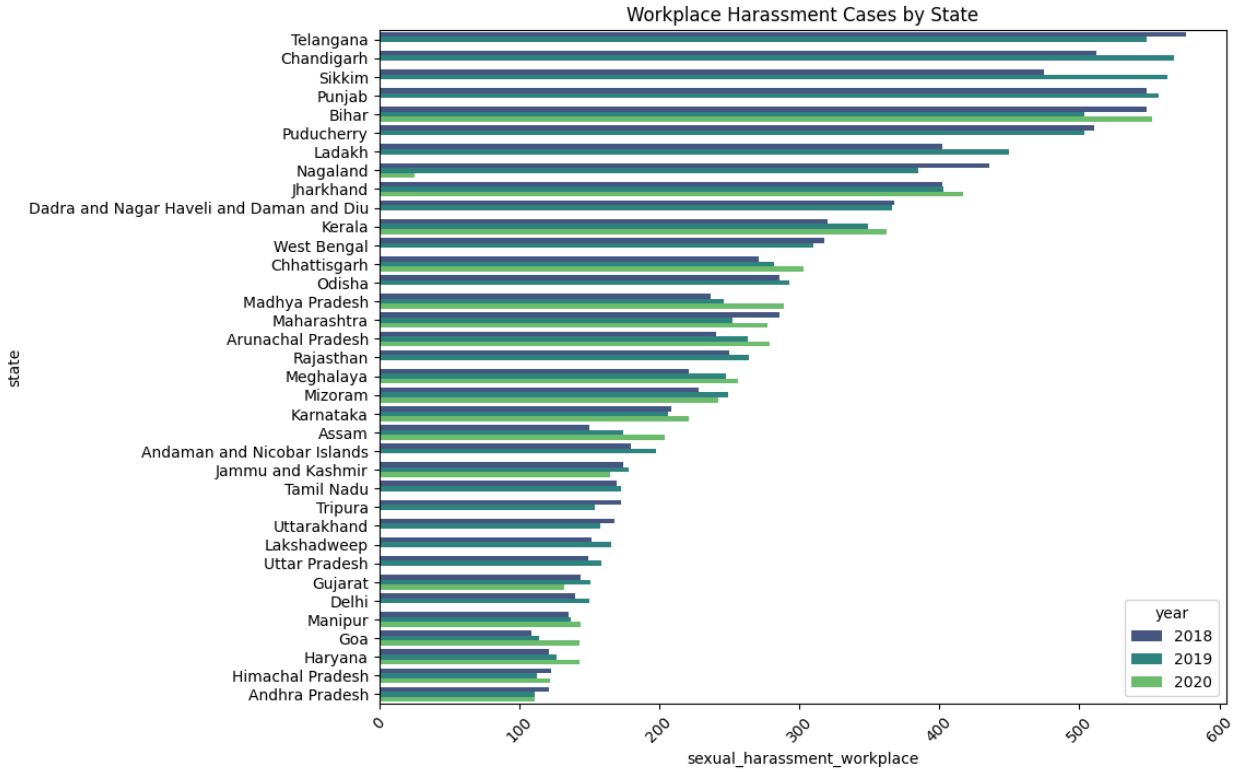
Key Insights

- The height of the bars shows the **magnitude of the problem** in each state.
- Comparing the bars for different years within a state helps understand the **short-term trend**.
- This plot can help **allocate resources** and design **awareness campaigns or policy changes** more effectively based on state-specific needs.

```
import seaborn as sns
import matplotlib.pyplot as plt

cases = df.groupby(['year',"state"])
["sexual_harassment_workplace"].sum().reset_index()
cases=cases[cases['year'].isin([2018,2019,2020])].sort_values(by='sexu
al_harassment_workplace',ascending=False)
# cases=cases.head(10)
plt.figure(figsize=(10,8))
sns.barplot(
    data=cases,
    y="state",
    x="sexual_harassment_workplace",
    hue="year",
    palette=sns.color_palette("viridis", n_colors=3)
)

plt.title("Workplace Harassment Cases by State")
plt.xticks(rotation=45)
plt.show()
```



#Finding Coorelation

```
numeric_df = df.select_dtypes(include='number')
numeric_df.head()
```

```
{"type":"dataframe","variable_name":"numeric_df"}
```

```
numeric_df.corr()
```

```
{"type":"dataframe"}
```

Relationship Between Rape Cases and Murder with Rape Cases (2001-2020)

This scatter plot visualizes the relationship between **Rape cases** and **Murder with Rape cases** across different districts for each year from 2001 to 2020.

What does the code do?

- It creates a **scatterplot** using the `numeric_df` DataFrame.
- The **x-axis** represents the number of **rape cases**.
- The **y-axis** represents the number of **murder with rape cases**.
- The **color** of each point on the scatter plot indicates the **year**, using the 'Set1' color palette to differentiate between years.

Why this visualization?

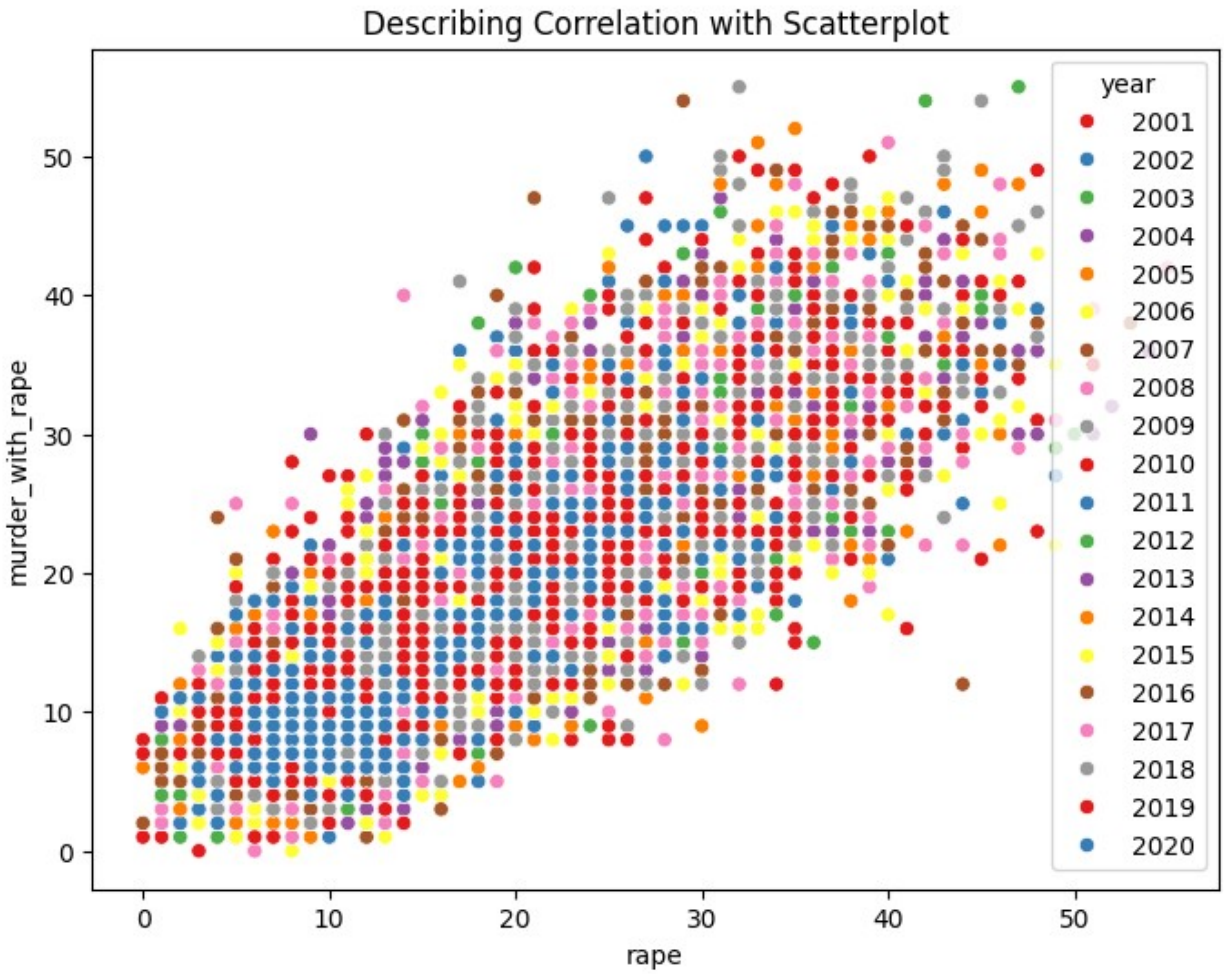
- **Visualize Correlation:** It helps to visually assess if there is a correlation between the number of rape cases and the number of murder with rape cases. A positive correlation would suggest that as rape cases increase, murder with rape cases also tend to increase.
- **Identify Trends Over Time:** By using 'year' as the hue, we can see if the relationship between these two crime types changes over the years. Different colored clusters or patterns could indicate shifts in the nature of these crimes.
- **Detect Outliers:** The scatter plot can reveal outliers, which are districts with unusually high or low numbers of either crime relative to the other.
- **Understand Distribution:** It provides a visual representation of how these two crime types are distributed across the data points (districts and years).

Key Insights

- The clustering or spread of the points can indicate the strength and direction of the relationship.
- The colors can highlight if certain years have a different distribution of cases compared to others.
- This plot is useful for understanding if these two severe crimes are linked and how this linkage might evolve over time.

```
plt.figure(figsize=(8,6))
sns.scatterplot(data=numeric_df,x='rape',y='murder_with_rape',hue='yea
r',palette='Set1')
plt.title("Describing Correlation with Scatterplot")
```

```
Text(0.5, 1.0, 'Describing Correlation with Scatterplot')
```


Describing Correlation with Scatterplot
year
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
murder_with_rape
rape
0
10
20
30
40
50