

## ▼ CRIME DATASET

In this CSV file, we have information about rape victims in India between 2000-2010, here rape victims separated into 2 groups, incest victims rape, and other victims of rape. There is information about every state in India on the same dates and their rape victims, their groups, and their age.

## ▼ Importing Libraries

As a first step, we will import all the necessary libraries that we think we will be requiring to perform the ANALYTICS.

```
# data analysis and wrangling
import pandas as pd
import numpy as np

# visualization
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
import sklearn.metrics as metrics
import plotly.express as px
from sklearn.metrics import confusion_matrix
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline
```

### Checking the Version

```
print('Numpy Version',np.__version__)
print('Pandas Version',pd.__version__)
print('Seaborn Version',sns.__version__)
print('Matplotlib Version',matplotlib.__version__)

Numpy Version 1.21.6
Pandas Version 1.3.5
Seaborn Version 0.11.2
Matplotlib Version 3.2.2
```

## ▼ Loading the data set

We will be loading the CRIME RATE DATASET using pandas.

```
# load the data
crime_path="CRIME_RATE_INDIA.csv"
df= pd.read_csv(crime_path)
df
```

	Area_Name	Year	Subgroup	Rape_Cases_Reported	Victims_Above_50_Yrs	Victims_
0	Andaman & Nicobar Islands	2001	Total Rape Victims	3	0	
1	Andaman & Nicobar Islands	2001	Victims of Incest Rape	1	0	
2	Andaman & Nicobar Islands	2001	Victims of Other Rape	2	0	
3	Andaman & Nicobar Islands	2002	Total Rape Victims	2	0	
4	Andaman & Nicobar Islands	2002	Victims of Incest Rape	0	0	
...	...	...	...	...	...	
1045	West Bengal	2009	Victims of Incest Rape	3	0	
1046	West Bengal	2009	Victims of Other Rape	2333	0	
1047	West Bengal	2010	Total Rape Victims	2311	0	
1048	West Bengal	2010	Victims of Incest Rape	4	0	
1049	West Bengal	2010	Victims of Other Rape	2307	0	

1050 rows × 11 columns



▼ Basic Data Exploration

In this step, we will perform the below operations to check what the data set comprises of. We will check the below things:

- info of the dataset
- head of the dataset
- summary of the dataset
- shape of the dataset

```
df.info()
```

#info() is used to check the Information about the data and the datatypes of each respecti

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1050 entries, 0 to 1049
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Area_Name                            1050 non-null   object
1   Year                                1050 non-null   int64
2   Subgroup                            1050 non-null   object
3   Rape_Cases_Reported                 1050 non-null   int64
4   Victims_Above_50_Yrs                1050 non-null   int64
5   Victims_Between_10-14_Yrs           1050 non-null   int64
6   Victims_Between_14-18_Yrs           1050 non-null   int64
7   Victims_Between_18-30_Yrs           1050 non-null   int64
8   Victims_Between_30-50_Yrs           1050 non-null   int64
9   Victims_of_Rape_Total                1050 non-null   int64
10  Victims_Upto_10_Yrs                 1050 non-null   int64
dtypes: int64(9), object(2)
memory usage: 90.4+ KB
```

```
df.head(10).T
```

\*\*\*head function will tell us the top records in the data set. Above head fucntion display

	0	1	2	3	4	5	
Area_Name	Andaman & Nicobar Islands	Andaman & Nicobar Islands	Andaman & Nicobar Islands	Andaman & Nicobar Islands	Andaman & Nicobar Islands	Andaman & Nicobar Islands	Andaman & Nicobar Islands
Year	2001	2001	2001	2002	2002	2002	2002
Subgroup	Total Rape	Victims of Incest	Victims of Other	Total Rape	Victims of Incest	Victims of Other	Victims of Other

```
df.describe(include='all').T
# The describe method will help to see how data has been spread for the numerical values.
#We can clearly see the minimum value, mean values, different percentile values and maximum values
```

	count	unique	top	freq	mean	std	min	max
Area_Name	1050	35	Andaman & Nicobar Islands	30	NaN	NaN	NaN	NaN
Year	1050.0	NaN	NaN	NaN	2005.5	2.87365	2001.0	2002.0
Subgroup	1050	3	Total Rape Victims	350	NaN	NaN	NaN	NaN
Rape_Cases_Reported	1050.0	NaN	NaN	NaN	361.92	592.180572	0.0	1050.0
Victims_Above_50_Yrs	1050.0	NaN	NaN	NaN	1.866667	4.640286	0.0	1050.0
Victims_Between_10-14_Yrs	1050.0	NaN	NaN	NaN	23.657143	50.677418	0.0	1050.0
Victims_Between_14-18_Yrs	1050.0	NaN	NaN	NaN	53.085714	115.127899	0.0	1050.0
Victims_Between_18-50_Yrs	1050.0	NaN	NaN	NaN	1.866667	4.640286	0.0	1050.0

```
print('The number of rows (observations) is',df.shape[0],'\n')
#Shape attribute tells us number of observations and variables we have in the data set.
#It is used to check the dimension of data.
```

```
The number of rows (observations) is 1050
The number of columns (variables) is 11
```

Checking for null value

```
df.isna().sum()
```

Area_Name	0
Year	0
Subgroup	0
Rape_Cases_Reported	0
Victims_Above_50_Yrs	0
Victims_Between_10-14_Yrs	0

```

Victims_Between_14-18_Yrs    0
Victims_Between_18-30_Yrs    0
Victims_Between_30-50_Yrs    0
Victims_of_Rape_Total         0
Victims_Upto_10_Yrs          0
dtype: int64

```

## ▼ DATA ANALYSIS


### ▼ Total case Yearwise

As we see in the graphs below, the number of rape victims is growing every year, and that makes us afraid about what will happen in the future, Is these numbers will stay growing or what?

```

total_case_yearwise= pd.DataFrame(df.groupby(['Year'])['Rape_Cases_Reported'].sum().reset_
total_case_yearwise

```

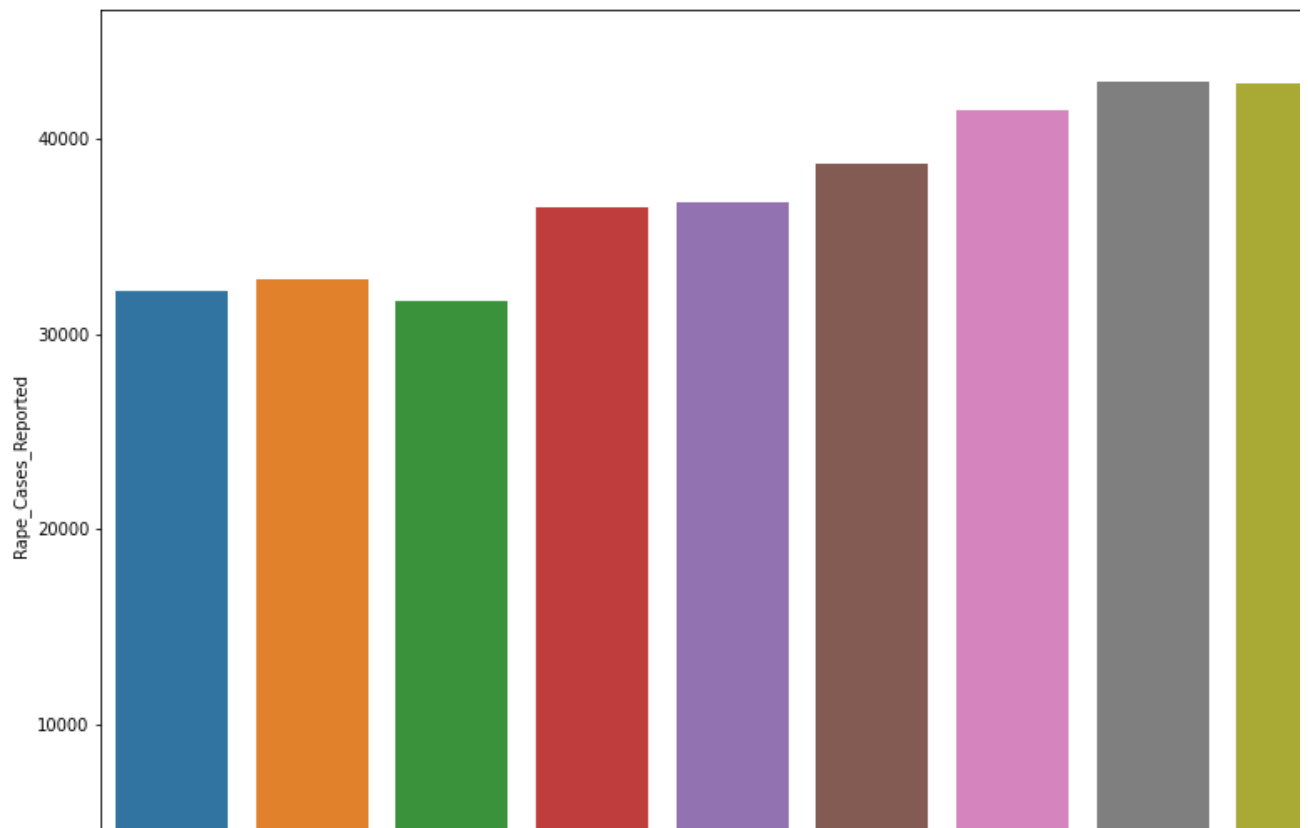
	Year	Rape_Cases_Reported	
0	2001	32150	
1	2002	32746	
2	2003	31694	
3	2004	36466	
4	2005	36718	
5	2006	38696	
6	2007	41474	
7	2008	42934	
8	2009	42794	
9	2010	44344	

```

plt.figure(figsize=(15, 10))
sns.barplot(x="Year", y="Rape_Cases_Reported", data=total_case_yearwise)

```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324e053310>




### ▼ Rape victims in each state

As we see in the graphs below, each state and its groups of victims we want to see the number of rape victims in the 2 groups. between 2000-2010 there is 190K reported cases to police, most of these reported cases were in Madhya Pradesh, West Bengal, Uttar Pradesh, and Maharashtra. And in Lakshadweep, Daman & Diu, Dadra & Nagar Haveli, and Puducherry, we saw low victims number, that is meaning in these states we have aware communities of the rape harm on women, not likes the other cites that have large victims of rape because reasons we don't know, maybe the high population and non-education in these states, or poverty, Actually, we can't know the real reasons of these high numbers until we do a careful study about these communities.

```
print('total number of victims: ',df.groupby('Area_Name')['Victims_of_Rape_Total'].sum()).s
total number of victims: 190154.0

case_reported= pd.DataFrame(df.groupby(['Area_Name'])['Rape_Cases_Reported'].sum().reset_i
case_reported
```

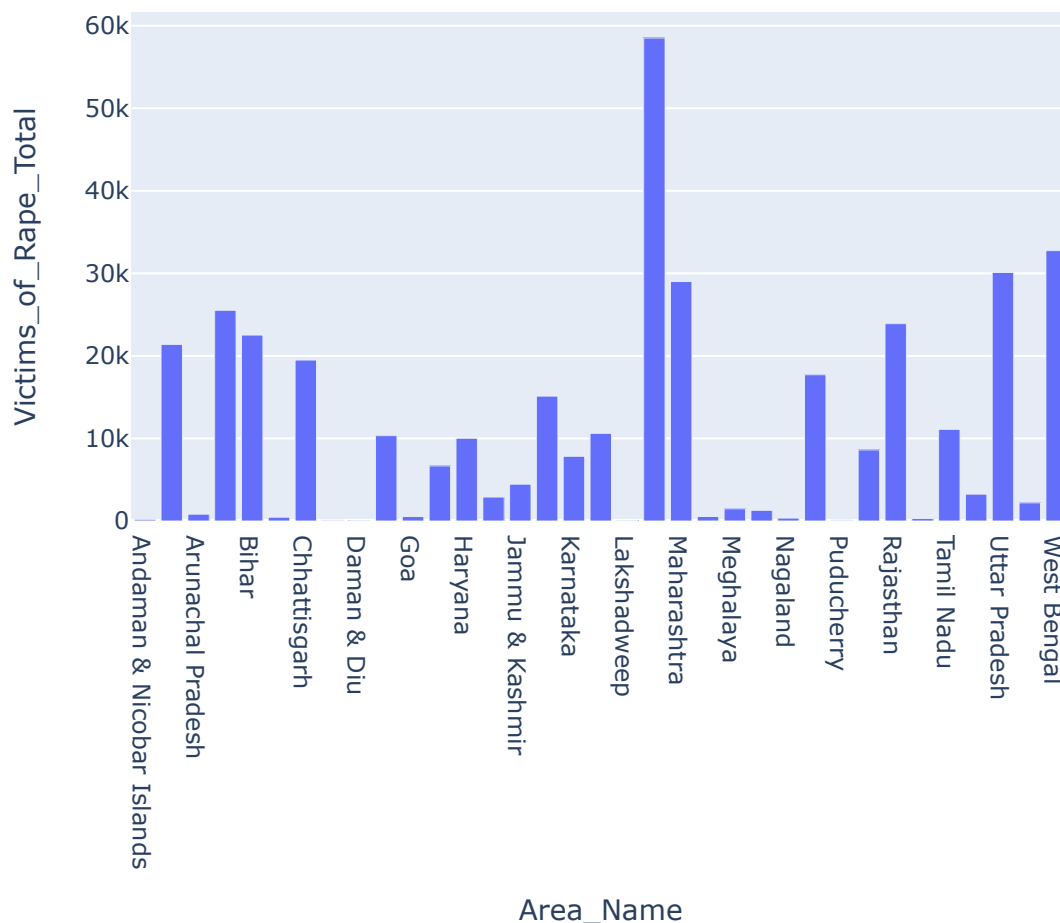
	Area_Name	Rape_Cases_Reported	
0	Andaman & Nicobar Islands	168	
1	Andhra Pradesh	21392	
2	Arunachal Pradesh	824	
3	Assam	25524	
4	Bihar	22526	
5	Chandigarh	454	
6	Chhattisgarh	19502	
7	Dadra & Nagar Haveli	98	
8	Daman & Diu	28	
9	Delhi	10292	
10	Goa	532	
11	Gujarat	6674	
12	Haryana	10018	
13	Himachal Pradesh	2906	
14	Jammu & Kashmir	4460	
15	Jharkhand	15126	
16	Karnataka	7834	
17	Kerala	10592	
18	Lakshadweep	14	
19	Madhya Pradesh	58512	
20	Maharashtra	28892	
21	Manipur	542	
22	Meghalaya	1452	
23	Mizoram	1292	
24	Nagaland	352	
25	Odisha	17720	
26	Puducherry	114	
27	Punjab	8600	
28	Rajasthan	23898	
29	Sikkim	290	
30	Tamil Nadu	11076	

```

a=df.groupby(['Area_Name', 'Year'])['Victims_of_Rape_Total'].sum().reset_index()
c= a.groupby(['Area_Name'])['Victims_of_Rape_Total'].sum().reset_index()

```

```
fig = px.bar(c, x = 'Area_Name', y = 'Victims_of_Rape_Total')
fig.show()
```



## ▼ Distribution of cases according to age

As we see in the pie chart, its the distribution of Victims according to their ages.

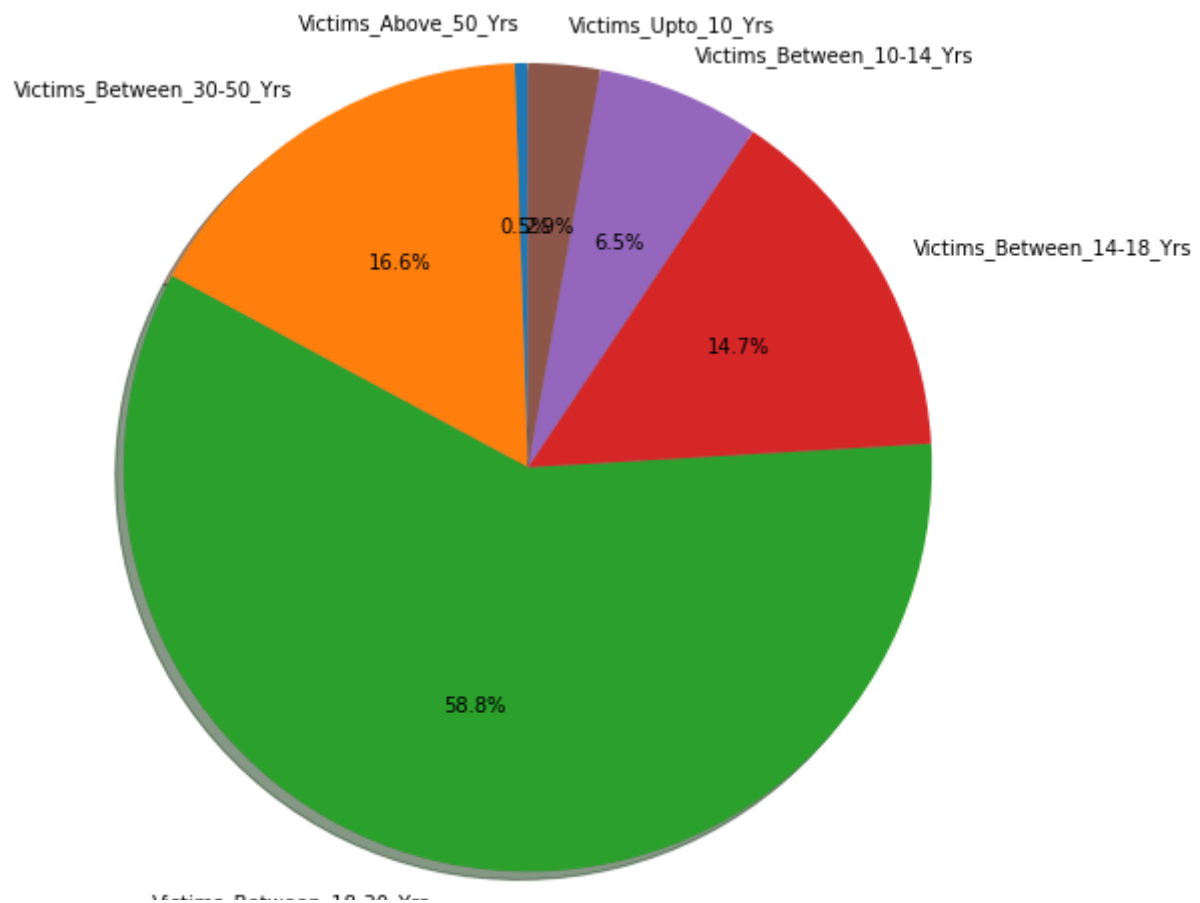
```
Victims_above_50 = df['Victims_Above_50_Yrs'].sum()
Victims_30_to_50 = df['Victims_Between_30-50_Yrs'].sum()
Victims_18_to_38 = df['Victims_Between_18-30_Yrs'].sum()
Victims_14_to_18 = df['Victims_Between_14-18_Yrs'].sum()
Victims_10_to_14 = df['Victims_Between_10-14_Yrs'].sum()
Victims_upto_10 = df['Victims_Upto_10_Yrs'].sum()

Age=['Victims_Above_50_Yrs','Victims_Between_30-50_Yrs','Victims_Between_18-30_Yrs','Victi
    'Victims_Upto_10_Yrs']
SUM=[Victims_above_50,Victims_30_to_50,Victims_18_to_38,Victims_14_to_18,Victims_10_to_14,'

fig1, ax1 = plt.subplots(figsize=(8,8))
ax1.pie(SUM, labels=Age, autopct='%1.1f%%',
        shadow=True, startangle=90)
ax1.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
```



```
plt.show()
```



## ▼ Age group wise distribution of victims

### Victims Above 50 Yrs

```
plt.figure(figsize=(10,5))
sns.barplot(x="Subgroup", y="Victims_Above_50_Yrs", data=df)
```

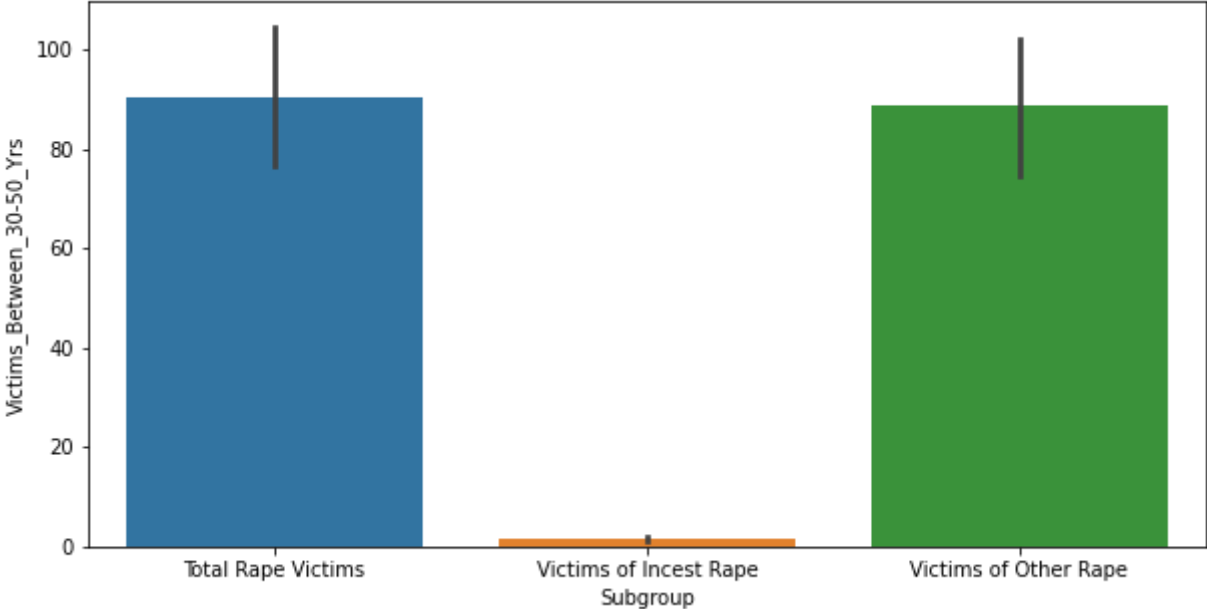
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324ce55610>

Victims Between 30-50 Yrs

30 | | |

```
plt.figure(figsize=(10,5))
sns.barplot(x="Subgroup", y="Victims_Between_30-50_Yrs", data=df)
```

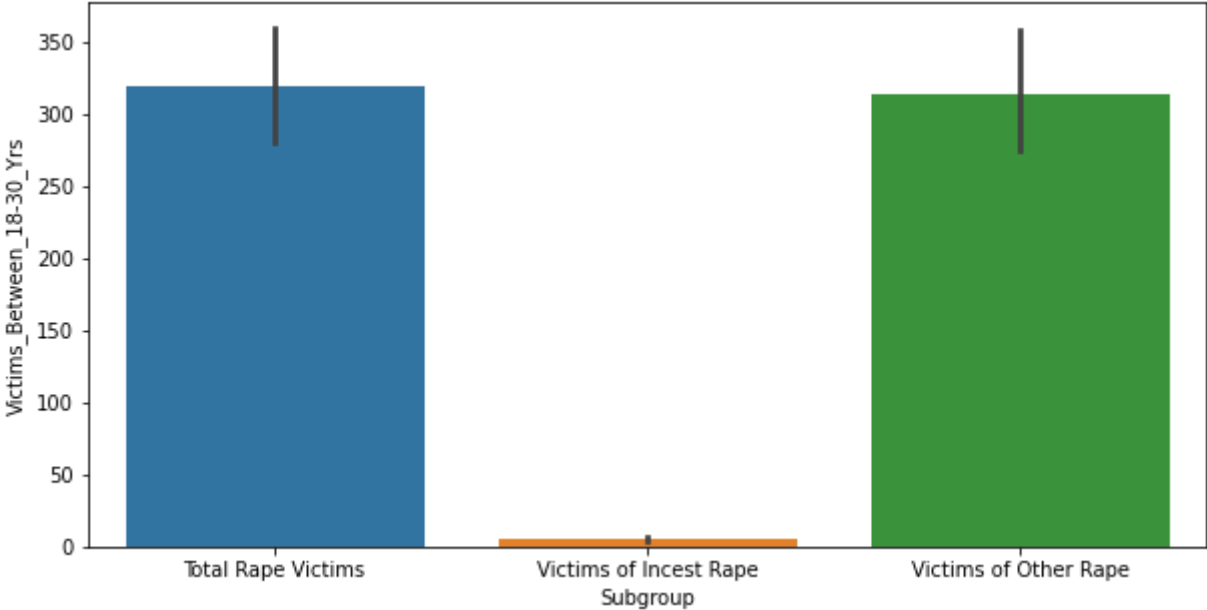
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324d214450>



Victims Between 18-30 Yrs

```
plt.figure(figsize=(10,5))
sns.barplot(x="Subgroup", y="Victims_Between_18-30_Yrs", data=df)
```

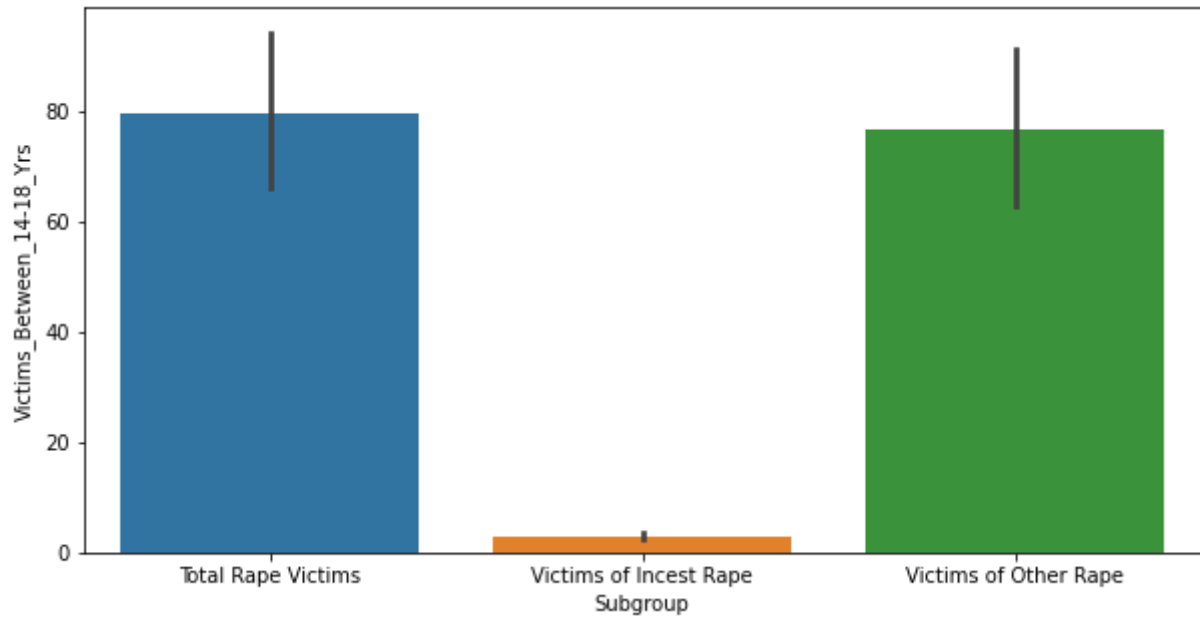
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324cd9dc50>



Victims Between 14-18 Yrs

```
plt.figure(figsize=(10,5))  
sns.barplot(x="Subgroup", y="Victims_Between_14-18_Yrs", data=df)
```

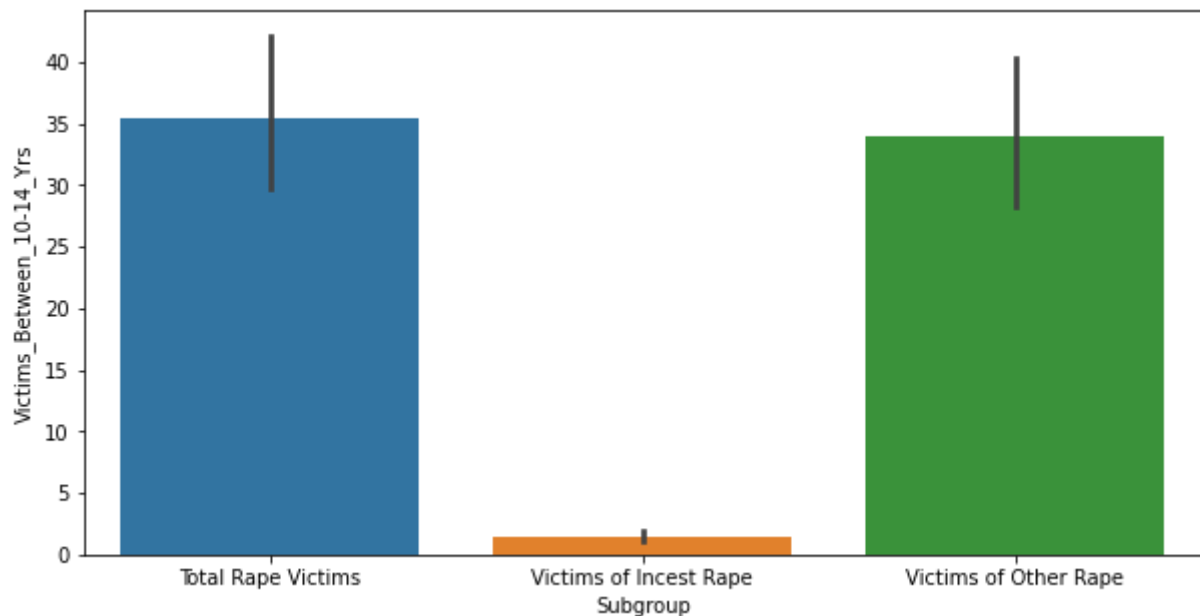
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324cda7ad0>



### Victims Between 10-14 Yrs

```
plt.figure(figsize=(10,5))  
sns.barplot(x="Subgroup", y="Victims_Between_10-14_Yrs", data=df)
```

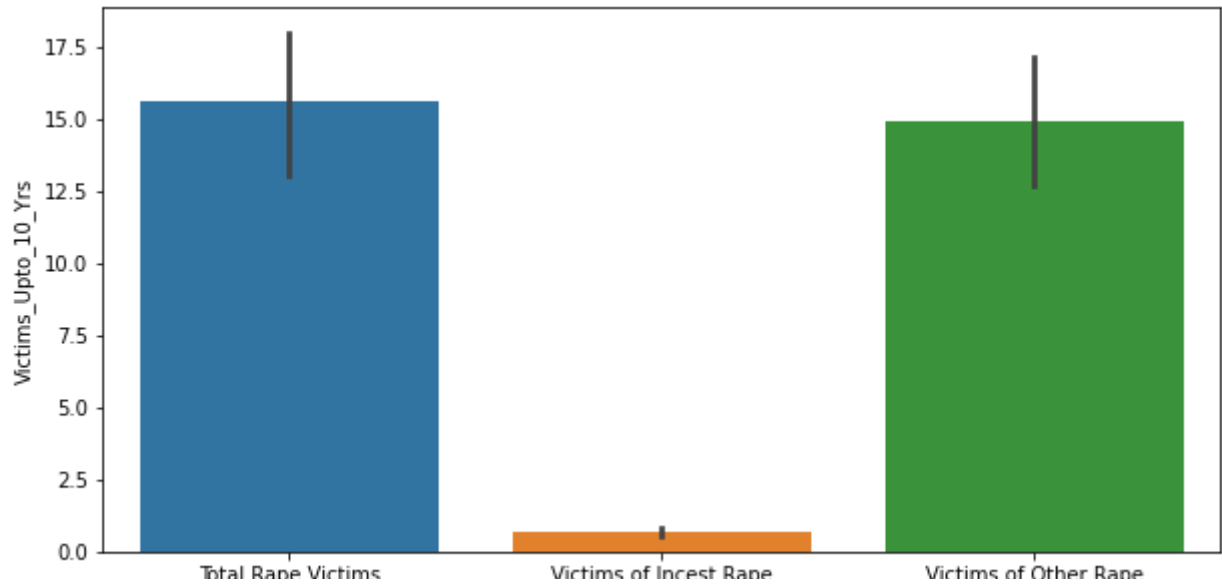
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324cd08450>



### Victims Between upto 10 Yrs

```
plt.figure(figsize=(10,5))  
sns.barplot(x="Subgroup", y="Victims_Upto_10_Yrs", data=df)
```

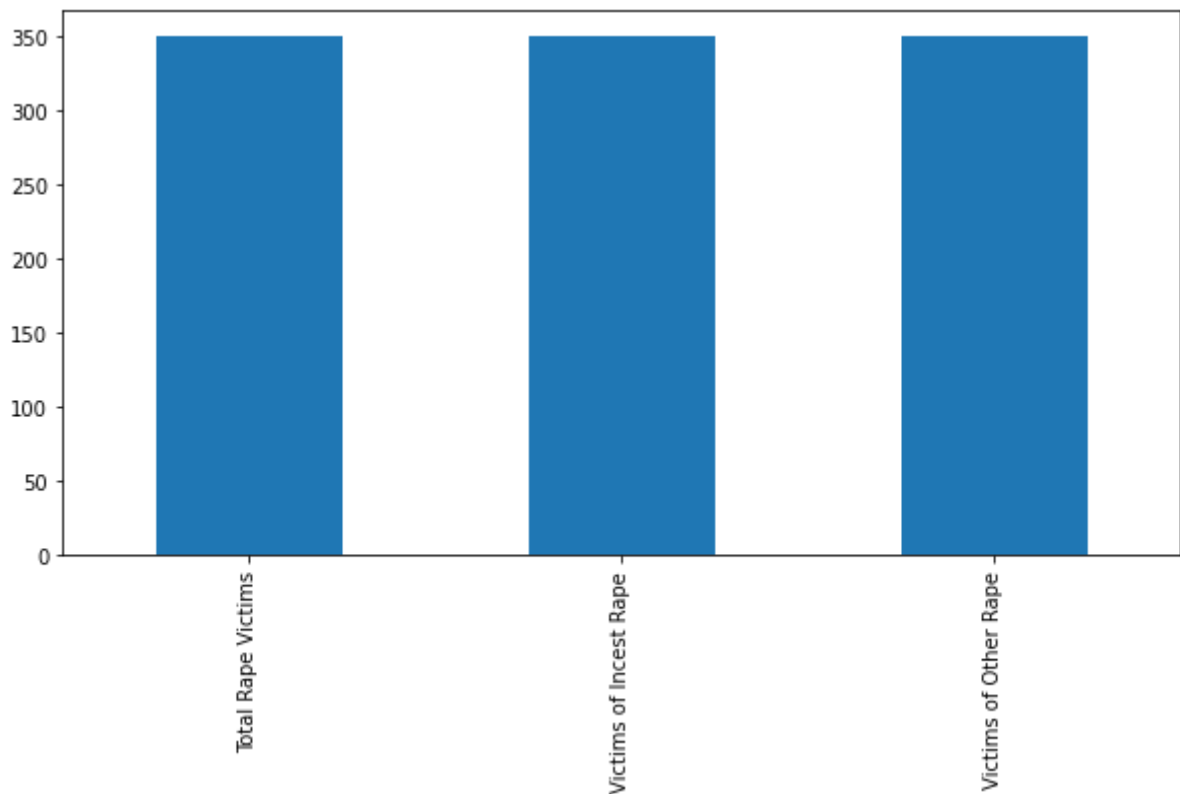
<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324cbe6e10>



### ▼ Total number of cases per Sub-category

```
plt.figure(figsize=(10,5))
pd.value_counts(df['Subgroup']).plot.bar()
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7f324cb4df10>



### ▼ Victims of incest rape

```
ab=df[df['Subgroup']=='Victims of Incest Rape']
bc=df[df['Subgroup']=='Victims of Other Rape']
```

```
fig = plt.figure()
```

```
ax = fig.add_axes([0,0,1,1])
```

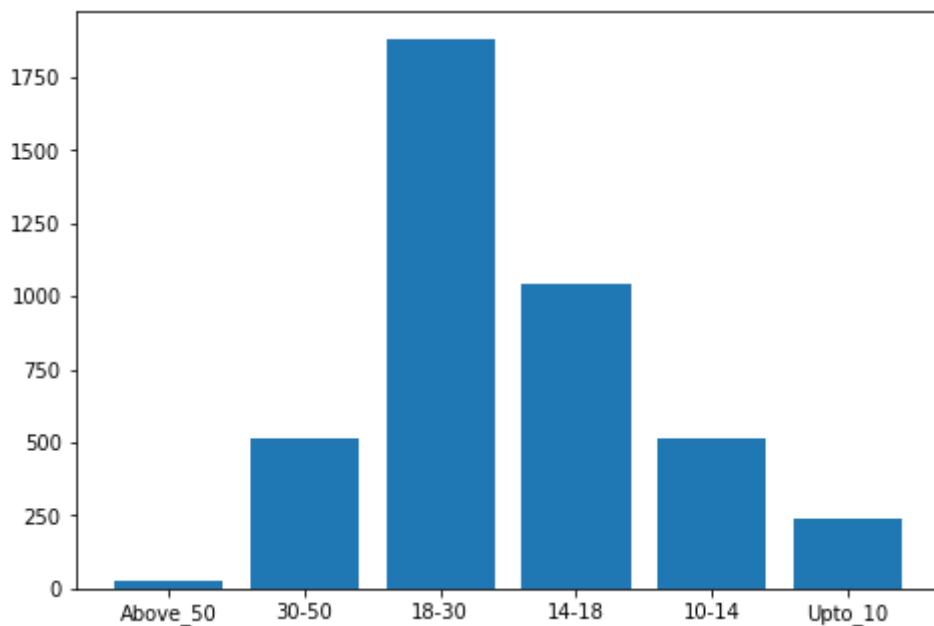
```
Victims_above_50 = ab['Victims_Above_50_Yrs'].sum()
Victims_30_to_50 = ab['Victims_Between_30-50_Yrs'].sum()
Victims_18_to_38 = ab['Victims_Between_18-30_Yrs'].sum()
Victims_14_to_18 = ab['Victims_Between_14-18_Yrs'].sum()
Victims_10_to_14 = ab['Victims_Between_10-14_Yrs'].sum()
Victims_upto_10 = ab['Victims_Upto_10_Yrs'].sum()
```

```
Range=['Above_50','30-50','18-30','14-18','10-14','Upto_10']
```

```
Total=[Victims_above_50,Victims_30_to_50,Victims_18_to_38,Victims_14_to_18,Victims_10_to_1
```

```
ax.bar(Range,Total)
```

```
plt.show()
```



## Victims of other rape

```
fig = plt.figure()
```

```
ax = fig.add_axes([0,0,1,1])
```

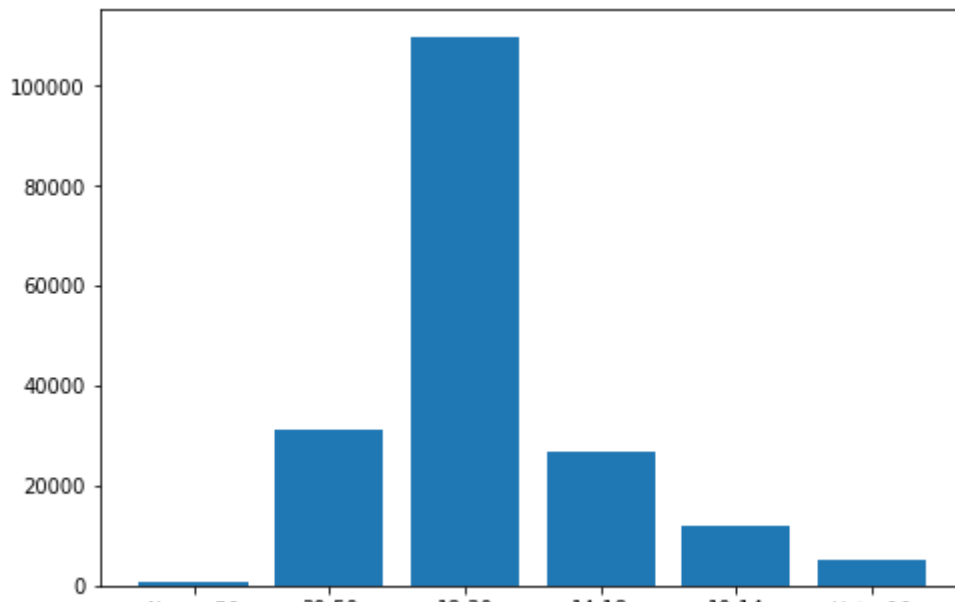
```
Victims_above_50 = bc['Victims_Above_50_Yrs'].sum()
Victims_30_to_50 = bc['Victims_Between_30-50_Yrs'].sum()
Victims_18_to_38 = bc['Victims_Between_18-30_Yrs'].sum()
Victims_14_to_18 = bc['Victims_Between_14-18_Yrs'].sum()
Victims_10_to_14 = bc['Victims_Between_10-14_Yrs'].sum()
Victims_upto_10 = bc['Victims_Upto_10_Yrs'].sum()
```

```
Range=['Above_50','30-50','18-30','14-18','10-14','Upto_10']
```

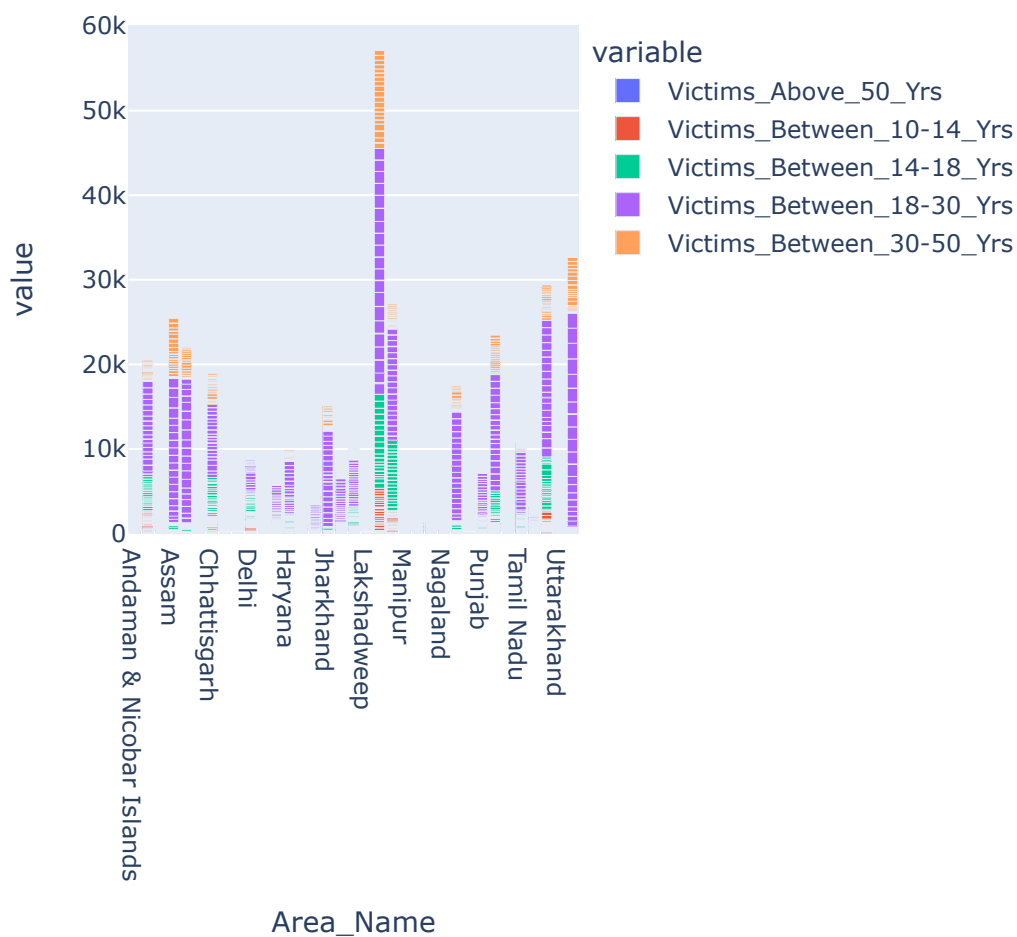
```
Total=[Victims_above_50,Victims_30_to_50,Victims_18_to_38,Victims_14_to_18,Victims_10_to_1
```

```
ax.bar(Range,Total)
```

```
plt.show()
```



```
fig = px.bar(df, x = 'Area_Name', y = ['Victims_Above_50_Yrs', 'Victims_Between_10-14_Yrs',
fig.show()
```



Finally, we finished our case study about rape victims in India between 2000-2010, we saw there are states that have huge numbers than others in rape victims and this number is growing every year. If India didn't do anything, that means it will be a disaster.

**The woman is half of the community, and we have to respect them and save them, they are our moms, sisters, daughter and our wives. Everyone in India has to move to change this situation, we have to save our women in India and in every country around the world.**

---

✓ 0s completed at 2:30 PM

