Unsupervised ML (Clustering)

CONSIDERING THE DATABASES THAT CONTAINS THE DISTRICTS !!

In [1]:
```
import pandas as pd
import warnings
warnings.filterwarnings('ignore')

district_wise=pd.read_csv(r"D:\vineet\012_District_wise_crimes_committed_IPC_20
```

In [2]:
```
district_wise
```

Out[2]:

	STATE/UT	DISTRICT	Year	MURDER	ATTEMPT TO MURDER	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	RAPE	CUSTODI RA
0	ANDHRA PRADESH	ADILABAD	2001	101	60	17	50	
1	ANDHRA PRADESH	ANANTAPUR	2001	151	125	1	23	
2	ANDHRA PRADESH	CHITTOOR	2001	101	57	2	27	
3	ANDHRA PRADESH	CUDDAPAH	2001	80	53	1	20	
4	ANDHRA PRADESH	EAST GODAVARI	2001	82	67	1	23	
...	...	...	...	...	...	...	...	
9012	LAKSHADWEEP	LAKSHADWEEP	2012	0	0	0	0	
9013	LAKSHADWEEP	TOTAL	2012	0	0	0	0	
9014	PUDUCHERRY	KARAIKAL	2012	5	6	2	6	
9015	PUDUCHERRY	PUDUCHERRY	2012	24	21	10	7	
9016	PUDUCHERRY	TOTAL	2012	29	27	12	13	

9017 rows × 33 columns

LET'S DROP SOME COLUMNS WHICH ARE NOT REQUIRED FOR OUR CLUSTERING ANALYSIS!!

In [3]:
```
district_wise=district_wise[['STATE/UT',
 'DISTRICT',
 'Year',
 'MURDER',
 'RAPE',
 'KIDNAPPING & ABDUCTION',
 'DACOITY',
 'ROBBERY',
 'BURGLARY',
 'RIOTS',
 'COUNTERFIETING',
 'ARSON',
 'HURT/GREVIOUS HURT']]
```

In [4]:
```
for i in range(len(district_wise['STATE/UT'])):
    if(district_wise['STATE/UT'][i]=='DELHI UT'):
        district_wise['STATE/UT'][i]='DELHI'
```

In [5]:
```
district_wise['STATE/UT'].unique()
```

Out[5]:
```
array(['ANDHRA PRADESH', 'ARUNACHAL PRADESH', 'ASSAM', 'BIHAR',
       'CHHATTISGARH', 'GOA', 'GUJARAT', 'HARYANA', 'HIMACHAL PRADESH',
       'JAMMU & KASHMIR', 'JHARKHAND', 'KARNATAKA', 'KERALA',
       'MADHYA PRADESH', 'MAHARASHTRA', 'MANIPUR', 'MEGHALAYA', 'MIZORAM',
       'NAGALAND', 'ODISHA', 'PUNJAB', 'RAJASTHAN', 'SIKKIM',
       'TAMIL NADU', 'TRIPURA', 'UTTAR PRADESH', 'UTTARAKHAND',
       'WEST BENGAL', 'A & N ISLANDS', 'CHANDIGARH', 'D & N HAVELI',
       'DAMAN & DIU', 'DELHI', 'LAKSHADWEEP', 'PUDUCHERRY'], dtype=object)
```

In [6]:
```
district_wise=district_wise[district_wise['DISTRICT']!='TOTAL']
```

In [7]:
```
district_wise_sc=pd.read_csv(r"D:\vineet\02_01_District_wise_crimes_committed_a
district_wise_sc
```

Out[7]:

	STATE/UT	DISTRICT	Year	Murder	Rape	Kidnapping and Abduction	Dacoity	Robbery	Arso
0	ANDHRA PRADESH	ADILABAD	2001	0	1	4	0	0	
1	ANDHRA PRADESH	ANANTAPUR	2001	0	4	0	0	0	
2	ANDHRA PRADESH	CHITTOOR	2001	3	3	0	0	0	
3	ANDHRA PRADESH	CUDDAPAH	2001	0	3	0	0	0	
4	ANDHRA PRADESH	EAST GODAVARI	2001	1	3	0	0	0	
...	...	...	...	...	...	...	...	...	.
9013	LAKSHADWEEP	LAKSHADWEEP	2012	0	0	0	0	0	
9014	LAKSHADWEEP	TOTAL	2012	0	0	0	0	0	
9015	PUDUCHERRY	KARAIKAL	2012	1	0	0	0	0	
9016	PUDUCHERRY	PUDUCHERRY	2012	1	0	0	0	0	
9017	PUDUCHERRY	TOTAL	2012	2	0	0	0	0	

9018 rows × 13 columns

In [8]:
```
district_wise_sc['STATE/UT'].unique()
```

Out[8]:
```
array(['ANDHRA PRADESH', 'ARUNACHAL PRADESH', 'ASSAM', 'BIHAR',
       'CHHATTISGARH', 'GOA', 'GUJARAT', 'HARYANA', 'HIMACHAL PRADESH',
       'JAMMU & KASHMIR', 'JHARKHAND', 'KARNATAKA', 'KERALA',
       'MADHYA PRADESH', 'MAHARASHTRA', 'MANIPUR', 'MEGHALAYA', 'MIZORAM',
       'NAGALAND', 'ODISHA', 'PUNJAB', 'RAJASTHAN', 'SIKKIM',
       'TAMIL NADU', 'TRIPURA', 'UTTAR PRADESH', 'UTTARAKHAND',
       'WEST BENGAL', 'A & N ISLANDS', 'CHANDIGARH', 'D & N HAVELI',
       'DAMAN & DIU', 'DELHI', 'LAKSHADWEEP', 'PUDUCHERRY'], dtype=object)
```

In [9]:
```
district_wise_sc['TOTAL CRIMES AGAINST SC']=district_wise_sc.iloc[:,3:].sum(axi
```

In [10]:
```
district_wise_sc_Total=district_wise_sc[['STATE/UT','DISTRICT','Year','TOTAL CR
```

In [11]:
```
district_wise_sc_Total
```

Out[11]:

	STATE/UT	DISTRICT	Year	TOTAL CRIMES AGAINST SC
0	ANDHRA PRADESH	ADILABAD	2001	55
1	ANDHRA PRADESH	ANANTAPUR	2001	127
2	ANDHRA PRADESH	CHITTOOR	2001	114
3	ANDHRA PRADESH	CUDDAPAH	2001	100
4	ANDHRA PRADESH	EAST GODAVARI	2001	89
...	...	...	...	...
9013	LAKSHADWEEP	LAKSHADWEEP	2012	0
9014	LAKSHADWEEP	TOTAL	2012	0
9015	PUDUCHERRY	KARAIKAL	2012	1
9016	PUDUCHERRY	PUDUCHERRY	2012	23
9017	PUDUCHERRY	TOTAL	2012	24

9018 rows × 4 columns

In [13]:
```
district_wise_sc_Total=district_wise_sc_Total[district_wise_sc_Total['DISTRICT'
```

In [14]:
```
district_wise_st=pd.read_csv(r"D:\vineet\02_District_wise_crimes_committed_agai
```

In [15]:
```
district_wise_st
```

Out[15]:

	STATE/UT	DISTRICT	Year	Murder	Rape	Kidnapping Abduction	Dacoity	Robbery	Arso
0	ANDHRA PRADESH	ADILABAD	2001	0	1	2	0	0	
1	ANDHRA PRADESH	ANANTAPUR	2001	0	0	0	0	0	
2	ANDHRA PRADESH	CHITTOOR	2001	0	0	0	0	0	
3	ANDHRA PRADESH	CUDDAPAH	2001	0	0	0	0	0	
4	ANDHRA PRADESH	EAST GODAVARI	2001	0	0	0	0	0	
...	...	...	...	...	...	...	...	...	.
9013	LAKSHADWEEP	LAKSHADWEEP	2012	0	0	0	0	0	
9014	LAKSHADWEEP	TOTAL	2012	0	0	0	0	0	
9015	PUDUCHERRY	KARAIKAL	2012	0	0	0	0	0	
9016	PUDUCHERRY	PUDUCHERRY	2012	0	0	0	0	0	
9017	PUDUCHERRY	TOTAL	2012	0	0	0	0	0	

9018 rows × 13 columns

In [28]:
```
district_wise_st['STATE/UT'].unique()
```

Out[28]:
```
array(['ANDHRA PRADESH', 'ARUNACHAL PRADESH', 'ASSAM', 'BIHAR',
       'CHHATTISGARH', 'GOA', 'GUJARAT', 'HARYANA', 'HIMACHAL PRADESH',
       'JAMMU & KASHMIR', 'JHARKHAND', 'KARNATAKA', 'KERALA',
       'MADHYA PRADESH', 'MAHARASHTRA', 'MANIPUR', 'MEGHALAYA', 'MIZORAM',
       'NAGALAND', 'ODISHA', 'PUNJAB', 'RAJASTHAN', 'SIKKIM',
       'TAMIL NADU', 'TRIPURA', 'UTTAR PRADESH', 'UTTARAKHAND',
       'WEST BENGAL', 'A & N ISLANDS', 'CHANDIGARH', 'D & N HAVELI',
       'DAMAN & DIU', 'DELHI', 'LAKSHADWEEP', 'PUDUCHERRY'], dtype=object)
```

In [29]:
```
district_wise_st=district_wise_st[district_wise_st['DISTRICT']!='TOTAL']
```

In [30]:
```
district_wise_st['TOTAL CRIMES AGAINST ST']=district_wise_st.iloc[:,3:].sum(axi
```

In [31]:
```
district_wise_st_Total=district_wise_st[['STATE/UT','DISTRICT','Year','TOTAL CR
```

In [32]:
```
district_wise_st_Total
```

Out[32]:

	STATE/UT	DISTRICT	Year	TOTAL CRIMES AGAINST ST
0	ANDHRA PRADESH	ADILABAD	2001	18
1	ANDHRA PRADESH	ANANTAPUR	2001	14
2	ANDHRA PRADESH	CHITTOOR	2001	2
3	ANDHRA PRADESH	CUDDAPAH	2001	4
4	ANDHRA PRADESH	EAST GODAVARI	2001	14
...	...	...	...	...
9010	DELHI	STF	2012	0
9011	DELHI	WEST	2012	0
9013	LAKSHADWEEP	LAKSHADWEEP	2012	0
9015	PUDUCHERRY	KARAIKAL	2012	0
9016	PUDUCHERRY	PUDUCHERRY	2012	0

8597 rows × 4 columns

In [33]:
```
district_wise_child=pd.read_csv(r"D:\vineet\03_District_wise_crimes_committed_a
```

In [34]:
```
district_wise_child
```

Out[34]:

	STATE/UT	DISTRICT	Year	Murder	Rape	Kidnapping and Abduction	Foeticide	Abetment of suicide	ab
0	ANDHRA PRADESH	ADILABAD	2001	0.0	0.0	0.0	0.0	0.0	
1	ANDHRA PRADESH	ANANTAPUR	2001	19.0	12.0	29.0	0.0	6.0	
2	ANDHRA PRADESH	CHITTOOR	2001	0.0	0.0	0.0	0.0	0.0	
3	ANDHRA PRADESH	CUDDAPAH	2001	0.0	0.0	0.0	0.0	0.0	
4	ANDHRA PRADESH	EAST GODAVARI	2001	0.0	0.0	0.0	0.0	0.0	
...	...	...	...	...	...	...	...	...	
9010	LAKSHADWEEP	LAKSHADWEEP	2012	0.0	0.0	0.0	0.0	0.0	
9011	LAKSHADWEEP	TOTAL	2012	0.0	0.0	0.0	0.0	0.0	
9012	PUDUCHERRY	KARAIKAL	2012	0.0	3.0	2.0	0.0	0.0	
9013	PUDUCHERRY	PUDUCHERRY	2012	0.0	4.0	16.0	0.0	0.0	
9014	PUDUCHERRY	TOTAL	2012	0.0	7.0	18.0	0.0	0.0	

9015 rows × 15 columns

In [35]:
```
district_wise_child['STATE/UT'].unique()
```

Out[35]:
```
array(['ANDHRA PRADESH', 'ARUNACHAL PRADESH', 'ASSAM', 'BIHAR',
       'CHHATTISGARH', 'GOA', 'GUJARAT', 'HARYANA', 'HIMACHAL PRADESH',
       'JAMMU & KASHMIR', 'JHARKHAND', 'KARNATAKA', 'KERALA',
       'MADHYA PRADESH', 'MAHARASHTRA', 'MANIPUR', 'MEGHALAYA', 'MIZORAM',
       'NAGALAND', 'ODISHA', 'PUNJAB', 'RAJASTHAN', 'SIKKIM',
       'TAMIL NADU', 'TRIPURA', 'UTTAR PRADESH', 'UTTARAKHAND',
       'WEST BENGAL', 'A & N ISLANDS', 'CHANDIGARH', 'D & N HAVELI',
       'DAMAN & DIU', 'DELHI', 'LAKSHADWEEP', 'PUDUCHERRY'], dtype=object)
```

In [36]:
```
district_wise_child=district_wise_child[district_wise_child['DISTRICT']!='TOTAL
```

In [37]:
```
district_wise_child.rename(columns={"Total":"TOTAL CRIMES AGAINST CHILDREN"},ir
```

In [38]:
```
district_wise_child.columns
```

Out[38]:
```
Index(['STATE/UT', 'DISTRICT', 'Year', 'Murder', 'Rape',
       'Kidnapping and Abduction', 'Foeticide', 'Abetment of suicide',
       'Exposure and abandonment', 'Procuration of minor girls',
       'Buying of girls for prostitution', 'Selling of girls for prostitutio
n',
       'Prohibition of child marriage act', 'Other Crimes',
       'TOTAL CRIMES AGAINST CHILDREN'],
      dtype='object')
```

In [39]:
```
district_wise_child_Total=district_wise_child[['STATE/UT','DISTRICT','Year','TC
```

In [40]:
```
district_wise_child_Total
```

Out[40]:

	STATE/UT	DISTRICT	Year	TOTAL CRIMES AGAINST CHILDREN
0	ANDHRA PRADESH	ADILABAD	2001	0
1	ANDHRA PRADESH	ANANTAPUR	2001	66
2	ANDHRA PRADESH	CHITTOOR	2001	0
3	ANDHRA PRADESH	CUDDAPAH	2001	0
4	ANDHRA PRADESH	EAST GODAVARI	2001	0
...	...	...	...	...
9007	DELHI	STF	2012	0
9008	DELHI	WEST	2012	542
9010	LAKSHADWEEP	LAKSHADWEEP	2012	0
9012	PUDUCHERRY	KARAIKAL	2012	7
9013	PUDUCHERRY	PUDUCHERRY	2012	25

8603 rows × 4 columns

In [41]:
```
district_wise_women=pd.read_csv(r"D:\vineet\42_District_wise_crimes_committed_a
```

In [42]:
```
district_wise_women
```

Out[42]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	C Hus Rela
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	
1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	
2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	
3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	
4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	
...	...	...	...	...	...	...	...	...	
9012	LAKSHADWEEP	LAKSHADWEEP	2012	0	0	0	1	0	
9013	LAKSHADWEEP	TOTAL	2012	0	0	0	1	0	
9014	PUDUCHERRY	KARAIKAL	2012	6	2	0	2	0	
9015	PUDUCHERRY	PUDUCHERRY	2012	7	14	0	7	2	
9016	PUDUCHERRY	TOTAL	2012	13	16	0	9	2	

9017 rows × 10 columns

In [43]:
```
district_wise_women=district_wise_women[district_wise_women['DISTRICT']!='TOTAL
```

In [44]:
```
district_wise_women['TOTAL CRIMES AGAINST WOMEN']=district_wise_women.iloc[:,3:
```

In [45]:
```
district_wise_women_Total=district_wise_women[['STATE/UT','DISTRICT','Year','TC
```

In [46]:
```
district_wise_women_Total
```

Out[46]:

	STATE/UT	DISTRICT	Year	TOTAL CRIMES AGAINST WOMEN
0	ANDHRA PRADESH	ADILABAD	2001	454
1	ANDHRA PRADESH	ANANTAPUR	2001	356
2	ANDHRA PRADESH	CHITTOOR	2001	456
3	ANDHRA PRADESH	CUDDAPAH	2001	278
4	ANDHRA PRADESH	EAST GODAVARI	2001	475
...	...	...	...	...
9010	DELHI	STF	2012	0
9011	DELHI	WEST	2012	739
9012	LAKSHADWEEP	LAKSHADWEEP	2012	2
9014	PUDUCHERRY	KARAIKAL	2012	11
9015	PUDUCHERRY	PUDUCHERRY	2012	35

8609 rows × 4 columns

MERGING THE ABOVE DATASETS!!

In [47]:
```
cluster=pd.merge(district_wise,district_wise_sc_Total,how='inner',left_on=['STA
```

GROUPING BY STATE/UT,DISTRICT AND SUMMING ALL THE CRIMES OVER THE YEARS!!

In [48]:
```
cluster=cluster.groupby(['STATE/UT','DISTRICT']).sum().reset_index().drop(['Yea
```

In [49]:
```
cluster
```

Out[49]:

	STATE/UT	DISTRICT	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY
0	A & N ISLANDS	A and N ISLANDS	15	3	12	5	9	90
1	A & N ISLANDS	ANDAMAN	118	92	73	4	63	708
2	A & N ISLANDS	CAR	2	0	1	1	0	0
3	A & N ISLANDS	NICOBAR	12	2	1	1	0	23
4	A & N ISLANDS	NORTH	5	4	0	0	1	6
...	...	...	...	...	...	...	...	...
819	WEST BENGAL	PURULIA	733	497	276	63	278	49
820	WEST BENGAL	SEALDAH G.R.P.	188	19	67	18	115	0
821	WEST BENGAL	SILIGURI G.R.P.	42	2	31	13	15	1
822	WEST BENGAL	SILIGURI_PC	32	21	123	3	29	3
823	WEST BENGAL	UTTAR DINAJPUR	638	833	1059	123	221	84

824 rows × 16 columns

In [ ]:
```

```

4.1 CLUSTERING USING THE KMEANS!!

In [50]:
```
x=cluster.iloc[:,2:]
```

In [59]:
```
import pandas as pd
from sklearn.cluster import KMeans

kmeans=KMeans(n_clusters=3,random_state=0)
y_kmeans=kmeans.fit_predict(x)
y_kmeans
```

Out[59]:
```
array([0, 0, 0, 0, 0, 0, 1, 1, 2, 2, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 2, 0, 2, 0, 2, 2, 0, 2, 2, 2, 1, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
       0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 2, 2, 0, 2, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
       0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2,
       2, 2, 2, 0, 0, 2, 2, 0, 2, 0, 0, 0, 2, 2, 0, 0, 0, 2, 0, 2, 0, 0,
       2, 2, 0, 0, 2, 2, 0, 0, 0, 2, 0, 0, 0, 2, 2, 2, 2, 1, 0, 0, 1, 0,
       2, 2, 2, 2, 2, 0, 2, 0, 0, 0, 2, 1, 2, 0, 0, 0, 2, 0, 0, 0, 2, 2,
       0, 0, 2, 2, 2, 0, 2, 2, 0, 2, 2, 0, 2, 1, 0, 1, 0, 0, 2, 0, 0, 0,
```

In [60]:
```
kmeans.labels_
```

Out[60]:
```
array([0, 0, 0, 0, 0, 0, 1, 1, 2, 2, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 2, 0, 2, 0, 2, 2, 0, 2, 2, 2, 1, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
       0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 2, 2, 0, 2, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0,
       0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2,
       2, 2, 2, 0, 0, 2, 2, 0, 2, 0, 0, 0, 2, 2, 0, 0, 0, 2, 0, 2, 0, 0,
       2, 2, 0, 0, 2, 2, 0, 0, 0, 2, 0, 0, 0, 2, 2, 2, 2, 1, 0, 0, 1, 0,
       2, 2, 2, 2, 2, 0, 2, 0, 0, 0, 2, 1, 2, 0, 0, 0, 2, 0, 0, 0, 2, 2,
       0, 0, 2, 2, 2, 0, 2, 2, 0, 2, 2, 0, 2, 1, 0, 1, 0, 0, 2, 0, 0, 0,
```

In [66]:
```
pd.value_counts(y_kmeans)
```

Out[66]:
```
0    627
2    159
1     38
dtype: int64
```

In [63]:
```
x[y_kmeans==0]
```

Out[63]:

	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY	RIOTS	COUNTERFIETING
0	15	3	12	5	9	90	17	1
1	118	92	73	4	63	708	115	9
2	2	0	1	1	0	0	0	0
3	12	2	1	1	0	23	0	0
4	5	4	0	0	1	6	0	0
...	...	...	...	...	...	...	...	..
818	707	555	1042	92	184	156	4082	18
819	733	497	276	63	278	49	920	1
820	188	19	67	18	115	0	128	9
821	42	2	31	13	15	1	26	3
822	32	21	123	3	29	3	9	0

627 rows × 14 columns

In [64]:
```
x[y_kmeans==1]
```

Out[64]:

	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY	RIOTS	COUNTERFIETI
6	1186	672	754	72	237	2519	1016	
7	1922	266	839	137	244	2536	1293	
10	1810	641	1398	141	765	11946	559	
11	1041	499	780	30	245	5069	768	2
13	1775	526	969	91	347	3509	1204	2
15	1455	712	1466	121	942	13435	2432	3
16	1674	853	1534	91	344	3258	2011	
17	1337	721	919	63	195	2717	1667	
18	842	528	650	22	126	2392	468	
19	1529	419	785	101	215	2494	1340	

In [65]:
```
x[y_kmeans==2]
```

Out[65]:

	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY	RIOTS	COUNTERFIETING
8	1270	368	907	96	207	2777	973	171
9	1196	298	653	39	106	1549	1326	48
27	1091	338	450	66	213	3266	392	16
29	405	274	217	24	66	1437	469	53
31	213	134	374	18	204	3101	62	57
...	...	...	...	...	...	...	...	..
753	1828	689	2914	186	1532	6663	2909	274
794	2329	1883	3311	248	739	387	5484	129
795	2114	2221	2209	194	392	556	8893	41
809	1345	1777	1853	132	477	507	791	60
823	638	833	1059	123	221	84	683	67

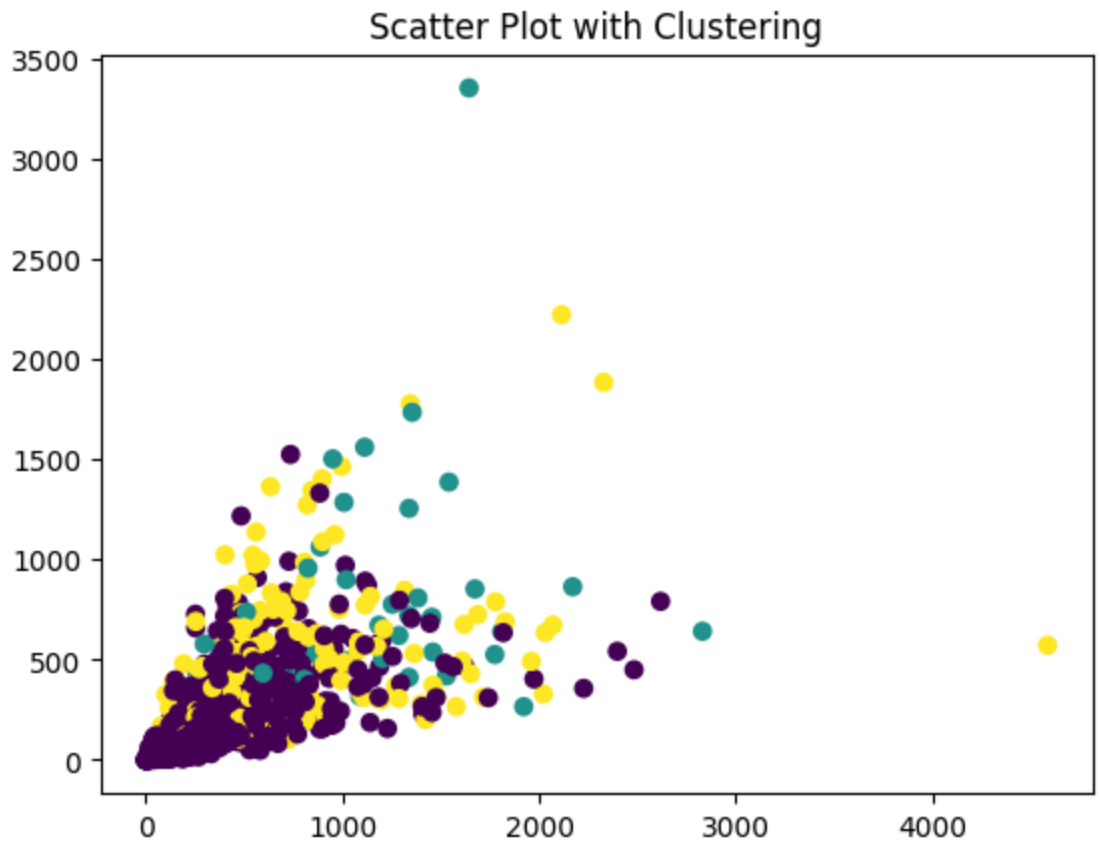
159 rows × 14 columns

In [62]:
```
import matplotlib.pyplot as plt

plt.scatter(x.values[:, 0], x.values[:, 1], c=kmeans.labels_, cmap='viridis')

plt.title('Scatter Plot with Clustering')

plt.show()
```



4.2 Create DataFrame for each cluster that shows data according to the areas.

REGION 1) Peaceful Area's

In [67]:
```
peaceful_areas=cluster[y_kmeans==0]
```

In [69]:
```
peaceful_areas.head(5)
```

Out[69]:

	STATE/UT	DISTRICT	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY	RIOT
0	A & N ISLANDS	A and N ISLANDS	15	3	12	5	9	90	
1	A & N ISLANDS	ANDAMAN	118	92	73	4	63	708	1
2	A & N ISLANDS	CAR	2	0	1	1	0	0	
3	A & N ISLANDS	NICOBAR	12	2	1	1	0	23	
4	A & N ISLANDS	NORTH	5	4	0	0	1	6	

Region 2. Moderate Area's

In [88]:
```
Moderate_Area = cluster[y_kmeans==1]
Moderate_Area.head(5)
```

Out[88]:

	STATE/UT	DISTRICT	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY	R
6	ANDHRA PRADESH	ADILABAD	1186	672	754	72	237	2519	
7	ANDHRA PRADESH	ANANTAPUR	1922	266	839	137	244	2536	
10	ANDHRA PRADESH	CYBERABAD	1810	641	1398	141	765	11946	
11	ANDHRA PRADESH	EAST GODAVARI	1041	499	780	30	245	5069	
13	ANDHRA PRADESH	GUNTUR	1775	526	969	91	347	3509	

Region 3. Vulnerable Area's

In [89]:
```
Vulnerable_area = cluster[y_kmeans==2]
Vulnerable_area.head(5)
```

Out[89]:

	STATE/UT	DISTRICT	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY
8	ANDHRA PRADESH	CHITTOOR	1270	368	907	96	207	2777
9	ANDHRA PRADESH	CUDDAPAH	1196	298	653	39	106	1549
27	ANDHRA PRADESH	RANGA REDDY	1091	338	450	66	213	3266
29	ANDHRA PRADESH	SRIKAKULAM	405	274	217	24	66	1437
31	ANDHRA PRADESH	VIJAYAWADA	213	134	374	18	204	3101

4.3.3 "Analyse your clusters and prepare a report that explains all your observations

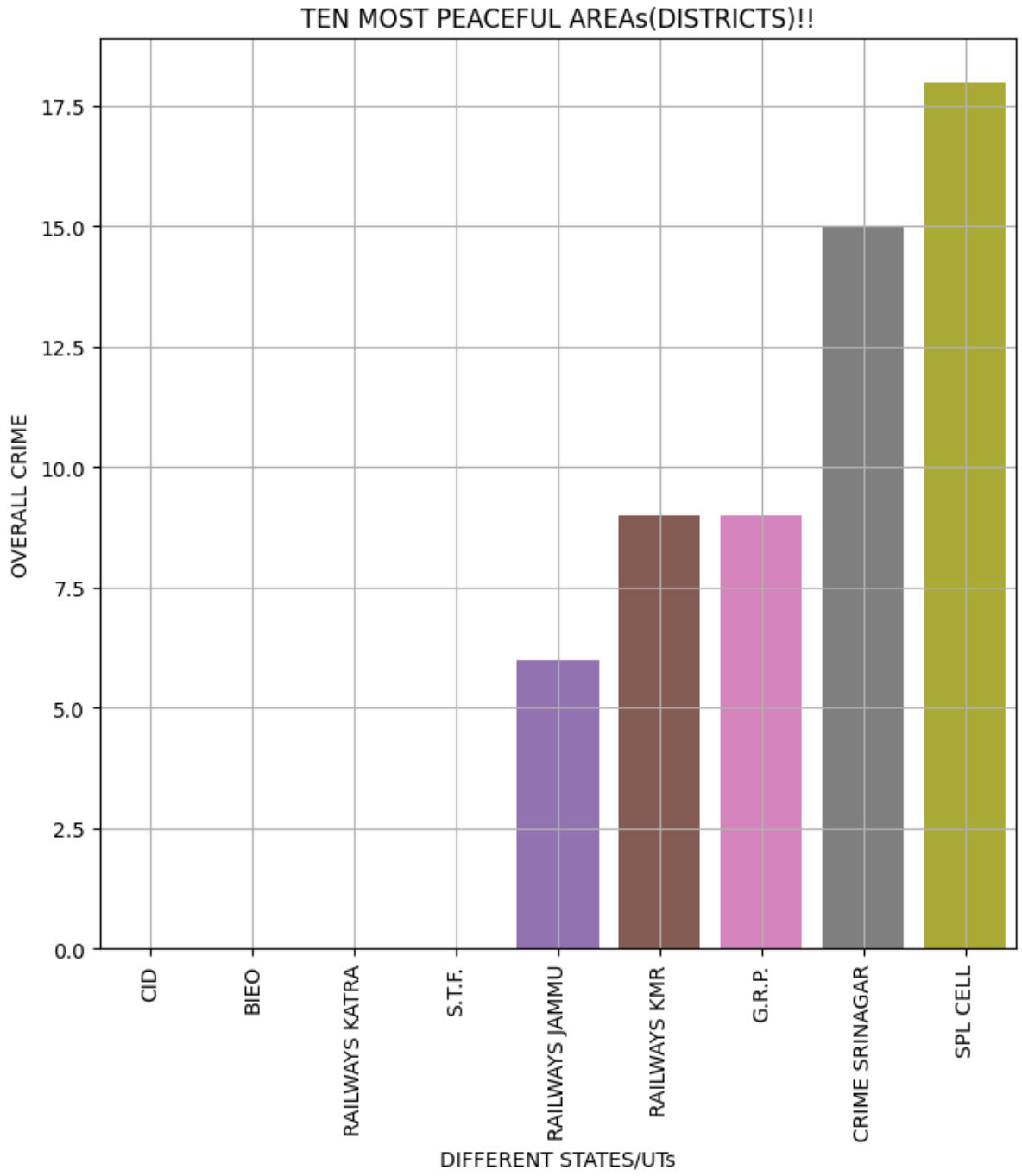
10 MOST SAFE AREAS.

In [90]:
```
peaceful_areas['OVERALL CRIME']=peaceful_areas.iloc[:,2:].sum(axis=1)
peaceful_areas.sort_values(by='OVERALL CRIME',ascending=True,inplace=True)
peaceful_areas.head(10).reset_index().drop(['index'],axis=1)
```

Out[90]:

	STATE/UT	DISTRICT	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLARY	RIC
0	HIMACHAL PRADESH	CID	0	0	0	0	0	0	
1	ASSAM	BIEO	0	0	0	0	0	0	
2	JAMMU & KASHMIR	RAILWAYS KATRA	0	0	0	0	0	0	
3	MANIPUR	CID	0	0	0	0	0	0	
4	DELHI	S.T.F.	0	0	0	0	0	0	
5	JAMMU & KASHMIR	RAILWAYS JAMMU	0	0	0	0	0	0	
6	JAMMU & KASHMIR	RAILWAYS KMR	0	0	0	0	0	1	
7	TRIPURA	G.R.P.	0	0	0	0	0	1	
8	JAMMU & KASHMIR	CRIME SRINAGAR	4	0	0	0	0	0	
9	DELHI	SPL CELL	0	0	0	0	0	0	

In [91]:
```
plt.figure(figsize=(8,8))
sns.barplot(x=peaceful_areas.head(10)['DISTRICT'],y=peaceful_areas.head(10)['OV
plt.xticks(rotation=90)
plt.xlabel("DIFFERENT STATES/UTs")
plt.title("TEN MOST PEACEFUL AREAs(DISTRICTS)!!")
plt.grid()
plt.show()
```



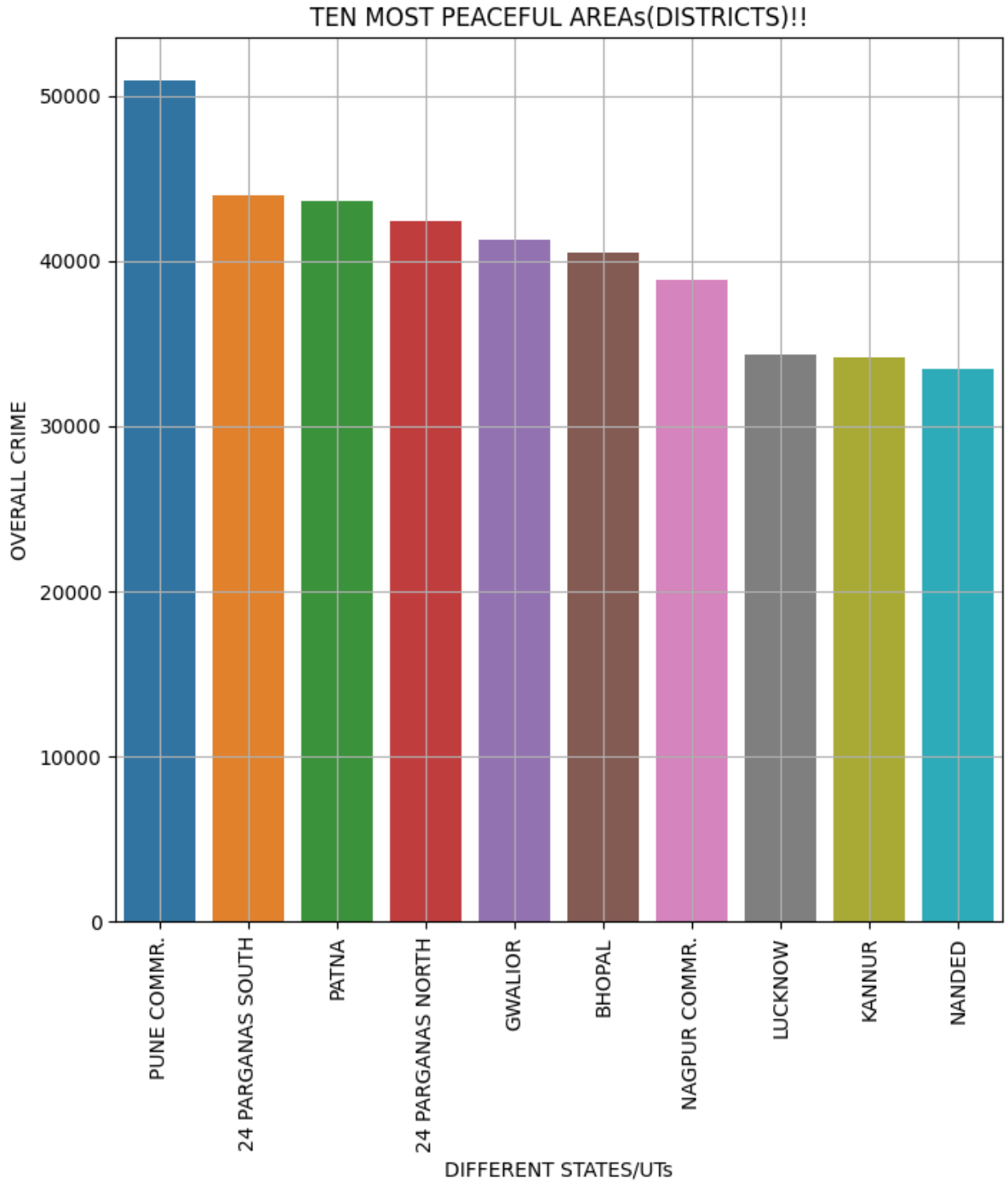
10 MOST VULNERABLE AREAS.

In [92]:
```
Vulnerable_area['OVERALL CRIME']=Vulnerable_area.iloc[:,2:].sum(axis=1)
Vulnerable_area.sort_values(by='OVERALL CRIME',ascending=False,inplace=True)
Vulnerable=Vulnerable_area.reset_index().drop(['index'],axis=1).head(10)
Vulnerable
```

Out[92]:

	STATE/UT	DISTRICT	MURDER	RAPE	KIDNAPPING & ABDUCTION	DACOITY	ROBBERY	BURGLA
0	MAHARASHTRA	PUNE COMMR.	1317	846	1087	309	3667	13
1	WEST BENGAL	24 PARGANAS SOUTH	2114	2221	2209	194	392	
2	BIHAR	PATNA	4586	572	3502	951	4901	6
3	WEST BENGAL	24 PARGANAS NORTH	2329	1883	3311	248	739	
4	MADHYA PRADESH	GWALIOR	1143	816	833	100	2055	11
5	MADHYA PRADESH	BHOPAL	821	1272	610	48	1839	11

In [93]:
```
plt.figure(figsize=(8,8))
sns.barplot(x=Vulnerable['DISTRICT'],y=Vulnerable['OVERALL CRIME'])
plt.xticks(rotation=90)
plt.xlabel("DIFFERENT STATES/UTs")
plt.title("TEN MOST PEACEFUL AREAs(DISTRICTS)!!")
plt.grid()
plt.show()
```

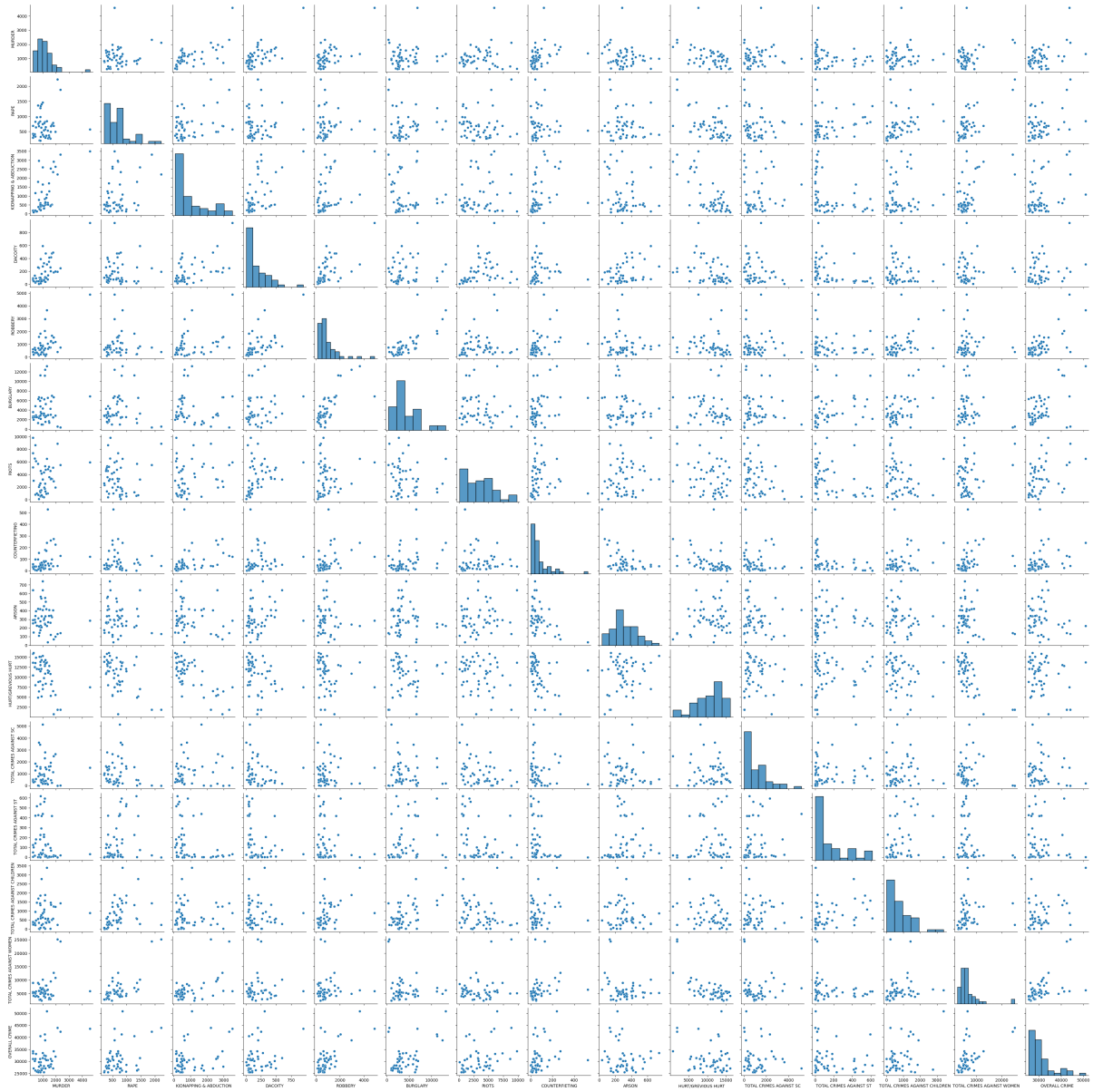


4.3.1 What is impacting more crimes in sensitive areas?

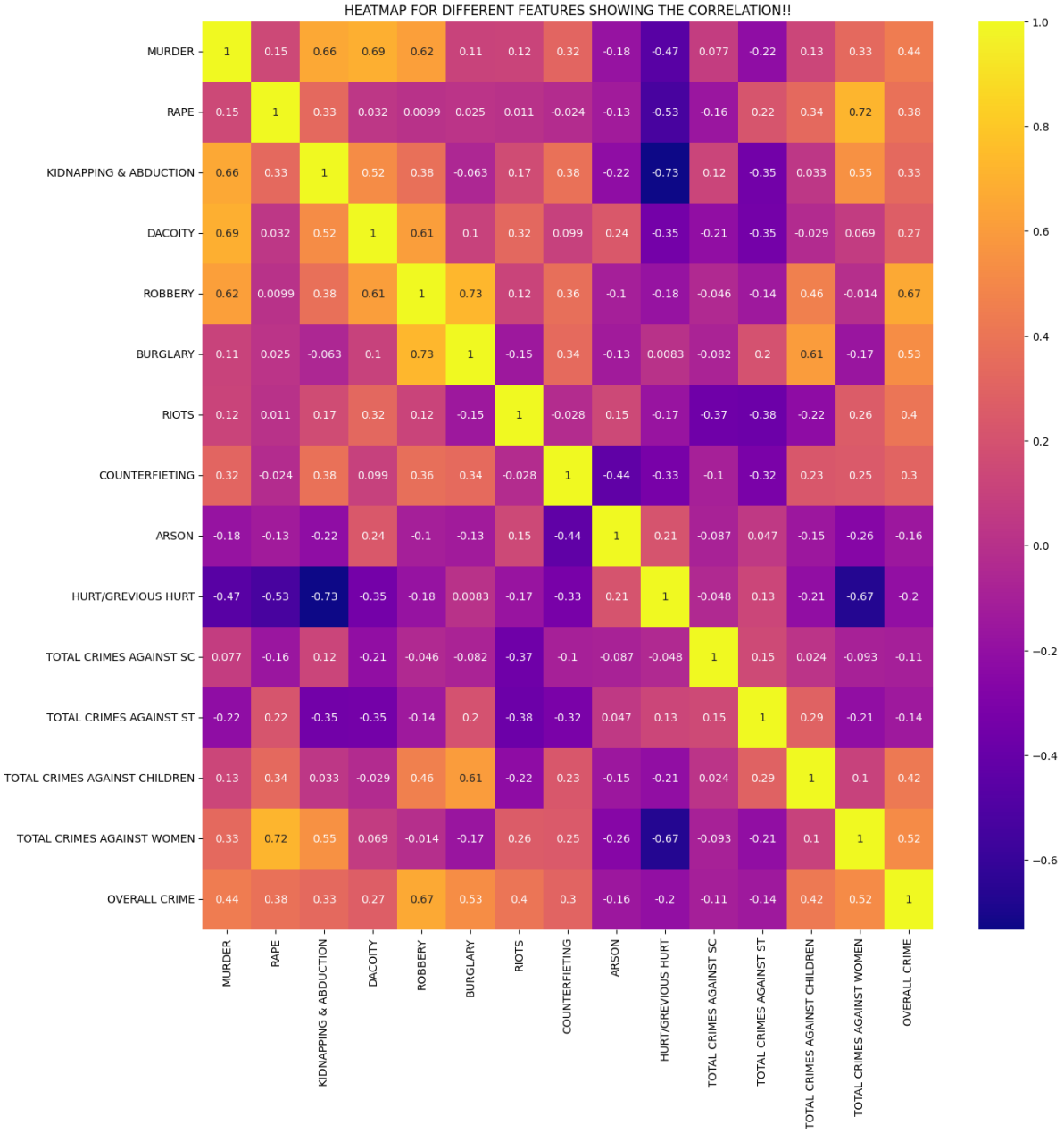
In [94]:
```
vulnerable_ar=Vulnerable_area.reset_index().drop(['index'],axis=1).head(50)
```

In [98]:
```
plt.figure(figsize=(25,90))
sns.pairplot(vulnerable_ar,palette='orange')
plt.xticks(size=20)
plt.show()
```

```
<Figure size 2500x9000 with 0 Axes>
```



In [100]:
```
plt.figure(figsize=(15,15))
sns.heatmap(vulnerable_ar.iloc[:,2:].corr(),annot=True,cmap='plasma')
plt.title("HEATMAP FOR DIFFERENT FEATURES SHOWING THE CORRELATION!!")
plt.show()
```



THE HEATMAP DESCRIBING THE CORRELATION FOR DIFFERENT FEATURES CAN BE SEEN FROM THE ABOVE PLOT!!

4.3.2 What needs to be done to reduce crime.

Reducing crime in India is a complex and multifaceted issue that requires a coordinated effort from various stakeholders, including the government, law enforcement agencies, civil society organizations, and the general public. Here are some measures that can help reduce crime in India:

Strengthen law enforcement: The government needs to allocate more resources and funds to law enforcement agencies to improve their capacity to prevent and investigate crimes. Police reforms should be implemented to make the force more accountable, efficient, and effective.

Address socio-economic factors: Poverty, unemployment, and lack of education are significant contributors to crime. The government needs to take measures to address these issues by creating job opportunities, improving access to education, and providing basic amenities like water, electricity, and healthcare.

Strengthen the justice system: The justice system in India is often slow and inefficient, which can lead to frustration and disillusionment among the public. Steps should be taken to streamline the justice system and make it more responsive to the needs of the people.

Raise public awareness: People should be made aware of the consequences of crime and the importance of reporting crime. Awareness campaigns should be conducted to educate people about their rights and responsibilities.

Improve community policing: Community policing can help build trust between the police and the community, leading to better crime prevention and investigation. The police should work closely with the community to identify and address local issues.

Use technology: Technology can play a significant role in preventing and solving crimes. The government should invest in modernizing the police force with advanced technology like CCTV cameras, drones, and digital forensics.

Reform the prison system: The prison system in India is overcrowded and inhumane, which can lead to the further radicalization of inmates. The government needs to take steps to improve the conditions in prisons and provide inmates with access to education and vocational training.

4.3.4 Capstone project overall story in your own words. Min 1000 words.

India is a diverse country with a population of over 1.38 billion people, and crime is a significant issue that affects its citizens. According to the data available from the National Crime Records Bureau (NCRB) for the year 2011, there were a total of 2.27 million reported cases of crime in India in 2011, a decrease of 4.54% from the previous year.

Crimes against women, including rape, sexual harassment, and domestic violence, continue to be a major concern in India. In 2011, there were 2,13,949 reported cases of crimes against women, and 11.04% of these cases were related to rape or sexual assault. Despite the enactment of stricter laws and increased public awareness, crimes against women remain a persistent problem in India.

Crimes against children including rape, kidnapping and abduction ,murder,traffacking,importation of girl are still the leading concerns in India. In 2011, there were 33,049 reported cases of crimes against children, and 21.51% of these cases were related to rape or sexual assault, and record 46.24% of the csaes are recorded of the murder only among differnt crimes against children.Even after various laws and stricter law enforcements being made to control the crimes against children a increase by 254.86% can be seen from the year 2001 to 2011.Madhya

Pradesh records the most cases counting to 40,217 over the years and is followed by Uttar Pradesh and Maharastra counting the numbers to 29,969 and 27,845.In the year 2011 33,718 cases were recorded against SC.

Involvement of juveniles can also be seen as a serious threat to the society.Often misguided,mishandled the juveniles tends towards to opt the wrong ways. As its easy to influence them stricter laws should be made to overcome this issue.Total of 23,597 cases were recorded in 2001 which then increased to 27,471 in the year 2011.Madhya Pradesh have the most cases recording to 61,939 and manipur being the least in the cases.

Human right violation by the police too offers significant amount of share in total ipc crimes.Year 2009 have the most cases recorded in any year , counting the crimes to 664.There is 82.815% decrease in the cases from the year 2001 to 2011.Chattisgarh have the most cases of human right violation across the states 2001 to 2011.

Its also important to note that those who have the duty of maintaining the law and order in the state i,e the police department have also the cases of getting assaulted by the Rioutous_mob,extremists,in counter insurgency operation,border opeartions and many more.Uttar pradesh recorded 2150 cases from the year 2001 to 2011, Jammu Kashmir also holds the 4 position with the total count of 1142.0 .Almost 5% increase in the cases can be seen from the year 2001 to 2011.In JAMMU & KASHMIR a decrease in the cases of assaults and killing of the police/army personnel acan be seen as the cases in 2001 were 300 which reduced to 50 in 2010 , 83.33 % direct decrease in the cases.

Maharastra, Uttar Pradesh and Madhya Pradesh had the most cases for DACOITY/ROBBERY/BURGALARY/THEFT combined together with counts being Maharastra-7,17,452 Madhya Pradesh-3,89,927 UTTAR PRADESH-3,66,166

Overall, while there has been a increase in the number of reported crimes in India hinting that there are still significant challenges to be addressed. The government needs to continue its efforts to improve law enforcement, raise public awareness, and address socio-economic factors that contribute to crime.