This data is collated from https://data.gov.in (https://data.gov.in). It has state-wise and district level data on the various crimes committed against women between 2001 to 2014. Some crimes that are included are Rape, Kidnapping and Abduction, Dowry Deaths etc.

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

readingb the csv file

In [3]:
```
crimes_df = pd.read_csv("crimes_against_women_2001-2014.csv")
```

In [4]:
```
crimes_df.head()
```

Out[4]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0

let us find total number of row and column

In [5]:
```
crimes_df.shape
```

Out[5]:
```
(10677, 10)
```

In [6]:
```
crimes_df.columns
```

Out[6]:
```
Index(['STATE/UT', 'DISTRICT', 'Year', 'Rape', 'Kidnapping and Abduction',
       'Dowry Deaths', 'Assault on women with intent to outrage her modesty',
       'Insult to modesty of Women', 'Cruelty by Husband or his Relatives',
       'Importation of Girls'],
      dtype='object')
```

In [7]:
```
crimes_df.dtypes
```

Out[7]:
```
STATE/UT                                             object
DISTRICT                                             object
Year                                                  int64
Rape                                                  int64
Kidnapping and Abduction                              int64
Dowry Deaths                                          int64
Assault on women with intent to outrage her modesty    int64
Insult to modesty of Women                            int64
Cruelty by Husband or his Relatives                   int64
Importation of Girls                                  int64
dtype: object
```

In [8]:
```
crimes_df.describe()
```

Out[8]:

	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
count	10677.000000	10677.000000	10677.000000	10677.000000	10677.000000	10677.000000	10677.000000	10677.000000
mean	2007.697949	57.989885	69.888358	20.181699	113.539196	27.419313	209.224314	0.175330
std	4.046874	214.230398	311.623450	98.276531	458.903951	167.806797	905.664362	2.228637
min	2001.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2004.000000	8.000000	6.000000	1.000000	10.000000	0.000000	11.000000	0.000000
50%	2008.000000	22.000000	20.000000	5.000000	34.000000	2.000000	50.000000	0.000000
75%	2011.000000	44.000000	49.000000	16.000000	85.000000	12.000000	144.000000	0.000000
max	2014.000000	5076.000000	10626.000000	2469.000000	10001.000000	4970.000000	23278.000000	83.000000

In [9]:
```
crimes_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10677 entries, 0 to 10676
Data columns (total 10 columns):
 #   Column                                          Non-Null Count  Dtype 
---  ------                                          --------------  ----- 
 0   STATE/UT                                        10677 non-null  object
 1   DISTRICT                                        10677 non-null  object
 2   Year                                            10677 non-null  int64 
 3   Rape                                            10677 non-null  int64 
 4   Kidnapping and Abduction                        10677 non-null  int64 
 5   Dowry Deaths                                    10677 non-null  int64 
 6   Assault on women with intent to outrage her modesty  10677 non-null  int64 
 7   Insult to modesty of Women                      10677 non-null  int64 
 8   Cruelty by Husband or his Relatives             10677 non-null  int64 
 9   Importation of Girls                            10677 non-null  int64 
dtypes: int64(8), object(2)
memory usage: 834.3+ KB
```

data preparation and cleaning

In [10]:
```
# is there null values in dataset

overall_crime = crimes_df.isna().sum()
overall_crime
```

Out[10]:
```
STATE/UT                                        0
DISTRICT                                        0
Year                                            0
Rape                                            0
Kidnapping and Abduction                        0
Dowry Deaths                                    0
Assault on women with intent to outrage her modesty    0
Insult to modesty of Women                      0
Cruelty by Husband or his Relatives             0
Importation of Girls                            0
dtype: int64
```

In [ ]:
```
# unique record from "DISTRICT" column
```

In [13]:
```
crimes_df.DISTRICT.nunique()
```

Out[13]:
```
1605
```

In [14]:
```
crimes_df["DISTRICT"].nunique()
```

Out[14]: 1605

In [15]:
```
crimes_df.head()
```

Out[15]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0

In [16]:
```
## for deleing the column do not required here
#crimes_df.drop(["DISTRICT"],axis=1,inplace=True)
```

Out[16]:

	STATE/UT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	ANDHRA PRADESH	2001	50	30	16	149	34	175	0
1	ANDHRA PRADESH	2001	23	30	7	118	24	154	0
2	ANDHRA PRADESH	2001	27	34	14	112	83	186	0
3	ANDHRA PRADESH	2001	20	20	17	126	38	57	0
4	ANDHRA PRADESH	2001	23	26	12	109	58	247	0
...	...	...	...	...	...	...	...	...	...
10672	Lakshadweep	2014	1	0	0	1	2	0	0
10673	Lakshadweep	2014	1	0	0	1	2	0	0
10674	Puducherry	2014	3	1	0	12	1	1	0
10675	Puducherry	2014	7	6	1	20	7	3	0
10676	Puducherry	2014	10	7	1	32	8	4	0

10677 rows × 9 columns

In [17]:
```
crimes_df.head()
```

Out[17]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0

In [21]:
```
crimes_df.drop_duplicates()
```

Out[21]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	34	175	0
1	ANDHRA PRADESH	ANANTAPUR	2001	23	30	7	118	24	154	0
2	ANDHRA PRADESH	CHITTOOR	2001	27	34	14	112	83	186	0
3	ANDHRA PRADESH	CUDDAPAH	2001	20	20	17	126	38	57	0
4	ANDHRA PRADESH	EAST GODAVARI	2001	23	26	12	109	58	247	0
...	...	...	...	...	...	...	...	...	...	...
10672	Lakshadweep	Lakshadweep	2014	1	0	0	1	2	0	0
10673	Lakshadweep	Total District(s)	2014	1	0	0	1	2	0	0
10674	Puducherry	Karaikal	2014	3	1	0	12	1	1	0
10675	Puducherry	Puducherry	2014	7	6	1	20	7	3	0
10676	Puducherry	Total District(s)	2014	10	7	1	32	8	4	0

10677 rows × 10 columns

In [24]:
```
# rename the columns

crimes_df.rename(columns={
    "Kidnapping and Abduction": "Kidnapping_Abduction",
    "Dowry Deaths":"Dowry_Deaths",
    "Assault on women with intent to outrage her modesty":"Hurting_of_women_modesty",
    "Insult to modesty of Women":"Insult_to_women_modesty",
    "Importation of Girls":"Importation_of_Girls",
    "Cruelty by Husband or his Relatives":"Domestic_Cruelty"},inplace=True
    )
```

In [25]:
```
crimes_df.head(1)
```

Out[25]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping_Abduction	Dowry_Deaths	Hurting_of_women_modesty	Insult_to_women_modes
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	

In [29]:
```
print(crimes_df["STATE/UT"].unique())
```

```
['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' '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 UT' 'Lakshadweep' 'Puducherry' 'Telangana'
 'A & N Islands']
```

In [30]:
```
crimes_df["STATE/UT"] = crimes_df["STATE/UT"].str.upper()
```

In [32]:
```
print(crimes_df["STATE/UT"].unique())
```

```
['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' 'A&N ISLANDS' 'D&N HAVELI' 'DELHI UT' 'TELANGANA']
```

In [35]:
```
# replace the values from columns
crimes_df["STATE/UT"].replace("D & N HAVELI",'D&N HAVELI',inplace=True)
crimes_df["STATE/UT"].replace('DELHI UT','DELHI',inplace=True)
crimes_df["STATE/UT"].replace('A & N ISLANDS','A&N ISLANDS',inplace=True)
```

In [36]:
```
print(crimes_df["STATE/UT"].unique())
```

```
['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' 'TELANGANA']
```

In [37]:
```
len(crimes_df["STATE/UT"].unique())
```

Out[37]: 36

In [39]:
```
crimes_df.head(1)
```

Out[39]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping_Abduction	Dowry_Deaths	Hurting_of_women_modesty	Insult_to_women_modes
0	ANDHRA PRADESH	ADILABAD	2001	50	30	16	149	

In [40]:
```
crimes_df.to_csv("cleaned_crime_data.csv")
```

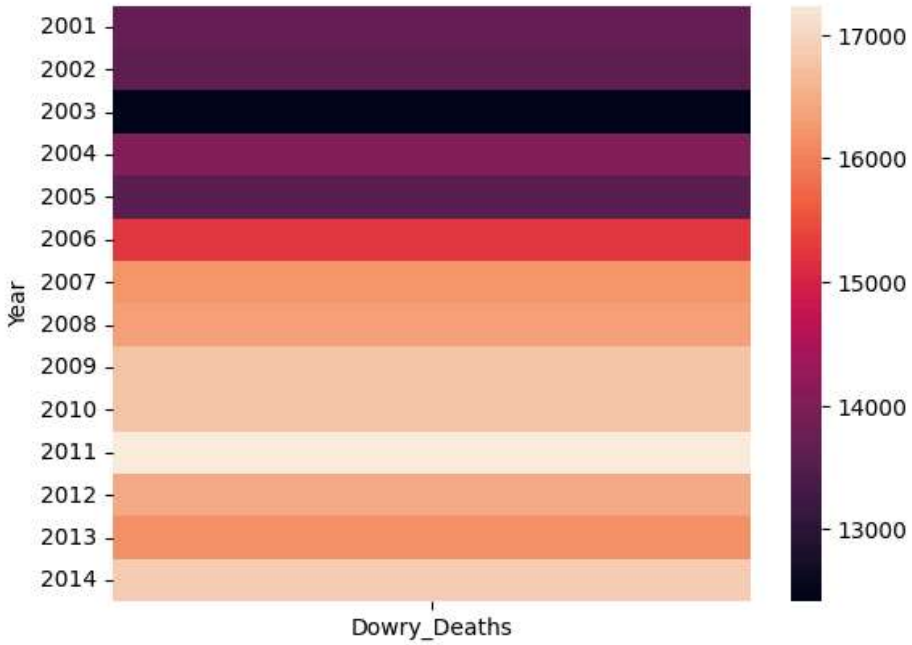
In [ ]:
```
# Q1: FInd out the how many total no of cases increase by year? for Dowry_Deaths and Kidnapping_Abductio
```

In [43]:
```
crimes_df.groupby("Year")[["Dowry_Deaths","Kidnapping_Abduction"]].sum()
```

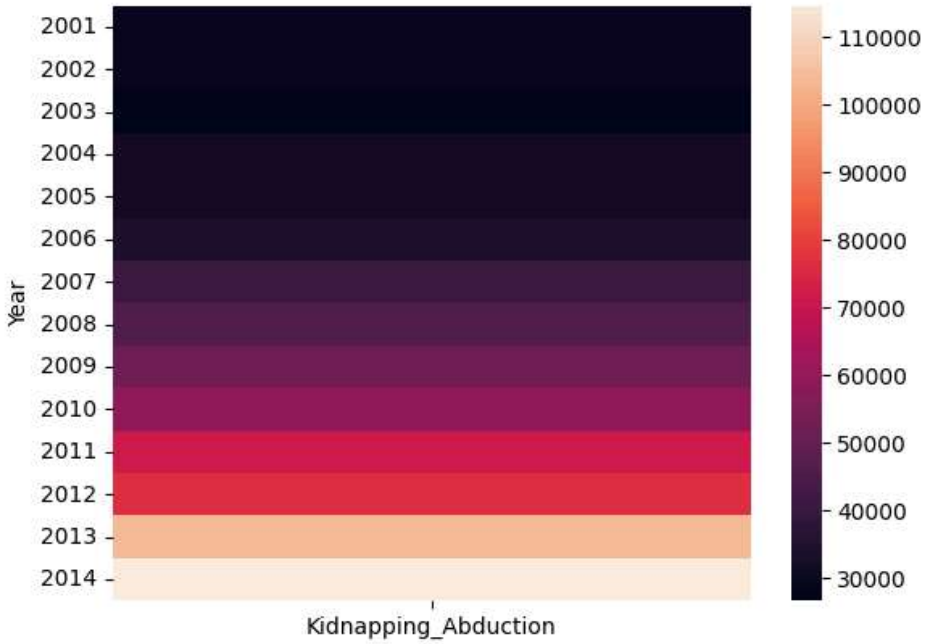
Out[43]:

Year	Dowry_Deaths	Kidnapping_Abduction
2001	13702	29290
2002	13644	29012
2003	12416	26592
2004	14052	31156
2005	13574	31500
2006	15236	34828
2007	16186	40832
2008	16344	45878
2009	16766	51482
2010	16782	59590
2011	17236	71130
2012	16466	76524
2013	16166	103762
2014	16910	114622

In [46]:
```
d =crimes_df.groupby("Year")[["Dowry_Deaths"]].sum()
sns.heatmap(d)
plt.show()
```



In [47]:
```
d =crimes_df.groupby("Year")[["Kidnapping_Abduction"]].sum()
sns.heatmap(d)
plt.show()
```



In [ ]:
```
## Q2 top 10 Higest cases for Rape in indiafrom 2001-214?
```

In [48]:
```
crimes_df.nlargest(10,"Rape")
```

Out[48]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping_Abduction	Dowry_Deaths	Hurting_of_women_modesty	Insult_to_wom
10244	MADHYA PRADESH	Total District(s)	2014	5076	5688	733	9609	
9426	MADHYA PRADESH	ZZ TOTAL	2013	4335	2873	776	8252	
10445	RAJASTHAN	Total District(s)	2014	3759	4421	408	5999	
10595	UTTAR PRADESH	Total District(s)	2014	3467	10626	2469	8605	
10291	MAHARASHTRA	Total District(s)	2014	3438	2457	279	10001	
8611	MADHYA PRADESH	TOTAL	2012	3425	1127	743	6655	
7810	MADHYA PRADESH	TOTAL	2011	3406	1088	811	6665	
9628	RAJASTHAN	ZZ TOTAL	2013	3285	4047	453	4829	
7025	MADHYA PRADESH	TOTAL	2010	3135	1030	892	6646	
9472	MAHARASHTRA	ZZ TOTAL	2013	3063	1874	320	8132	

In [49]:
```
## SHow me the highest reported death cases because of dowary?

crimes_df.sort_values("Dowry_Deaths",ascending=False).head(10)
```

Out[49]:

	STATE/UT	DISTRICT	Year	Rape	Kidnapping_Abduction	Dowry_Deaths	Hurting_of_women_modesty	Insult_to_women_m
10595	UTTAR PRADESH	Total District(s)	2014	3467	10626	2469	8605	
9760	UTTAR PRADESH	ZZ TOTAL	2013	3050	9737	2335	7303	
8132	UTTAR PRADESH	TOTAL	2011	2042	7525	2322	3455	
8938	UTTAR PRADESH	TOTAL	2012	1963	7910	2244	3247	
5796	UTTAR PRADESH	TOTAL	2008	1871	4439	2237	2955	
6563	UTTAR PRADESH	TOTAL	2009	1759	5078	2232	2782	
7342	UTTAR PRADESH	TOTAL	2010	1563	5468	2217	2793	
650	UTTAR PRADESH	TOTAL	2001	1958	2879	2211	2870	
5040	UTTAR PRADESH	TOTAL	2007	1648	3363	2076	2522	
1366	UTTAR PRADESH	TOTAL	2002	1415	2298	1893	2145	

Q4: I wan to know the all state and UT wise total cases?

In [51]:
```
df = crimes_df.groupby("STATE/UT")[["Rape","Kidnapping_Abduction","Dowry_Deaths"]].sum()
```

In [52]:
```
df
```

Out[52]:

STATE/UT	Rape	Kidnapping_Abduction	Dowry_Deaths
A&N ISLANDS	336	212	20
ANDHRA PRADESH	32150	34504	13844
ARUNACHAL PRADESH	1316	1470	6
ASSAM	40190	62074	3268
BIHAR	30758	57086	32206
CHANDIGARH	770	1682	90
CHHATTISGARH	29308	11808	2758
D&N HAVELI	132	224	2
DAMAN & DIU	60	44	6
DELHI	20312	46586	3758
GOA	1062	640	38
GUJARAT	11644	34670	1108
HARYANA	17110	20016	7372
HIMACHAL PRADESH	4674	4116	112
JAMMU & KASHMIR	7038	21164	294
JHARKHAND	22826	14186	7896
KARNATAKA	15056	16262	7016
KERALA	20030	4452	700
LAKSHADWEEP	20	2	0
MADHYA PRADESH	90996	35608	21090
MAHARASHTRA	48974	30368	9696
MANIPUR	1068	2606	6
MEGHALAYA	2642	670	36
MIZORAM	2070	30	8
NAGALAND	562	190	2
ODISHA	30480	25588	10782
PUDUCHERRY	208	306	56
PUNJAB	14656	15096	3524
RAJASTHAN	45684	66278	11854
SIKKIM	570	180	4
TAMIL NADU	16660	30908	5060
TELANGANA	1958	1422	578
TRIPURA	5060	2202	752
UTTAR PRADESH	51150	135906	57256
UTTARAKHAND	3752	6484	1974
WEST BENGAL	47876	61158	12308

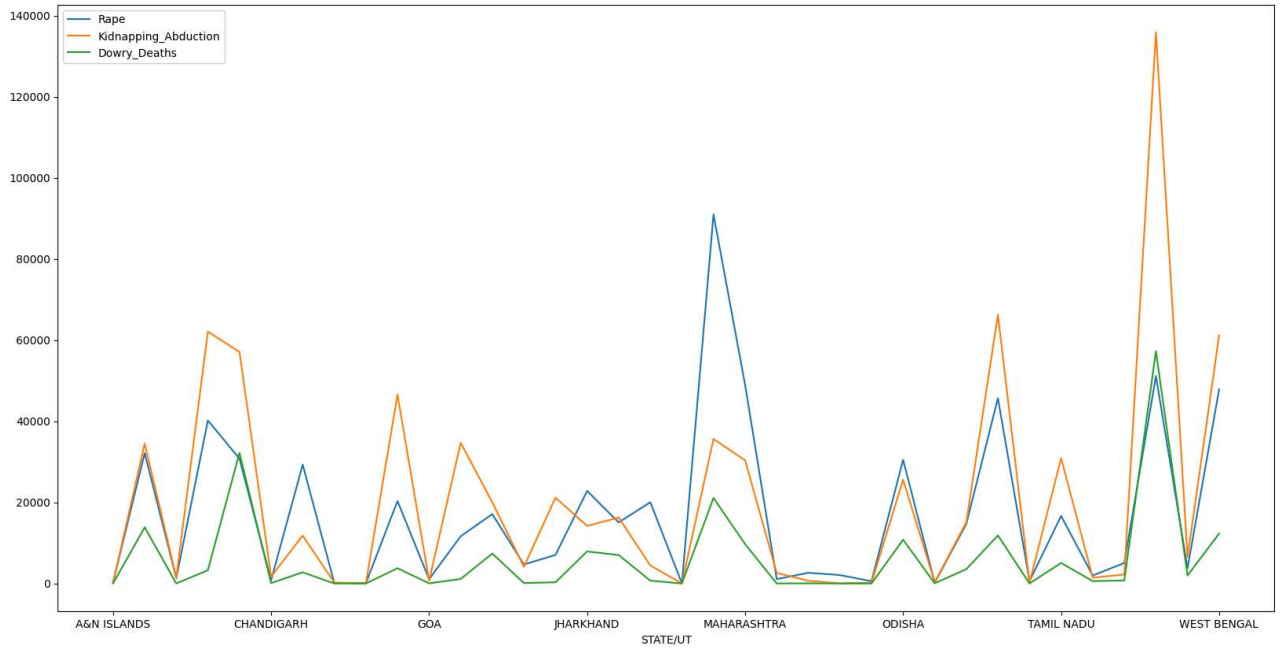
In [54]:
```
df.sort_values("Dowry_Deaths",ascending=False).head(5)
```

Out[54]:

STATE/UT	Rape	Kidnapping_Abduction	Dowry_Deaths
UTTAR PRADESH	51150	135906	57256
BIHAR	30758	57086	32206
MADHYA PRADESH	90996	35608	21090
ANDHRA PRADESH	32150	34504	13844
WEST BENGAL	47876	61158	12308

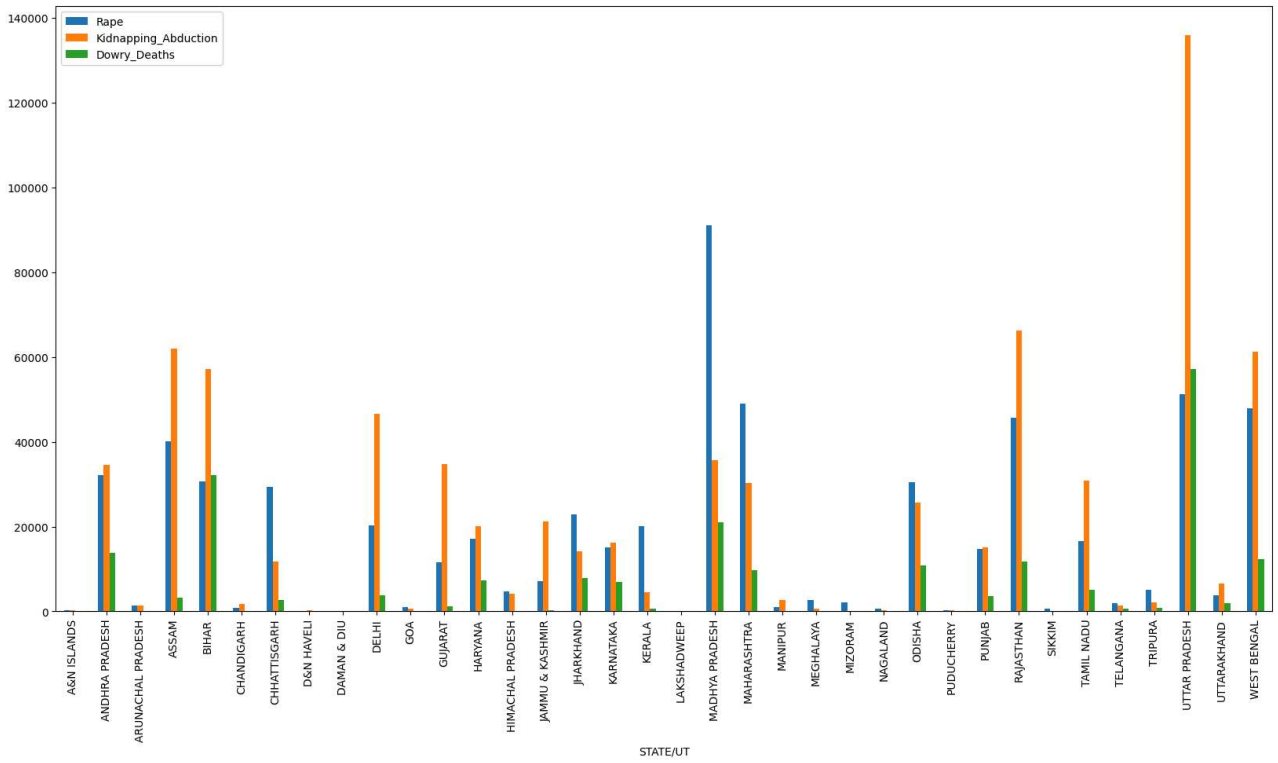
In [56]:
```
df.plot(kind='line',figsize=(20,10))
```

Out[56]: <Axes: xlabel='STATE/UT'>



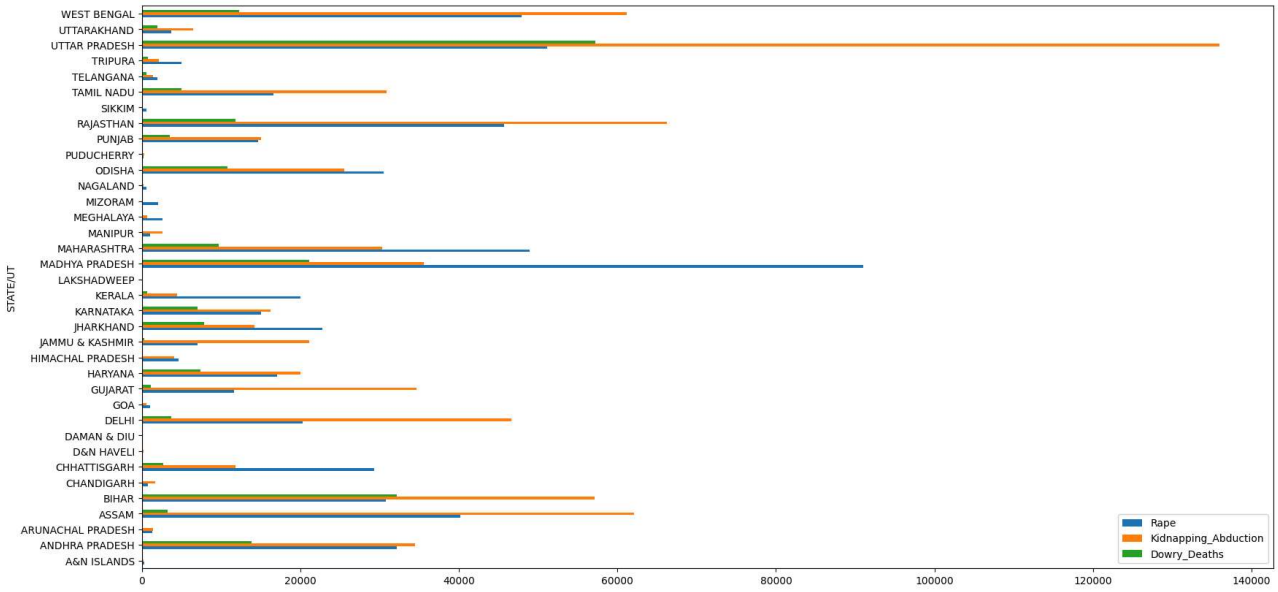
In [57]:
```
df.plot(kind='bar',figsize=(20,10))
```

Out[57]: <Axes: xlabel='STATE/UT'>



In [58]:
```
df.plot(kind='barh',figsize=(20,10))
```

Out[58]: <Axes: ylabel='STATE/UT'>



In [ ]:
```

```