EDA Project

EDA -CSE3040
Submitted by

> Saptharishee M
> 20MIA1150

> Chenchu Aravind V
> 20MIA1126

> Sarveswaran
> 20MIA1128

> Shiva Sindhu Perla
> 20MIA1104

Visit our googlesite : [EDA-Suicide in India]

Motivation

Each suicide is a personal tragedy that prematurely takes the life of an individual and has a continuing ripple effect, affecting the lives of families, friends and communities. Every year, more than 1,00,000 people commit suicide in our country. There are various causes of suicides like professional/career problems, sense of isolation, abuse, violence, family problems, mental disorders, addiction to alcohol, financial loss, chronic pain etc.

ADSI 2020 (Accidental Deaths & Suicides in India 2020) states that there was 10% increase in suicide rates that is on an average more than 17 died because of suicide every single hour. Reports suggested that covid-19 pandemic has increased the number of suicide so we decided to investigate the reason behind suicide on a normal year. We found State/UT-wise professional profile of suicide victim during 2001-2012 dataset from datagov.in .

Abstract

Problem uses a government provided real dataset of Suicide between 2001 to 2012 and analyses the cause based on gender, age-group, location of the person (in India). Apart from causes were there any other similarities of between the data? What is the major cause of suicide? which group of people died of suicide more? It also tries to

get a visual idea and inference about the data. The aim of the project is to find the major reason of suicide and accidental death that happened a decade ago.

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

Dataset that we use is taken from datagov.in.Tittle of the dataset is "State/UT-wise professional profile of suicide victim during 2001-2012".

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

```
data=pd.read_csv('/content/drive/MyDrive/EDA Dataset/Final_dataset.csv',index_col="CAUSE")
#dataset is read using pandas framework
```

```
data
```

	STATE/UT	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Female upto 14 years	F
CAUSE										

We found few part of data that can mislead the analysis so we figured out what are those rows and tried to seperate it

(Government)	PRADESH	2001	0	34	37	44	0	135	0

```
totalSeperated=data.loc['Total'].reset_index()
totalSalaried=data.loc['Total Salaried'].reset_index()
TotalSelfEmployed=data.loc['Total Self-employed'].reset_index()
```

Public Sector	ANDHRA	2001	0	32	36	29	13	110	0

```
totalSeperated
```

	CAUSE	STATE/UT	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Female upto 14 years	Fem 15 ye
0	Total	ANDHRA PRADESH	2001	173	2058	2213	1566	369	6379	185	1
1	Total	ANDHRA PRADESH	2002	101	2173	2831	1935	531	7571	139	1
2	Total	ANDHRA PRADESH	2003	155	2270	2846	1745	497	7513	197	1
3	Total	ANDHRA PRADESH	2004	95	2562	3448	2175	802	9082	142	1
4	Total	ANDHRA PRADESH	2005	143	2444	3178	2156	944	8865	161	1
...	...	...	...	...	...	...	...	...	...	...	
451	Total	TOTAL (ALL INDIA)	2008	1165	24396	29519	18819	6645	80544	1216	20
452	Total	TOTAL (ALL INDIA)	2009	1501	23746	29335	19596	7293	81471	1450	20
453	Total	TOTAL (ALL INDIA)	2010	1640	26387	30444	20768	7941	87180	1490	21

totalSalaried

	CAUSE	STATE/UT	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Female upto 14 years	Fe 1 y
0	Total Salaried	ANDHRA PRADESH	2001	1	316	368	211	23	919	5	
1	Total Salaried	ANDHRA PRADESH	2002	0	299	351	196	10	856	0	
2	Total Salaried	ANDHRA PRADESH	2003	2	186	298	219	35	740	2	
3	Total Salaried	ANDHRA PRADESH	2004	4	275	441	232	30	982	0	
4	Total Salaried	ANDHRA PRADESH	2005	2	265	438	298	39	1042	0	
...	...	...	...	...	...	...	...	...	...	...	
451	Total Salaried	TOTAL (ALL INDIA)	2008	11	3525	5496	3512	458	13002	11	
452	Total Salaried	TOTAL (ALL INDIA)	2009	3	3771	5510	3211	549	13044	3	
453	Total Salaried	TOTAL (ALL INDIA)	2010	32	3857	5290	3274	439	12892	10	

TotalSelfEmployed

	CAUSE	STATE/UT	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Female upto 14 years
0	Total Self-employed	ANDHRA PRADESH	2001	83	906	1218	835	194	3236	75
1	Total Self-employed	ANDHRA PRADESH	2002	35	956	1412	1099	351	3853	19
2	Total Self-employed	ANDHRA PRADESH	2003	26	849	1478	942	330	3625	47
3	Total Self-employed	ANDHRA PRADESH	2004	13	1250	1885	1332	446	4926	35

```
#Removing three rows from CAUSE column
dropRow=data.drop(['Total','Total Salaried','Total Self-employed'])
```

```
#reseting CAUSE column
index_reset=dropRow.reset_index()
```

451	Self-	(ALL	2008	255	11264	16225	10533	3943	42220	206

```
#Making STATE/UT column as index column

setIndex=index_reset.set_index('STATE/UT')
```

```
setIndex
```

STATE/UT	CAUSE	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Femal upt 1 year
ANDHRA PRADESH	House Wife	2001	0	0	0	0	0	0	
ANDHRA PRADESH	Service (Government)	2001	0	34	57	44	0	135	
ANDHRA PRADESH	Service (Private)	2001	1	250	275	138	10	674	
ANDHRA PRADESH	Public Sector Undertaking	2001	0	32	36	29	13	110	

```
print(data["STATE/UT"].value_counts())
"""While examining the unique value of colums of dataset we found that rows with data in "Sta
column like "TOTAL (UTs),TOTAL (ALL INDIA),TOTAL (STATES)" were actually the sum of other row
 to make a clear analysis we seprated it """
```

```
ANDHRA PRADESH       180
TOTAL (STATES)       180
RAJASTHAN            180
SIKKIM               180
TAMIL NADU           180
TRIPURA              180
UTTAR PRADESH        180
UTTARAKHAND          180
WEST BENGAL          180
A & N ISLANDS        180
ARUNACHAL PRADESH    180
CHANDIGARH           180
```

```
#This is totally cleaned data.

totalCleaned=setIndex.drop(['TOTAL (ALL INDIA)','TOTAL (UTs)','TOTAL (STATES)']).reset_index(
```

```
TOTAL (UTs)          180
```

```
totalCleaned
```

	STATE/UT	CAUSE	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	To M
0	ANDHRA PRADESH	House Wife	2001	0	0	0	0	0	
1	ANDHRA PRADESH	Service (Government)	2001	0	34	57	44	0	
2	ANDHRA PRADESH	Service (Private)	2001	1	250	275	138	10	
3	ANDHRA PRADESH	Public Sector Undertaking	2001	0	32	36	29	13	
4	ANDHRA PRADESH	Student	2001	56	202	19	18	0	
...	...	...	...	...	...	...	...	...	
5035	PUDUCHERRY	Professional Activity	2012	0	23	47	29	6	
5036	PUDUCHERRY	Farming/Agriculture Activity	2012	0	0	0	0	0	
5037	PUDUCHERRY	Others (Please Specify)	2012	0	22	52	48	7	

```
totalCleaned["STATE/UT"].value_counts()
#after cleaning we have 144 value counts in each type where as before we had 180
```

```
ANDHRA PRADESH       144
UTTARAKHAND          144
PUNJAB               144
RAJASTHAN            144
SIKKIM               144
TAMIL NADU           144
TRIPURA              144
UTTAR PRADESH        144
WEST BENGAL          144
NAGALAND             144
A & N ISLANDS        144
CHANDIGARH           144
D & N HAVELI         144
DAMAN & DIU          144
DELHI (UT)           144
LAKSHADWEEP          144
ODISHA               144
MIZORAM              144
ARUNACHAL PRADESH    144
HIMACHAL PRADESH     144
ASSAM                144
BIHAR                144
CHHATTISGARH         144
GOA                  144
GUJARAT              144
HARYANA              144
JAMMU & KASHMIR      144
MEGHALAYA            144
JHARKHAND            144
KARNATAKA            144
KERALA               144
MADHYA PRADESH       144
MAHARASHTRA          144
MANIPUR              144
PUDUCHERRY           144
Name: STATE/UT, dtype: int64
```

```
totalCleaned.describe()
```

	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and

```
totalCleaned.isnull()
```

	STATE/UT	CAUSE	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Female upto 14 years	Fem 15 ye
0	False	False	False	False	False	False	False	False	False	False	F
1	False	False	False	False	False	False	False	False	False	False	F
2	False	False	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	False	False	F
4	False	False	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	...	...	
5035	False	False	False	False	False	False	False	False	False	False	F
5036	False	False	False	False	False	False	False	False	False	False	F
5037	False	False	False	False	False	False	False	False	False	False	F
5038	False	False	False	False	False	False	False	False	False	False	F

```
totalCleaned.isnull().sum()
#there are no "Nan"/Null values in the dataseet
```

```
STATE/UT                      0
CAUSE                         0
Year                          0
Male upto 14 years            0
Male 15-29 years              0
Male 30-44 years              0
Male 45-59 years              0
Male 60 years and above       0
Total Male                    0
Female upto 14 years          0
Female 15-29 years            0
Female 30-44 years            0
Female 45-59 years            0
Female 60 years and above     0
Total Female                  0
Grand Total                   0
dtype: int64
```

```
totalCleaned.columns
#available columns in the dataset
```

```
Index(['STATE/UT', 'CAUSE', 'Year', 'Male upto 14 years', 'Male 15-29 years',
       'Male 30-44 years', 'Male 45-59 years', 'Male 60 years and above',
       'Total Male', 'Female upto 14 years', 'Female 15-29 years',
       'Female 30-44 years', 'Female 45-59 years', 'Female 60 years and above',
       'Total Female', 'Grand Total'],
      dtype='object')
```

```
totalCleaned.shape
#the dataset has 5040 rows and 16 columns
```

```
(5040, 16)
```

```
totalCleaned.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5040 entries, 0 to 5039
Data columns (total 16 columns):
 #   Column                     Non-Null Count  Dtype 
---  ------                     --------------  ----- 
 0   STATE/UT                   5040 non-null   object
 1   CAUSE                      5040 non-null   object
 2   Year                       5040 non-null   int64 
 3   Male upto 14 years         5040 non-null   int64 
 4   Male 15-29 years           5040 non-null   int64 
 5   Male 30-44 years           5040 non-null   int64 
 6   Male 45-59 years           5040 non-null   int64 
 7   Male 60 years and above    5040 non-null   int64 
 8   Total Male                 5040 non-null   int64 
 9   Female upto 14 years       5040 non-null   int64 
 10  Female 15-29 years         5040 non-null   int64 
 11  Female 30-44 years         5040 non-null   int64 
 12  Female 45-59 years         5040 non-null   int64 
 13  Female 60 years and above  5040 non-null   int64 
 14  Total Female               5040 non-null   int64 
 15  Grand Total                5040 non-null   int64 
dtypes: int64(14), object(2)
memory usage: 630.1+ KB
```

```
duplicate = totalCleaned[totalCleaned.duplicated()]
duplicate
```

	STATE/UT	CAUSE	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Fema up yea
155	ARUNACHAL PRADESH	Others (Please Specify)	2001	0	0	0	0	0	0	
167	ARUNACHAL PRADESH	Others (Please Specify)	2002	0	0	0	0	0	0	
179	ARUNACHAL PRADESH	Others (Please Specify)	2003	0	0	0	0	0	0	
191	ARUNACHAL PRADESH	Others (Please Specify)	2004	0	0	0	0	0	0	
203	ARUNACHAL PRADESH	Others (Please Specify)	2005	0	0	0	0	0	0	
2195	MANIPUR	Others (Please Specify)	2003	0	0	0	0	0	0	
2351	MEGHALAYA	Others (Please Specify)	2004	0	0	0	0	0	0	
2711	NAGALAND	Others (Please Specify)	2010	0	0	0	0	0	0	
3851	UTTARAKHAND	Others (Please Specify)	2009	0	0	0	0	0	0	
4031	WEST BENGAL	Others (Please Specify)	2012	0	0	0	0	0	0	
4067	A & N ISLANDS	Others (Please Specify)	2003	0	0	0	0	0	0	
4247	CHANDIGARH	Others (Please Specify)	2006	0	0	0	0	0	0	
4271	CHANDIGARH	Others (Please Specify)	2008	0	0	0	0	0	0	

4391	D & N HAVELI	Others (Please Specify)	2006	0	0	0	0	0	0

```
corelation=totalCleaned.corr()
corelation
```

	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Fema upto yea
Year	1.000000	0.001099	0.034479	0.033621	0.035200	0.025278	0.034510	-0.007
Male upto 14 years	0.001099	1.000000	0.478972	0.270396	0.242803	0.236641	0.355764	0.880
Male 15-29 years	0.034479	0.478972	1.000000	0.908447	0.845308	0.732547	0.942788	0.485
Male 30-44 years	0.033621	0.270396	0.908447	1.000000	0.943425	0.827476	0.982674	0.282
Male 45-59 years	0.035200	0.242803	0.845308	0.943425	1.000000	0.892688	0.965853	0.257
Male 60 years and above	0.025278	0.236641	0.732547	0.827476	0.892688	1.000000	0.875868	0.263

```
totalCleaned.skew()
#skewness
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping
  """Entry point for launching an IPython kernel.
Year                          0.000000
Male upto 14 years            6.336851
Male 15-29 years              3.495508
Male 30-44 years              4.100567
Male 45-59 years              4.204126
Male 60 years and above       5.268641
Total Male                    3.982215
Female upto 14 years          6.520858
Female 15-29 years            7.111149
Female 30-44 years            6.777747
```

```
Female 45-59 years           7.286008
Female 60 years and above    6.923788
Total Female                 6.621926
Grand Total                  3.388685
dtype: float64
```

```
totalCleaned.kurt()
#Kurtosis
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping
  """Entry point for launching an IPython kernel.
Year                         -1.216800
Male upto 14 years           51.373042
Male 15-29 years             15.039606
Male 30-44 years             21.533892
Male 45-59 years             21.243609
Male 60 years and above      34.241158
Total Male                   19.757975
Female upto 14 years         56.116966
Female 15-29 years           61.287294
Female 30-44 years           53.960684
Female 45-59 years           73.730618
Female 60 years and above    56.698368
Total Female                 52.802501
Grand Total                  13.248276
dtype: float64
```

```
totalCleaned.var()
#variance
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping
  """Entry point for launching an IPython kernel.
Year                            11.919032
Male upto 14 years             144.064179
Male 15-29 years             13746.625549
Male 30-44 years             24579.632098
Male 45-59 years             12046.113168
Male 60 years and above       2615.029687
Total Male                  176120.752688
Female upto 14 years           146.586526
Female 15-29 years           26468.333952
Female 30-44 years           14000.039179
Female 45-59 years            3412.980110
Female 60 years and above      674.377274
Total Female                126346.211937
Grand Total                 363346.274442
dtype: float64
```

```
totalCleaned.std()
```

```
#standard deviation
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping
  """Entry point for launching an IPython kernel.
Year                          3.452395
Male upto 14 years           12.002674
Male 15-29 years            117.246004
Male 30-44 years            156.778927
Male 45-59 years            109.754787
Male 60 years and above      51.137361
Total Male                  419.667431
Female upto 14 years         12.107292
Female 15-29 years          162.690915
Female 30-44 years          118.321761
Female 45-59 years           58.420716
Female 60 years and above    25.968775
Total Female                355.452123
Grand Total                 602.782112
dtype: float64
```

```
totalCleaned.cov()
#covarince
```

	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male years a abo
Year	11.919032	0.045545	13.956440	18.197757	13.337765	4.4626
Male upto 14	0.045545	144.064179	674.040893	508.821529	319.857099	145.2469

```
totalCleaned.mean()
#mean of each column
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping
  """Entry point for launching an IPython kernel.
Year                         2006.500000
Male upto 14 years              3.225198
Male 15-29 years               54.914484
Male 30-44 years               65.539484
Male 45-59 years               43.079365
Male 60 years and above        16.423611
Total Male                    183.182143
Female upto 14 years            3.259921
Female 15-29 years             46.231548
Female 30-44 years             31.427381
Female 45-59 years             15.320040
Female 60 years and above       6.486508
Total Female                  102.725397
Grand Total                   285.907540
dtype: float64
```

upto

```
totalCleaned.median()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping
  """Entry point for launching an IPython kernel.
Year                         2006.5
Male upto 14 years              0.0
Male 15-29 years                5.0
Male 30-44 years                5.0
Male 45-59 years                2.0
Male 60 years and above         0.0
Total Male                     18.0
Female upto 14 years            0.0
Female 15-29 years              2.0
Female 30-44 years              1.0
Female 45-59 years              0.0
Female 60 years and above       0.0
Total Female                    6.0
Grand Total                    35.0
dtype: float64
```

```
totalCleaned.mode()
```

	STATE/UT	CAUSE	Year	Male upto 14 years	Male 15-29 years	Male 30-44 years	Male 45-59 years	Male 60 years and above	Total Male	Fema up yea
0	A & N ISLANDS	Others (Please Specify)	2001.0	0.0	0.0	0.0	0.0	0.0	0.0	(
1	ANDHRA PRADESH	NaN	2002.0	NaN	NaN	NaN	NaN	NaN	NaN	N
2	ARUNACHAL PRADESH	NaN	2003.0	NaN	NaN	NaN	NaN	NaN	NaN	N

```
totalCleaned.isnull().any()
#there are no  null values
```

```
STATE/UT                    False
CAUSE                       False
Year                        False
Male upto 14 years          False
Male 15-29 years            False
Male 30-44 years            False
Male 45-59 years            False
Male 60 years and above     False
Total Male                  False
Female upto 14 years        False
Female 15-29 years          False
Female 30-44 years          False
Female 45-59 years          False
Female 60 years and above   False
Total Female                False
Grand Total                 False
dtype: bool
```

11	JAMMU &	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

```
cv = lambda x: np.std(x) / np.mean(x)
#labda function to calculate coeff of variance
```

12	KARNATAKA	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

```
cv(totalCleaned)
#coefficient of variance
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/fromnumeric.py:3579: FutureWarning: Dr
  return std(axis=axis, dtype=dtype, out=out, ddof=ddof, **kwargs)
/usr/local/lib/python3.7/dist-packages/numpy/core/fromnumeric.py:3438: FutureWarning: Dr
  return mean(axis=axis, dtype=dtype, out=out, **kwargs)
Year                        0.001720
Male upto 14 years          3.721161
Male 15-29 years            2.134853
Male 30-44 years            2.391892
Male 45-59 years            2.547482
Male 60 years and above     3.113340
```

```
Total Male                   2.290757
Female upto 14 years         3.713615
Female 15-29 years           3.518696
Female 30-44 years           3.764552
Female 45-59 years           3.812974
Female 60 years and above    4.003109
Total Female                 3.459873
Grand Total                  2.108102
dtype: float64
```

30	TAMIL NADU	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

```
totalCleaned.groupby('Year').agg(['mean','median'])
#mean and meadian after grouping by year
```

	Male upto 14 years		Male 15-29 years		Male 30-44 years		Male 45-59 years		Ma an
	mean	median	mean	median	mean	median	mean	median	me
Year									
2001	3.566667	0.0	47.573810	4.0	55.959524	4.0	36.685714	1.0	14
2002	3.109524	0.0	49.802381	4.0	59.073810	5.0	37.988095	2.0	15
2003	3.042857	0.0	51.659524	4.0	58.759524	4.0	38.359524	2.0	15
2004	3.492857	0.0	51.469048	4.0	61.547619	4.0	40.630952	2.0	15
2005	3.161905	0.0	51.959524	5.0	61.742857	5.0	41.038095	2.0	15
2006	2.842857	0.0	54.183333	5.0	66.211905	5.0	41.297619	2.0	15
2007	2.819048	0.0	55.823810	5.0	67.978571	5.0	45.507143	2.0	16
2008	2.773810	0.0	58.085714	5.0	70.283333	5.0	44.807143	3.0	15
2009	3.573810	0.0	56.538095	6.0	69.845238	4.0	46.657143	2.0	17
2010	3.904762	0.0	62.826190	6.0	72.485714	5.0	49.447619	3.0	18
2011	3.747619	0.0	63.342857	6.0	74.761905	5.0	48.826190	3.0	18
2012	2.666667	0.0	55.709524	4.5	67.823810	5.0	45.707143	2.0	18

```
totalCleaned.groupby('Year').agg(['mean','median'])["Grand Total"]
#displaying "Grand Total" column mean and meadian after grouping by year
```

Year	mean	median
2001	258.347619	29.5
2002	262.897619	32.5
2003	263.930952	30.5
2004	270.707143	26.5
2005	271.223810	34.0
2006	281.219048	34.0
2007	291.992857	33.0
2008	297.659524	39.5
2009	302.740476	41.0

```
df_crosstab = pd.crosstab(totalCleaned['Total Male'],totalCleaned['Total Female'],margins = F
print(df_crosstab)
print("mean \n",df_crosstab.mean())
```

```
Total Female  0     1     2     3     4     5     6     7     8     9     ...  \
Total Male                                                                ...
0              749    28     8    16    10    11     6     4     4     6  ...
1              139    17     3     3     5     1     1     0     0     0  ...
2               98    18     7     7     4     2     1     0     2     0  ...
3               81    17     9     7     3     0     0     0     1     1  ...
4               72    26    10     4     1     4     3     1     1     2  ...
...            ...   ...   ...   ...   ...   ...   ...   ...   ...   ...  ...
3638              0     0     0     0     0     0     0     0     0     0  ...
3763              0     0     0     0     0     0     0     0     0     0  ...
3799              0     0     0     0     0     0     0     0     0     0  ...
3968              0     0     0     0     0     0     0     0     0     0  ...
4111              0     0     0     0     0     0     0     0     0     0  ...

Total Female  3501  3522  3563  3582  3639  3653  3716  4083  4155  5020
Total Male
0                1     1     1     1     1     1     1     1     1     1
1                0     0     0     0     0     0     0     0     0     0
2                0     0     0     0     0     0     0     0     0     0
3                0     0     0     0     0     0     0     0     0     0
4                0     0     0     0     0     0     0     0     0     0
...            ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
3638             0     0     0     0     0     0     0     0     0     0
3763             0     0     0     0     0     0     0     0     0     0
3799             0     0     0     0     0     0     0     0     0     0
3968             0     0     0     0     0     0     0     0     0     0
4111             0     0     0     0     0     0     0     0     0     0

[873 rows x 595 columns]
mean
 Total Female
```

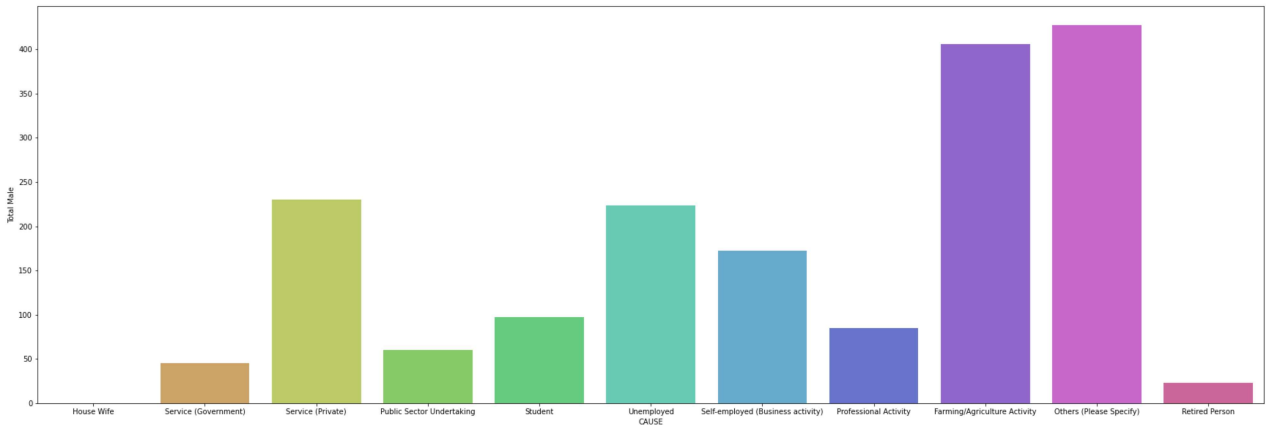
```
0       1.847652
1       0.342497
2       0.216495
3       0.187858
4       0.164948
          ...
3653    0.001145
3716    0.001145
4083    0.001145
4155    0.001145
5020    0.001145
Length: 595, dtype: float64
```

```
totalCleaned['CAUSE'].value_counts()
```

```
Others (Please Specify)             840
House Wife                          420
Service (Government)                420
Service (Private)                   420
Public Sector Undertaking           420
Student                             420
Unemployed                          420
Self-employed (Business activity)   420
Professional Activity               420
Farming/Agriculture Activity        420
Retired Person                      420
Name: CAUSE, dtype: int64
```

```
#Visualizing total males in each cause using barplot
fig, ax = plt.subplots(figsize=(30, 10))
sns.barplot(x = "CAUSE", y = "Total Male",palette = 'hls',ax=ax, data = totalCleaned,ci = Non
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ade2f610>
```



Therefore bar plot shows that under number of "Total Males" the cause has much higher rate in Others and Farming/agriculture activity and least in Retired person and null in house wife.

```
#Visualizing total females in each cause using barplot
fig, ax = plt.subplots(figsize=(30, 10))
sns.barplot(x = "CAUSE", y = "Total Female",palette = 'hls',ax=ax, data = totalCleaned,ci = N
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33adc580d0>
```

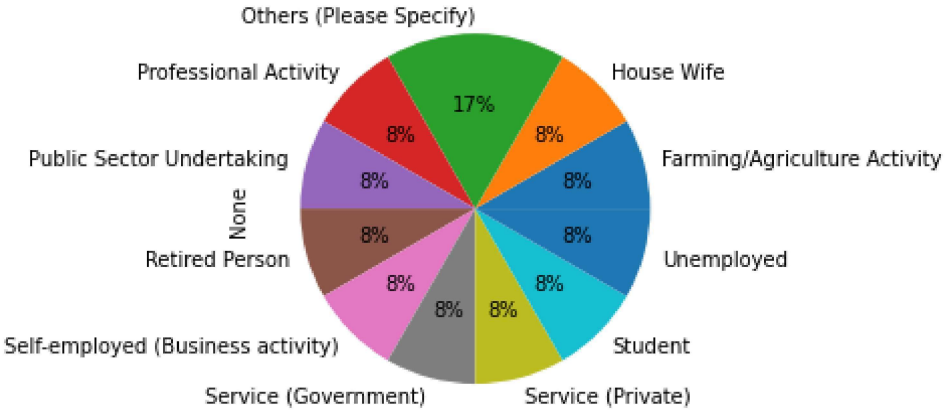
Therefore by using bar graph it shows that under number of "Total Females" the cause has much higher rate in house wife and lower rate in Retired person

```
print("Total cases from 2001-12: \n",totalCleaned.groupby("Year")["Grand Total"].sum())
```

```
Total cases from 2001-12:
 Year
2001    108506
2002    110417
2003    110851
2004    113697
2005    113914
2006    118112
2007    122637
2008    125017
2009    127151
2010    134599
2011    135585
2012    120488
Name: Grand Total, dtype: int64
```

```
totalCleaned.groupby('CAUSE').size().plot(kind='pie', autopct='%1.0f%%')
```

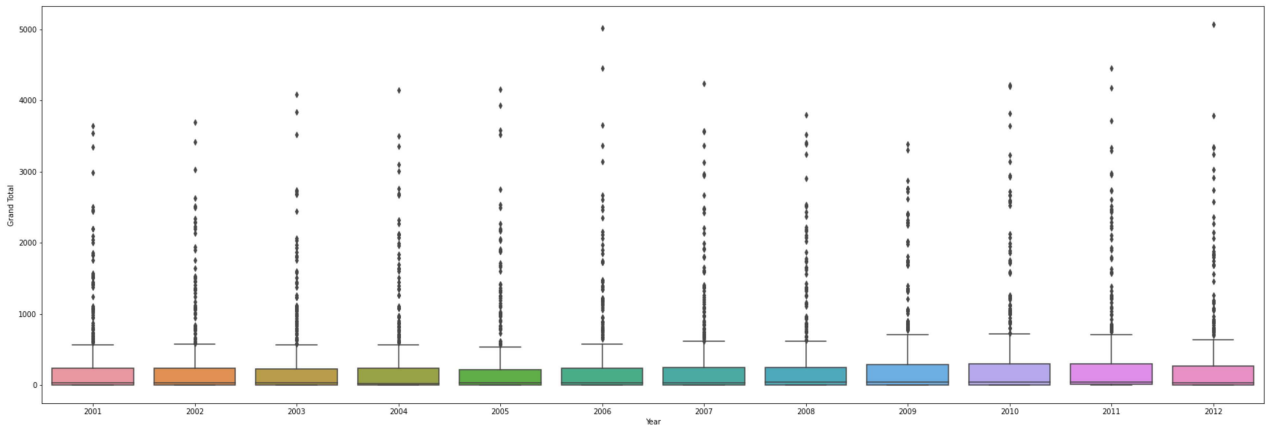
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ad770fd0>
```



This Pie chart shows the percentage in total by cause-wise where every cause has equal rate/percentage of 8% and under others its 18% which is the highest rate.

```
fig, ax = plt.subplots(figsize=(30, 10))
sns.boxplot(x='Year',y='Grand Total',data=totalCleaned)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ad692390>
```



Double-click (or enter) to edit

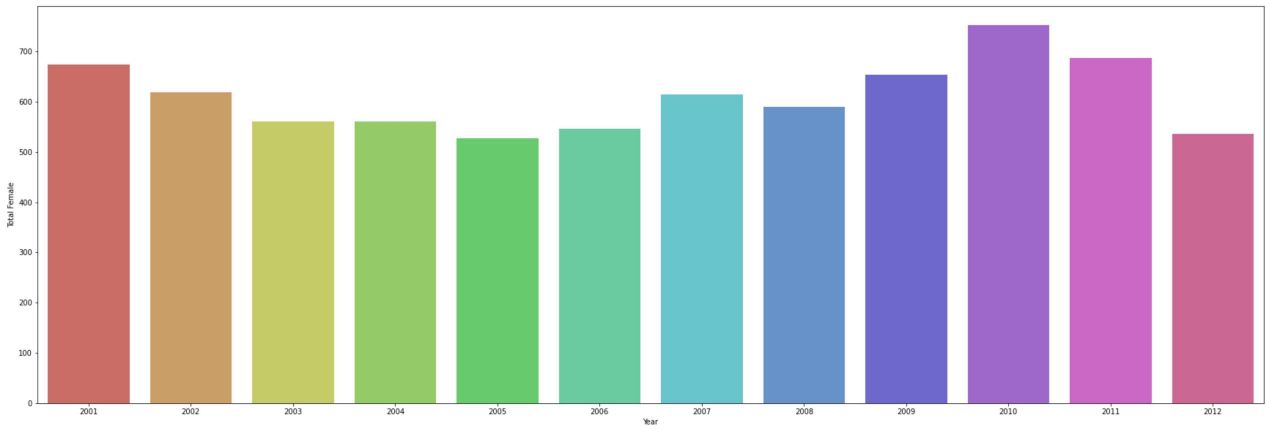
```
#Visualizing Total from 2001 to 2012 using line plot
totalCleaned.groupby("Year")["Grand Total"].sum().plot(kind="line")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ad515410>
```

by using line plot we can see that there is a continuous increase from 2001 to 2011 and sudden drop down in 2012

```
#Visualizing total Selfemployed women in each year
fig, ax = plt.subplots(figsize=(30, 10))
sns.barplot(x = "Year", y = "Total Female",palette = 'hls',ax=ax, data =TotalSelfEmployed,ci
```

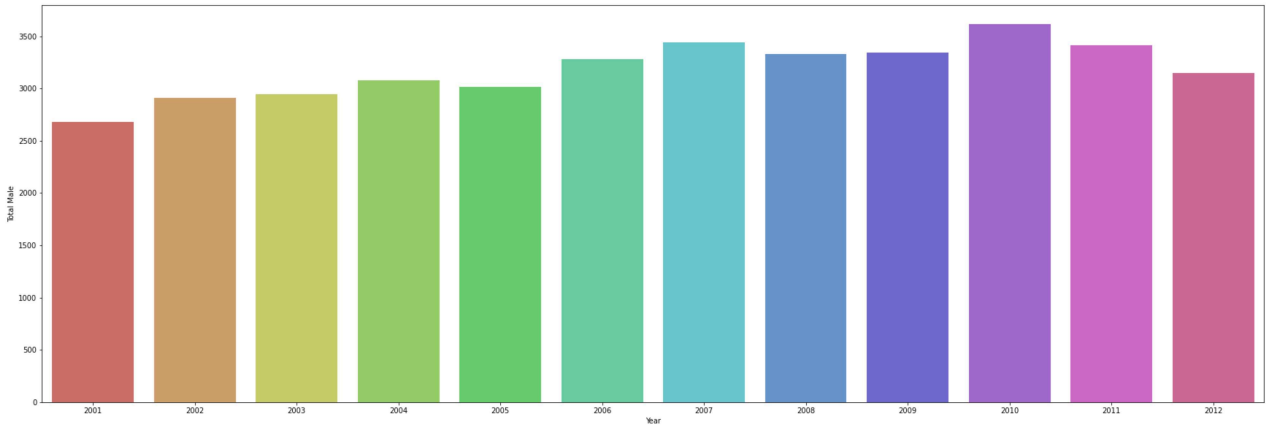
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ad492190>
```



This bar plot shows the self employed womens are more in 2010 and least in 2012

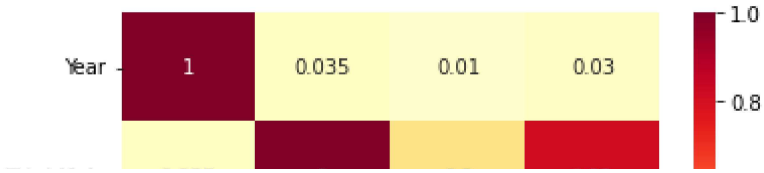
```
#Visualizing total Selfemployed women in each year
fig, ax = plt.subplots(figsize=(30, 10))
sns.barplot(x = "Year", y = "Total Male",palette = 'hls',ax=ax, data =TotalSelfEmployed,ci =
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ad42b590>
```

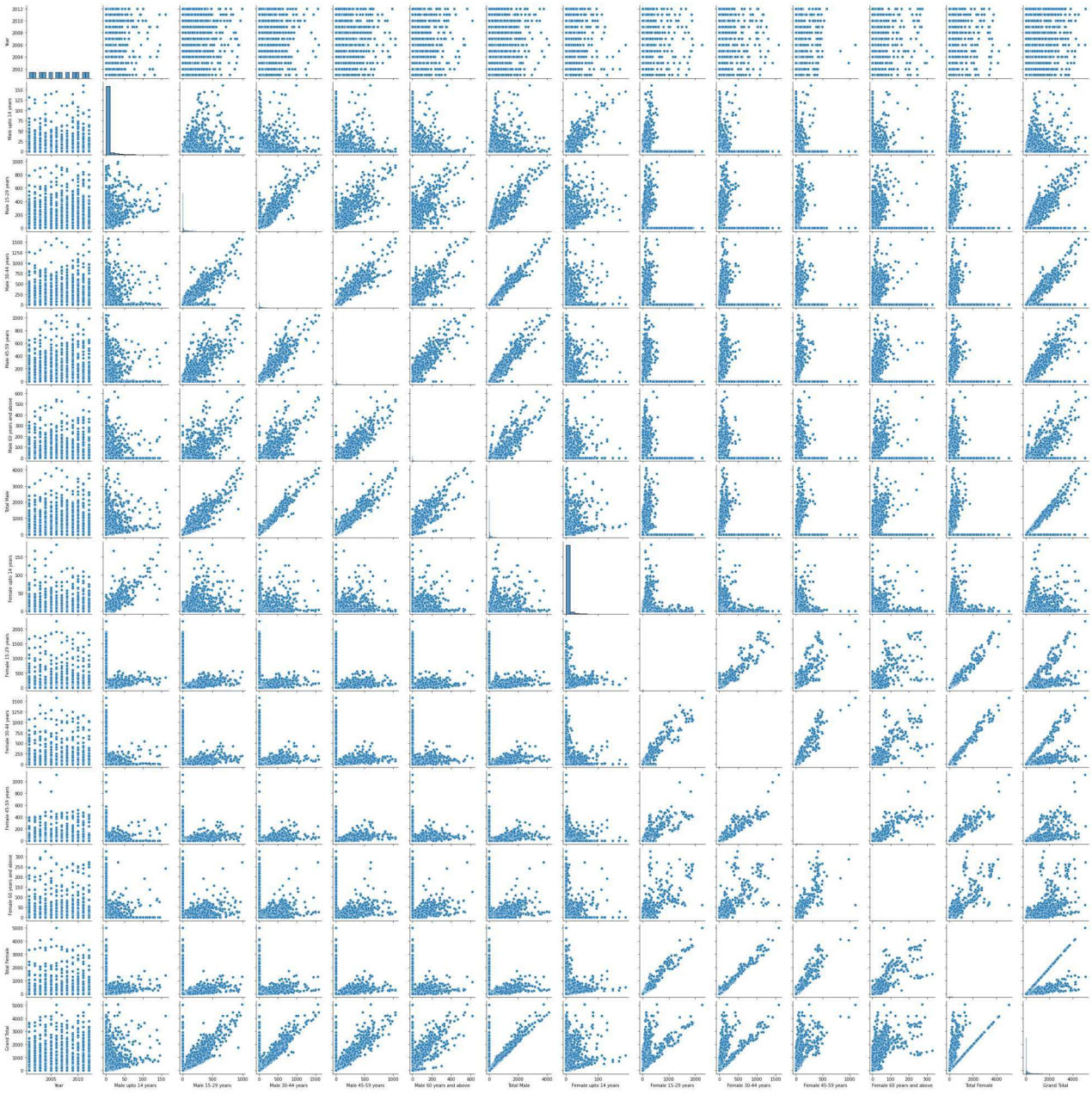


The Bar plot shows Total Male suicide in various years

```
#plotting the correlation matrix of Year,Total Male,Total Female and Grand Total
sns.heatmap(totalCleaned[['Year','Total Male','Total Female','Grand Total']].corr(), annot=Tr
plt.show()
#spurious hit
```




```
sns.pairplot(data = totalCleaned)
plt.show()
#plot of all columns
#spurious hit
```

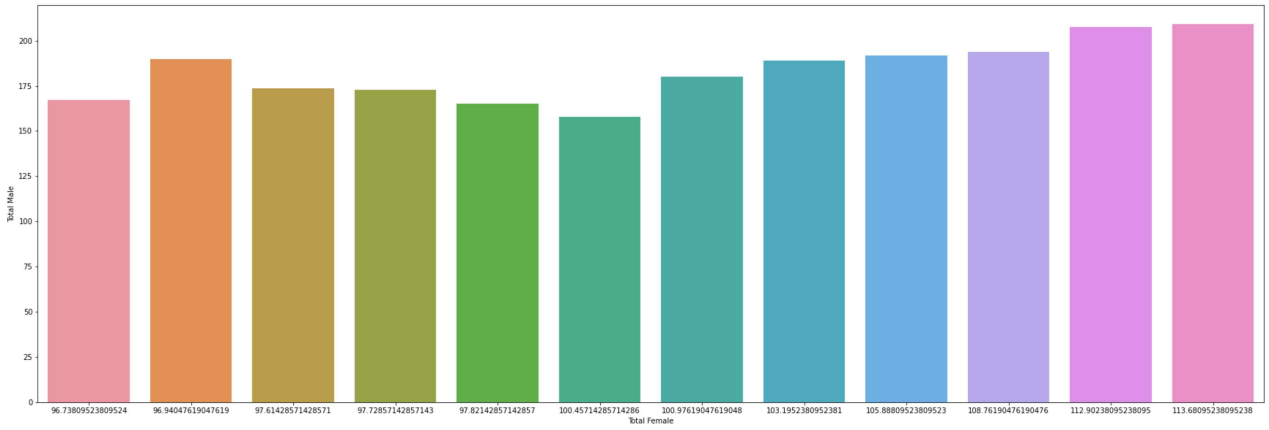

Year
Male upto 14 years
Male 15-29 years
Male 30-44 years
Male 45-59 years
Male 60 years and above
Total Male
Female upto 14 years
Female 15-29 years
Female 30-44 years
Female 45-59 years
Female 60 years and above
Total Female
Grand Total

```
totalCleaned.columns
```

```
Index(['STATE/UT', 'CAUSE', 'Year', 'Male upto 14 years', 'Male 15-29 years',
       'Male 30-44 years', 'Male 45-59 years', 'Male 60 years and above',
       'Total Male', 'Female upto 14 years', 'Female 15-29 years',
       'Female 30-44 years', 'Female 45-59 years', 'Female 60 years and above',
       'Total Female', 'Grand Total'],
      dtype='object')
```

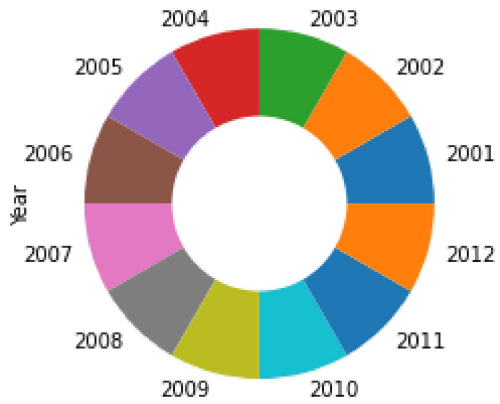
```
fig, ax = plt.subplots(figsize=(30, 10))
df_sample=totalCleaned.groupby("Year").agg("mean")
sns.barplot(x='Total Female',y='Total Male',data=df_sample,ax=ax)
#
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f339fa59090>
```



Ploting mean of Total male and Toatal Female after grouping them by year

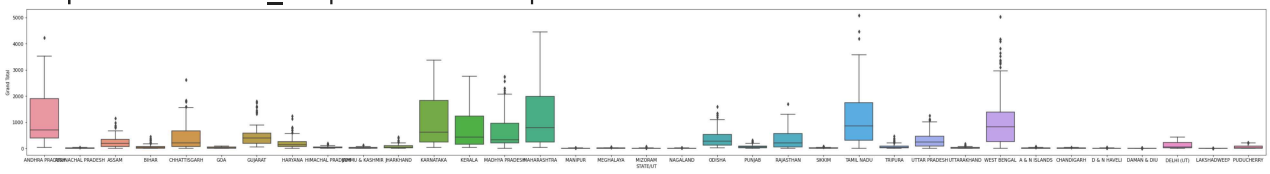
```
totalCleaned.Year.value_counts().plot.pie().add_artist(plt.Circle((0, 0), 0.50, fc='white'))
plt.show()
#
```



Doughnut plot based on years

```
fig, ax = plt.subplots(figsize=(50, 6))
sns.boxplot(y='Grand Total',x='STATE/UT',ax=ax,data=totalCleaned)
#
```

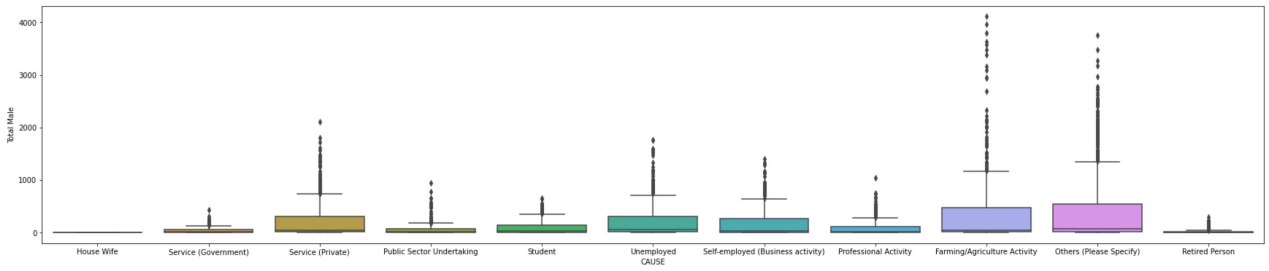
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33a9f39a50>
```



Box plot on Grand total and State/UT

```
fig, ax = plt.subplots(figsize=(30, 6))
sns.boxplot(y='Total Male',x='CAUSE',data=totalCleaned)
#
```

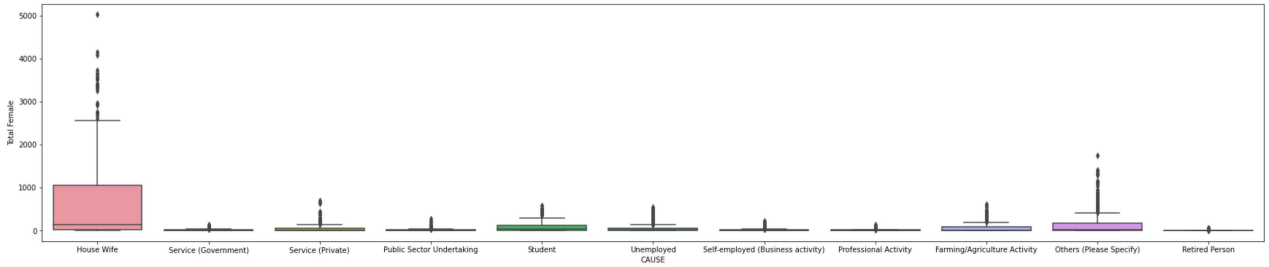
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33aa3f5090>
```



Ploting the cause of total male

```
fig, ax = plt.subplots(figsize=(30, 6))
sns.boxplot(y='Total Female',x='CAUSE',data=totalCleaned)
#
```

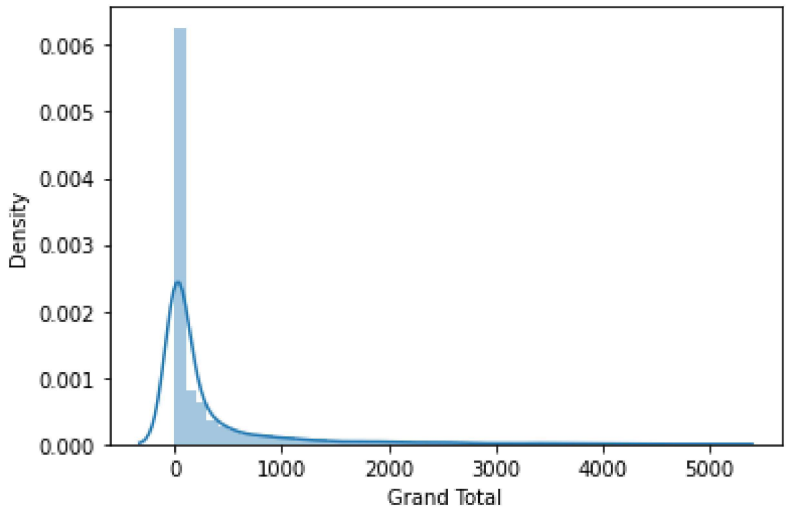
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f33a66c6a90>
```



Ploting the cause of total Female

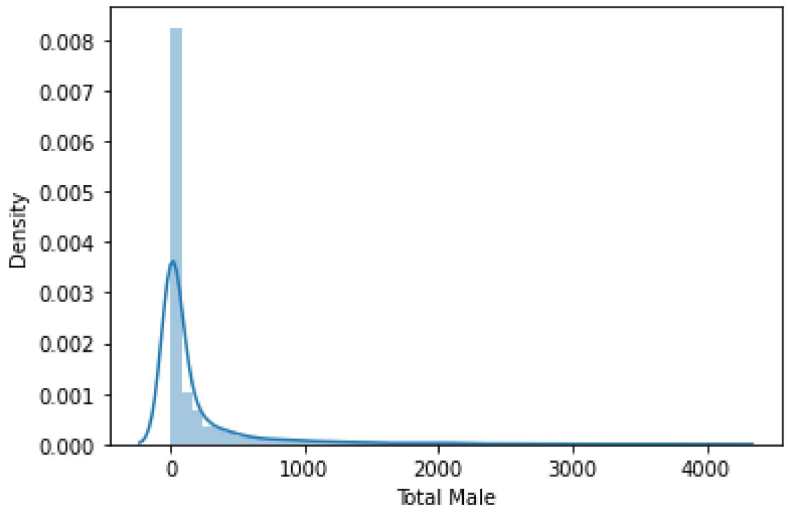
```
sns.distplot(totalCleaned["Grand Total"])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarnin
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f33ad28e410>
```




```
sns.distplot(totalCleaned["Total Male"])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarnin
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f33aa5adfd0>
```

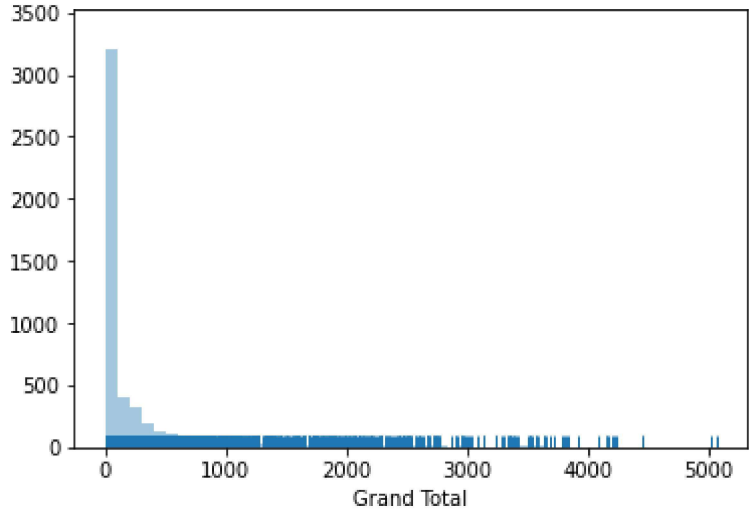



```
sns.distplot(totalCleaned["Total Female"])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarnin
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f33a5e00b50>
```


```
sns.distplot(totalCleaned["Grand Total"], kde=False, rug=True);
```

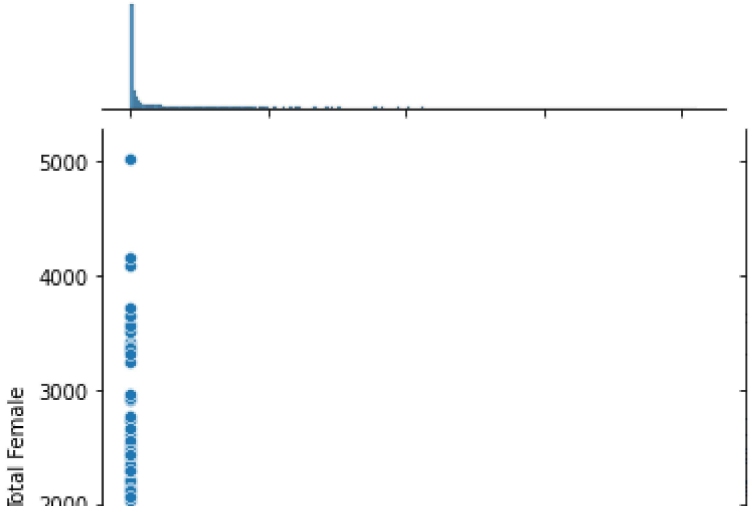
```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarnin
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2103: FutureWarnin
  warnings.warn(msg, FutureWarning)
```



ScatterPlot

```
sns.jointplot(totalCleaned["Total Male"], totalCleaned["Total Female"])
```

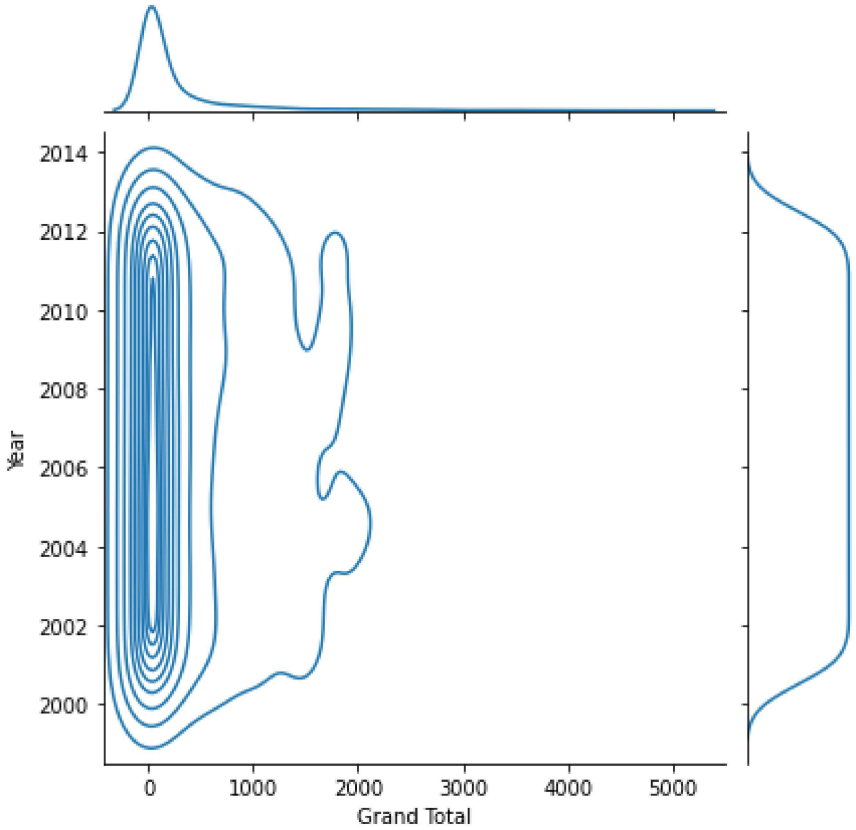
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: P
  FutureWarning
<seaborn.axisgrid.JointGrid at 0x7f33a9603e50>
```



Kernel Density Estimation

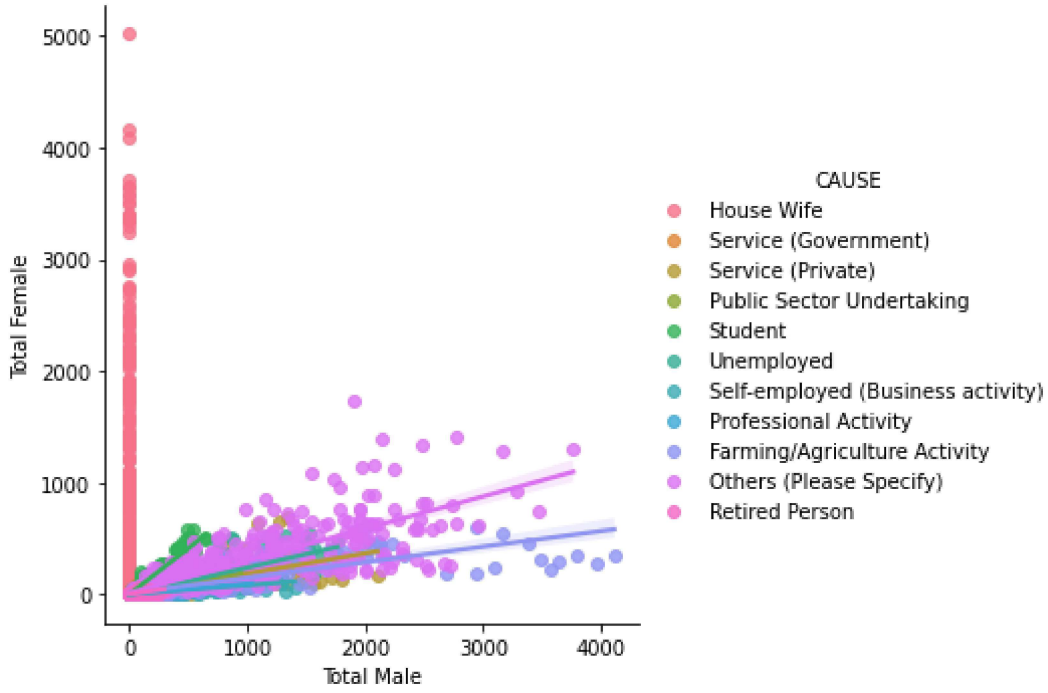
```
sns.jointplot(totalCleaned['Grand Total'], totalCleaned['Year'], kind="kde")
# kernel density estimation (KDE) is a non-parametric way to estimate the probability density
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: P
  FutureWarning
<seaborn.axisgrid.JointGrid at 0x7f33aa213690>
```



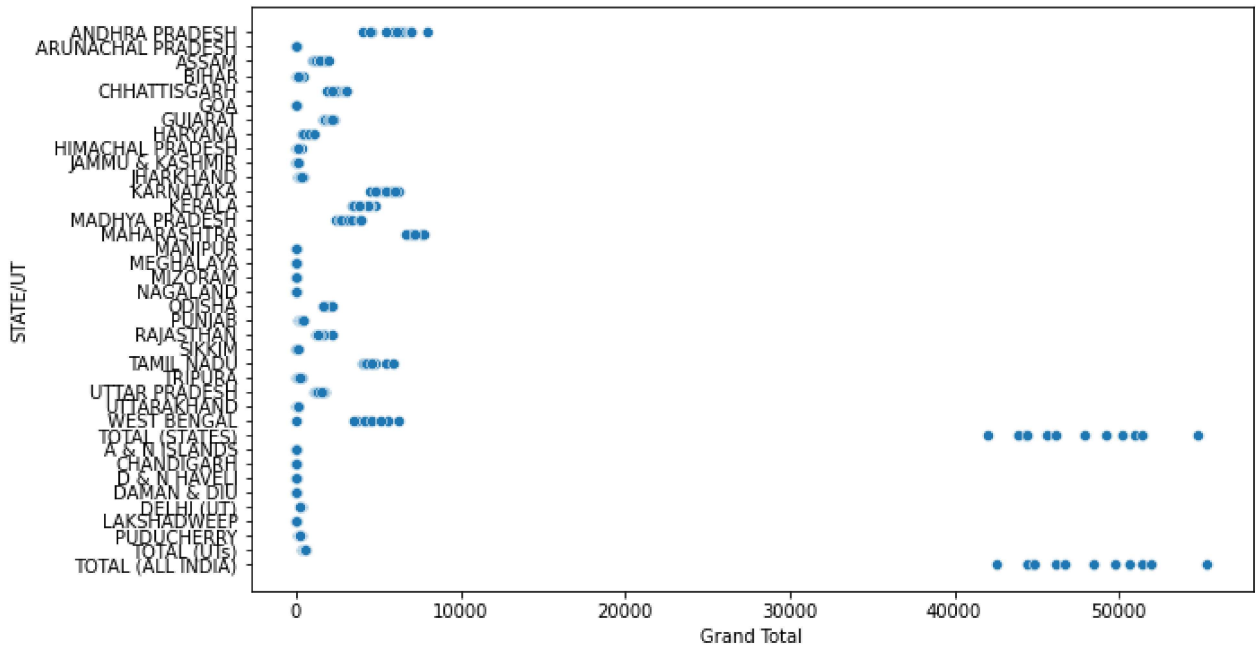
```
# sns.pairplot(totalCleaned[['CAUSE', 'Total Male', 'Total Female']])

sns.lmplot(x="Total Male", y="Total Female",hue="CAUSE", data=totalCleaned)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f9572228950>
```




```
#ploting the self employed people who suicided in each state
fig, ax = plt.subplots(figsize=(10, 6))
sns.scatterplot(x=TotalSelfEmployed["Grand Total"],y=TotalSelfEmployed['STATE/UT'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f95715603d0>
```



```
TotalSelfEmployed["CAUSE"].value_counts()
```

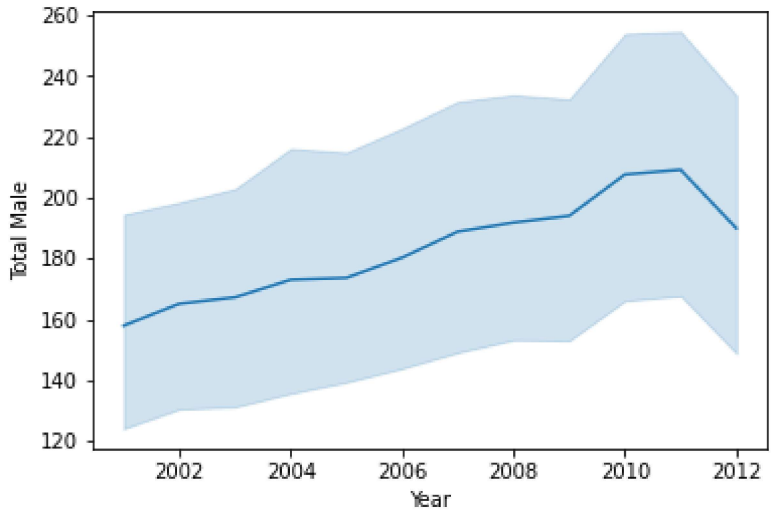
```
Total Self-employed    456
Name: CAUSE, dtype: int64
```

```
TotalSelfEmployed.value_counts()
```

```
STATE/UT           CAUSE                     Year  Male upto 14 years  Male 15-29 years
ARUNACHAL PRADESH  Others (Please Specify)   2003  0                   0
LAKSHADWEEP        Others (Please Specify)   2008  0                   0
NAGALAND           Others (Please Specify)   2010  0                   0
CHANDIGARH         Others (Please Specify)   2006  0                   0
A & N ISLANDS      Others (Please Specify)   2003  0                   0

GUJARAT            Service (Government)      2007  0                   29
                                             2006  0                   8
                                             2005  0                   21
                                             2004  0                   15
WEST BENGAL        Unemployed                2012  0                   0
Length: 5016, dtype: int64
```

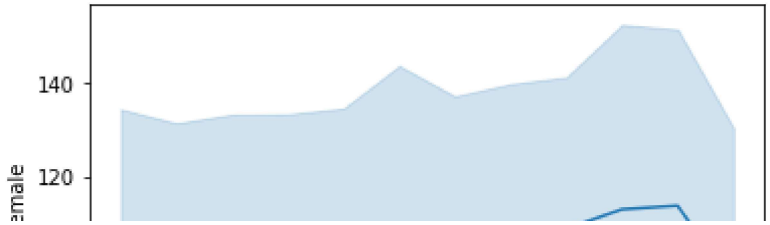
```
sns.lineplot(x=TotalSelfEmployed['Year'],y=TotalSelfEmployed['Total Male'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f956d090150>
```




```
sns.lineplot(x=TotalSelfEmployed['Year'],y=TotalSelfEmployed['Total·Female'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f957156e210>
```



Conclusion

"Total Males" the cause has much higher rate in Others and Farming/agriculture activity and least in Retired person and null in house wife. there is a continuous increase from 2001 to 2011 and sudden drop down in 2012 self employed womens are more in 2010 and least in 2012. The postive trend during 2007-2008 can be due to Global Ecomnomic crisis . Andra Pradesh has been the state with most number of Suicides.