A DSBDA Mini Project Report

on

"Suicides In India"

Submitted to the

Pune Institute of Computer

Technology,Pune

In partial fulfillment for the award of the

Degree of

Bachelor of Engineering

in

Information Technology

By

Anmol koul   :T190058508
Prasad gawade   :T190058570
Bhagyashree Karnavat   :T190058501
Shreyas Khandait   :T190058606

Under the Guidance of

Prof. Ravi Murumkar

Department Of Information Technology

Pune Institute of Computer Technology College of Engineering

Sr. No 27, Pune- Satara Road,   Dhankawadi, Pune - 411 043.

A.Y. 2021-2022

CERTIFICATE

This is to certify that the mini project report entitled **"Suicides In India"**

Submitted by

Anmol koul : T190058508

Prasad Gawade: T190058570

Bhagyashree Karnavat: T190058501

Shreyas khandait: T190058606

from TE10 is a record of bonafide work carried out by him/her under the supervision and guidance of Prof. Ravi Murumkar and it is approved for the partial fulfillment of the requirement of TE (Information Technology B Engineering) – 2019 course of Savitribai Phule Pune University, Pune in the academic year 2021-2022.

Prof. Ravi Murumkar   Dr. A. S. Ghotkar

Internal guide   Head of the
                                                                              Department

Prof. Ravi murumkar

Internal guide

Date:

Place:
Date:

Acknowledgement

I would like to express my special thanks of gratitude to my teacher (Ravi Murumkar) as well as our Head of department (Dr. A. S. Ghotkar) who gave me the golden Opportunity to do this wonderful project on the topic (suicides in India), which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them.

Secondly I would also like to thank my senior colleagues and friends who helped me a lot in finalizing this project within the limited time frame.

I would like to appreciate my team members who individually worked a lot and successfully completed our mini-project.

Resources that we have used for completing our project stackoverflow, tutorialspoint, javatpoint, we are glad to get advice from seniors and lab teacher.

Thank you !

Abstract

The suicide rate in India is 10.3. In the last three decades, the suicide rate has increased by 43% but the male female ratio has been stable at 1.4 : 1. Majority (71%) of suicide in India are by persons below the age of 44 years which imposes a huge social, emotional and economic burden. Fifty four articles on suicides have been published in IJP. Several studies reveal that suicidal behaviours are much more prevalent than what is of cially reported. Poisoning, hanging and self immolation (particularly women) were the methods to commit suicide. Physical and mental illness, disturbed interpersonal relationships and economic difficulties were the major reasons for suicide. The vulnerable population was found to be women, students, farmers etc. A social and public health response in addition to a mental health response is crucial to prevent suicidal behaviour in India.

Key words: Suicide, India, Risk factors

Index

Introduction

Suicide is the final outcome of complex interactions of biological, genetic, psychological, sociological and environmental factors. It is an increasingly important public health issue: from 1990 to 2010 the number of global suicides increased by 32%. It is particularly important among young adults 15 to 49 years of age among whom it accounts for 4.8% of all female deaths and 5.7% of all male deaths.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120287/ - B1 Eighty-four percent of global suicides occur in low and middle-income countries (LMICs); India and China alone account for 49% of global suicides. There is substantial variability both in the prevalence of suicide and in the factors that influence the occurrence of suicide between geographic regions, cultures, and over time, so country-specific analyses are needed to develop targeted suicide prevention efforts.

The World Health Organization (WHO) estimates that of the nearly 900,000 people who die from suicide globally every year, 170,000 are from India. However, India's National Crime Records Bureau (NCRB) – which report official suicide rates based on police reports – estimated only 135,000 suicides in 2011. One possible reason for under-reporting of suicide by the NCRB is that suicide remains a crime in India.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120287/ - B6 Detailed independent verbal autopsy investigations of all unnatural deaths in several rural areas of the country report suicide rates that are up to five-fold higher than the official national average. Despite the importance of suicide both as a health problem and as a social problem, the issue receives little attention from Indian policy makers. This neglect of the issue is reinforced by cultural influences, religious sanctions, stigmatization of the mentally ill, competing political imperatives, and socio-economic factors

Literature Review

3.1. Prevalence of suicide

A community-based surveillance system using verbal autopsy covering a population of 108,000 in a rural development block of Tamil Nadu (one of the more highly developed states in southern India) has been operating for several decades. Several reports have appeared using these data. The prevalence of suicide for the period from 1994 to 1999 was 95 per 100,000 population (annual rates ranged from 84 to 106 per 100,000) accounting for 8 to 12% of all deaths. From 2000 to 2002 the average suicide rate was 92 per 100,000 (112 in men and 72 in women) accounting for 9.8 to 11.4% of all deaths. For the period from 1998 to 2004 the average suicide rate was 82 per 100,000. In young people 10 to 19 years of age, over the 10-year period from 1992 to 2001 the average male suicide rate was 58 per 100,000 (accounting for 25% of all deaths) and the average suicide rate for women was 148 per 100,000 (accounting for 50 to 75% of all deaths).https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120287/ - B26 Over the same 10-year period, the mean suicide rate in persons 55 years of age or older was 189 per 100,000.

3.2. Methods of suicide

Hanging was the most frequently reported method of suicide in most of the studies, accounting for 10 to 72% of all suicides. The second most frequently reported method was self-poisoning (often ingestions of organophosphate pesticides), which accounted for 16 to 49% of all suicides. The proportion of all suicides attributed to drowning ranged from 3 to 39% and the proportion attributed to burning or self-immolation ranged from 6 to 57%. Other reported methods of suicide include jumping off heights (0.5 to 2% of all suicides being run over by a train (6 to 13% of all suicides)and using a firearm (3% of all suicides).

Some studies report gender-based differences in method preference. A community-based surveillance study by Prasad and colleagues reported in 2006 found that significantly more women chose drowning and burning as modes of suicide than men, while significantly more men chose hanging.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120287/ - B36 A similar pattern of gender-based method preference was reported by Abraham and colleagues among persons 55 years of age and older. Other studies report higher rates of suicides by hanging in males than females, a predominance of males in suicide decedents who use other violent methods(e.g., jumping in front of a train), and a predominance of females among suicides by self-immolation.

3.3 Discussion

Our review clearly indicates that the young in India are at a high risk of suicide, a finding that is consistent with the Registrar General mortality study and the NCRB data. In the Registrar General mortality survey, 40% of suicide deaths in men and 56% in women occurred between the ages of 15 and 29. A 15-year-old individual in India had a cumulative risk of 1.3% of dying by suicide before the age of 80.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4120287/ - B6

Although our review could not demonstrate a clear gender difference in risk for suicide, many of the reviewed studies reported an overall male excess when combining suicides across all age groups. However, several of the studies also reported a higher prevalence of suicide in females than males in the younger age groups. This gender difference is in keeping with that seen in other Asian countries. Unlike in most European countries and the United States, where male suicides far outnumber female suicides (3:1 to 4:1),in most Asian countries the ratio is much lower, and in China the ratio actually falls below 1:1 (i.e., there are more female suicides than males suicides).The higher rates of female suicide in Asian countries may be linked to the position of women in the traditionally patriarchal societies of Asia. Niaz and Hassan argue that "women's lack of empowerment and both financial and emotional dependence have restricted their self-expression and choices in life. This, along with family, social and work pressures, has a definite impact on women's mental health."

The quality of the information about suicide in India is quite limited, but the picture drawn from the eclectic mix of studies identified in this review indicate that it is an important and growing public health problem that is not being given sufficient attention by the government or the society at large. Compared to the pattern of suicides in high-income countries, suicide in India is more prevalent in women (particularly young women), is much more likely to involve ingestion of pesticides, and is more closely associated with poverty and less

closely associated with mental illness. Initial targets for suicide prevention in India should include: (a) decriminalization of suicide; (b) restricting access to pesticides; and (c) improving the monitoring of fatal and non-fatal suicidal behavior.

3. OBJECTIVIVE

In this mini-project we wanted to develop an efficient system that analyses the data given from the dataset.

The main objective of the project is to make a detailed analyses by studying the graphical interface and draw a conclusion about what age group, which area, what are the main causes that really causes suicides in India .

4. DESIGN AND IMPLEMENTATION

a) Dataset Description :

"Suicide In India 2020-2021.csv" This dataset can be used for analysis of suicides that took place in India during the time period of 2001-2012 .

Context
This data set contains yearly suicide detail of all the states/u.t of India by various parameters from 2001 to 2012. Content
Time Period: 2001 - 2012 Granularity: Yearly Location: States and U.T's of India.

Parameters:
a) Suicide causes
b) Education status
c) By means adopted
d) Professional profile
e) Social status

Columns:
   1 State

   2 Year

   3 Type_code

   4 Type

   5 Gender

6 Age_Group

7 Total

Implementaion

Importing Libraries and Dataset

```
In [27]: #IMPORTING LIBRARIES
         import pandas as pd  #data analysis
         import numpy as np   #multi-dimensional arrays
         import matplotlib.pyplot as plt   #embedding plots
         import seaborn as sns    #data visualisation lib to plot graphs
         import warnings
         import os   #directory manipulation

         warnings.filterwarnings('ignore')   #ignores warnings
```

```
In [28]: data=pd.read_csv('Suicides_in_India_2001-2012 (1).csv')
```

Pre-Processing The Data

PRE-PROCESSING THE DATA AVAILABLE

```
In [29]:  #SIZE OF DATA:
          print("Available data size (rows,columns):")
          data.shape   #The shape of an array is the number of elements in each dimension

          Available data size (rows,columns):

Out[29]:  (237519, 7)
```

Display The Dataset:

```
In [30]:  data.head()   #displays top 5 rows
Out[30]:
```

	State	Year	Type_code	Type	Gender	Age_group	Total
0	A & N Islands	2001	Education_Status	Diploma	Female	NaN	0.0
1	A & N Islands	2001	Education_Status	No Education	Female	NaN	4.0
2	A & N Islands	2001	Education_Status	Post Graduate and Above	Female	NaN	0.0
3	A & N Islands	2001	Education_Status	Middle	Female	NaN	33.0
4	A & N Islands	2001	Education_Status	Graduate	Female	NaN	0.0

Random Sample of data

```
In [31]:  data.sample(10)
Out[31]:
```

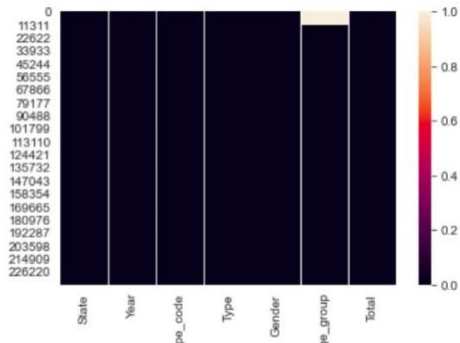|	State	Year	Type_code	Type	Gender	Age_group	Total
172459	Madhya Pradesh	2007	Causes	Fall in Social Reputation	Female	52.0	0.0
104596	Andhra Pradesh	2012	Causes	Physical Abuse (Rape/Incest Etc.)	Female	40.0	8.0
192919	A & N Islands	2005	Causes	Ideological Causes/Hero Worshipping	Male	74.0	0.0
181913	Puducherry	2011	Causes	Cancer	Male	49.0	1.0
133176	Nagaland	2001	Professional_Profile	Public Sector Undertaking	Male	33.0	0.0
11578	West Bengal	2002	Education_Status	Graduate	Male	NaN	180.0
94294	Rajasthan	2012	Means_adopted	By Jumping from (Building)	Female	29.0	1.0
233974	Uttar Pradesh	2004	Causes	Other Prolonged Illness	Female	60.0	13.0
8233	Puducherry	2005	Social_Status	Married	Female	NaN	124.0
199546	Chandigarh	2007	Causes	Fall in Social Reputation	Male	70.0	0.0

Checking For Missing Values

Checking Dataset For Missing values

```
In [33]: print(data.isna().sum())      #count NaN values in data frame

         sns.heatmap(data.isna())      #Plot rectangular data as a color-encoded matrix
```

```
State          0
Year           0
Type_code      0
Type           0
Gender         0
Age_group  11856
Total          8
dtype: int64
```

Out[33]: <matplotlib.axes._subplots.AxesSubplot at 0x1bf1c0f5fd0>


```
In [34]: age_med = data['Age_group'].median()
         age_med
```

Out[34]: 37.0

```
In [35]: data.head(10)
```

Out[35]:

	State	Year	Type_code	Type	Gender	Age_group	Total
0	A & N Islands	2001	Education_Status	Diploma	Female	NaN	0.0
1	A & N Islands	2001	Education_Status	No Education	Female	NaN	4.0
2	A & N Islands	2001	Education_Status	Post Graduate and Above	Female	NaN	0.0
3	A & N Islands	2001	Education_Status	Middle	Female	NaN	33.0
4	A & N Islands	2001	Education_Status	Graduate	Female	NaN	0.0
5	A & N Islands	2001	Education_Status	Hr. Secondary/Intermediate/Pre-Universit	Female	NaN	0.0
6	A & N Islands	2001	Education_Status	Primary	Female	NaN	7.0
7	A & N Islands	2001	Education_Status	Matriculate/Secondary	Female	NaN	6.0
8	A & N Islands	2001	Education_Status	Matriculate/Secondary	Male	NaN	10.0
9	A & N Islands	2001	Education_Status	Post Graduate and Above	Male	NaN	0.0

```
In [36]: data['Age_group'].fillna(age_med,inplace = True)
```

```
In [38]:  print(data.isna().sum())      #count NaN values in data frame

          sns.heatmap(data.isna())      #Plot rectangular data as a color-encoded matrix
```

```
State         0
Year          0
Type_code     0
Type          0
Gender        0
Age_group     0
Total         8
dtype: int64
```

Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x1bf1ed31bb0>
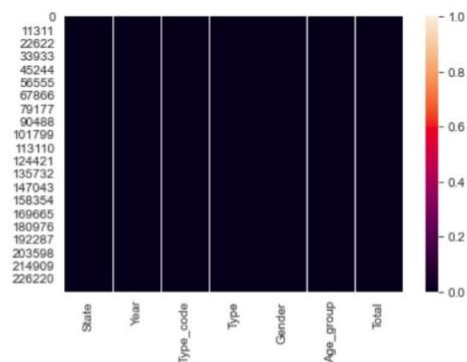
WE CAN SEE FROM ABOVE RESULT THAT THERE ARE NO MISSING VALUES IN THE DATASET

Visualization and Analysis

Bar Graph and Histograms

"ggplot()" initializes a ggplot object. It is used to declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all subsequent layers unless specifically overridden.

SUICIDES RATE OVER YEARS

```
In [41]: data.groupby('Year').sum()['Total'].plot(kind='bar',figsize=(13,7),title ="Year wise suicides frequency");
         # summing up the amounts year wise
```
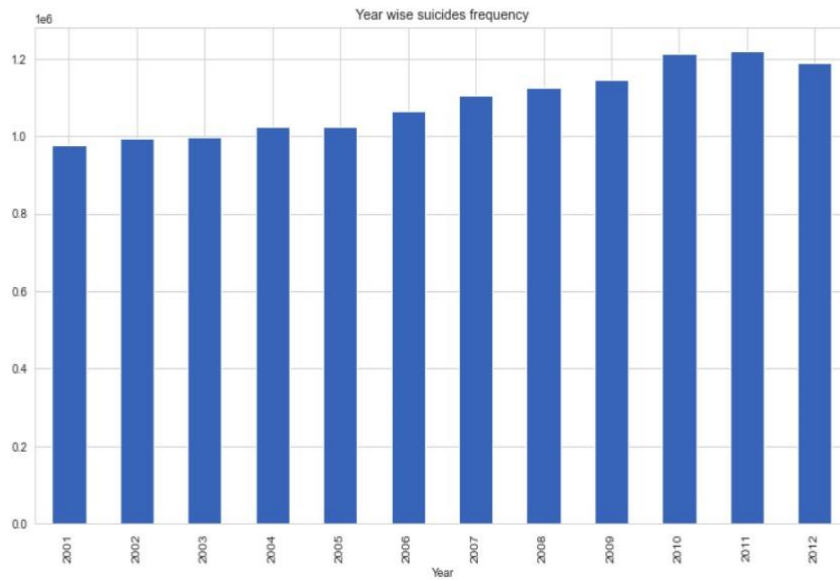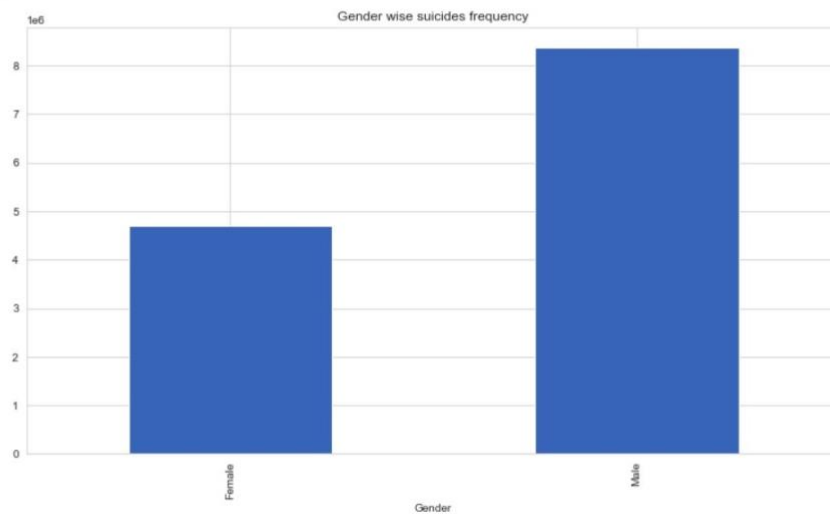


```
In [42]: #LINE CHART
         year = data.groupby('Year').sum()['Total'].plot(kind='line', figsize=(6,6), title = 'Suicide rate per year');
         year.set_xlabel('\nYear')
         year.set_ylabel('Count\n')
         sns.set_style('whitegrid');
```
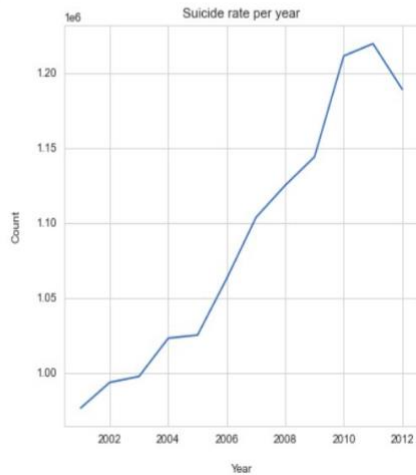

Gender wise suicides frequency

Gender wise suicides frequency

In [43]: `data.groupby('Gender').sum()['Total'].plot(kind="bar",figsize=(13,7),title ="Gender wise suicides frequency")`


Gender wise suicides frequency

Educational background vs Total Suicides vs Gender

Analysis of new data frames
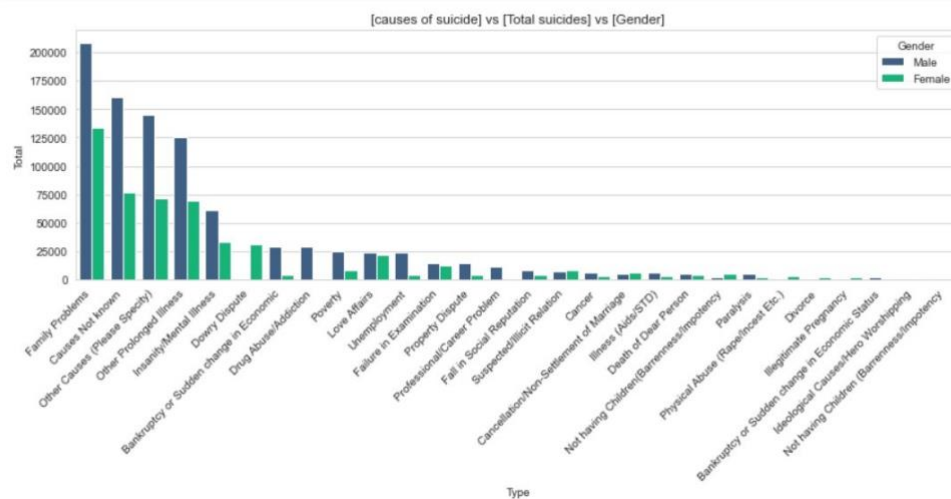
Educational background vs Total suicides vs Gender

```
In [47]: plt.figure(figsize=(12,6))
         sns.barplot(x='Type',y='Total',hue='Gender',data=eduSort,palette='viridis')
         plt.xticks(rotation=45,ha='right')
         plt.title('[Educational background] vs [Total suicides] vs [Gender]')
         plt.tight_layout()
```
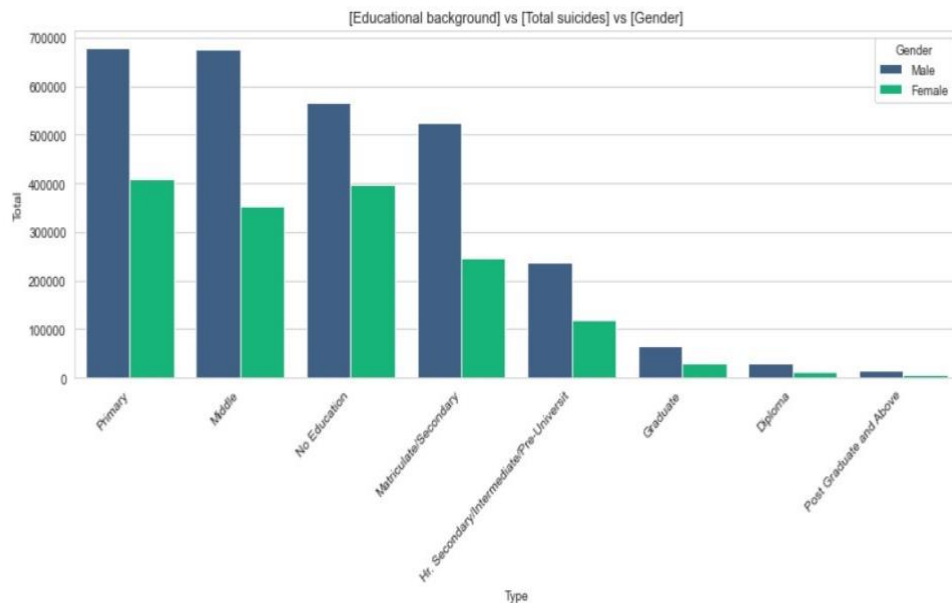

Causes of suicide vs total suicides vs Gender

Causes of suicide vs Total suicides vs Gender

```
In [48]: plt.figure(figsize=(12,6))
         sns.barplot(x='Type',y='Total',hue='Gender',data=causeSort,palette='viridis')
         plt.xticks(rotation=45,ha='right')
         plt.title('[causes of suicide] vs [Total suicides] vs [Gender]')
         plt.tight_layout()
```

Professional Status vs Total suicides vs Gender

Professinal Status vs Total suicides vs Gender

```
In [49]: plt.figure(figsize=(12,6))
         sns.barplot(x='Type',y='Total',hue='Gender',data=profSort,palette='viridis')
         plt.xticks(rotation=45,ha='right')
         plt.title('[Professional Status] vs [Total suicides] vs [Gender]')
         plt.tight_layout()
```


[Professional Status] vs [Total suicides] vs [Gender]
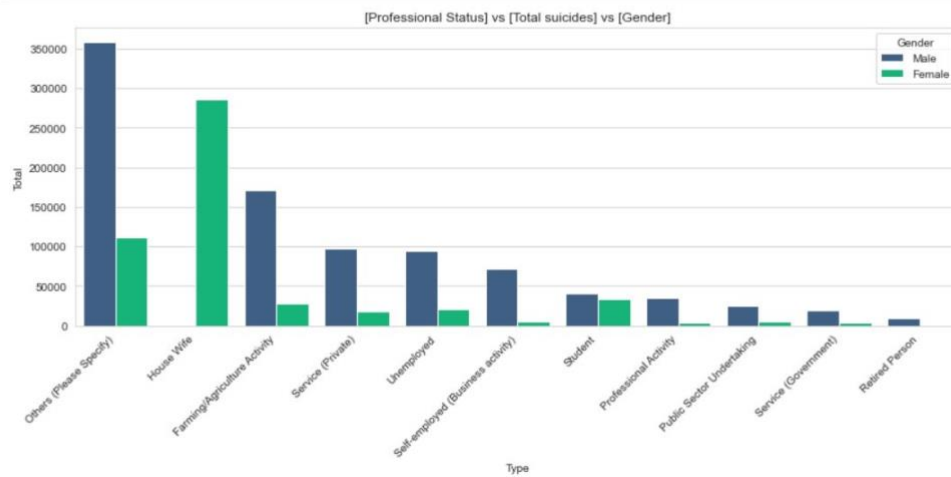
Suicides due to Addiction

Suicides due to addictions

```
In [52]: data1 = data.loc[(data.Type == 'Drug Abuse/Addiction')].groupby(['State'])['Total'].sum().reset_index().sort_va
         ('Total',ascending=False).head(60)
         data1.set_index(['State'])
         data1=data1.set_index(['State'])
         plt.subplots(figsize=(15,10))
         g = sns.barplot(x='Total',y=data1.index,data=data1,palette="Reds_d",).set_title('Drug Abuse/Addiction')
         plt.xticks(rotation=30)
         plt.xlabel('Total number of suicides')
```

Out[52]: Text(0.5, 0, 'Total number of suicides')
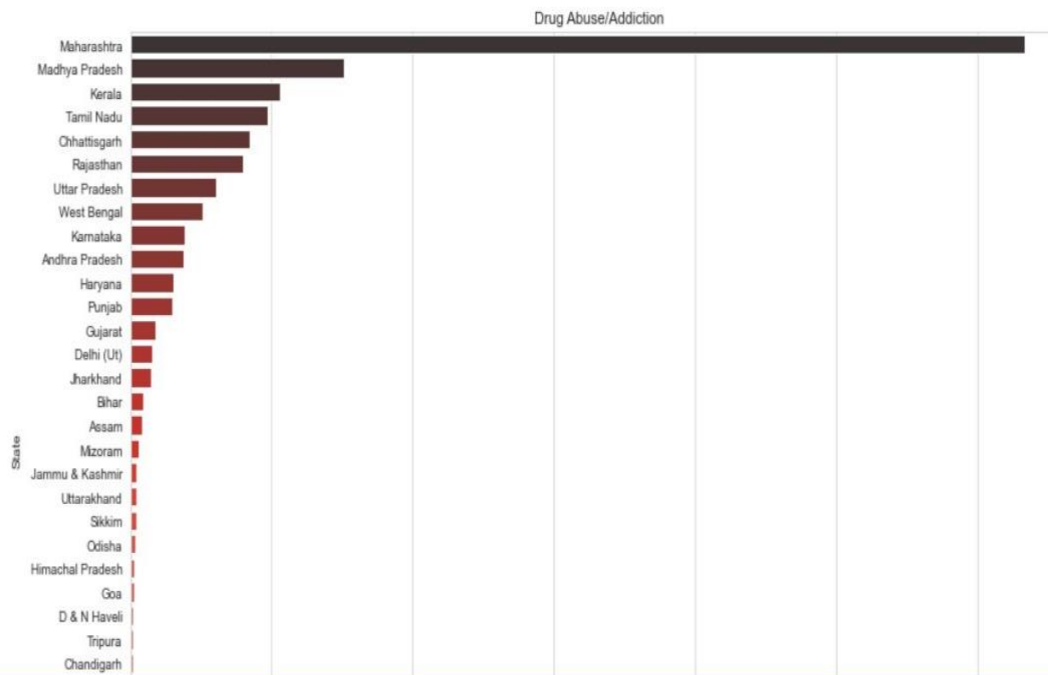
Suicides due to family problems

```
In [55]: causesType=data[data['Type_code']=='Causes']
         causeTypeGrp=causesType.groupby(['State','Type'],as_index=False).sum()
         causeTypeGprpvt=causeTypeGrp.pivot(index='State',columns='Type',values='Total')
```

```
In [56]: plt.figure(figsize=(12,6))
         plt.xticks(rotation=45,ha='right')
         sns.heatmap(causeTypeGprpvt,cmap="YlGnBu")
```

Out[56]: <matplotlib.axes._subplots.AxesSubplot at 0x1bf1c78ff70>
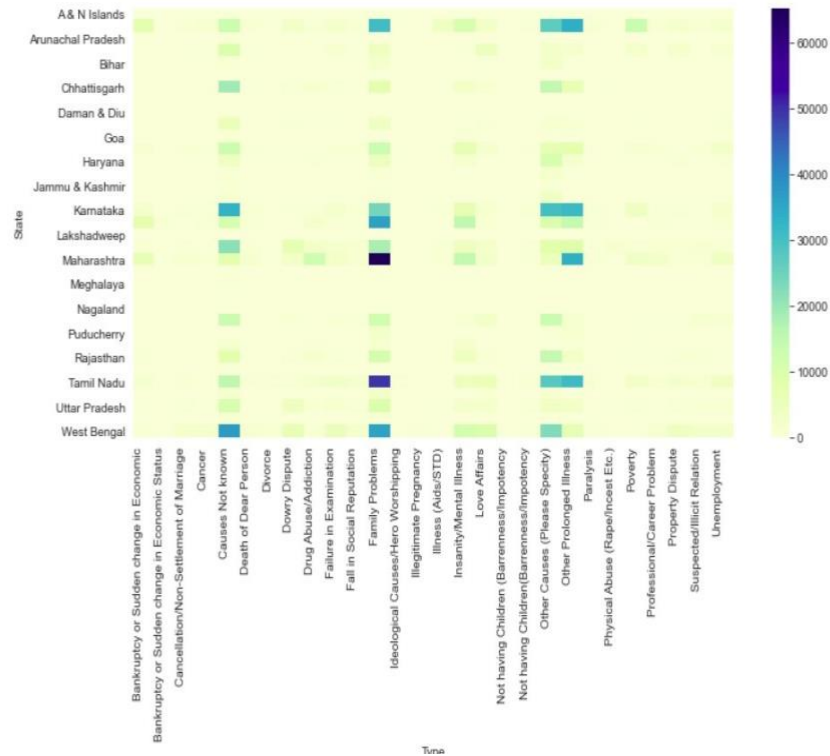
Conclusion

1 Age group between **15-44** has the highest number of suicides.

2 **Maharashtra, West Bengal and Tamil Nadu** have the highest suicide rate this might also be because of the high population in these states.

3 **Family Problems, Illness, Mental Illness** are some of the main reason while most of the reasons are still unknown.

4 **Hanging, Consuming Insecticides** is more common but, most of other means are unknown.

5 Most of the **professional profile** is unknown but **house wife** comes in second, while **Farming** comes in at third.

6 The reason for suicide at the age group between **0-14** is because of **Failure in Examination.**

References

www.kaggle.com www.researchgate.com

www.upgrad.com