Shikshana Prasaraka Mandali's Accredited with A+ Grade by NAAC

SIR PARASHURAMBHAU COLLEGE (AUTONOMOUS)

TILAK ROAD, PUNE – 411 030.

Department of Computer Science

Certificate

This is to certify that **Mr. Harishchandra Devadatta Bansode** has presented a mini project titled **Suicide Analysis In India** in partial fulfillment of the requirements of M.Sc. (Computer Science) Part I (semester I) course.

Date: / /20

Teacher In-Charge

Head,
Department of Computer Science

INDEX

Sr. No	Topic Name	Page No.
1	Description of the data	3
2	Data observation techniques	5
3	Technique used for preprocessing data	6
4	Python Code	9
5	Data Visualization and Interpretation	15
6	Bibliography	35

Description of the data

Context

This data set contains yearly suicide detail of all the States and union territories of India by various parameters from 2001 to 2012.

Data Source

data.world.com(https://data.world/rajanand/suicides-in-india)

Data Shape

237519 entries and 7 columns.

State

Name of the state in India.

We have data of 35 states.

Year

The year in which suicides were committed.

Type code

It has 5 values which are:

- Suicide causes
- Education status
- By means adopted
- Professional profile
- Social status

Type

It consists of various factors and causes of death.

Gender

- o Male
- o Female

Age Group

It consists of various age groups.

- **>** 0-100+
- **>** 0-14
- > 15-29
- > 30-44
- > 45-59
- **>** 60+

Total

It consists of total number of suicides in that particular row.

Data observation techniques

To data analysis I used following data visualization techniques: -

- 🖶 Bar chart
- Line chart
- Pie chart

Technique used for preprocessing data

1. Import libraries

As main libraries, I am using Pandas, NumPy and time;

Pandas: Use for data manipulation and data analysis.

NumPy: a fundamental package for scientific computing with Python. As for the visualization I am using Matplotlib and Seaborn.

```
# import lib
import numpy as np #linear algebra
import pandas as pd #data processing
import seaborn as sns #for making statistical graphics
import os #for interacting with the operating system.
import matplotlib.pyplot as plt #collection of functions that make matplotlib work
%matplotlib inline
#sets the backend of matplotlib to the 'inline' backend
```

2. Read Data

Reading CSV file

```
[ ] 1 data=pd.read_csv("Suicides in India 2001-2012.csv.zip")
```

After importing standard libraries and loading the data in a pandas Data Frame, we have a look at the data.

	State	Year	Type_code	Туре	Gender	Age_group	Total
0	A & N Islands	2001	Causes	Illness (Aids/STD)	Female	0-14	0
1	A & N Islands	2001	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0
2	A & N Islands	2001	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0
3	A & N Islands	2001	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0
4	A & N Islands	2001	Causes	Dowry Dispute	Female	0-14	0
5	A & N Islands	2001	Causes	Family Problems	Female	0-14	0
6	A & N Islands	2001	Causes	Ideological Causes/Hero Worshipping	Female	0-14	0
7	A & N Islands	2001	Causes	Other Prolonged Illness	Female	0-14	0
8	A & N Islands	2001	Causes	Property Dispute	Female	0-14	0
9	A & N Islands	2001	Causes	Fall in Social Reputation	Female	0-14	0
]	-		e of Dat shape	aset			
	(23751	9, 7)				

3. Checking for missing values

```
#Checking missing values
    data.isna().sum()

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

Since there are no missing values found in the dataset, I didn't use any missing values handling techniques.

As part of data cleaning, we replace some states which are mentioned separately in the dataset like A & N Islands (Ut), A & N Islands, as both these entries are for Andaman and Nicobar Islands and there is no point in having separate entries for them. Similarly, other multiple entries for states and type codes have been replaced as in the below screenshot.

```
[ ] 1  # rename states
2  data.replace('A & N Islands (Ut)', 'A & N Islands', inplace=True)
3  data.replace('Chandigarh (Ut)', 'Chandigarh', inplace=True)
4  data.replace('D & N Haveli (Ut)', 'D & N Haveli', inplace=True)
5  data.replace('Daman & Diu (Ut)', 'Daman & Diu', inplace=True)
6  data.replace('Lakshadweep (Ut)', 'Lakshadweep', inplace=True)
7  data.replace('Delhi (Ut)', 'Delhi', inplace=True)

[ ] 1  # rename Type
2  data.replace('Bankruptcy or Sudden change in Economic', 'Bankruptcy or Sudden change in Economic Status', inplace=True)
3  data.replace('By Other means (please specify)', 'By Other means', inplace=True)
4  data.replace('Not having Children(Barrenness/Impotency', 'Not having Children (Barrenness/Impotency', inplace=True)
```

Python Code

- Total Suicides in India Year wise

```
[ ] 1 data_year_total = data.groupby(data["Year"]).agg({"Total":"sum"})

1 data_year_total.reset_index(inplace=True)

[ ] 1 data_year_total.head(12)
```

```
plt.figure(figsize=(15,6))
plt.title("Suicides in india \n Total Vs Year\n")
sns.barplot(x="Year" , y="Total" , data=data_year_total , palette="viridis" )
```

```
1  D = data.groupby('Year').Total.sum()
2  x = D.index.values
3  y = D.values
4  plt.figure(figsize = (12,7))
5  year = range(2001,2013)
6  plt.plot(D.index, D.values)
7  plt.xticks(year, rotation = 45)
8  plt.title("Suicides in india\n Total vs Years \n")
9  plt.ylabel("Total")
10  plt.xlabel("Years")
11  plt.tight_layout()
12  plt.show()
```

- Total Suicides In India Gender Wise

A quick look at the gender-wise distribution of people committing suicide tells us that we have more males committing suicide than women.

```
[ ] 1 data_gender = data.groupby("Gender").agg({"Total":"sum"})
2 data_gender.reset_index(inplace=True)
3 data_gender
4
```

```
plt.figure(figsize=(12,6))
plt.title("Suicides in India \n Male vs Female \n")
g = sns.barplot(x="Gender" , y="Total" , data=data_gender, palette="muted" )
```

Total Suicides In India Age Wise

```
[ ] 1 cause=data[data['Type_code']=='Causes'].reset_index(drop=True)
      2 cause=cause.drop("Type_code", axis=1)
      1 def Age_group(age):
            sum=cause[cause["Age_group"]==age]['Total'].sum()
return sum
      3
      5 age_group=["0-14", "15-29", "30-44", "45-59", "60+"]
      6 value=[]
      7 for age in age_group:
      8
              #dict[age]=Age_group(age)
          value.append(Age_group(age))
     10
     11
     #age=pd.DataFrame.from_dict(dict, orient='index')
     13 age=pd.DataFrame({"age":age_group,"count":value})
14 sns.catplot(x="age", y="count", data=age, kind="bar")
     15 age
```

- Total Suicides In India State Wise

we look at the state-wise suicide numbers to find out that the states with the highest number of suicides are Maharashtra, West Bengal, Andhra Pradesh, Tamil Nadu, and Karnataka.

```
[ ] 1 grp = data2.groupby('State')['Total'].sum()
2 total_suicides = pd.DataFrame(grp).reset_index().sort_values('Total',ascending=False)
3 total_suicides
4 total_suicides = total_suicides[2:]
```

```
fig , ax = plt.subplots(figsize=(20,6))
g=sns.barplot(x='State',y='Total',data=total_suicides)
var=g.set_xticklabels(g.get_xticklabels(),rotation=90)
```

- Distribution By Type code

```
data_Type_code = data.groupby("Type_code").agg({"Total":"sum"})
data_Type_code.reset_index(inplace=True)
data_Type_code
4
```

```
import seaborn as sns
plt.figure(figsize=(12,6))
sns.set(rc={'figure.figsize':(11.7,8.27)})
sns.countplot(data['Type_code'])
```

- Total Suicides In India Social Status Wise

```
[ ] 1   data_by_code = data[data['Type_code'] == 'Social_Status']
2   data_by_code['Type'].unique()
3
4   data_by_social = data_by_code.groupby('Type')['Total'].sum()
5   data_by_social
```

```
1  data_by_social_type = pd.DataFrame(data_by_social).reset_index().sort_values('Total')
2  print('Total people who have their social status mentioned in the data {}'.format(data_by_code.shape[0]))
3  labels = data_by_social_type['Type']
4  explode = (0,0,0,0,0.1)
5  plt.figure(figsize = (13,7))
6  plt.pie(data_by_social_type['Total'],labels = labels,autopct='%1.1f%%',explode = explode,shadow=True)
7  plt.show()
```

Total Suicides In India Professional Profile Wise

```
data_by_code = data[data['Type_code'] == 'Professional_Profile']
data_by_code['Type'].unique()

data_by_proffession = data_by_code.groupby('Type')['Total'].sum()
data_by_proffession
```

```
data_by_proffession_type = pd.DataFrame(data_by_proffession).reset_index().sort_values('Total')
plt.figure(figsize=(12,6))
print('Total people who have their professional mentioned in the data {}'.format(data_by_code.shape[0]))
sns.barplot(x = data_by_proffession_type['Total'], y = data_by_proffession_type['Type'])
```

Total Suicides In India Means adopted Wise

```
data_by_code = data[data['Type_code'] == 'Means_adopted']
data_by_code['Type'].unique()

data_by_means = data_by_code.groupby('Type')['Total'].sum()
data_by_means
```

```
data_by_means_type = pd.DataFrame(data_by_means).reset_index().sort_values('Total')
plt.figure(figsize=(12,6))
fig , ax = plt.subplots(figsize=(18,6))
g = sns.barplot(x = data_by_means_type['Total'], y = data_by_means_type['Type'])
```

Total Suicides In India Education Status Wise

```
data_by_code = data[data['Type_code'] == 'Education_Status']
data_by_code['Type'].unique()

data_by_education = data_by_code.groupby('Type')['Total'].sum()
data_by_education
```

```
data_by_education_type = pd.DataFrame(data_by_education).reset_index().sort_values('Total')
print('Total people who have their social status mentioned in the data {}'.format(df_by_code.shape[0]))
labels = data_by_education_type('Type')
plt.figure(figsize = (13,7))
plt.pie(data_by_education_type['Total'],labels = labels,autopct='%1.1f%%')
plt.show()
```

Total Suicides In India Causes Wise

```
data_by_code = data[data['Type_code'] == 'Causes']
data_by_code['Type'].unique()

data_by_causes = data_by_code.groupby('Type')['Total'].sum()
data_by_causes

...

[] data_by_causes_type = pd.DataFrame(data_by_causes).reset_index().sort_values('Total')
fig , ax = plt.subplots(figsize=(13,7))
g = sns.barplot(x = data_by_causes_type['Total'], y = data_by_causes_type['Type'])
```

- Analysing suicide rates in Maharashtra

```
1  state_data = data.groupby(['State'])
2  result = state_data['Total'].sum().sort_values()
3  result

1  result = mh_data.groupby(['Year'])['Total'].sum()
2  result

1  mh_data = state_data.get_group('Maharashtra')
2  mh_data
```

SUCIDE DATA ANALYSIS IN INDIA

```
publication of suicides place of suicides p
```

Cause of suicides

```
filt = mh_data["Type_code"] == 'Causes'
mh_data.loc[filt].groupby('Type')['Total'].sum().sort_values(ascending = False)
```

Social status

```
filt = mh_data["Type_code"] == 'Social_Status'
    result = mh_data.loc[filt].groupby('Type')['Total'].sum().sort_values(ascending = False)
    result
```

```
1 explode = [0.1,0,0,0,0]
2 plt.figure(figsize = (12,6))
3 plt.pie(result.values, explode = explode, shadow= True, startangle = 90,radius=1.2)
4 plt.tight_layout()
5 plt.legend(result.index, loc = 2, bbox_to_anchor=(1, 0., 0.5, 0.5))
6 plt.show()
```

- Education status

```
[ ] 1 filt = mh_data["Type_code"] == 'Education_Status'
2    result = mh_data.loc[filt].groupby('Type')['Total'].sum().sort_values()
3    result
```

```
plt.figure(figsize = (12,6))
plt.style.use("seaborn")
plt.barh(result.index,result.values)
plt.xlabel("No. of suicides")
plt.show()
```

Professional status

```
filt = mh_data["Type_code"] == 'Professional_Profile'
result = mh_data.loc[filt].groupby('Type')['Total'].sum().sort_values()
result
```

```
plt.figure(figsize = (12,6))
plt.style.use("seaborn")
plt.barh(result.index,result.values)
plt.xlabel("No. of suicides")
plt.show()
```

Methods used for suicide

```
filt = mh_data["Type_code"] == 'Means_adopted'
result = mh_data.loc[filt].groupby('Type')['Total'].sum().sort_values()
result
```

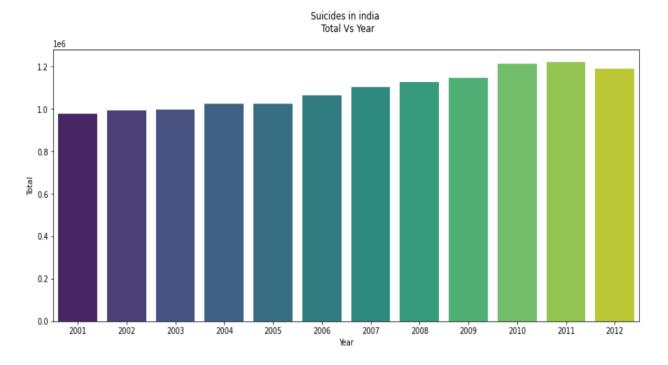
```
plt.figure(figsize = (12,6))
plt.style.use("seaborn")
plt.barh(result.index,result.values)
plt.xlabel("No. of suicides")
plt.show()
```

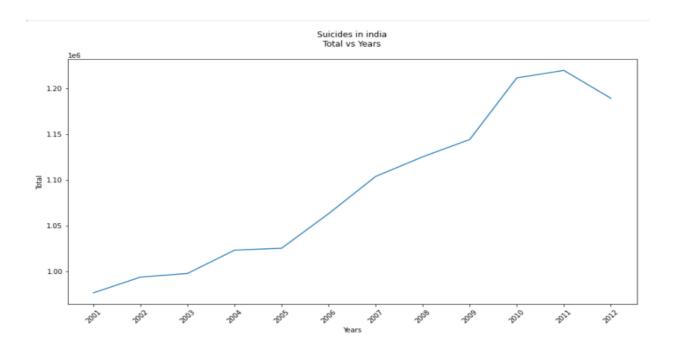
Data Exploration and Visualization

We see a gradual increase in the number of suicides every year, however, in 2012 we have a dip in the numbers as compared to 2011.

Total Suicides in India Year wise

₽		Year	Total
	0	2001	976464
	1	2002	993648
	2	2003	997622
	3	2004	1023137
	4	2005	1025201
	5	2006	1062991
	6	2007	1103667
	7	2008	1125082
	8	2009	1144033
	9	2010	1211322
	10	2011	1219499
	11	2012	1189068





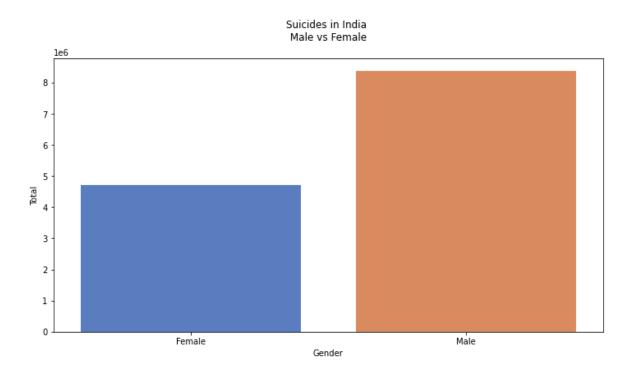
We *see* a gradual increase in the number of suicides every year, however, in 2012 we have a dip in the numbers as compared to 2011.

Total Suicides in India Gender Wise

Tabular representation

	Gender	Total
0	Female	4702974
1	Male	8368760

graphical representation



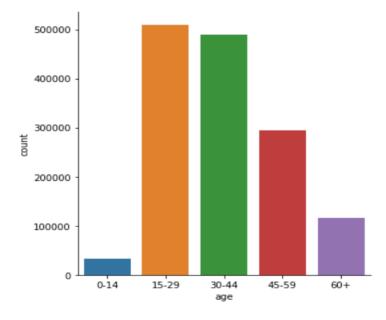
A quick look at the gender-wise distribution of people committing suicide tells us that we have more males committing suicide than women.

Distribution of cases as per the age group

Tabular representation

	age	count
0	0-14	32685
1	15-29	509776
2	30-44	488713
3	45-59	294333
4	60+	115467

graphical representation



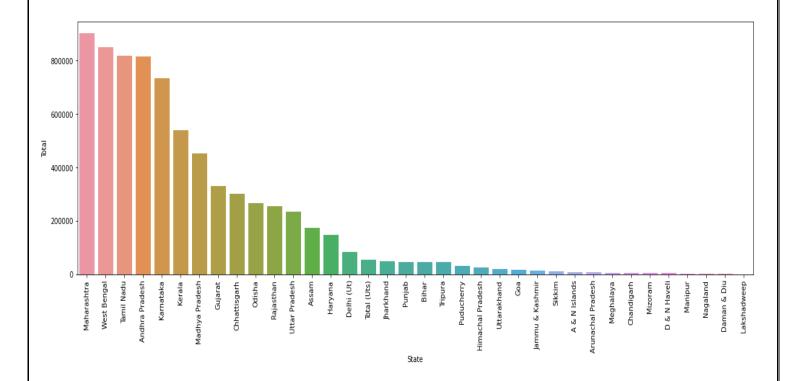
From the below visualization it is clear that youngsters (15-29 age) and middle age (30-44) tend to commit the maximum number of suicides

It can be due to several reasons like:

- unemployment
- academic stress
- bad friend circle
- farmers (since they have to be young and strong enough to do farming)
- addictions

Total Suicides in India State wise

	State	Total
20	Maharashtra	901945
27	West Bengal	849936
30	Tamil Nadu	818691
1	Andhra Pradesh	814059
18	Kamataka	734825
17	Kerala	538946
19	Madhya Pradesh	451535
11	Gujarat	330858
8	Chhattisgarh	302354
25	Odisha	267234
28	Rajasthan	255134
35	Uttar Pradesh	233352
3	Assam	172276
12	Haryana	147176
8	Delhi (Ut)	84272
33	Total (Uts)	53836
15	Jharkhand	49720
27	Punjab	46350
4	Bihar	46214
34	Tripura	45965
28	Puducherry	32144
13	Himachal Pradesh	26562
38	Uttarakhand	18496
10	Goa	17363
14	Jammu & Kashmir	14821
29	Sildkim	9606
0	A & N Islands	8109
2	Arunachal Pradesh	6633
22	Meghalaya	5415
6	Chandigarh	5164
23	Mizoram	4154
7	D & N Haveli	3430
21	Manipur	2102
24	Nagaland	1728
8	Daman & Diu	1391



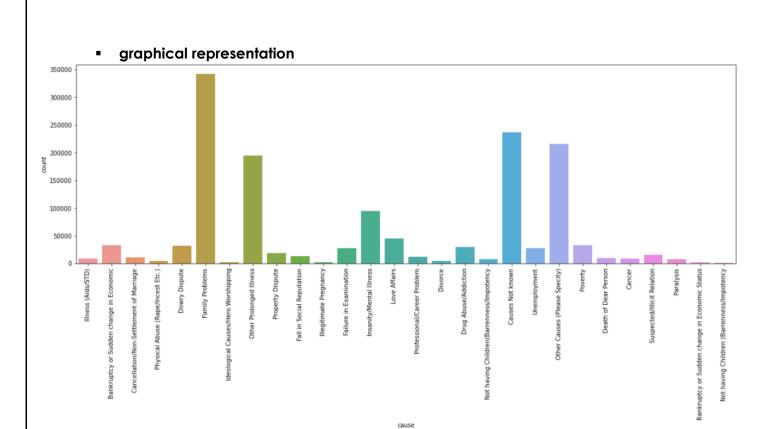
States with Higher Suicide cases

- Maharashtra
- West Bengal
- Tamil Nadu
- Andhra Pradesh

we look at the state-wise suicide numbers to find out that the states with the highest number of suicides are Maharashtra, West Bengal, Tamil Nadu, Andhra Pradesh, and Karnataka.

Total Suicides in India Cause wise

0	Illness (Aids/STD)	8723
1	Bankruptcy or Sudden change in Economic	32755
2	Cancellation/Non-Settlement of Marriage	11298
3	Physical Abuse (Rape/Incest Etc.)	3992
4	Dowry Dispute	31970
5	Family Problems	341952
6	Ideological Causes/Hero Worshipping	2118
7	Other Prolonged Illness	194565
8	Property Dispute	18652
9	Fall in Social Reputation	13464
10	Illegitimate Pregnancy	2494
11	Failure in Examination	27005
12	Insanity/Mental Illness	94229
13	Love Affairs	45039
14	Professional/Career Problem	12554
15	Divorce	4133
16	Drug Abuse/Addiction	30046
17	Not having Children(Barrenness/Impotency	7822
18	Causes Not known	237069
19	Unemployment	27365
20	Other Causes (Please Specity)	216050
21	Poverty	32684
22	Death of Dear Person	10321
23	Canoer	9058
24	Suspected/Illicit Relation	14911
25	Paralysis	7286
26	Bankruptcy or Sudden change in Economic Status	2655
27	Not having Children (Barrenness/Impotency	766

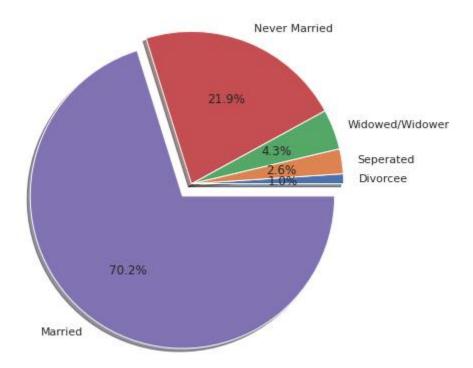


Total Suicides in India Social Status Wise

Tabular representation

Туре	
Divorcee	45816
Married	3065322
Never Married	954903
Seperated	115413
Widowed/Widower	186339
Name: Total, dtype:	int64

graphical representation



It is clear that married people are more Suicides.

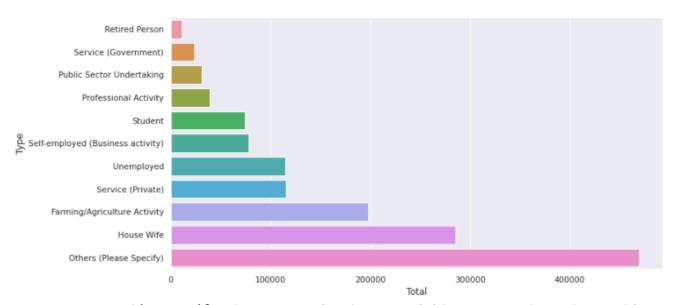
Which makes sense because marriage issues may cause conflict between the couple and as a result, they might be prone to commit suicide

Total Suicides In India Professional Profile Wise

Tabular representation

Туре	
Farming/Agriculture Activity	197923
House Wife	285243
Others (Please Specify)	469147
Professional Activity	39204
Public Sector Undertaking	30786
Retired Person	11334
Self-employed (Business activity)	78112
Service (Government)	23325
Service (Private)	115472
Student	74323
Unemployed	114374
Name: Total, dtype: int64	

graphical representation



Farmers and housewife's have committed more suicide compared to others. This makes sense because most of the Indian farmers have debt and their life depends on the yield of their crops, if the yield is not good then they will not be able to clear their debt and in the worst case they might commit suicide. Global warming, monsoon delay, drought etc. can lead to bad yield. Housewife's might have issues in their marriage which this might be a reason for such a high number of cases.

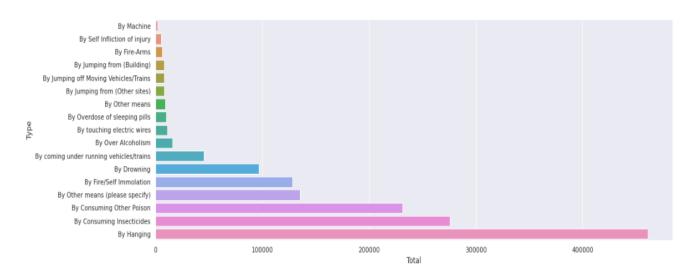
Domestic violence, dowry, gender discrimination, etc. might be some of the reasons for housewives to commit suicide.

Total Suicides in India Means adopted Wise

Tabular representation

Туре	
By Consuming Insecticides	275501
By Consuming Other Poison	231178
By Drowning	96711
By Fire-Arms	6294
By Fire/Self Immolation	128006
By Hanging	460955
By Jumping from (Building)	7871
By Jumping from (Other sites)	8127
By Jumping off Moving Vehicles/Trains	8116
By Machine	1661
By Other means	9238
By Other means (please specify)	135132
By Over Alcoholism	15973
By Overdose of sleeping pills	9960
By Self Infliction of injury	5093
By coming under running vehicles/trains	45299
By touching electric wires	10816
Name: Total, dtype: int64	

graphical representation



Hanging, Consuming Insecticides and poison, self-immolation are the top means adopted by people who suicide.

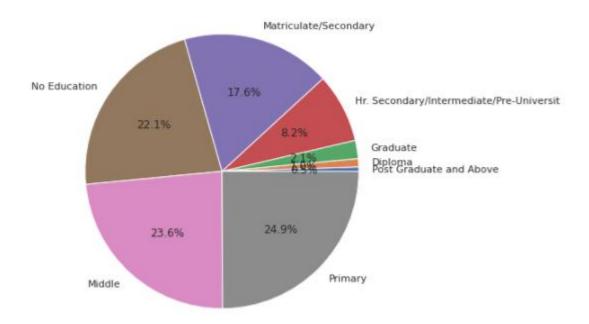
Total Suicides in India Education Status Wise

Tabular representation

Туре	
Diploma	42459
Graduate	93822
Hr. Secondary/Intermediate/Pre-Universit	356724
Matriculate/Secondary	769698
Middle	1028913
No Education	965271
Post Graduate and Above	22425
Primary	1088481

Name: Total, dtype: int64

Graphical representation

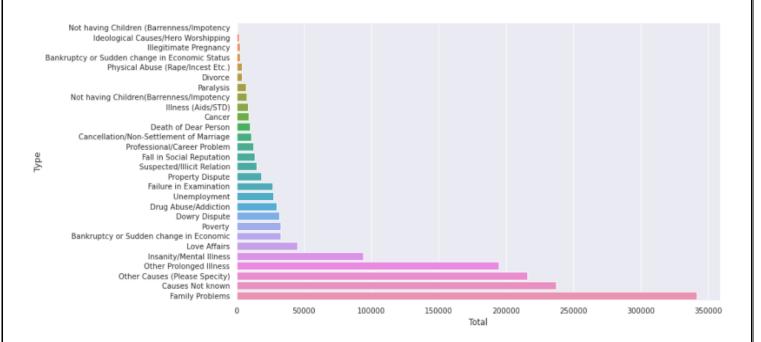


People with low education are committing more suicide.

People with Diploma and Graduate tend to commit least no. of suicide

Total Suicides in India Causes Wise

Type	
Bankruptcy or Sudden change in Economic	32755
Bankruptcy or Sudden change in Economic Status	2655
Cancellation/Non-Settlement of Marriage	11296
Cancer	9058
Causes Not known	237069
Death of Dear Person	10321
Divorce	4133
Dowry Dispute	31970
Drug Abuse/Addiction	30046
Failure in Examination	27005
Fall in Social Reputation	13464
Family Problems	341952
Ideological Causes/Hero Worshipping	2118
Illegitimate Pregnancy	2494
Illness (Aids/STD)	8723
Insanity/Mental Illness	94229
Love Affairs	45039
Not having Children (Barrenness/Impotency	766
Not having Children(Barrenness/Impotency	7822
Other Causes (Please Specity)	216050
Other Prolonged Illness	194565
Paralysis	7286
Physical Abuse (Rape/Incest Etc.)	3992
Poverty	32684
Professional/Career Problem	12554
Property Dispute	18652
Suspected/Illicit Relation	14911
Unemployment	27365
Name: Total, dtype: int64	



Top causes of suicide are:

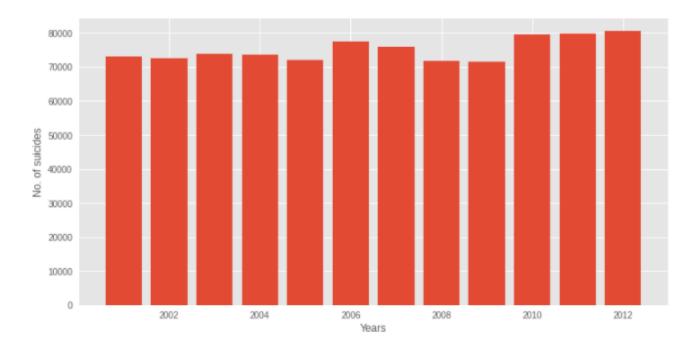
- 1) Family Problems
- 2) Prolonged Illness
- 3) Mental Illness
- 4) Love affairs
- 5) Bankruptcy or sudden change in economic status

Analyzing suicide rates in Maharashtra

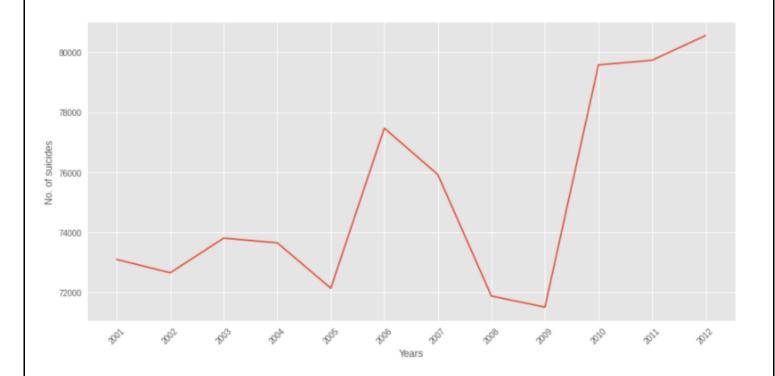
Tabular representation

Year		
2001	73090	
2002	72645	
2003	73800	
2004	73645	
2005	72130	
2006	77470	
2007	75920	
2008	71870	
2009	71500	
2010	79580	
2011	79735	
2012	80560	
Name:	Total, dtype:	int64

Graphical representation



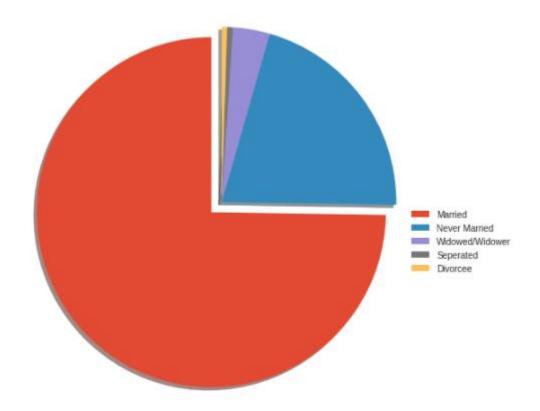
SUCIDE DATA ANALYSIS IN INDIA



Year 2012 has the highest number, 80,560 suicides. Year 2009 has the lowest number of suicides, 71,500.

Social status

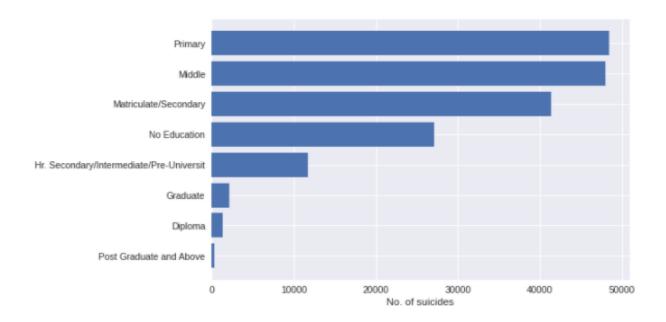
Type			
Marrie	ed		134843
Never	Marrie	d	37539
Widowe	ed/Widow	ver	6157
Sepera	ated		958
Divor	cee		892
Name:	Total,	dtype:	int64



Most suicides were among married people.

Education status

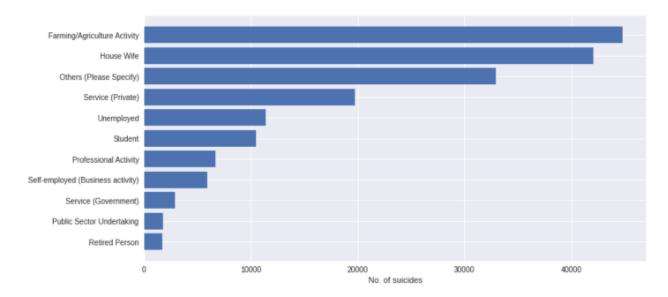
Туре	
Post Graduate and Above	358
Diploma	1377
Graduate	2146
Hr. Secondary/Intermediate/Pre-Universit	11715
No Education	27088
Matriculate/Secondary	41304
Middle	47927
Primary	48474
Name: Total, dtype: int64	



This stat is very interesting: People with no education have lower rates of suicide than people with half education.

Professional status

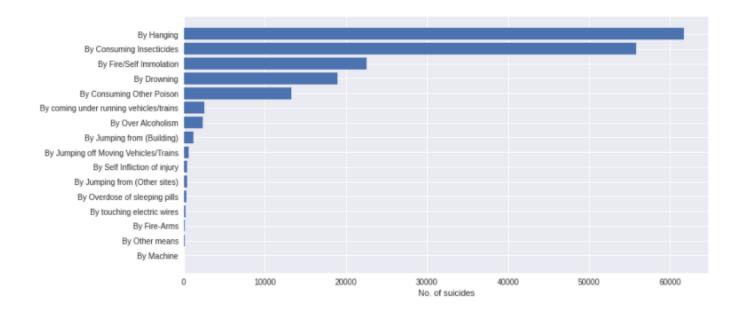
Туре	
Retired Person	1727
Public Sector Undertaking	1771
Service (Government)	2913
Self-employed (Business activity)	5932
Professional Activity	6665
Student	10441
Unemployed	11389
Service (Private)	19754
Others (Please Specify)	32969
House Wife	42059
Farming/Agriculture Activity	44769
Name: Total, dtype: int64	



This stat is more disheartening, suicide rates among farmers is much higher than any other profession.

Methods used for suicide

Туре	
By Machine	21
By Other means	140
By Fire-Arms	157
By touching electric wires	213
By Overdose of sleeping pills	325
By Jumping from (Other sites)	362
By Self Infliction of injury	414
By Jumping off Moving Vehicles/Trains	550
By Jumping from (Building)	1164
By Over Alcoholism	2331
By coming under running vehicles/trains	2565
By Consuming Other Poison	13248
By Drowning	18923
By Fire/Self Immolation	22552
By Consuming Insecticides	55773
By Hanging	61651
Name: Total, dtype: int64	



CONCLUSION

Nowadays suicide is very common in India. Male commits more suicide from the above analysis. People who commit more suicide are mostly married, farmers, housewife's, or youngsters.

- Maharashtra, West Bengal, and Tamil Nadu are the top 3 states with the highest rate of suicide
- Nagaland, Lakshadweep, Daman & Diu are the top 3 states with the lowest rate of suicide
- From 2006 to 2011 there was increase in the suicide rate but in 2012 it decreased.
- Hanging was the most used method for committing suicide.
- Male from age group 30-44 committed most suicide.
- Female from age group 15-29 committed most suicide.



I have successfully designed and implemented this project by using python programming. I found following websites very helpfully to complete this mini project.

- Website: -
 - ➤ Wikipedia
 - ► https://ourworldindata.org/suicide
 - ► https://www.community.data.gov.in/
 - >https://www.kaggle.com/rajanand/suicides-in-india
- Coding language: -
 - **>** Python.