



**FE - 520 : Introduction to Python for Financial Applications**

## **Topic: Suicide Prevention Analysis**

**Team 23:**

**Nishit Rao**

**Dilpreet Saini**

**Samruddhi Zemse**

**Table of Contents**

1. Introduction	3
2. Problem Statement	3
3. Objectives	4
4. Dataset	4
5. Methodology	6
6. Results	7
7. Challenges	15
8. Recommendation & Future Scope	15
9. Conclusion	16
10. References	17

## 1. Introduction

Suicide is a highly sensitive and complex phenomenon. Understanding its roots has always been inconsistent throughout the history. Conventional beliefs characterize suicide as a choice of free-willed individuals acting in personal despair. However, social scientists argue that a large systemic forces stronger than the individuals exist behind the suicides. It is also believed that suicides tend to produce a chain-reaction effect, i.e., a greater number of individuals at a given time in a given geographic area committing suicides could produce more suicides.

According to the World Health Organization (WHO), more than nine lakh people die from suicide every year globally. It is one of the leading causes of death among teenagers, and three-fourths of the suicides globally occur in middle- and low-income nations. Every one in hundred deaths is by suicide across the world. For every individual dying by suicide, almost 10-20 times more people attempt suicide. These statistics point out the glaring invisibility of the suicide, which is often perceived as a private concern. Suicide is death caused by injuring oneself with the intent to die. A suicide attempt is when someone harms themselves with any intent to end their life, but they do not die despite of their actions. Having a thorough and reliable database of suicides is highly essential, as it enables us to understand the underlying causes that push people towards suicide & its spread. In India, the "[Accidental Deaths and Suicides in India](#)", hereafter referred to as 'ADSI' is such a comprehensive data source published as an annual report by the National Crime Records Bureau (NCRB). In this project, we look at some of the key statistics relating to suicides and its trends in India. There is a subsequent increase in the suicide rate in the last decade in India and hence we have decided to analyze data related to this information such that we can help in creating awareness.

## 2. Problem Statement

In 2021, there were an estimated 900,000 suicides globally, and approximately 18,000,000 suicide attempts. Despite reports of almost 33% decrease in the global age-standardized mortality rate from suicide between 1992 and 2022, suicide remains one of the leading causes of preventable deaths in both developing and developed countries.

### 3. Objectives

To work on suicide prevention, we need to have enough information and analyze it to design effective strategies. Our main objective was to find out any trends or linkage between factors like gender, age, educational status, or professional profile can help in building strategies. In this project we intend to:

- To analyze role of age in suicides
- To analyze role of gender in suicides
- To analyze educational and social background in suicides
- To analyze cause and means of suicide
- To analyze correlation and derive insights from this exploratory data analysis

The main agenda is to analyze this data to create awareness such that we can prevent this from happening in the future by creating awareness. The analysis was based on accurate data provided by the most comprehensive, databank available with the Government of India on the subject. This can be used by policy makers, NGOs, researchers and public at large. Keeping in view the extensive, and increasing, dependence of various stakeholders on the information contained in the report, we have limited our analysis to finding out the above information.

### 4. Data Set

The dataset consists of data related to suicide attempt cases in India from 2011 to 2022. It is a combination of multiple datasets. It is freely available on Kaggle and [www.macrotrends.com](http://www.macrotrends.com). This dataset has 237519 rows and 7 columns. These 7 columns are State, Year, Type Code, Type, Gender, Age Group, Total

State: It indicates the States of India where the cases are from. The data also includes union territories within the State column which was put into one column during data cleaning.

Year: 2011-2022 is the 11-year period considered for this dataset.

Type Code: It is divided into Causes, Education Status, Means Adopted, Professional Profile and Social Status.

Type: This indicates the type of type code category such as for Causes they are illness, abuse, dowry, economic factors, family problems etc. while for Education

status it is primary, graduate, middle-school, or no-education etc. and so on for other categories mentioned in type code.

Gender: It indicates the binary gender Male or Female.

Age Group: It indicates the age from 0-100 divided into age groups: 0-14, 15-29, 30-44, 45-59, and 60+

Total: This gives the number of suicide cases.

#### Reading the dataset

```
In [4]: df = pd.read_csv('/Users/nishitrao/Downloads/SuicidesInIndia2011-2012.csv')
df.head()
```

Out[4]:

	State	Year	Type_code	Type	Gender	Age_group	Total
0	A & N Islands	2011	Causes	Illness (Aids/STD)	Female	0-14	0
1	A & N Islands	2011	Causes	Bankruptcy or Sudden change in Economic	Female	0-14	0
2	A & N Islands	2011	Causes	Cancellation/Non-Settlement of Marriage	Female	0-14	0
3	A & N Islands	2011	Causes	Physical Abuse (Rape/Incest Etc.)	Female	0-14	0
4	A & N Islands	2011	Causes	Dowry Dispute	Female	0-14	0

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

Out[5]: (237519, 7)

We further break down this dataset by listing out the columns, type\_code, state, age group etc. just to get a clear picture of how vast the dataset really is. This made us realize that the search for the perfect dataset hasn't ended as there are still a lot of redundant data which needs to be cleaned and pre-processed in order to proceed with exploratory data analysis. Hence, we removed the redundant data and brought down the number of rows to 236583 and columns remained 7.

```
In [6]: #Getting rows and columns
Rows = df.shape[0]
Columns = df.shape[1]
print ("Number of rows are", Rows)
print ("Number of columns are", Columns)

Number of rows are 236583
Number of columns are 7
```

```
In [8]: df.columns
```

```
Out[8]: Index(['State', 'Year', 'Type_code', 'Type', 'Gender', 'Age_group',
   'Total'], dtype='object')
```

```
In [9]: df['Type_code'].value_counts()
```

```
Out[9]: Causes          109200
Means_adopted      67200
Professional_Profile 49263
Education_Status     6720
Social_Status        4200
Name: Type_code, dtype: int64
```

```
In [10]: df["State"].value_counts()
```

```
Out[10]: Karnataka      6792
Madhya Pradesh      6792
Maharashtra        6792
Andhra Pradesh     6791
Odisha              6791
Rajasthan           6791
Chhattisgarh       6790
Bihar                6790
Haryana             6790
Kerala              6788
Uttar Pradesh       6787
Tamil Nadu          6786
Gujarat              6786
Assam                6786
Jharkhand            6785
Tripura              6782
Delhi (Ut)           6782
West Bengal          6780
```

## 5. Methodology

The project was carried out in several phases by our three members. We first sought out for data that was relevant for our project. We found out that dataset related to India had several factors such as Age group, gender, states, social and educational status, cause and means adopted for suicide. Hence this data felt apt for analysis. We started with data cleaning, data pre-processing and then proceeded to do exploratory data analysis and eventually drew conclusion based on these results.

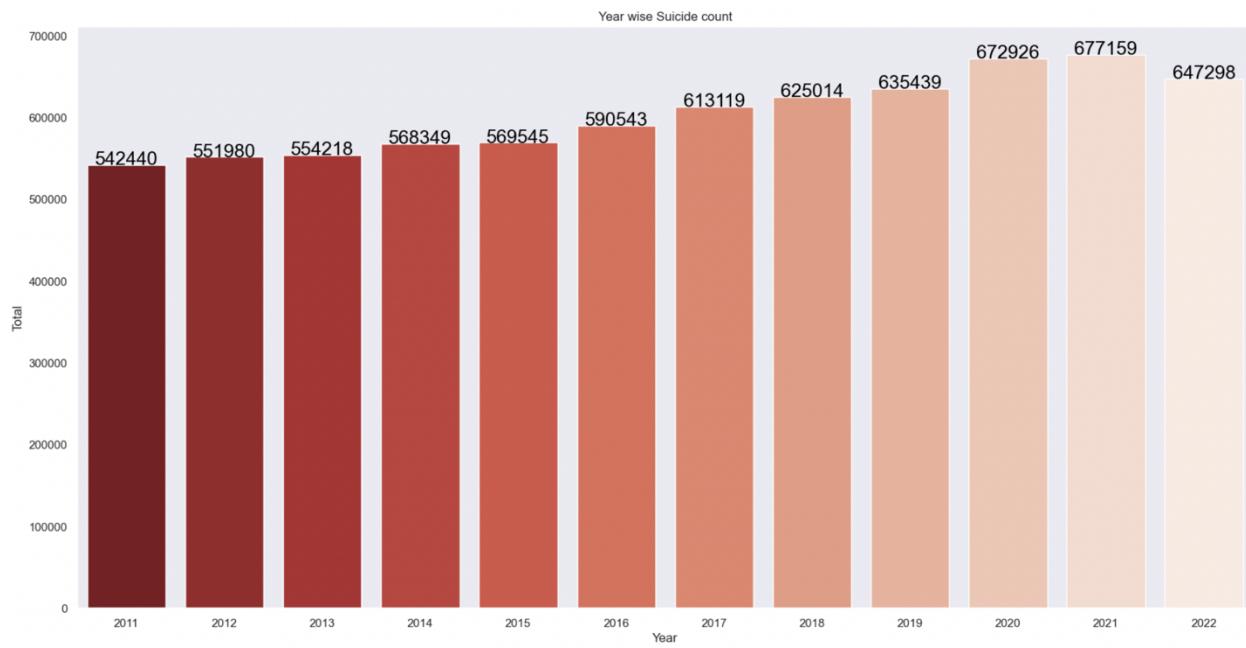
Starting out, we tried to first list all the columns and causes so we knew how many trends we could analyze against each other. The common denominator for each trend

was kept as year, gender, age, or state. We started with reading the dataset, cleaning it, pre-processing it and then went on to do exploratory data analysis. We found out the numerical data for suicide rate in the past decade based on gender, state, age group, education status, professional profile, and social status. We also found out the major causes of suicide and means adopted to commit it. We plotted this data using bar plots, cat plots, pie graphs and presented the numerical data in tables. We leveraged libraries Pandas, Matplotlib, Seaborn and NumPy in Python for our analysis. We used pandas to get information on the type of data. We have performed analysis to understand relationship between suicides in India and effect of factors like state, year, gender, social status on the suicide rate.

We used `.value_counts()`, `.sort_values()`, `.sum()`, `.reset_index()`, `.groupby()` as some of the functions that brought out the required results.

## 6. Results

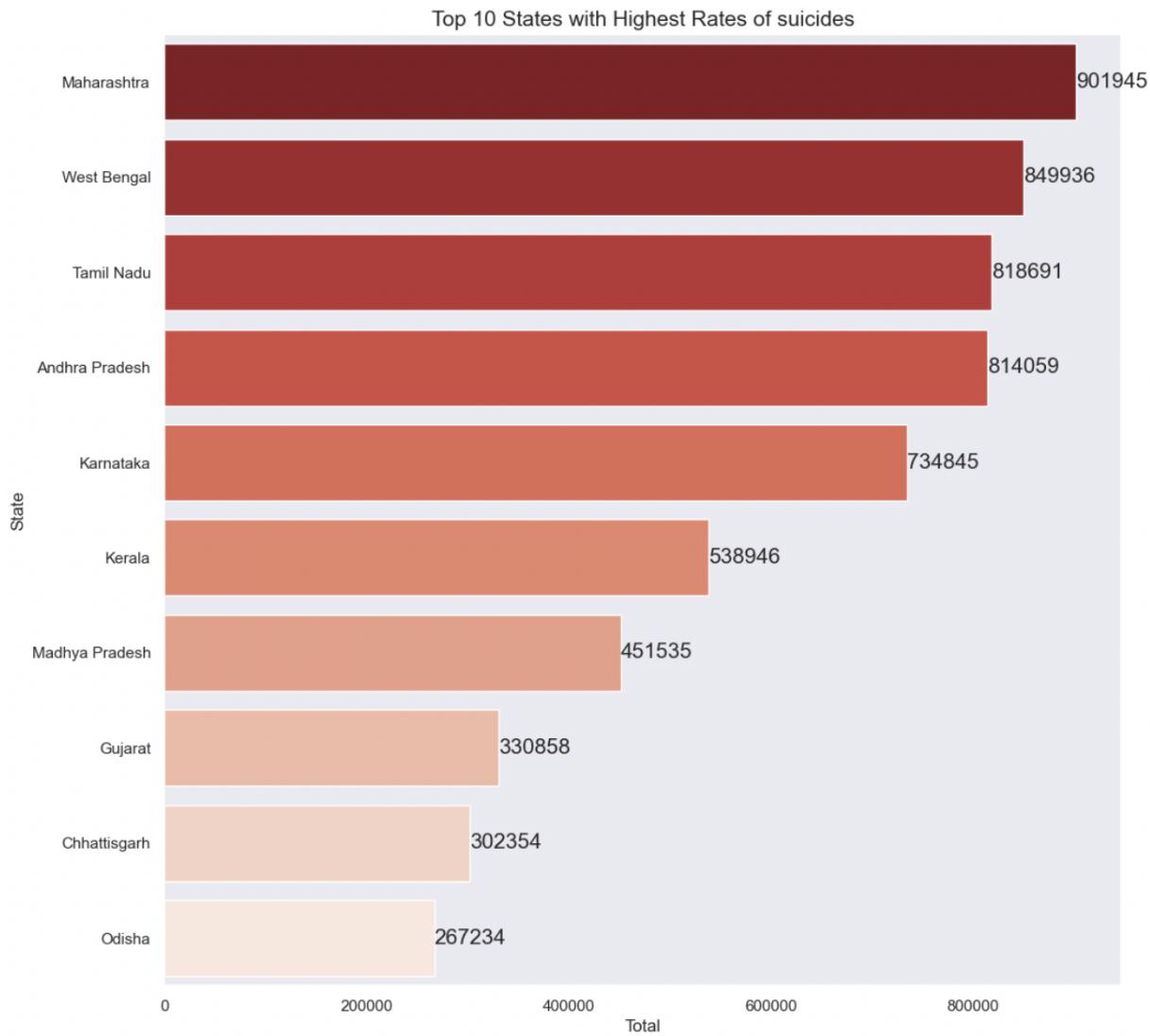
### 6.1 Increase in Year-wise suicide count



It is noted that the suicide rate has been increasing over the years in the past decade. It dips in the final year of 2022 but that maybe because the data for 2022 is incomplete as of now. It is a worrying cause and one that needs immediate attention

as going by the statistics, we need a formulaic strategy that can help narrow down the people who need to be given immediate attention or care.

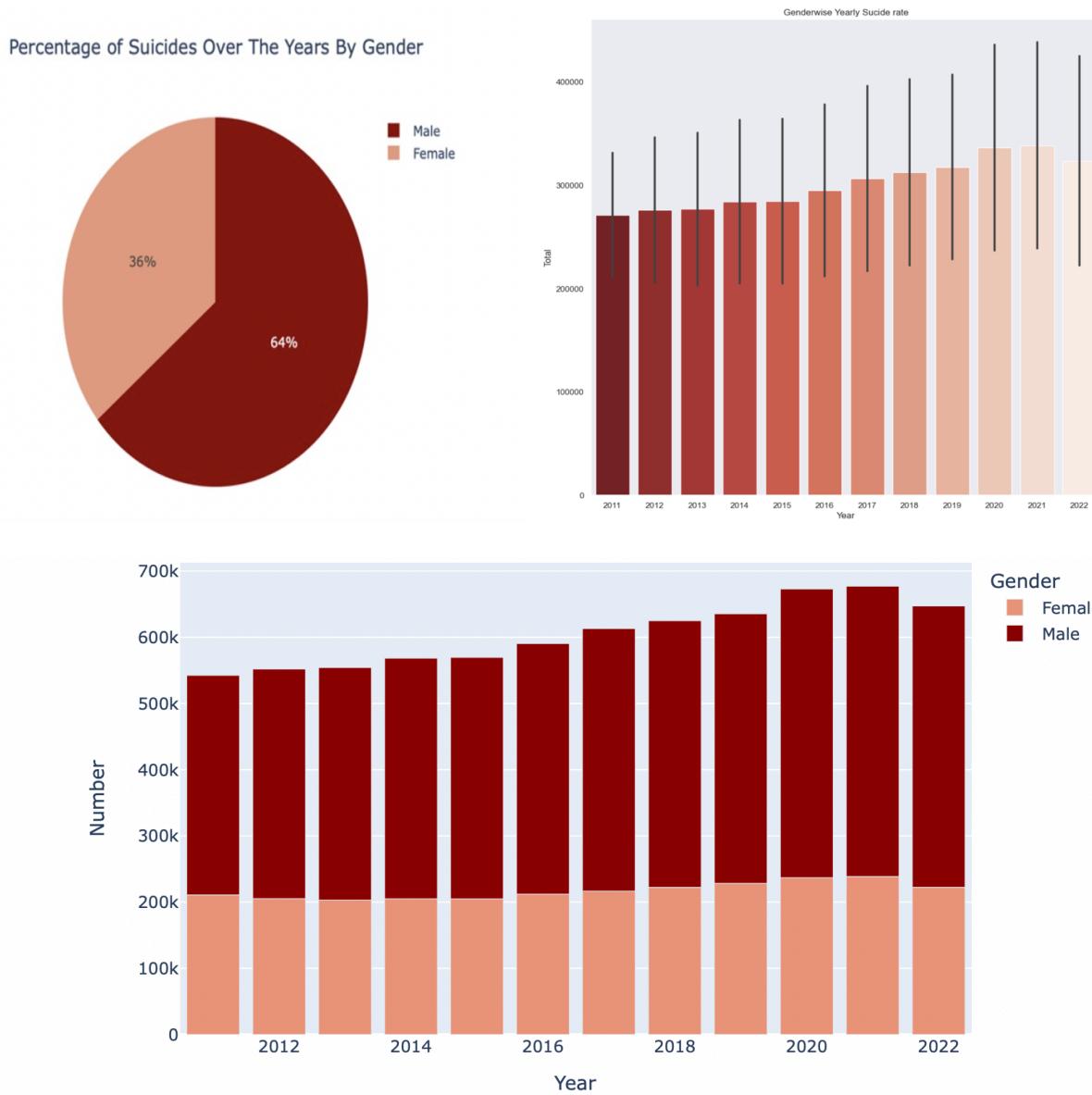
## 6.2 Five states account for half of the suicides in India



As stated earlier, many systemic factors also play a role in the incidences of suicides. The role of geographical location, the legislations, and the governance process, become crucial in terms of access to better mental health care, restricting access to lethal means of death, and understanding the interpersonal relationships and sense of community building. Suicides in India are not uniform across these factors. Five states account for approximately half of the total suicides in India. These five states have consistently been contributing significantly to the suicide burden in India. Even

among these five states, Maharashtra, and Tamil Nadu account for half of the share of five states. Other states like Madhya Pradesh, West Bengal, and Karnataka together account for another 25% of the suicides in India. The states with the highest literacy rate were seen to have the most suicides.

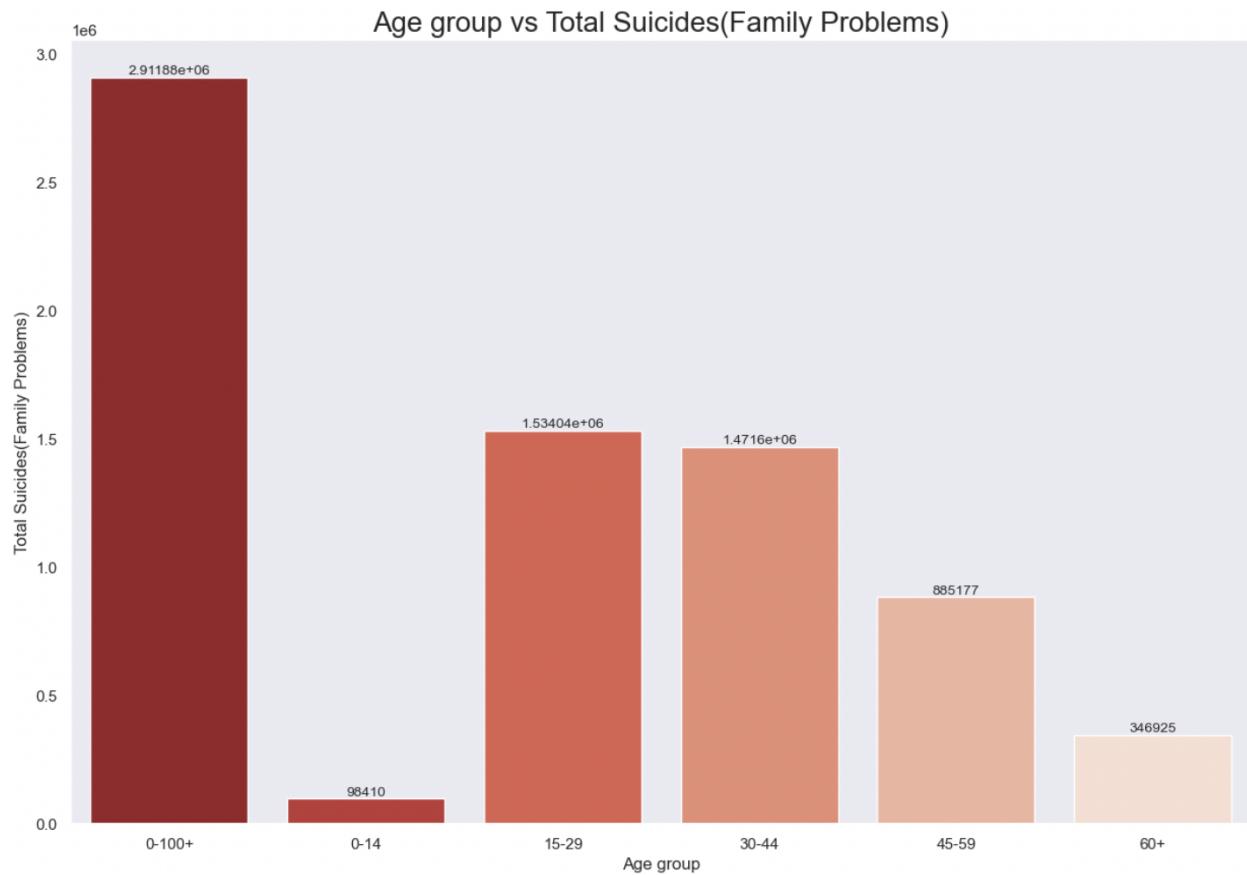
### 6.3 Men are more prone to committing suicide than women



We tried to check if gender affects or is linked to suicide rates in India. We can infer that the suicide rate is higher for men than it is for women. Gender wise suicides

over the years have become more and more lopsided. It was seen that the number of women remained more or less in the same range however the number of men went up by more than 100K. We leveraged seaborn and matplotlib library in pandas to derive a line plot analysis of suicide rate across the given years with the total number of suicides for each year in the y axis and year represented in the x-axis. We can see also see from the below graph that the rate of increase in suicides for men is much higher than rate of increase in suicides for women. We established this analysis .groupby(), .sum() operations and seaborn library in python.

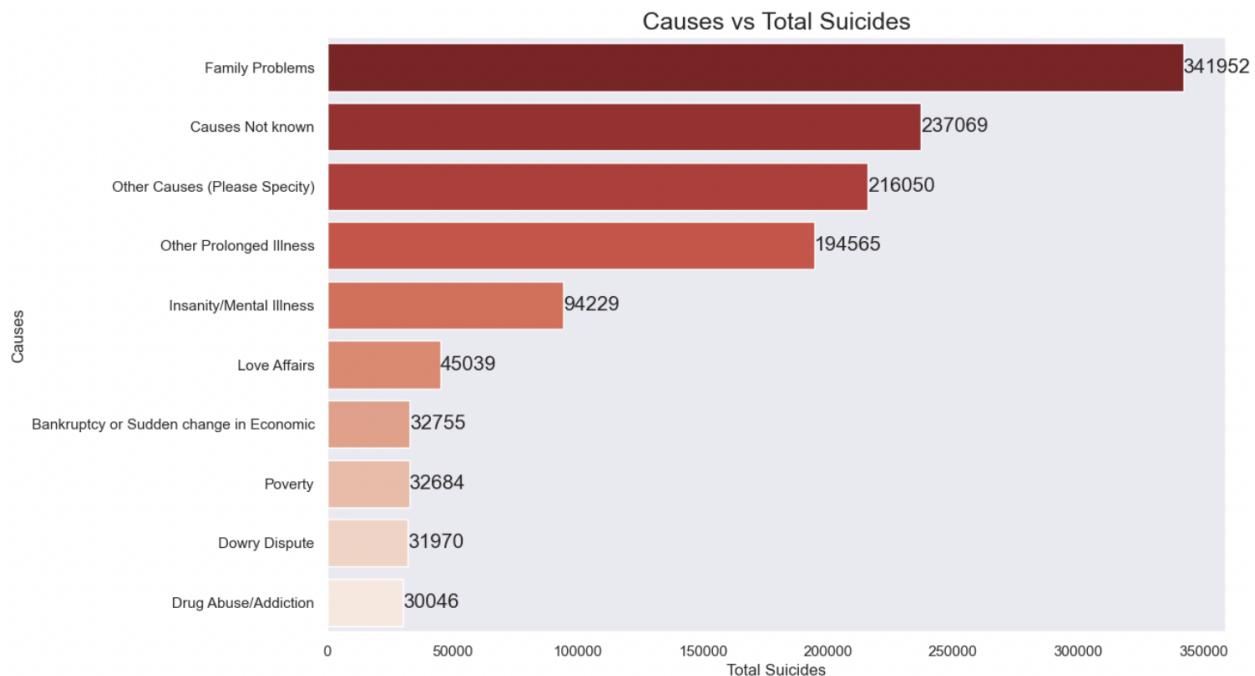
#### 6.4 Ages 15-29 recorded the highest number of suicides



While the first column refers to all the cases, it can be seen that over the decade the total number of suicides are highest in the age group of 15-29 years. It is a young and impressionable age when teenagers and young adults who are vulnerable get bogged down by intense pressure of keeping up with adversities and decide to take drastic actions. This tells us that NGOs, lawmakers, and influencers should target

school and college going children as the center of their campaigns and laws to protect their mental health and create awareness programs catering to them.

## 6.5 Family Problems followed by Mental Illness is the major cause of suicides in the country

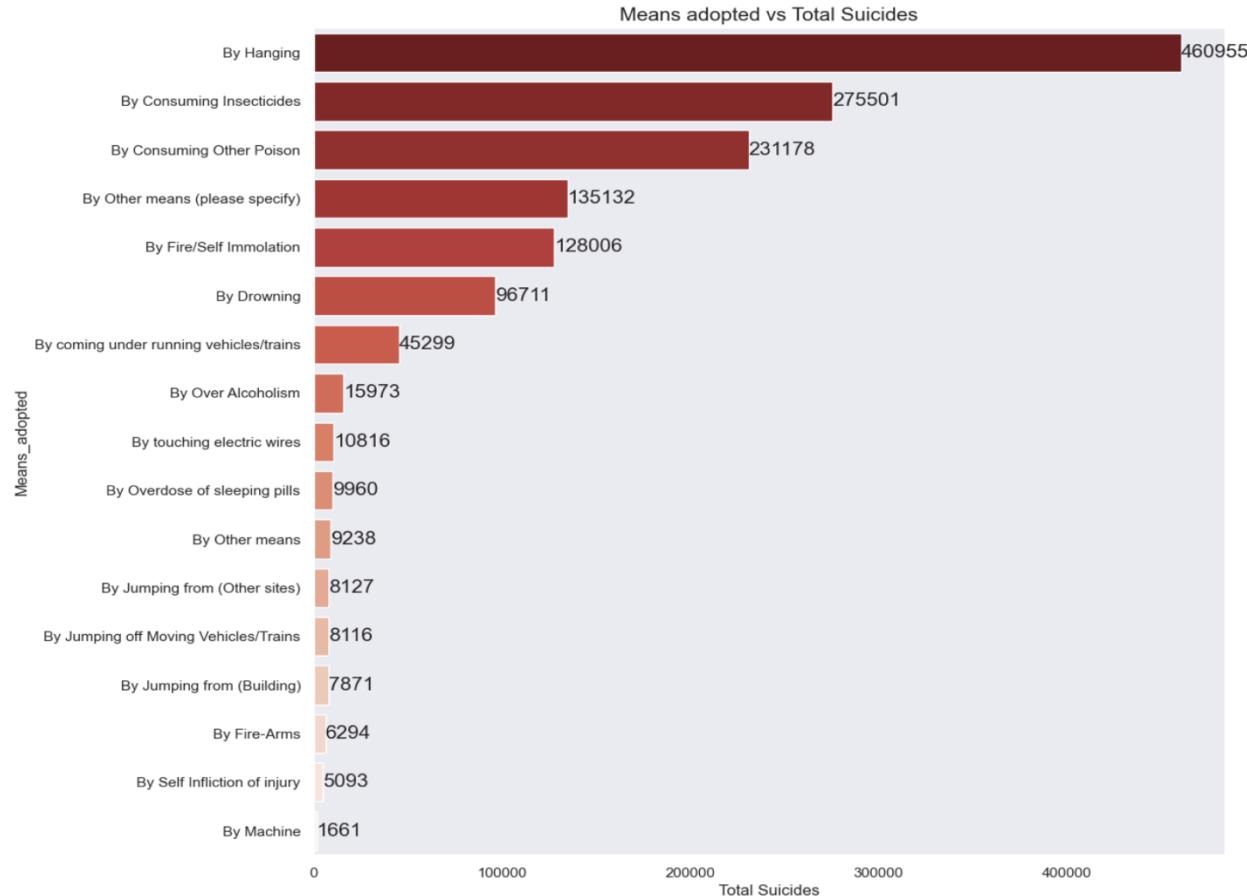


Analyzing causes that push people to commit suicide led us to sort the type\_code category with multiple causes. Family problems is the reason for the highest number of suicides among both men and women followed by mental and physical illness. We can also see from the analysis that for many suicide cases the causes are unknown which could possibly be because of lack of data or lack of insight into mental health of individuals. It is a very important parameter as it gives insight into the unknown and is a possible extension for future research in this subject.

## 6.6 Suicide by Hanging was the most common means adopted

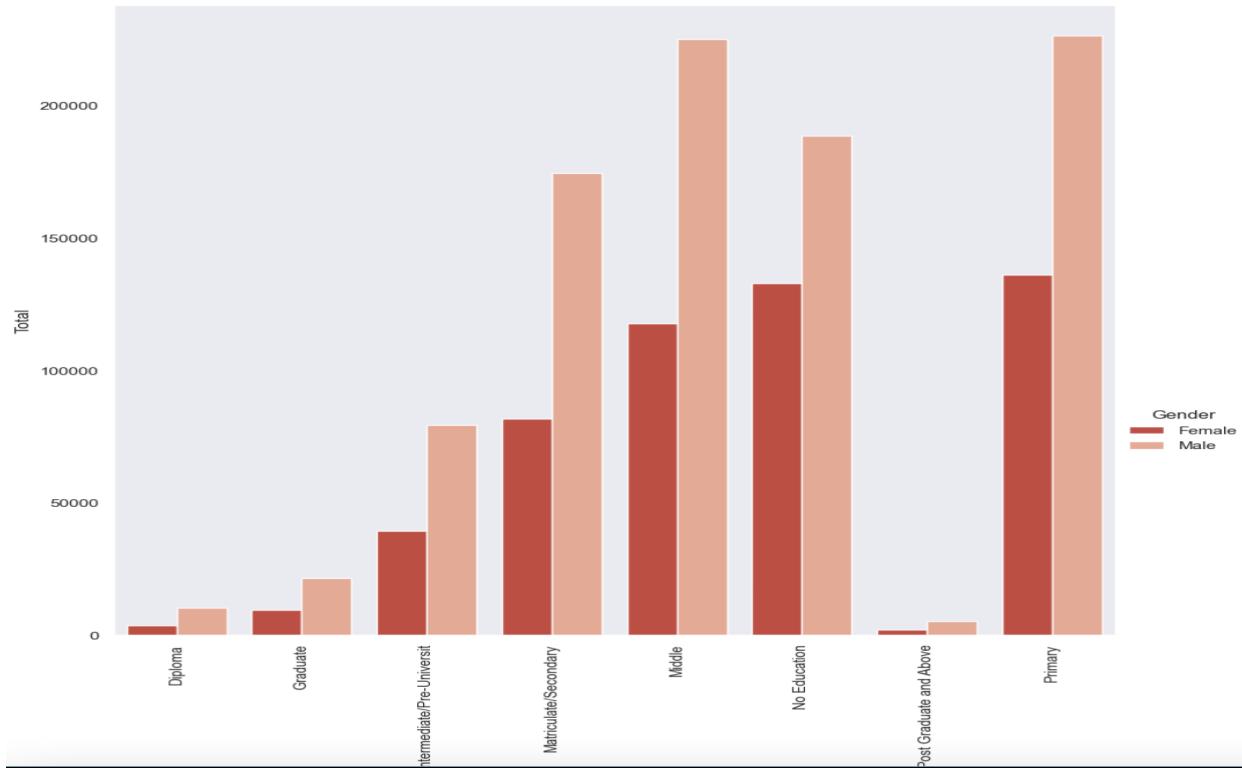
A very important clause in suicide prevention would always be to understand means and methods used to commit suicide. The data set contains information on different methods people used to commit suicide we mapped the mean count against each method of committing suicide and plotted it. We can see that maximum number of

people committed suicide by hanging themselves followed by consuming insecticides and consuming poison.

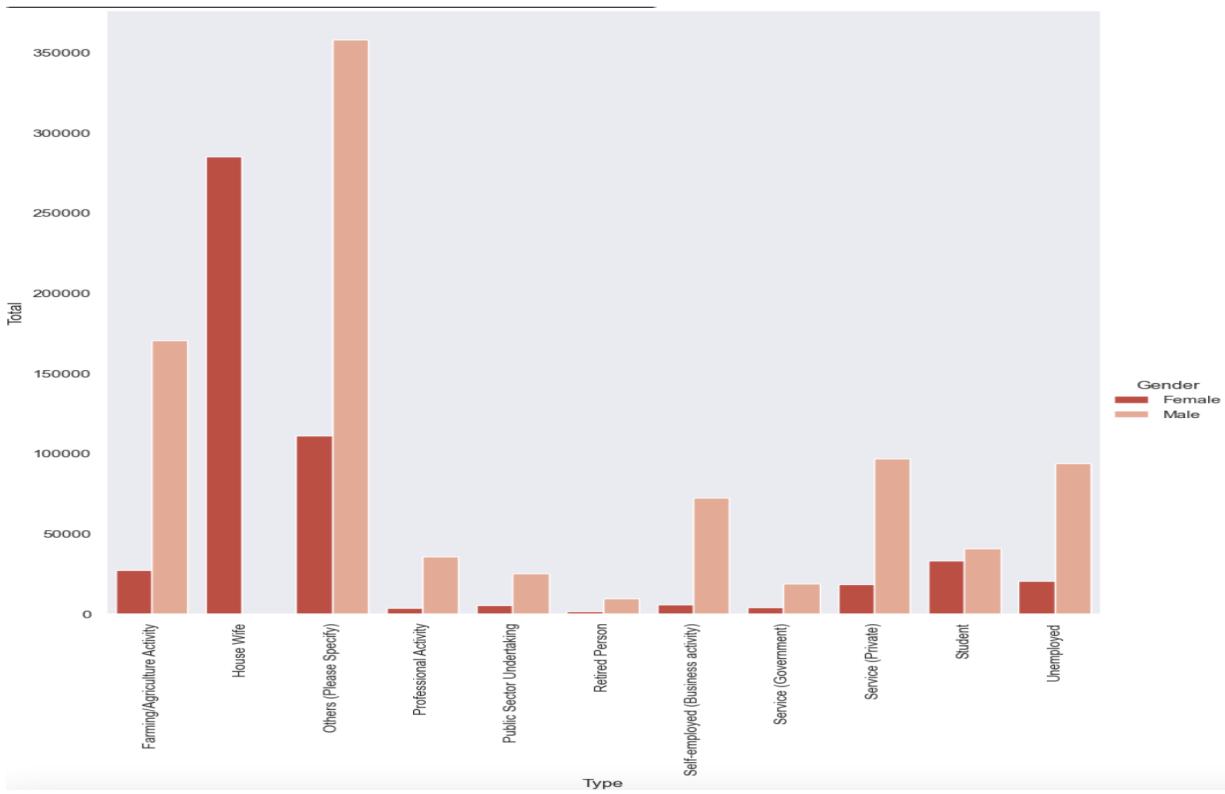


## 6.7 Primary school education was the educational status of most cases

It was seen that those that have no education, primary education and middle school education are most prone to suicide among both men and women. This highly contradicts the finding of states with highest literacy having most cases in the past decades. This also shows that the population of India is very vast and deeper level of analysis needs to be carried out with further geographical breakdown into cities and towns since a of states are a mix of urban and rural towns.

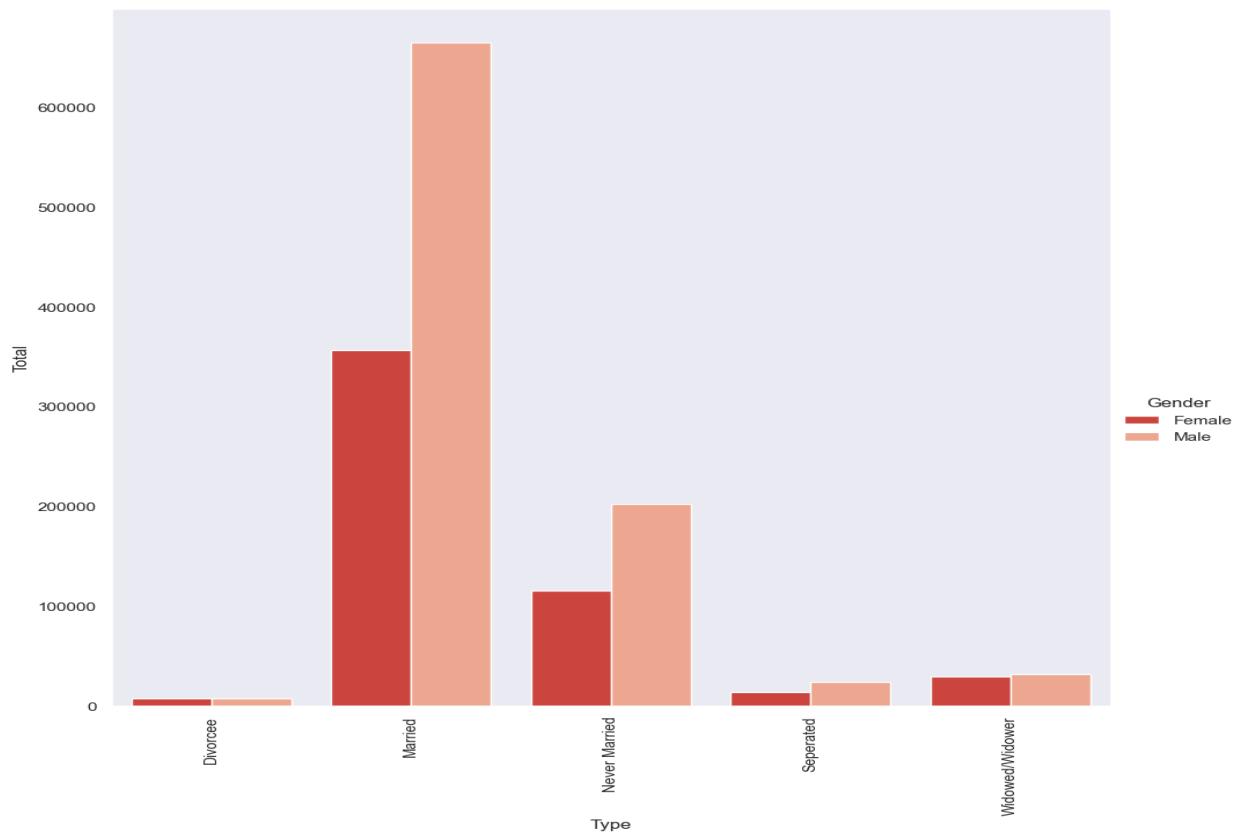


## 6.8 Few professions see majority of the share in total suicides



Economic reasons are one of the most important causes of suicidal tendencies. The type of employment, the social status attached to it, the demands of working places, and other similar factors contribute to the incidence of suicides. Accordingly, not all professions are equally vulnerable to suicide. Some categories such as self-employed, students, and unemployed persons are more vulnerable to committing suicide. However, housewives, and farmers / the people engaged in the agriculture sector – both cultivators, and agricultural laborers account for the highest share among men and women who commit suicide. The share of suicides by daily wage-earners, professionals/salaried persons, and unemployed persons is showing an increasing trend. In these increasing categories, one can observe a sudden spike in the years 2017, and 2020, when demonetization and COVID-19 happened.

### 6.9 Married men followed by married women are most prone to suicide.



There exists a strong correlation between suicides and interpersonal relationships, particularly marriage. Marriage became the biggest tendency to commit any act of self-destruction. It can be seen that married men account for the highest number of total suicide cases followed by married women. Almost half of the count of married women is the count of single men that committed suicide. This can be further investigated if we divide the married cases into arranged and love marriages since India has a custom of arrange marriages and many of them happen against the will of the involved individuals.

## 7. Challenges

It was challenging to find a dataset that had all the features that could help us make our analysis meaningful. We started with our project proposal which included dataset related to USA states. Our projected analysis was much more ambitious and did not have the data that was required inside the dataset. However, professor asked us to cover those points and hence we extended our research to other factors which led us to a whole new dataset. The new dataset included data about suicide cases in India. However, this dataset had a lot of redundant data which needed to be cleaned to give a clean visualization. Quality of data was the biggest challenge. Also, we felt the time for this project was very less and we couldn't reach the level of analysis we wished to do. There wasn't enough time to analyze the plenty of trends that can be analyzed by pitting one factor against the other and keeping the denominators as common. Most datasets did not have explicit data hence we had to combine multiple datasets and some states and causes that were on low priority were left out in the process.

## 8. Recommendation & Future Scope

- In the future we can analyze using machine learning algorithms and classifiers to better understand the data set and build a model that can state probable causes based on history of a particular state.
- Sentiment analysis via social media posts can be done by searching for keywords using linguistic
- Extending the gender column to include different genders other than binary of male, female such as transgender, non-binary etc.

- Dataset needs to be broken down further in terms of geographical area to have better analytics on the data.
- We can tie up with mental health organizations to provide analysis on this dataset and help them narrow down their targeted audience
- More awareness can be spread using a bigger dataset of a particular age group, area, profession, or gender to carry out further exploratory analysis for an organization that panders to awareness of a specific sector.

## 9. Conclusion

Suicide rate is on the rise in recent years in India which was assumed to be due to several factors like demonetization and COVID-19, but it happens to be family problems. Maybe the former factors are indirectly causing family problems or physical/mental illness which were the leading causes. The inferences were all based on the latest data, and we believe it can be used to build laws that can help bring down the numbers drastically. However, this study does not end here. Having a thorough and reliable database of suicides is highly essential, as it enables us to understand the underlying causes that push people towards suicide & its spread. Taking the analysis from the above graphs into consideration, we can infer and conclude that culture and religion has a huge impact on a country like India where with males being more prone to suicides with a high increase in deaths year on year.

People in the age group of 15-29 and 30-44 are most prone to committing suicide. Suicide is a complex phenomenon whose underlying roots are difficult to comprehend. Though there might be several reasons behind the commission of such acts, one or two reasons dominate over the others. A critical review of suicide prevention methods is facilitated by an exponential increase in effectiveness data. If we look at the data on the causes behind the suicides, the share of suicides due to family problems, love affairs, and substance abuse has been rising. This points out to the diminishing interpersonal relationships, and sense of solidarity among the individuals as well as the households. The study of suicides is of utmost importance, and it is high time that it should move away from an excessive focus on individuals, and focus on structural factors like poverty, inequality, mental health issues, etc., to understand the vulnerability of the people. And strategies for suicide prevention must extend beyond the formal organizational approaches to reach the vulnerable.

Adherence of suicide prevention advice can be inculcated using this data in depression management and suicide prevention apps to evidence-based clinical guideline recommendations such as mood and suicidal thought tracking, safety plan development, recommendation of activities to deter suicidal thoughts, information and education, access to support networks, and access to emergency counseling. Suicide prevention programs targeting one or more of these factors successfully decrease the number of suicides. An important and widespread component of suicide prevention strategies are crisis helplines, which provide timely and anonymous advice to callers at current risk of suicide and are effective in deterring active suicidal thoughts.

## 10. References

- <https://ncrb.gov.in/en/accidental-deaths-suicides-india-adsi>
- [www.macrotrends.com](http://www.macrotrends.com)
- <https://factly.in>
- <https://www.kaggle.com/code/iphylkelvin1993/global-suicide-indicator>
- <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1461-z>



Contact the 988 Suicide and Crisis Lifeline if you are experiencing mental health-related distress or are worried about a loved one who may need crisis support.

- Call or text 988
- Chat at [988lifeline.org](https://988lifeline.org)

Connect with a trained crisis counselor. 988 is confidential, free, and available 24/7/365.

Visit the 988 Suicide and Crisis Lifeline for more information at [988lifeline.org](https://988lifeline.org).