Institute for mental health

Suicide Rates in India Report February 2021

Michelle Michalowski

Spark Individual Assignment Feb 2021

1. Abstract

The Institute for Mental Health from Madrid has been dealing with a very sad and sensitive topic in recent weeks, which is nevertheless very important to talk about in our society: suicide and mental health issues.

Emerging countries, such as India or China definitely have strong societal problems to solve and with it one of the biggest problems, mental health and suicides. South East Asia stands out as a leader in trying to reduce suicide rates by one third until 2023. As India's suicide rates cover 18% of the worldwide official suicides, tackling the mental health in India would make a difference. This study attempts to look deeper into India's society and tries to better understand the different groups being in danger of committing suicides.

The following of this study is structured as follows: The first part of this study looks at the current situation and explains it based on already available studies. Next, the data set used for this study will be presented. In the main part, we will be discussing the main insights gathered through an extensive data exploration and finally we will be giving some insightful recommendation for future policies.

2. Scenario Information

Every 40 seconds a human dies by committing suicide worldwide and each suicide brings 20 more suicide attempts with it. Even though, there is evidence for a clear relationship between mental health issues and suicide rates, there are many more suicides committed impulsively in times of special crisis. Risk factors can include every kind of discrimination, isolation, problems in relationships or just financial issues.

Similar to other emerging countries, suicide and mental health disorders are an evident problem in India. With about 800,000 suicides committed yearly worldwide, 135,000 are Indian citizens which makes in total 17% of all deaths. Obviously those estimates vary among different sources. As an example the medical Journal Lancet estimated the number of total deaths through suicide at 187,000 in 2010, while the official number provided by the government was only at about 134,600 deaths. What is clear for other studies is that there certain states in India which have significantly higher suicide rates, there are more man committing suicide than women and that the suicide rates are high but there are countries which struggle with higher rates as Japan.

3. Background information

Having described the importance of this issue and the dimension in numbers, we would like be a little more specific about the underlying data. The data for this study is official data from the National Crime Records Bureau (NCRB), Government of India. This agency, based in New Delhi, is responsible for collecting crime-related data and analysing it in order to better coordinate police work and make it more data-based.

Even though this data is also published on the official site of this police authority, we have decided to choose the dataset provided by Rajanand Ilangovan on Kaggle. The dataset can be found at the following link: https://www.kaggle.com/rajanand/suicides-in-india

In total the quality oft the data set is high and it offers a good size of around 240 thousand rows. The data set covers annually data with a period of 11 years, starting in 2001 and ending in 2012. From a geographic point of view, the dataset is also very diversified, as it includes all the states of India, making it very interesting for subsequent analysis.

Although we will go into this in more detail later, we would like to briefly introduce the various features. Even if the dataset only has seven different columns, a lot of information can be derived from it. The columns are divided as follows: State, Year, Type_code, Type, Gender, Age_group, Total. Given the type columns, which contain various information on socio-economic characteristics, it was possible to find out very precisely what kind of people are at risk. More on this later in the exploratory data analysis section.

4. Goal for the analysis

The goal of this analysis is to better understand the development, current state and the people committing suicide. Obviously, we cannot change the past, but it is important to address this sensitive and dramatic issue and to better understand the exact reasons behind it. With this improved understanding, prevention measures can be better designed and become more data-based than emotions-based.

Since this study is analysing socio-demographic data (such as age-range, social status, income, education) we are seeking to understand the groups of people suffering from mental health issues better. With these findings, more concrete recommendations for prevention can be provided. Such recommendations could be, for example, that less educated people in particular are at high risk in specific regions of India and that these people should be targeted by preventive measures. On the other hand, one can also find out which people are less affected and thus better allocate the various resources available.

5. Analysis and results

Exploratory Data Analysis (EDA)

Typically, we start our data analysis with an exploratory data analysis. This analysis serves to get a better overview of the data and to gain a basic understanding of it.

The entire analysis presented here was done using Pyspark, which is a python API for spark and a framework which is efficient for analysing large datasets like the one presented here.

To start with, we would like to explain the key facts of the dataset. The dataset has 237,519 observations and a total of seven features. Following a presentation of the schema:

```
root
 |-- State: string (nullable = true)
 |-- Year: integer (nullable = true)
 |-- Type_code: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Gender: string (nullable = true)
 |-- Age_group: string (nullable = true)
 |-- Total: integer (nullable = true)
```

This DataFrame has **237519 rows**.

Those seven columns are categorized as follows: State, Year, Type_code, Type, Gender, Age_group and the Total:

- **State:**
  - Identifying the State inside India
- **Year:**
  - Identifying the year of this specific observation
- **Type_code:**
  - Identifying the type of observation, includes: Education status, Professional Profile, Means adopted (how person committed suicide), social status and the causes
- **Type:**
  - The type basically presents the same as the type_code but specifies a bit further
- **Gender:**
  - Identifying the gender split into female and male
- **Age_group:**
  - Identifying the age of the person while not giving the exact number rather the range
- **Total:**
  - The total number of suicides committed

5. Analysis and results

Exploratory Data Analysis (EDA)

Overall, the quality of the data set was very good with no missing values and only some mistakes within the categorical variable of States. After removing some wrongly assigned sates the data was already ready to work with. Checking for distinct values within each feature, we obtained the following result:

State	Type_code	Type	Age_group
38	5	69	6

After going through basic quality checks, we looked a bit closer into the actual content of the data. Since the main purpose of this analysis was to look into the drivers for mental issues and committing suicide we worked on some basic aggregations. Results of those were that:

- The **most frequent cause** for committing suicide was family problems with a total of 341952.
- The **least frequent cause** for committing suicide is not having children with a total of 766.
- The **most frequent professional profil**e was Others which means that the actual profession could not be identified with a total of 469147.
- The **least frequent professional profile** was being retired with a total of 11334.
- The **most frequent education status** was finished education in the primary school with a total of 362827.
- The **least frequent education status** was finished education with post graduate studies with a total of 7475.
- The **most frequent mode of committing suicide** was by hanging with a total of 460955.
- The **least frequent mode of committing suicide** was by machine with a total of 1661.
- The **most frequent social status** is being married with a total of 1021774
- The **least frequent social status** is being divorced with a total of 15272.

However, those values should be interpreted with caution as for example, the share of people who are married in India is higher than those living divorced.

5. Analysis and results

Questions relevant in the context to better understand the causes and groups

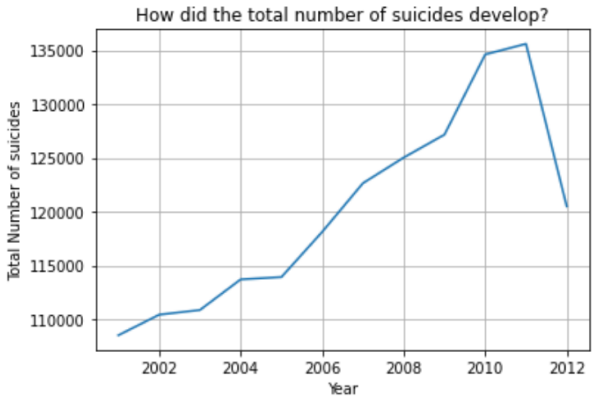
During our analysis we focused on the following main questions as a guideline for our extensive research:

How did the yearly sum of suicides develop and how did they develop compared to other countries/worldwide?

Calculating the yearly sum of suicides for the given time period, it becomes evident that there is an increase in the suicides committed for each year except for 2012. Below the concrete numbers obtained:

Year	Sum
2001	108506
2002	110417
2003	110851
2004	113697
2005	113914
2006	118112
2007	122637
2008	125017
2009	127151
2010	134599
2011	135585
2012	120488

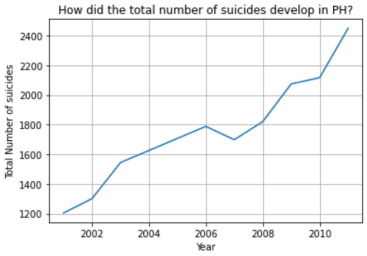
Since it is hard to compare absolute numbers, we decided to create a basic plot giving a better overview of the findings:

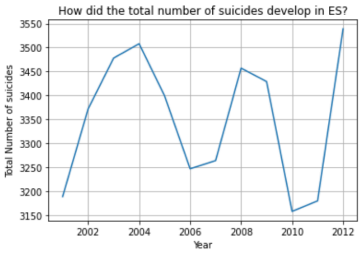


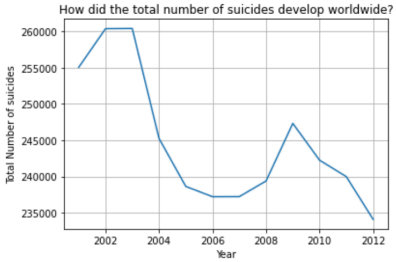
Similar to what we could see by analysing the numbers, we see a peak in the year 2011 and a steady increase of cases for the years from 2001 and 2011.

5. Analysis and results

As we are convinced that those numbers themselves have limited explanatory power, we took into consideration a second data source. The data set used here is provided by the WHO and basically includes the same information as our data set except that it covers a variety of countries all over the world.







Starting with the Philippines as comparison one can say that similar to what we have seen for India, there is a steady increase of number of suicides. The difference here is that the relative increase of suicides in the Philippines is definitely higher. As we can see in 2001 there were around 1200 suicides, while in 2011 the number doubled with around 2400.

Looking closer at the development of suicides committed in Spain it can be seen that in the given time frame the number is relatively stable, not like we have seen for India and the Philippines. Finally, looking at the worldwide development of suicides, we can see the overall trend is towards less suicides committed.

What was the yearly % change of suicide cases?

After looking into the absolute yearly change of suicides committed, we want to look closer on the relative development. As can been in the following tables there is a positive % change for all years except for 2012. What stands out is that the change was extremely high in 2010 which could be an after-effect of the global economic crisis that has plunged many Indian families into financial ruin. At the same time, it is very pleasing that the measures taken to reduce mental health issues have taken root and we can observe a negative growth in the year. On average, the growth was 2 % per year.

Year	Sum	Previous	diff	perc_change
2001	542440	null	0	null
2002	551980	542440	9540	2.0
2003	554218	551980	2238	0.0
2004	568349	554218	14131	3.0
2005	569545	568349	1196	0.0
2006	590543	569545	20998	4.0
2007	613119	590543	22576	4.0
2008	625014	613119	11895	2.0
2009	635429	625014	10415	2.0
2010	672926	635429	37497	6.0
2011	677159	672926	4233	1.0
2012	647288	677159	-29871	-4.0

min_perc_change	max_perc_change	avg_perc_change
-4.0	6.0	2.0

5. Analysis and results

Do the numbers of suicides committed change among age groups and gender?

After getting a better sense of the scale of the problem and growth, let's take a closer look at the groups that are particularly affected, starting with gender and age. The results of the analysis show that there is a significant difference between age groups, with mainly older people committing suicide and a relatively small number of children. Looking at the number of suicides committed by gender, the result is really surprising with almost twice as many men as women. This means that men are twice as likely to commit suicide as women.

Age_group	Sum
Very old adult	2911862
Young adult	1534037
Medium aged adult	1471599
Older aged adult	885177
Retired adult	346925
Child	98410

Gender	Sum
Female	2606922
Male	4641088

Is there a linear relationship between education status and number of suicides committed?

Moving on to the next interesting feature, the education of the people, we decided to look at a correlation analysis to see if there is a direct evident linear relationship and the result obtained is very significant: The resulting correlation coefficient with a value of -0.93 is strongly negative and very close to one, were one could already speak of a perfect negative correlation. The concrete interpretation of this would then be to say there higher the school-leaving qualification acquired, the less likely it is that this person will commit suicide. On the other hand, this also means that the lower the level of education, the higher the risk of committing suicide.

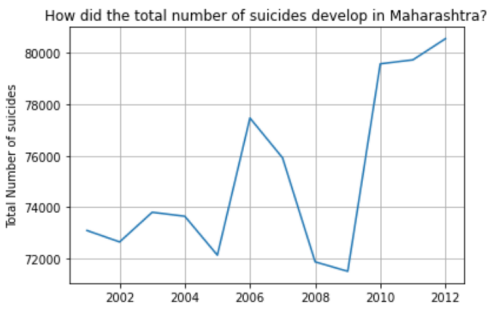
5. Analysis and results

How do suicide rates change among states?

To better differentiate regionally, we have also looked at the aggregate total per state in India. It is striking that it differs significantly from state to state. The states Maharashtra, West Bengal and Tamil Nadu are the most affected by the problem. We then looked at how the numbers developed in the most affected state Maharashtra and obtained the following result:

State	Sum
Maharashtra	901945
West Bengal	849936
Tamil Nadu	818691
Andhra Pradesh	814059
Karnataka	734825
Kerala	538946
Madhya Pradesh	451535
Gujarat	330858
Chhattisgarh	302354
Odisha	267234
Rajasthan	255134
Uttar Pradesh	233352
Assam	172276
Haryana	147176
Delhi (Ut)	84272
Jharkhand	49720
Punjab	46350
Bihar	46214
Tripura	45965
Puducherry	32144

only showing top 20 rows



Here one can see that the number of people committing suicide has also increased in the state with the highest rate. However, the increase is not particularly strong and the number of people committing suicide is at a constant high level.

How did the number for each of the causes develop over the years?

Since there are quite many cases, we decided to focus on the two main causes: family problems and insanity/mental illness. Looking closer at the growth rates the analysis unfortunately does not show a clear trend for those causes. In some years both show a negative growth in some years a positive growth.

Year	Total	Previous	diff	perc_change_family_problems
2001	24162	null	0	null
2002	25085	24162	923	4.0
2003	26308	25085	1223	5.0
2004	25531	26308	-777	-3.0
2005	25441	25531	-90	0.0
2006	30771	25441	5330	21.0
2007	29238	30771	-1533	-5.0
2008	29777	29238	539	2.0
2009	30082	29777	305	1.0
2010	31856	30082	1774	6.0
2011	32909	31856	1053	3.0
2012	30792	32909	-2117	-6.0

Year	Total	Previous	diff	perc_change_illness
2001	15947	null	0	null
2002	17296	15947	1349	8.0
2003	15636	17296	-1660	-10.0
2004	16341	15636	705	5.0
2005	14820	16341	-1521	-9.0
2006	16497	14820	1677	11.0
2007	16452	16497	-45	0.0
2008	16651	16452	199	1.0
2009	16028	16651	-623	-4.0
2010	17044	16028	1016	6.0
2011	16154	17044	-890	-5.0
2012	15699	16154	-455	-3.0

After checking the metrics for overall percentage change looking at both causes categories, there is a clear difference. For family problems we can see an average increase of 3% while for illness the average is only at 0%. Lastly, the rather neutral development for illness as a cause could be due to the increased wealth and better medicare infrastructure.

min_perc_change	max_perc_change	avg_perc_change_family
-6.0	21.0	3.0

min_perc_change	max_perc_change	avg_perc_change_illness
-10.0	11.0	0.0

6. Conclusion & Recommendations

Overall, the dataset used here has been shown to be useful in understanding mental health issues. Based on the questions addressed, it can be concluded that India does have a problem with rising suicide rates compared to the global average, but at the same time does not do worse than comparable countries such as the Philippines, which is also an emerging country.

In terms of annual growth, there was a clear growth trend with an average of 2%, which hopefully will decrease in the coming years with appropriate policy programmes.

If one were to define the most severely affected segments of society as a group, older men with little schooling and family problems living in the states Maharashtra, West Bengal and Tamil Nadu would be the most vulnerable.

Based on these findings, it can be concluded that this group should be particularly targeted by the measures. These could look as follows. Improved schooling would ensure a more stable future for these people and thus reduce emotional stress. In addition, it would also be important to introduce education and awareness around mental health so that the issue is openly addressed in society and perceived as a real and combatable problem.

Last but not least, as a mental health institute, we think that it would help to provide special psychological services for poorer families with problems.

7. References

En.wikipedia.org. 2021. *Suicide in India*. [online] Available at: <https://en.wikipedia.org/wiki/Suicide_in_India#:~:text=In%202016%20the%20number%20of,with%2017.5%25%20of%20world%20population.> [Accessed 20 February 2021].

Who.int. 2021. *Suicide - India*. [online] Available at: <https://www.who.int/india/health-topics/suicide> [Accessed 20 February 2021].

Snowdon, John. "Indian suicide data: What do they mean?." *The Indian journal of medical research* 150.4 (2019): 315.

Kaggle.com. 2021. *Suicides in India*. [online] Available at: <https://www.kaggle.com/rajanand/suicides-in-india> [Accessed 12 February 2021].