Analysis on Accidental Deaths and Suicide In India

CS685 Project Report

Group 20

- Manjyot Singh Nanra - 21111038 - manjyots21@iitk.ac.in
- Akshay Kumar Chittora - 21111007 - akshay21@iitk.ac.in
- Jeet Sarangi - 21111032 - jeets@iitk.ac.in
- Akash Dayanand Chavan - 180054 akashh@iitk.ac.in
- Rohit Kulhari - 180623- rkulhari@iitk.ac.in

1 Abstract

Loss of lives due to accident and suicides result in irreparable damages to the families and society at large. Exhaustive information on accidents and suicide is of utmost importance to carry out any meaningful study and devise appropriate intervention measure.

In this project we aim to perform analysis on the accidental deaths and suicide data for India to with a goal that in future, policy makers, law enforcement agencies, government departments, researchers, NGOs, etc find the information useful.

2 Problem Statement

We have performed exploratory analysis on suicide and accidental deaths in India. We have tried to analyse trend of suicide numbers over the last 5-6 years for India in general and states specifically as well. We have also explored on problems on following subtopics.

- Analysing suicide numbers for **literacy groups** to find which groups are more vulnerable.
- Analysing suicides based on **profession**.
- Analysing **causes** of suicide to find out leading causes of suicide.
- Analysing different **means** by which suicide happens.
- Analysing suicides based on **economic status** of the people.
- Analysing suicides based on the **social status** of people
- Analysing **overall trend** of the suicides.
- Analysing **overall trend of the accidental deaths.**

For this purpose, we have used data of Accidental Deaths and Suicides in India(ADSI) published by National Crime Records Bureau(NCRB).
We have performed this analysis so that we can get an insight of which which groups of people are more vulnerable to suicides and get overall trend on the topic so that necessary intervening actions can be taken.

3 Introduction

Suicide and accidental deaths ends up taking precious lives and does a irreparable damage to the families and society. We have performed analysis on data of Accidental Deaths and Suicide in India to find out patterns or vulnerable groups that are more susceptible towards suicide. Analysis is performed on data available by NCRB which is in both PDF and CSV formats. Data Mining techniques are used to convert the PDF format database into required CSV formats. Different format of data has been cleaned so that exploratory analysis could be performed on them.

4 Motivation

Approximately 1.5% of all deaths worldwide are by suicide. India itself recorded around 419 suicide deaths daily for the year of 2020. The effects of suicide are tragic and felt long after the individual has taken their own life. Every step that can help towards prevention of precious lives is helpful. Hence we are motivated to do analysis on suicide data so that we can get more insights on groups or states regarding on suicide numbers and take appropriate action. Being aware of the issue is the first step to solving it. We will also be doing analysis on accidental death data to get insight this issue. Loss of death due to accidental deaths should be avoided at any cost hence we are motivated to get insights on what are the major causes. Goal of this is that in future, required authorities can use this analysis to take appropriate decisions to mitigate these problems.

5 Dataset

We have obtained Accidental death and Suicide in India(ADSI) data from two sources. Data from 2015-2019 is obtained from data.gov.in website [1] [2]. Data for year 2020 is obtained from official ncrb website[2]. Data from 2015-2019 is available in csv format, but data for 2020 is available in pdf format, so pre-processing is required data is used for analysis. All the data is available in State-wise and UT wise format. We have also made utilization of state census data in few analysis. Various dataset used for analysis are as follows.

- Educational Status wise Distribution of Suicides
- Economic Status wise Distribution of Suicides
- Social Status wise Distribution of Suicides
- Profession wise Distribution of Suicides
- Distribution of Suicides by Means/Mode Adopted
- Causes – wise Distribution of Suicides
- Census Dataset

6 Methodology

6.1 Pre Processing

Pre-processing was essential for our project because of the various sources our data is coming from. Few from pdf files and a few from CSV. We had done processing that all files are in CSV format. We have used Camelot Library from python, which helps in extracting tables from pdf files. However, this library wasn't working well for a few pdf files, so we manually converted the pdf files to desired CSV files. All data files from 2020 have been converted from pdf to CSV. Distribution of Suicides by Means/Mode Adopted 2011 pdf file has also been converted to CSV. The majority of the project effort was spent on data collection and pre-processing.

6.2 Data Cleaning

After obtaining CSV files for the year 2015 to 2020, data cleaning was done. Every file had different column headers and was of different format. So each file of same format had to be modified to have uniform columns and rows. Union Territory Dadra and Nagar Haveli and Daman and Diu were dropped from analysis because of them being merged into single Union Territory caused inconsistencies. After all files from 2015-2020 were cleaned they were used for exploratory analysis.

6.3 Data Analysis

6.3.1 Education Data Analysis

Performed Education data analysis on data from 2016-2020. 2020 data was converted from pdf format to csv using camelot library. Data cleaning was performed then to make data ready for analysis. Analysis has been done to visualize top education group suicide numbers. We have also taken ratio of number of suicides in education group with total education population obtained from census dataset. Pie chart has been plotted to see percentage shares and normal line graphs to visualize suicide trend for each education group.

6.3.2 Cause Wise Data Analysis

Performed analysis on Cause Wise Suicide data available from year 2016-2020. Data from 2020 year was in csv format which was manually converted to csv format. After the data cleaning, analysis was performed by plotting pie charts for causes year wise as well as taking aggregation to give a percent wise distribution and visualization on leading causes of suicides.

6.4 Social Data Analysis

In this analysis we took the suicides numbers from the social data for each year and compared the data for male and female accordingly. The data is extracted from the csv file and 2020 from the pdf file of nrcb, after the data preprocessing phase the dataset us used to create data frames for male and females and total separately, we can infer from the data that which year has how many females unmarried or males unmaaried have committed suicide and also the same can be done with married people also.Also data is used to infer the number of suicides of different sates for both married as well as unmarried people.

6.5 Economic Data Analysis

Performed analysis on the economic data wise Suicide numbers for different years. In this analysis data is being collected for year 2015 to 2020 from the government website, used that data to collect the information for different years as to which income base has how many numbers of suicides and how that rate is increasing or decreasing. The data is being plote for years wise and for each income group like less than 1 lakh, 1 to 5 lakh, 5 to 10 lakhs and more than 10 lakhs, for each group data is represented to give a visualization on the income factor consideration in the number of suicides.

6.6 Profession Data Analysis

Performed analysis on division of suicides Professions wise from year 2015-2020.Data for year 2020 was in pdf format which we converted manually to csv format.After obtaining all the data for all years we built a summary of all trends by plotting the division of suicides cases profession wise to find the percentage distribution in all then tried finding most vulnerable group.For that group then we tried finding statewise highest where the suicide cases are highest for that group and then finally plotted the male to female counts for each professional group.

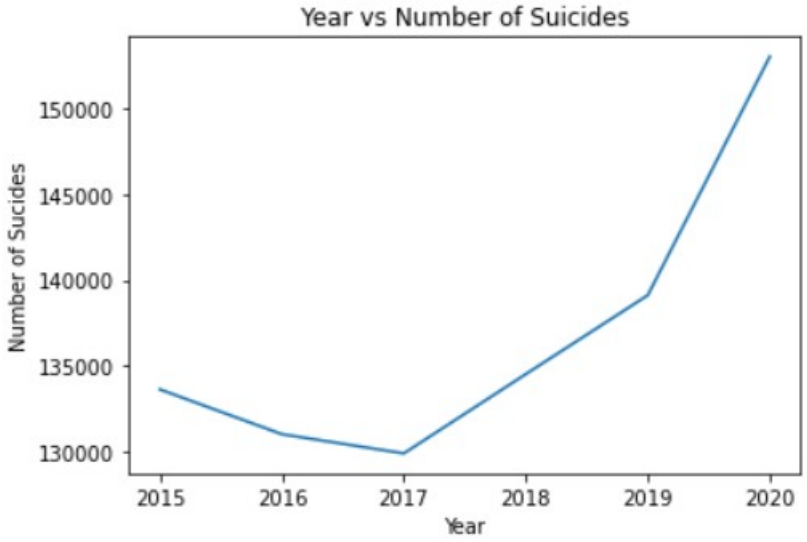
6.7 Means Adopted Data Analysis

Performed Data analysis on Data on means adopted for suicides using manual method obtained 2020 data from from pdf then using these tried finding for all state wise distribution of all suicide cases by means or modes they'ev adopted for comminting suicides.

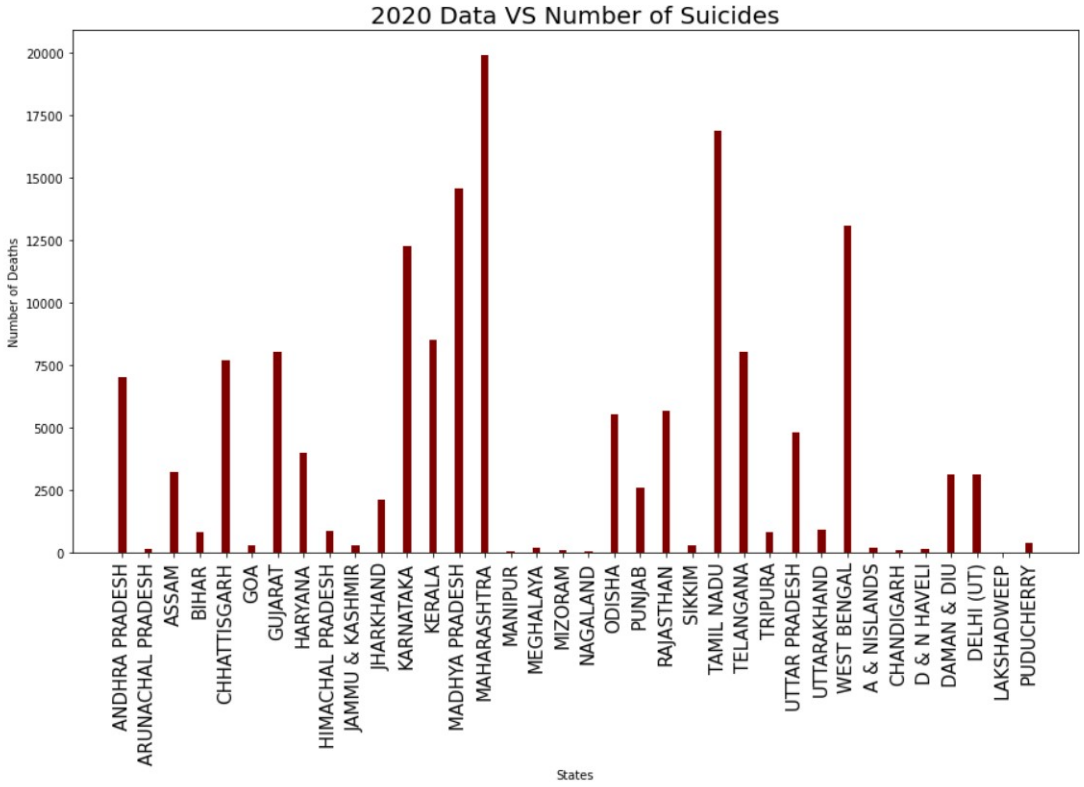
6.8 Accident Data Analysis

Performed analysis on Accident data from the Dataset published by NCRB(National Crime Report Bureau form year 2015 to 2020,for year 2020 obtained the csv from pdf manually.Then using this plotted all year wise Accidental Deaths and then tried finding states where deaths are highest and lowest.Then Statewise plotted for male and female Death cases across all 6 years.Then using pie chart of total male-female cases we came to know that male Deaths are high.Finally ended with plotting male-females deaths reported for all the states of India.

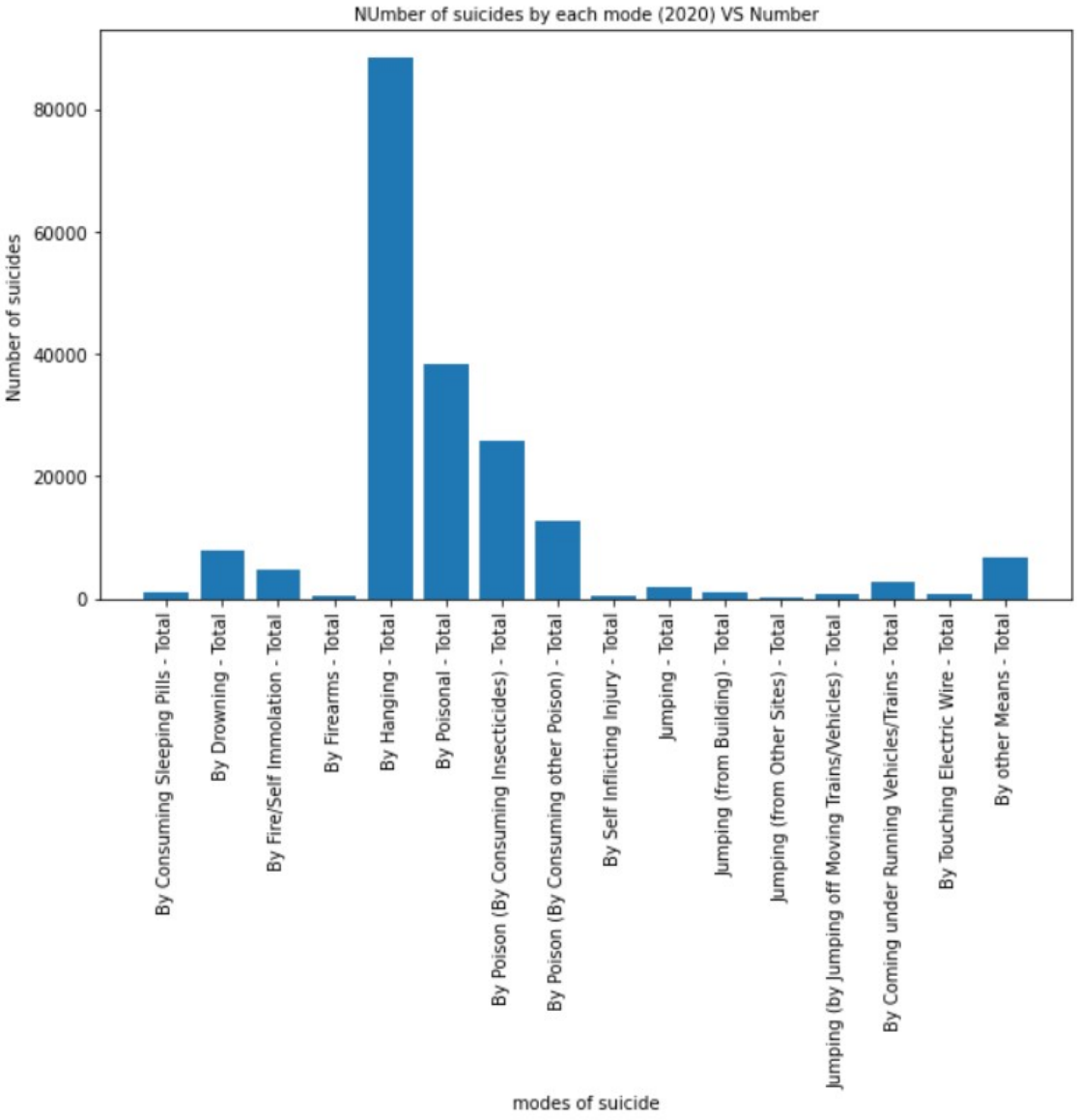
7 Results



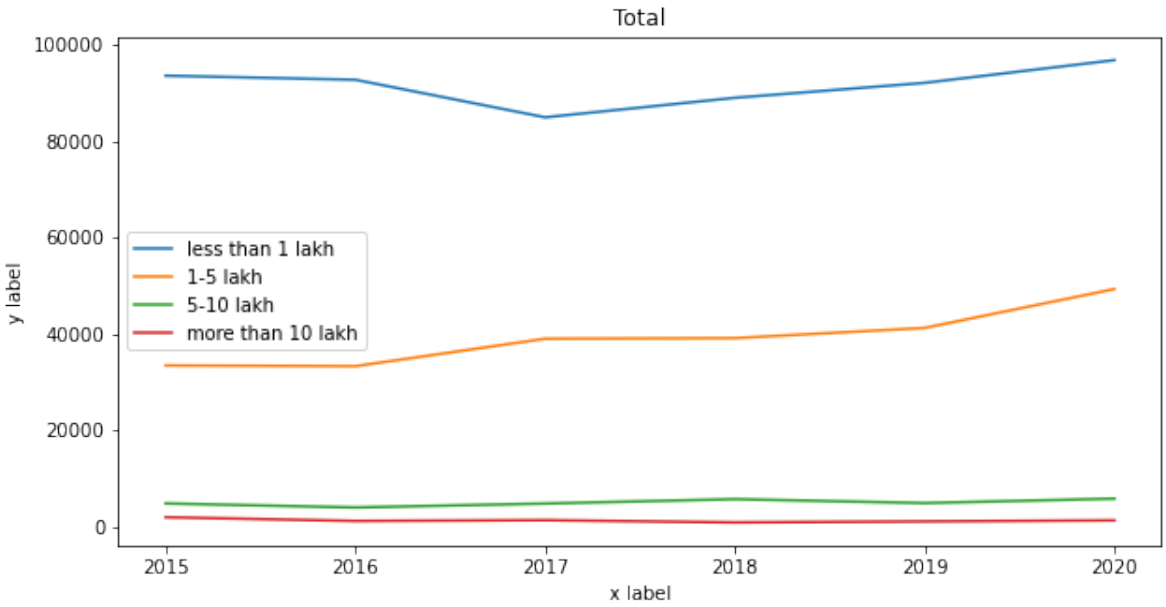
This resulting graph plot shows how the trend of number of cases for suicides got decreased for year 2017 and the shows an increasing trend for all years after that and was highest in 2020 may be due to rise in number of Covid-19 and rising unemployment could be a reason because from profession wise analysis we had obtained result showing Daily wage worker being most vulnerable group for suicides and during covid-19 many of this group lost there jobs.



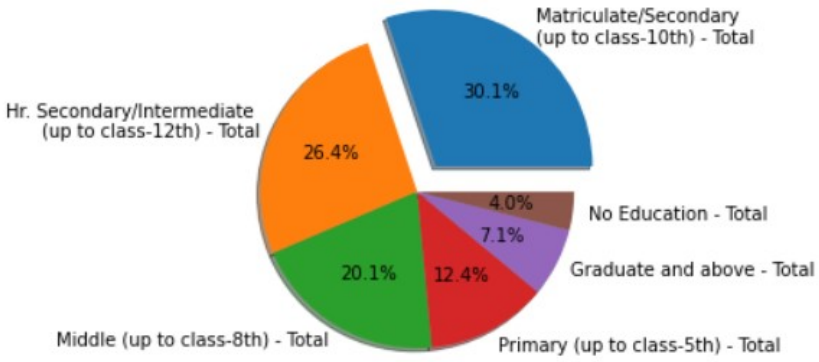
In year 2020 Maharastra and Tamilnadu were the states where the number of suicides were the most may be due to being so much crowded and as many people from small town come to these places in search of works and during 2020 due to covid 19 unemployment was on peak.



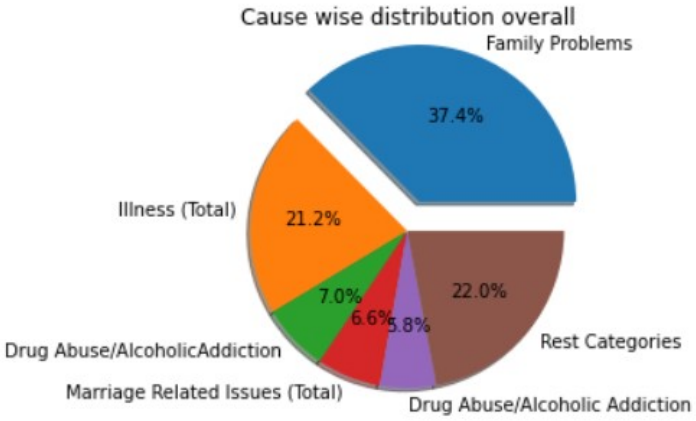
By seeing the above we can infer that people mostly use the mode of hanging for suicide and then by poison then by Drowning.Hence if somehow we can restrict sell of poisons it might reduce the suicides as well



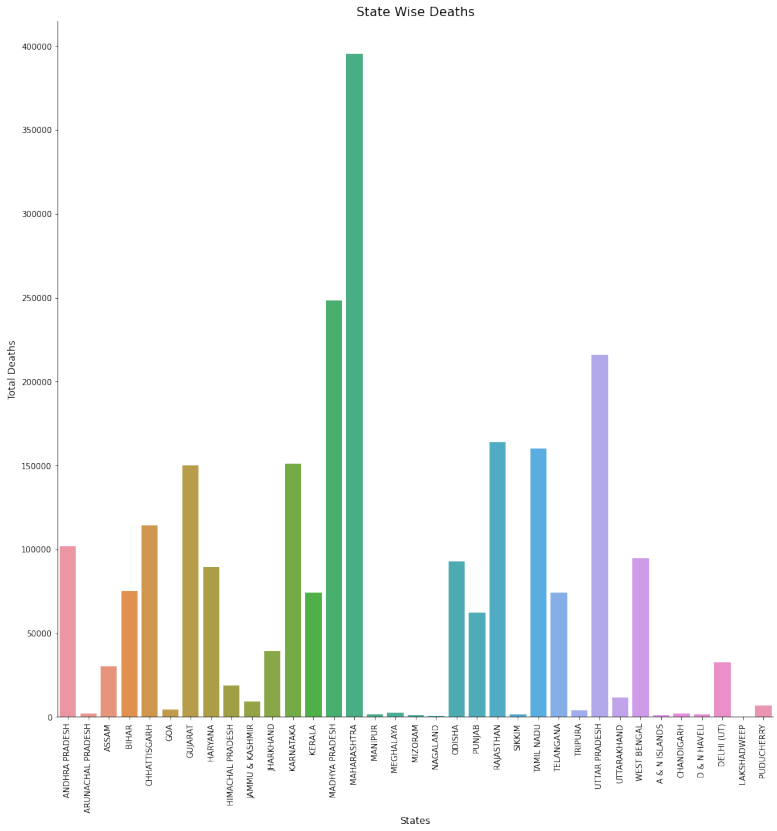
In this graph economic data wise distribution of numbers is shown, we can see that as the for all every year number of suicides is more with people having less income i.e.,less than 1 Lakh, more and more fluctuations can be seen with people with income ranging from Less than 1 and 1 to 5 lakhs and for the rest it is remaining almost constant for all years.



This is the pie-chart for analysis on education group. For this, we have aggregated suicide numbers from 2016 to 2020 for each group, and divide each of them with total population of that age group according to the census data. Here we can observe that Matriculate/Secondary(upto class 10th) education group people are more likely to commit suicide followed by people who have studied till class 12th then class8th. Education group Secondary, higher secondary and middle occupy 75% of the percentage share.



This is pie chart for analysis on causes of suicide. We have aggregated the suicide numbers from 2016-2020 for this. As we can see we get insight the 37.4% share among the suicides were due to family problems. 21.2% was attributed by illness. Among causes around 60% share was due to illness and family problems. Drug abuse, alcohol addiction and marriage related issues also had a significant contribution towards suicides.



On this above analysis we have tried to show how the number of cases of Accidental Deaths for each states are varying and states like Maharastra and Madhya Pradesh has highest share in this and states of North-eastern India like Manipur mizoram India has lowest.

8 Discussion

In this project we performed analysis on the accidental deaths and suicide data for India with a goal that in future, policy makers, law enforcement agencies, government departments, researchers, NGOs, etc find the information useful. We used the data from 2015 to 2020 to do the analysis and get the insights of the data.

- For the country as a whole, 2020 has shown a greater increase in number of suicides, however its not been very significant increase for other years but the increasing number of cases every year should work as an alarm to the authorities to take appropriates steps in this issue.
- Profession wise it is clear that Daily wage earners are the most vulnerable to suicides this too is high in TamilNadu and Maharastra.
- Modes wise we can infer thhat most of the suicides people attempt are by hanging followed by different poisons.
- Education data shows that Matriculate/Secondary education group people are more likely to commit suicide followed by 12th, Secondary/Higher Secondary occupies 75 percent of the total share in numbers which increases demand for increasing awareness in the school education policies.
- If we go for causes of suicides we can infer that most of the numbers almost 37 percents are beacuse of the family problems or the internal family issues.
- Economic wise also we can infer that people with less incomes are more incline toward suicide and numbers are increasing for less incomes group as well.
- This resulting graph plot shows how the trend of number of cases for suicides got decreased for year 2017 and the shows an increasing trend for all years after that and was highest in 2020 may be due to rise in number of Covid-19 and rising unemployment could be a reason.
- For accidents we can see that number of cases for accidental deaths has reduced in 2020, it was at max in 2016 and least in 2015.
- We can also infer from Accidents data that Maharashtra is the top state with most number of accidents this may be due to over crowding in this state.

Future Direction In our project, we were only able to analyse data from 2015-2020. Future works include analysing data from previous years too and more data about other factors that influence the deaths numbers would addd

more to the work. Data before 2015 would required a lots of pre-processing work but it would be good for analysis purpose. Also we can give extract lot and lot more info as the data keeps adding up.

References

[1] "Open government data (ogd) platform india."

[2] "Ncrb adsi dataset."