INFORMATICS FOR ENGINEERING MANAGEMENT

EM 624- Final Project

Suicides in India Data Analysis- Year: 2001-2012

Submitted By: Nishit Parmar

Guided by: Dr. Carlo Lipizzi

Date: 27th April 2018

INDEX:

I) Project Overview and Research Question:

The cases of suicides are increasing over the years. There are several reasons which contribute to it. According to a 2012 Lancet report, India has one of the world's highest suicide rates for youth aged 15 to 29. The dataset used was released by National Crime Records Bureau (NRCB), Government of India. The target to analyze this data is to find out the trend in suicides over the years 2001-2012, to find the relation between the event of suicide and several parameters like:

The state in which a suicide occurs

1) Age
2) Gender
3) Social status
4) Education/Literacy level
5) Cause of suicide
6) Means adopted

After analyzing all these parameters in detail, a conclusion must be drawn.

II) Research Motivation:

Ending one's life is scary, a newspaper article headline read "Every hour, one student commits suicide in India". Enough research and analysis are not done on the topic which drove me to analyze this data set. The result and conclusions will clarify the different trends and reasons in suicide to concerned organizations like: National Suicide Prevention Helplines, the police and other related bodies and will help them prevent it and drastically reduce the numbers.

III) Dataset Description:

The data set is a collection of the number of suicides in India in from the duration of 2001 till 2012 based on several parameters

On the dataset we can analyze:

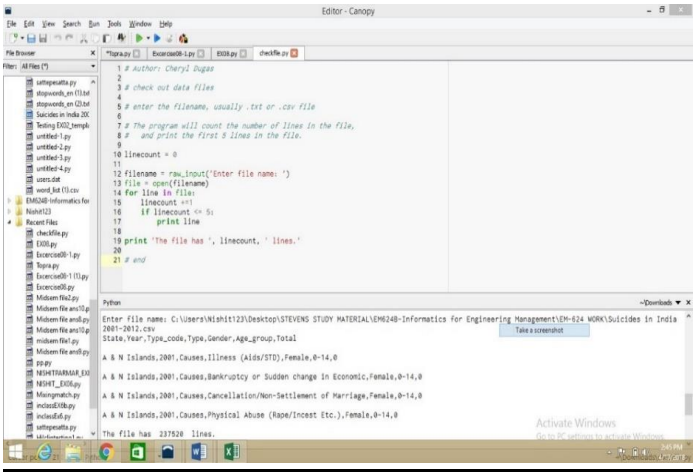
1) The states in which the maximum suicides take place.
2) In which age groups is suicides most prevalent.
3) To find the correlation between factors like literacy, social class, gender etc. with suicide
4) The most common cause for the suicides in India, across the country.
5) The most opted means to commit suicide.
6) Try to analyze the above and minimize the conditions which lead to suicides

Source of the data – National Crime Records Bureau (NRCB), Government of India. (https://data.gov.in/government-open-data-license-india)

Types of data items – The parameters include:

1) Years
2) State.
3) Social class
4) Education (Literacy Level)
5) Gender
6) Age group
7) Causes
8) Means adopted
9) Professional profile

Number of records and sample:



DATA- 237520 lines.

IV) Methodology:

Data cleaning:

1) Total suicides and Total state wise suicides were deleted as they were calculated in the program.
2) There were spelling mistakes in the data which were found and replaced to make the data more representable and understandable (14962 entries)
3) Other than that, the data was quite clean and interpretable, therefore did not need data cleaning tools.

Python program overview: Python was used to visualize the given data. The data becomes much more interpretable after being visualized. The steps followed to build the program are:

1) The libraries used are :

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

2) Pandas is used for reading the csv file into a Pandas data structure .
3) Under the Type_code- the different categories are: Causes , Education_Status, Means_adopted, Professional_profile, Social_status.
4) Read.loc[] is used as a label based indexer for selecting it by the label in which : .groupby() to group according to columns, .sum() to return the sum of values, reset_index() to sort by the values, ascending =False(as we want descending plots), and .head() represents the number of variables which will be portrayed on the graph

```
data1 = read.loc[(read.Type_code == 'Education_Status')].groupby(['Type',])['Total'].sum().reset_index().sort_values('Total',ascending=Fals
data1.set_index(['Type',])
data1=data1.set_index(['Type',])
plt.subplots(figsize=(10,6))
plot = sns.barplot(y='Total',x=data1.index,data=data1,palette="PuOr",).set_title('Relation of Education status with suicides in India')
plt.xticks(rotation=90)
plt.xlabel('Education Level')
plt.ylabel('Total no. of suicides')
plt.tight_layout()
```

5) 8 graphs are plotted with the help of this template.

6) plt.tight_layout is used because the graph x axis was very descriptive and to fit in the prescribed axis and space

V) Results and inferences :

After analyzing the data set for Suicides in India for the years 2001 till 2012, we derive the following conclusions:

A) *Relationship between the education status and the number of suicides.*

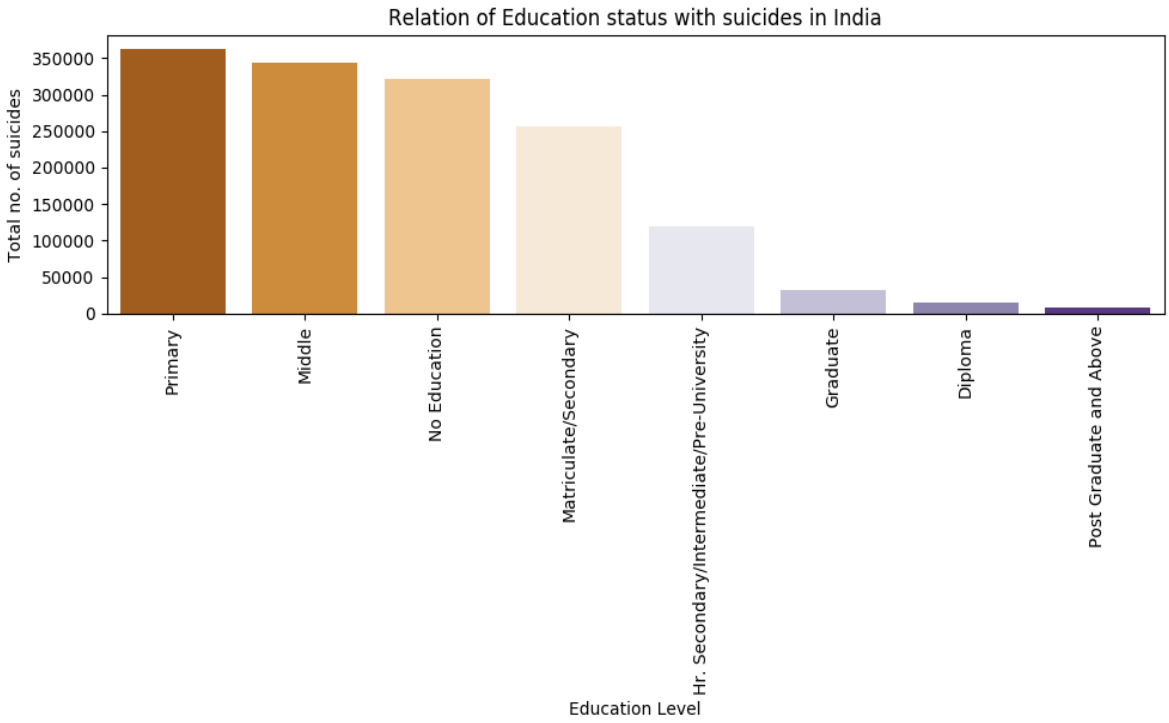


Figure 1: Relationship between Education status and suicides in India

According to Figure 1, it can be drawn that vast number of individuals who committed suicides belong to the category of low education levels or no education at all. It is possible that this might be related to pressure of academic performances among primary and middle education levels. Majority of the people who ended their own lives were below the Graduate level of education. By this we can conclude that to prevent suicides, one measure is to educate the people more, better schooling facilities should be provided to the mass. There should be incentives provided for higher education. People should be motivated to enroll themselves in educational institutes. Also, proper mentoring should be done in the schools to avoid such events of suicides.

B) *Major Causes of Suicide in India*

According to Figure 2. , It can be seen that while a huge proportion of reasons for committing suicide are unknown. However it can be seen that the five major causes which have led to suicides are:

1) Family Problems
2) Prolonged illness
3) Insanity/ Mental Disorders
4) Love affairs
5) Sudden changes in economic status

Other cause which contribute towards suicide are Poverty, Dowry disputes, Drug abuse/addictions, unemployment and many more…The action that can be taken to avoid suicides are: Make the police and concerned authorities more approachable to the people to share their problems. Provide better healthcare facilities, conduct regular public healthcare checkups and educate the people more about National suicide helpline numbers.

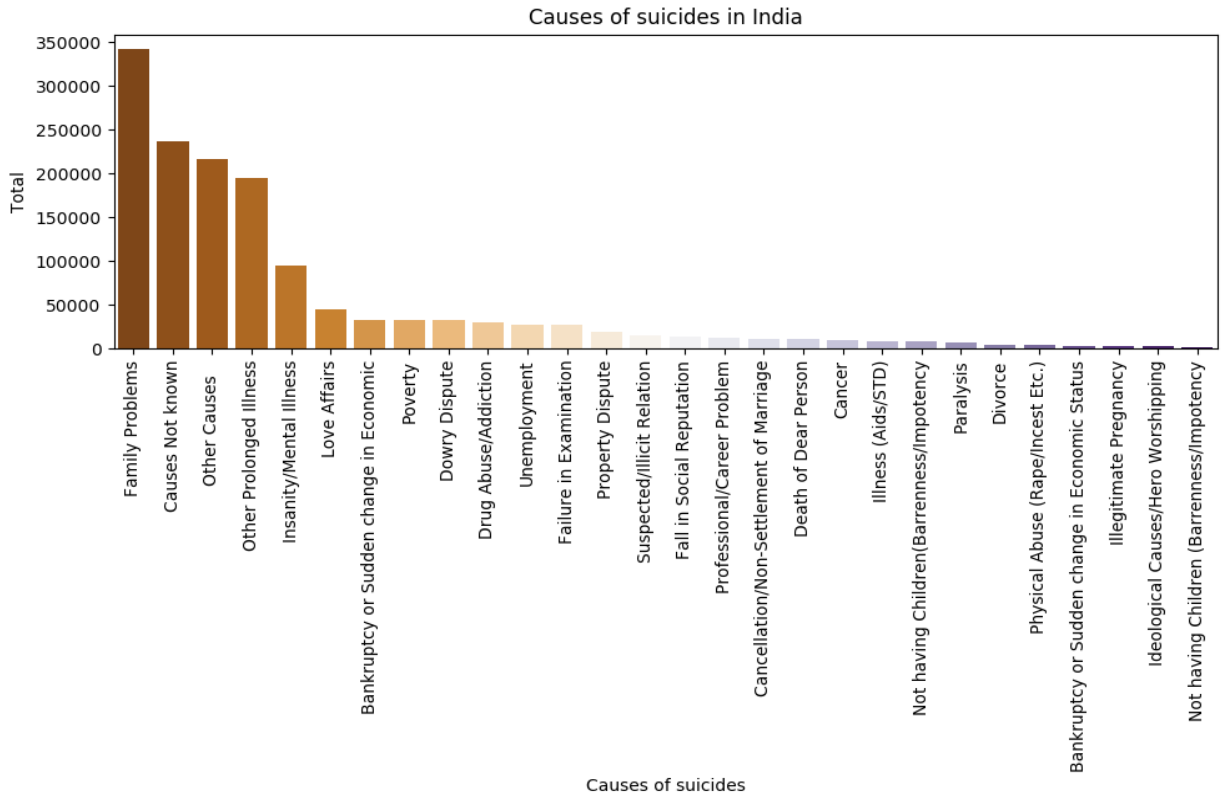


Figure 2. Causes of suicides in India

C) *Relationship between Professional status and suicides*

Depending on the professional profile, figure 3 depicts that females i.e. housewives have committed most number of suicides over the years. This might be an indication that women in professional careers are mentally much stronger and have engaged less in such events.

The second most affected are the Farmers or people involved in agricultural activities. This is because of the low incomes in the sector, ineffective government policies which have failed to shelter this professional class efficiently. The government should provide more incentives to the farmers and people engaged in the other agricultural activities.

The third most affected are in the Service (Private Sector), this might be because of the work pressure and no work satisfaction.

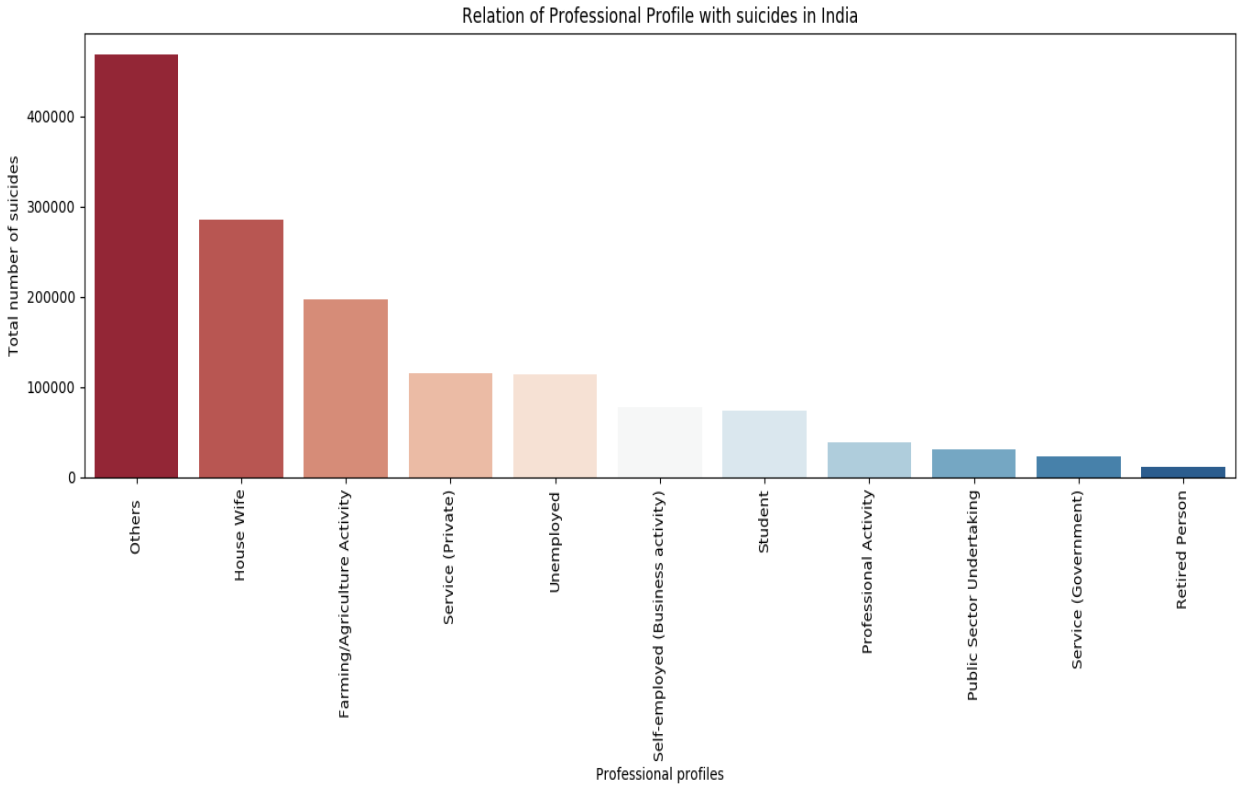


Figure 3. Relationship between Professional Profiles and suicide

D) *Most frequent means of adopting suicides in India*

The most frequent means adopted for committing suicide is by hanging, followed by consuming insecticides or by consuming other poisons. The people should be educated that suicide is never the solution to any of their problems

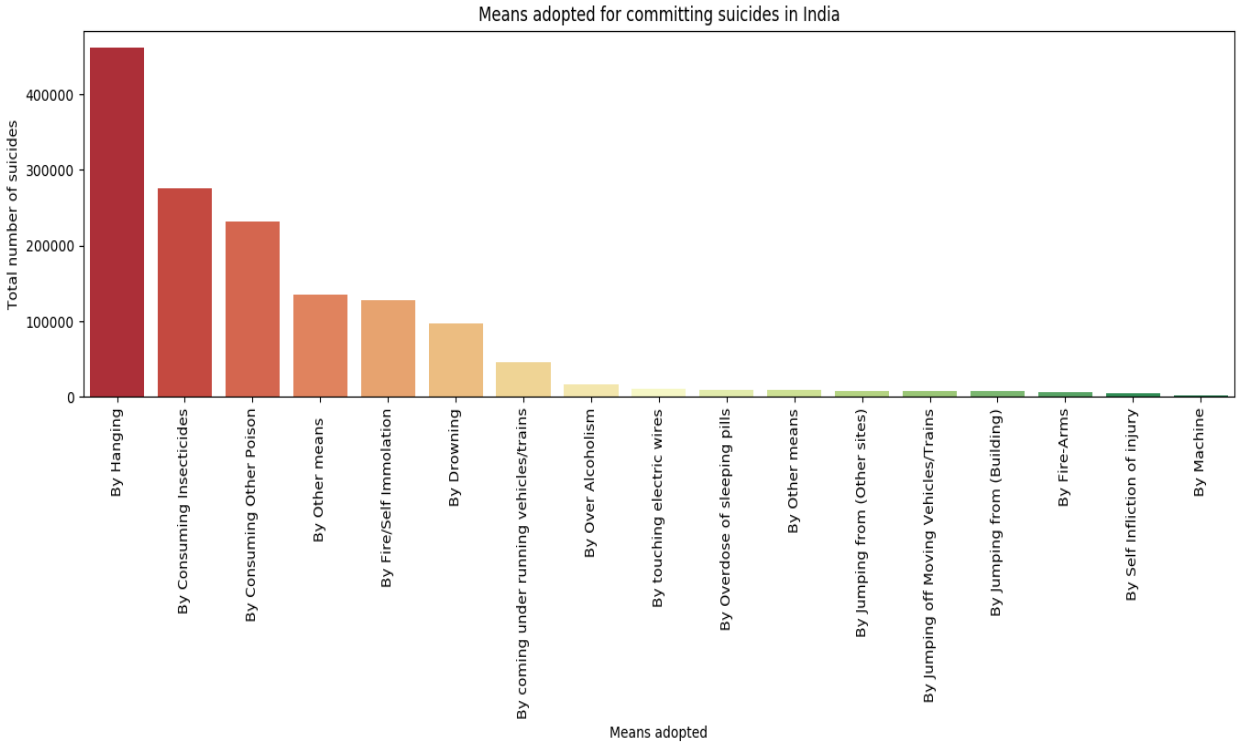


Figure 4. Means adopted to commit suicides in India

E) *Relationship of social status with suicides in India*

It can be seen from figure 5 that the social status has a very interesting pattern to observe, the married people have committed the highest number of suicides.

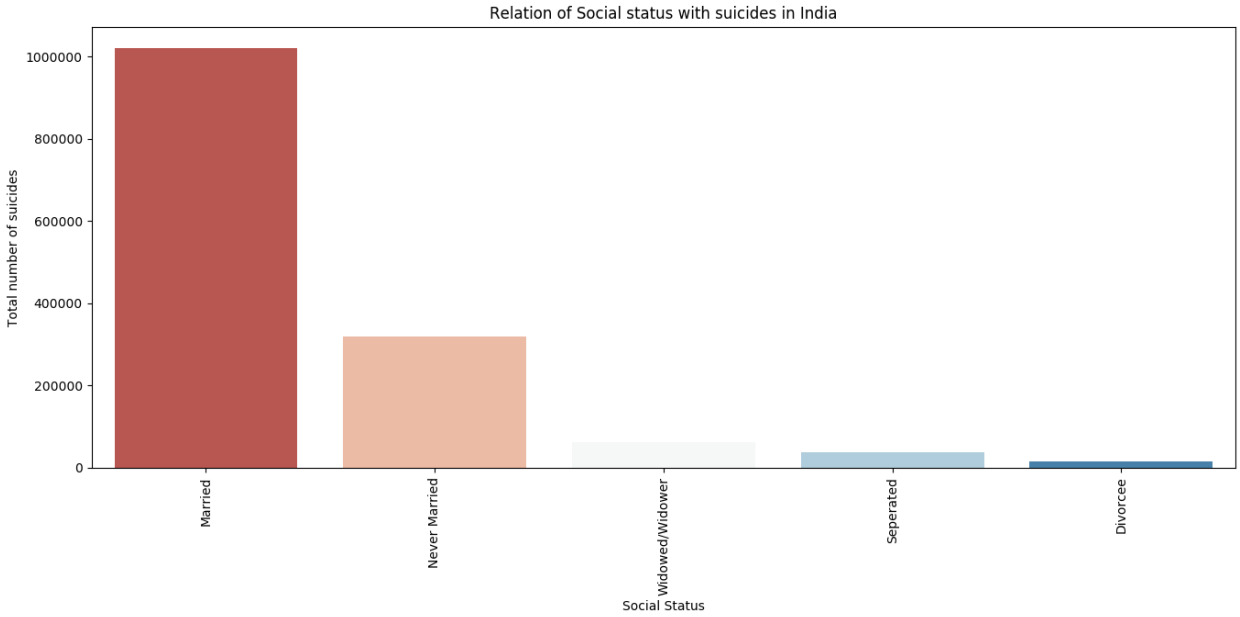


Figure 5. Relationship of social status with suicides.

F) *Gender wise bifurcation*

After analyzing figure 6, it can be seen that the ratio of Male vs Female suicides is almost 60:30. Male gender is much likely to perform suicides than females.

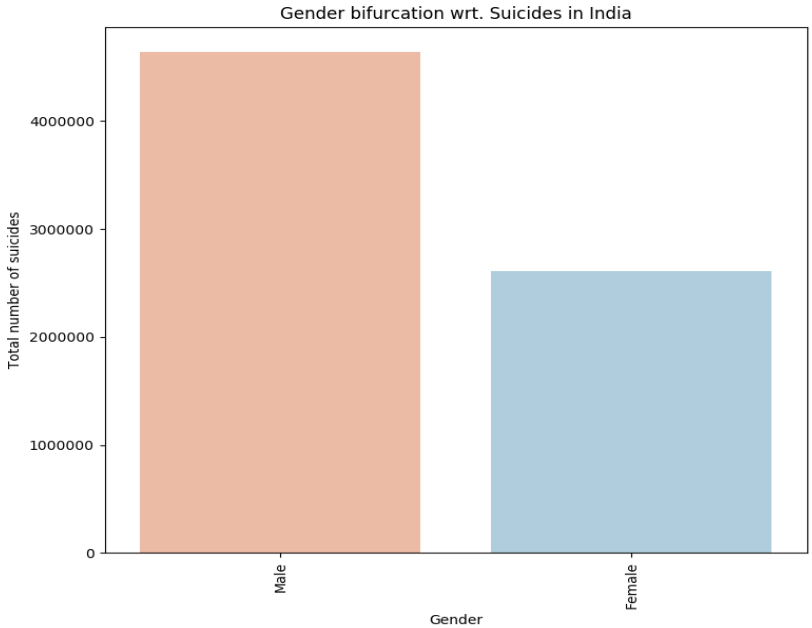


Figure 6. Gender wise bifurcation of suicides in India

G) *Age wise analysis*

The age group of 15-29 is the most vulnerable group to suicides followed by 30-44 age group. They come under the category of adolescents and the adults, workforce of the nation. This is the future of the country, they should be mentored well in avoid such lethal consequences.

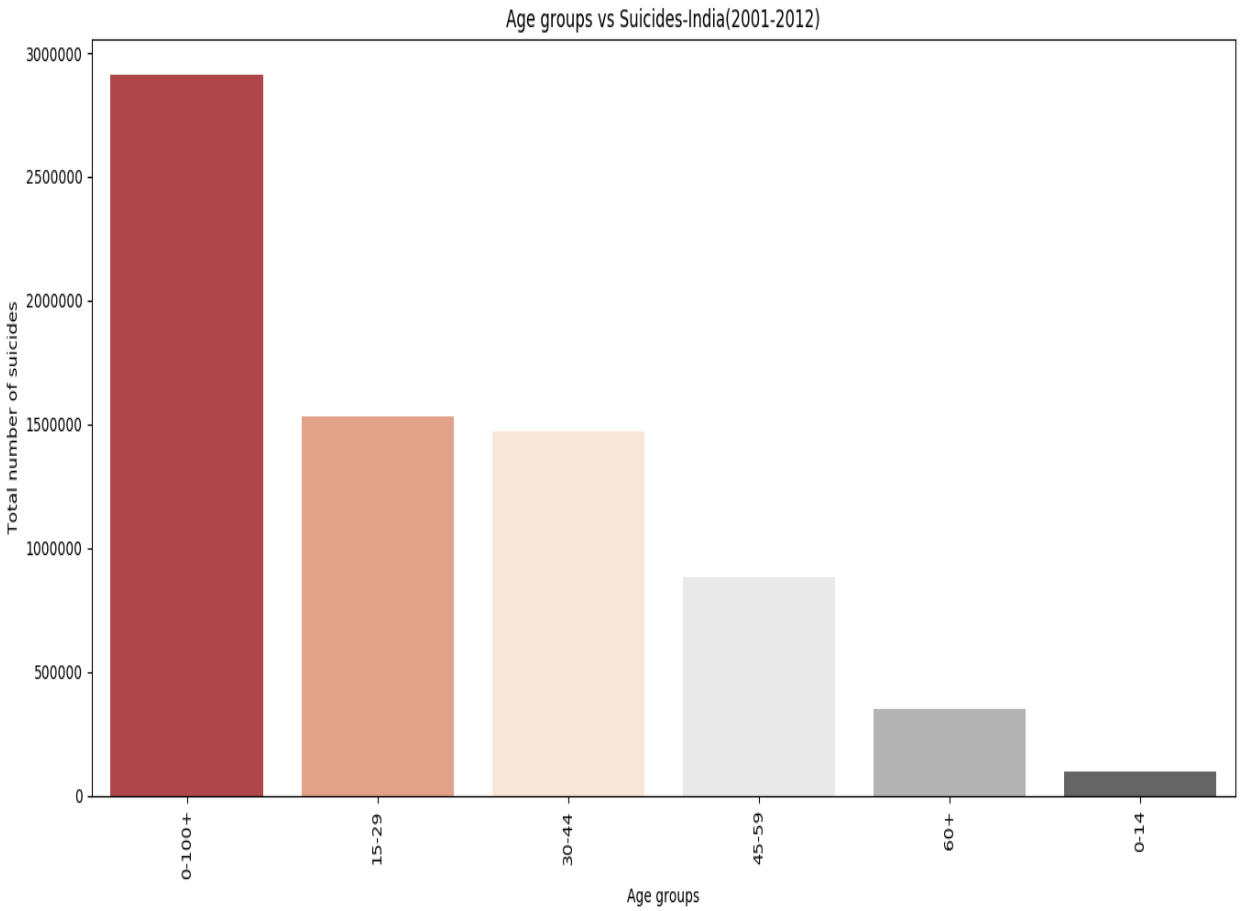


Figure 7. Age wise analysis

H) *State wise analysis*

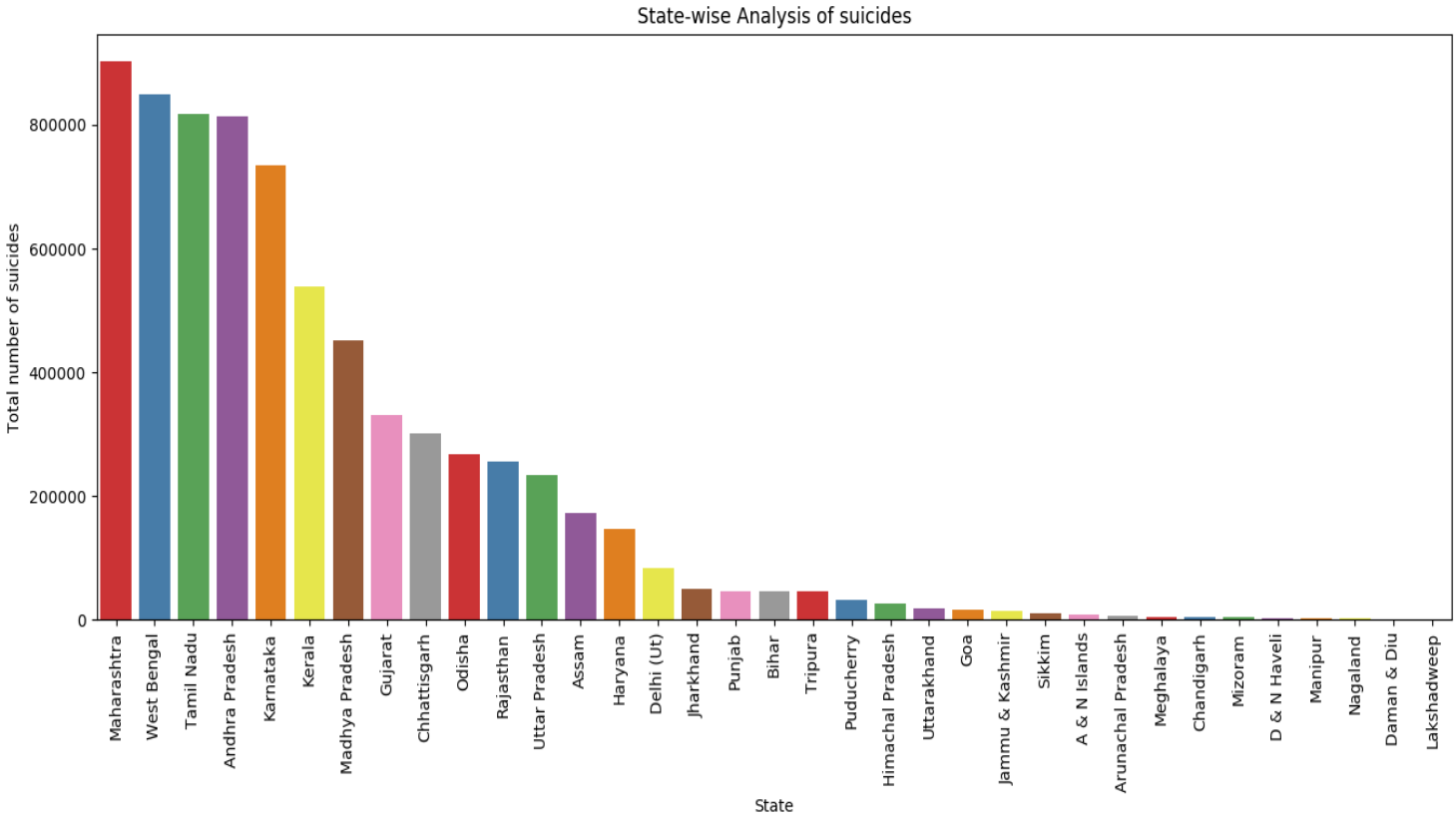


Figure 8. State wise analysis

India according to states, Maharashtra is the state which has highest number of suicides over the years. West Bengal, Tamil Nadu and Andhra Pradesh follow. Vigilance should be made stricter in the states, the Police should be more careful. There should be public awareness camps organized to fight this. This can be linked to the occupation as well, Maharashtra has a huge population of farmers.

VI) References:

[1] https://www.hindustantimes.com/health-and-fitness/every-hour-one-student-commits-suicide-in-india/story-7UFFhSs6h1HNgrNO60FZ2O.html

[2] https://bokeh.pydata.org/en/latest/docs/reference/palettes.html

[3] http://pandas.pydata.org/pandas-docs/version/0.13/visualization.html

[4] https://seaborn.pydata.org/

[5] https://www.datacamp.com/community/tutorials/seaborn-python-tutorial

[6] https://www.kaggle.com/rajanand/suicides-in-india

[7] https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.sort_values.html

[8] https://seaborn.pydata.org/generated/seaborn.barplot.html