**Project Report**

**On**

**Analysis of Suicide in India Who Why and How?**



**Submitted by**

**Shruti Mukherjee**

**Roll No - 1828203**

**School of Computer Engineering , KIIT Deemed to be University,**

**Bhubaneswar, India**

**Content**

|  |  |  |
| --- | --- | --- |
| Topics No | Topics | Page No |
| 1 | Abstract | 3 |
| 2 | Keywords | 3 |
| 3 | Introduction | 4 |
| 4 | Literature Review | 7 |
| 5 | Database creation | 9 |
| 6 | Existing System and Drawback | 10 |
| 7 | Research Methodology | 10 |
| 8 | Technology Used | 12 |
| 9 | Results and discussions | 13 |
| 10 | Future work | 46 |
| 11 | Conclusion | 46 |
| 12 | References | 47 |

**Abstract:**

Worldwide, suicide rate is considered one of the most significant issue. With each passing year, the number of suicide is getting increased phenomenally and because of this reason, this research is carried out to predict the causes of suicide in India by using the machine learning algorithms and data mining techniques in order to identify the root causes behind the suicide so that the authorities can take advantage in order to prevent the suicide cases by creating awareness and by rectifying the predicted causes of suicides. According to a research, about 800,000 people commit suicide worldwide every year. Out of these, 135,000 (17%) are residents of India, a nation with 17.5% of world population. In this research, we have analyzed the pattern of suicide cases and predict the causes of future suicides by using machine learning algorithms, the Artificial Neural Network and Support Vector Machine

A study is presented at analyzing the major factors that affect the number of suicides in different parts of India from year 2000 to 2012 and subsequently using them to predict the number of suicides in the future in different parts of India. By analyzing the data and predicting the major causes of suicides it can help government to know which part of population is most affected by this problem so that government can take the required steps to reduce them. The Indian government keeps the database of each suicides that happens in India. Along with the age-group, cause of death, state of victim. This data was made public by crime branch bureau of the data analytics purpose. Relationship will be made between the different features of suicide so that a linear relationship can be formed and then linear regression will be used to develop a model for the prediction of number of suicides in rear future. Through this study the Indian government will come to know that which part of population is most affected by suicides so that government can work on preventive measures for different parts of the country.

Keywords:

Machine learning algorithm, Prediction, Data Analysis.

**Introduction:**

Suicides is one of the major problem that Indian government is facing. About 800000 people commit suicide worldwide every year, of these 17% are residents of India. The male to female suicide ratio has been about 2:1 in India. On an average a total number of suicides in India per day is 300. According to the Suicides reports in India and National Crime Records Bureau, the total number of suicides in India as per 2014 statistics is 1,09,456. Suicide is the act of deliberately killing oneself or, more specifically, an act deliberately initiated and performed by the person concerned in the full knowledge, or expectation, of its fatal outcome. The Indian Government classifies a death as suicide if it meets the following three criteria: it is an unnatural death, the intent to die originated within the person, there is a reason for the person to end his or her life. In most cases the person writes the reason in a suicide note or it remains unspecified. Suicide prevention is a term for the collective efforts of local citizen organizations, health professionals and related professionals to reduce the incidence of suicide. Other than direct interventions to stop an impending suicide, methods also involve treating the psychological symptoms of depression,providing counselling to the person, improving the coping strategies of persons who would otherwise seriously consider suicide, reducing the prevalence of conditions believed to constitute risk factors for suicide, and giving people hope for a better life by resolving current problems. The first step in public health approach to suicide prevention is to identify those who are at the risk of suicide attempts. To identify this that which population is at greater risk this study would be useful. This can be done by making a correlation between different features and the number of people committing suicides. The relationship can be used to develop a linear relationship so that number of suicides due to a particular cause in particular age group within a particular state can be predicted.

What are the main causes of increasing suicide deaths and what are its statistics in comparison with other causes?

1. Hypothesis: To analyze the suicidal trend and explanatory association and relationship between suicide rate and economic changes.
2. Limitation of Work: The purpose of this research is to predict the causes of suicide in general irrespective of the age group or gender.

In short, this research is not predicting the causes independently for the every age group or to classify the causes according to the male and female separately.

Background:

Suicide is the 10th leading cause of death worldwide. More than one million people commit suicide every year, representing an annual global suicide mortality rate of 16 per 100,000 (Nock et al. ([2012](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR17" \o "Nock MK, Borges G, Ono Y (2012) Suicide: global perspectives from the WHO world mental health surveys. Cambridge University Press, Cambridge, UK))). World Health Organization (WHO) reports that suicide attempts are up to 20 times more frequent than completed suicides. According to recent statistics, among more than a million suicidal deaths worldwide, 20% are Indians while India is 17% of the world population Singh and Singh ([2003](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR19" \o "Singh AR, Singh SA. (2003) Preface, towards a suicide free society: identify suicide prevention as public health policy. Mens Sana Monographs, II (2))). As per NCRB report, the total number of suicides reported in [2014](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR1" \o "Accidental deaths and suicides in India 2013. A 2013 report by NCRB, Ministry of Home Affairs, The Government of India, New Delhi.                     http://ncrb.nic.in/StatPublications/ADSI/ADSI2013/ADSI-2013.pdf                                                              ) are 131,666 out of which 89,129 are males, 42,521 are females and 16 are transgender. According to WHO reports, India ranks 43rd in descending order of rates of suicide with a rate of 10.6 per 100,000 in 2009 (Radhakrishnan and Andrade ([2012](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR18" \o "Radhakrishnan R, Andrade C (2012) Suicide: an Indian perspective. Indian J Psychiatry 54(4):304–319))).

Suicide is a major public and mental health problem which demands urgent action. Suicide is the act of intentionally terminating one’s own life (Nock et al. ([2008b](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR16" \o "Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S (2008b) Suicide and suicidal behavior. Epidemiol Rev 30:154))). It is often carried out as a result of despair, the cause of which is frequently attributed to a mental disorder such as depression, borderline personality disorder, alcoholism or drug abuse, stress factors such as financial difficulties or troubles with interpersonal relationships. A suicide attempt possesses self-initiated, potentially injurious behavior, the presence of intent to die and non-fatal outcome (Levi et al. ([2008](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR12" \o "Levi Y, Horesh N, Fischel T, Treves I, Or E, Apter A (2008) Mental pain and its communication in medically serious suicide attempts: an \“impossible situation. J Affect Disord 111:244–250))). The costs of suicide are not only loss of life, but the mental, physical and emotional stress imposed on family members and friends. Other costs are for the public resources, as people who attempt suicide often require help from health care and psychiatric institutes. Suicide is a final act of behavior that is probably the result of interactions of several different factors. Predictors of suicidal behavior and risk factors include a history of previous suicide attempts, particular demographic variables, clinical symptoms and issues related to medical and social support (Hawton and Heeringen ([2009](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR6" \o "Hawton K, Heeringen KV (2009) Suicide. Lancet 373(9672):1372–1381))).

An estimated 804,000 suicides occurred worldwide in 2012, representing an annual global age-standardized suicide rate of 11.4 per 100,000 population (15.0 for males and 8.0 for females). With regard to age, the suicide rate is highest in persons aged 70 or over for both men and women in almost all regions of the world. In some countries, the suicide rate is highest in the young age. According to WHO report, suicide is the second leading cause of death in 15–29 year-olds. The countries of the Eastern Europe and East Asia have the highest suicide rate in the world. While the region with the lowest suicide rate is Latin America. Asian countries account for approximately 60% of the world’s suicides (Chen et al. ([2012](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR4" \o "Chen Y, Wu KC, Yousuf W, Yip P (2012) Suicide in Asia: opportunities and challenges. Epidemiol Rev 34(1):129–144))). Compared with Western countries, Asian countries have a higher average suicide rate, lower male-to-female suicide gender ratio, and higher elderly-to-general-population suicide ratios. Vijayakumar et al. (2005[a](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR22" \o "Vijayakumar L, John S, Pirkis J, Whiteford H (2005a) Suicide in developing countries (2) risk factors. Crisis 26(3):112–119),[b](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR23" \o "Vijayakumar L, Nagaraj K, Pirkis J, Whiteford H (2005b) Suicide in developing countries (1): frequency, distribution, and association with socioeconomic indicators. Crisis 26(3):104–111),[c](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR24" \o "Vijayakumar L, Pirkis J, Whiteford H (2005c) Suicide in developing countries (3) prevention efforts. Crisis 26(3):120–124)) studied suicides in developing countries. Maher et al. ([2011](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR13" \o "Maher MT, Aly GEAA, Abla AEA, Mohamed AM, Mamdouh KZ (2011) Suicide mortality in Cairo city, Egypt: a retrospective study. Egypt J Forensic Sci 1:30–34)) studied suicide mortality in Cairo city during 1998 to 2002.

According to WHO, an estimate of number of suicides for the year 2020 is approximately 1.53 million and ten to twenty times more people are estimated to attempt suicide worldwide. These figures do not include the suicide attempts, which are up to 20 times more frequent than completed suicide (Kumar et al. ([2013](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR10" \o "Kumar S, Verma AK, Bhattacharya S, Rathore S (2013) Trends in rates and methods of suicides in India. Egypt J Forensic Sci 3:75–80))). The estimates for the year 2020 represent on an average one death every 20 sec and one attempt every one to two seconds (Gvion and Apter ([2012](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR5" \o "Gvion Y, Apter A. (2012) Suicide and Suicidal behavior. Public Health Reviews, 34(2):1–20))). In most of the countries, suicides are under-reported. Even in some countries, suicides are treated as illegal act and it is very likely that it is unreported. In countries with good vital registration data, suicide may often be misclassified as an accident or other cause of death. Registering a suicide is a complicated procedure involving several different authorities, often including law enforcement. In countries without reliable registration of deaths, suicides simply dies uncounted.

We observe that, gender difference plays a significant role among all age groups in India as well as across the world. According to Nock et al. ([2008a](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR15" \o "Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A (2008a) Cross-national prevalence and risk factors for suicidal ideation, plans, and attempts. Br J Psychiatry 192:98–105)), suicide is more prevalent among men, whereas nonfatal suicidal behaviours are more prevalent among women and persons who are young, unmarried, or have a psychiatric disorder. Tousignant et al. ([1998](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR21" \o "Tousignant M, Seshadri S, Raj A (1998) Gender and suicide in India: a Multiperspective approach. Suicide Life Threat Behav 28(1):50–61)) reported that the gap between male and female suicide rates in India is relatively small. But since 2009, this gap has shown continuous increase. The overall male-female ratio of suicide victims in India for the year 2014 was 68:32 while it was 59:41 in 1998. Steen and Mayer ([2004](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR20" \o "Steen DM, Mayer P (2004) Modernization and the male-female suicide ratio in India 1967–1997: divergence or convergence? Suicide Life Threat Behav 34(2):147–159)) studied the effect of modernization on male-female suicide ratio in India during 1967–1997.

Suicide is a leading cause of death among teenagers and young people under 35 years of age across the world. Even in India, 66.28% (87, 252 out of *n* = 1, 31, 650 male and female suicides) of the suicide victims are between the age group 18–45 years according to NCRB report for [2014](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR2" \o "Accidental deaths and suicides in India 2014. A 2014 report by NCRB, Ministry of Home Affairs, The Government of India, New Delhi.                     http://ncrb.nic.in/StatPublications/ADSI/ADSI2014/adsi-2014%20full%20report.pdf                                                              ). Specifically, in India, the suicide victims’ boy-girl ratio (below 18 years of age) is 51:49. Mayer and Ziaian ([2002](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR14" \o "Mayer P, Ziaian T (2002) Suicide, gender, and Age variations in India. Crisis 23(3):98–103)) studied gender and age variations in suicides in India. Lasrado et al. ([2016](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR11" \o "Lasrado RA, Chantler K, Jasani R, Youn A (2016) Structuring roles and gender identities within families explaining suicidal behavior in south India. Crisis 37(3):205–211)) studied suicidal behavior in South India whereas Issa et al. ([2016](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR8" \o "Issa SY, Dossary ME, Salam MA, Madani OA, AlMazroua MK, Alsowayigh K, Hamd MA, AboZayed AM, Kharoshah M (2016) Suicidal deaths in depth – eastern Provience- Saudi Arabia. Egypt J Forensic Sci 6:240–247)) studied suicidal deaths in Saudi Arabia. Hobson and Leech ([2014](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR7" \o "Hobson MF, Leech NL (2014) The impact of exposure to peer suicidal self-directed violence on youth suicidal behavior: a critical review of the literature. Suicide Life Threat Behav 44(1):58–77)) studied the youths’ suicidal behavior and noted that there is a significant relationship between media coverage and youth suicide.

The consumption of insecticides (poisoning) (Argo et al. ([2010](https://ejfs.springeropen.com/articles/10.1186/s41935-017-0007-9" \l "ref-CR3" \o "Argo A, Bongiorno D, Bonifacio A, Pernice V, Liotta R, Indelicato S, Zerbo S, Fleres P, Ceraulo L, Procaccianti P (2010) A fatal case of a paint thinner ingestion: comparison between toxicological and histological findings. Am J Forensic Med Pathol 31(2):186–191))), hanging and firearms are the most common means of suicide globally, but many other methods are used with the choice of method, often varying according to population group such as age-group, gender, profession, social status, educational status, etc. In India hanging, poisoning, firearm/self-immolation, and drowning are the prominent means of suicides. During 2014, almost 51.12% (67,303 out of *n*) of the total male suicides are committed by hanging, poisoning and drowning while near about 24.56% (32, 333  out of *n*) of the total female suicide are committed by hanging, poisoning and fire/self-immolation. Overall, 67.83% (89, 295  out of *n*) of the total suicides are committed by hanging and poisoning.

There is no single reason why someone may try to take its own life, but certain factors can increase the risk, such as illness, family problems, financial loss, harmful use of alcohol, act cumulatively to increase a person’s vulnerability to suicidal behavior, etc. According to NCRB report 2014, ‘Family Problems (other than marriage related problems)’ and ‘Illness’ have together accounted 39.76% (52, 341  out of *n*) of the total suicides.

Knowledge about suicidal behavior has increased greatly in recent decades. Research at different levels, has shown the importance of the interplay between biological, psychological, social, environmental and cultural factors in determining suicidal behaviors. At the same time, literature has helped to identify many risks and protective factors for suicide, both in the general population and in vulnerable groups.

In this paper, we study the recent trends in the number of suicides in India and briefly review various risk factors for suicide. The main objective of this study is to explore the association between various attributes such as gender, age of suicide victims, the social, economic, educational and professional status of suicide victims, means of suicides, risk factors for suicides, etc. We study the association between various attributes through the Correspondence Analysis (CA). The detail of the methodology to study the suicidal behavior of Indians is discussed in the next section.

**Literature Review:**

Worldwide, Suicide rate is one of the most important problems. The total number of individuals who committed suicide is increasing with each passing year. It is projected that because of the various causes, around eight hundred thousand individuals expires while attempting suicide.

Suicide is considered as a disease and according to the report of WHO (World Health Organization), 17 percent residents of the global suicide sufferers belongs to India.

According to the CDC-2015, in the last few years, researchers have focused on recognizing, understanding, curing and impediment of suicidal patterns and behavior. Regardless of all the efforts and studies, the rate of suicide is not decreasing.

Majority of the people who attempted suicide does not plan or strategize to attempt a suicide.

For that reason, it is very important to make better prediction about the individuals who are expected to take action on their thoughts of attempting suicide.

A researcher projected an integrated framework of machine learning for the prediction of suicide risks. Basically, the proposed structure has three components.

1. Temporal characteristic extraction
2. Risk Regulation
3. An ensemble loop for feature selection and ordinal categorization.

Globally, suicide is measured as one the most important issue which leads to the mental health as it is one of the major reason of death. Hence, it is one of the main challenges for the detection and the prevention of suicidal consideration.

For the estimation of suicide rates, the likelihood or probability could be forecasted surrounded by a specified forthcoming era of sentinel measures which are as follows :

1. Low-risk proceedings mean suicide risks are not detected.
2. Moderate-risk measures are self-damage or injuries that does not direct towards the significant consequences.
3. High-risk proceedings are those with major consequences such as deaths.

A research has been published by a researcher which intended to find out the major features that have an effect on the amount of suicide in some particular districts of India and later utilizes those features to estimate the quantity of suicides to be held in future. This suicide estimation can assist or help the authorities in forming leading decisions related to the regions which are affected by high number of suicide. The characteristics in the research represent the fraction of the populace which are distress mainly as a result of suicides. The government of India keeps a record by maintaining a database of the registered cases of suicides for each and every state of India. Database records are made accessible for the public with the intention of analytics of the information present in registered data

Besides, with the amount of suicide cases for every region, the demographical information of that particular state were also considered while developing the estimation model.

There were three basic groups that were considered while developing a model and those categories are educational level, martial stage, and census information of the region.

Researcher applied a Karl Pearson’s coefficient of correlation to verify the association of the features and to identify the correlation amongst them. After identifying the strength of association a regression model was applied to for estimating the amount of suicide rate in future.

The conclusive results were significantly important as there as there were nine features which reportedly acquire a significant linear association with the amount of registered suicides. Estimation model which was developed by utilizing those nine attributes predicted a linear relationship by providing the 99% of estimation accuracy.

Another researcher recommends a technique for estimating the suicides. He proposes to utilize the data available for the registered suicides in order to estimate the suicidal behavior amongst individuals. According to him, Sentiment Investigation can play an important role as it is one of the latest experiments developed in machine learning as social networking systems present substantial amount of information and is being gathered and created by the clients/users of the social networking sites. He is in opinion of to extract benefit from the information available at the social networking sites by analyzing the mechanism of the thought procedure which is based upon the opinion, view and the sentiments provided by the user. Social networking platforms are progressively more associated or linked with multiple phenomena like harassment, depression or even suicide cases and because of this it is very important to make an effort to discover the possible sufferers as early as possible so that the prevention of such incidents like suicides would be achievable.

To summarize, author of the research particularly suggest to concentrate on the required terminological sources associated to suicide by means of developing a method for assembling a vocabulary which is correlated with the terminologies of suicide. In this study, Weka Software was utilized which is one of the data mining tools and supports the algorithms based upon the machine learning to investigate and to extract out the meaningful information from the data or the information presented by a Twitter platform.

Thus, as a result an algorithm is proposed along with the mechanism of processing the semantic investigation involving the training data set which were the collection of tweets along with data group established by the tweets on WordNet. Investigational conclusion depicts that the process established on the machine learning technique along with the sentiment investigation can obtain the information of suicidal thoughts or behavior by utilizing the data available at the twitter platform.

Additionally, this study authenticates the helpfulness and efficiency of performance in predicting the suicidal behavior in an individual.

**Database Creation:**

A dataset is a collection of data. Mostly a data set refers to the contents of a single database table, or a single statistical data matrix, where every table column represents a particular variable, and each table row corresponds to a given member of the data set. The objective of project is to find the relations between the dataset to predict the future dataset.

The data set used in this project contains yearly suicide detail of all the states/u.t of India by various parameters from 2001 to 2012. This data is real and National Crime Records Bureau (NCRB), Govt of India has shared this dataset under Govt. Open Data License - India. NCRB has also shared the historical data on their website. The data contains various fields like State, year, reason, gender, age group and total.

There are various reasons due to which a person may commit suicide. It may be due to education burden, family problem, financial problem, health status, etc. There may be certain states where particular reason for suicide is higher than other reason. We need to find relation between reason of suicide and the state in which it is committed. So that government will focus on particular reason resulting in minimizing that factor of suicide. For example, if in particular state the maximum suicides are committed by farmers due to low agriculture production or higher input and lower output in agriculture than the government can focus more on particular field thus making more schemes that would help farmers of particular state resulting in reduced number of suicides in that state.

Another feature is age group. The dataset is divided in different age groups. So that we can find relationship between age and number of suicides which helps us to know that particular age group population is mostly affected by which suicide reason. For example, if in particular age group the major reason of suicide is failure in examination than more counselling institutes should be open in that state to do counselling of all people belonging to that age group resulting in reduction of suicides

The existing considered these states: - 

1. Maharashtra 
2. West Bengal 
3. Tamil Nadu 
4. Andhra Pradesh 
5. Karnataka 
6. Kerala 
7. Madhya Pradesh 
8. Gujarat 
9. Rajasthan 
10. Uttar Pradesh 
11. Punjab 
12. Bihar

While the proposed system considers some extra places too: - 

1. Chhattisgarh 
2. Odisha 
3. Assam 
4. Haryana 
5. Delhi (Ut) 
6. Jharkhand 
7. Tripura 
8. Puducherry 
9. Himachal Pradesh 
10. Uttarakhand 
11. Goa 
12. Jammu & Kashmi
13. Sikkim 
14. A & N Islands 
15. Arunachal Pradesh 
16. Meghalaya 
17. Chandigarh 
18. Mizoram 
19. D & N Haveli 
20. Manipur 
21. Nagaland 
22. Daman & Diu 
23. Lakshadweep

**Existing system and Drawback:**

Existing System: - 

* Analysis of Suicide Victim Data for the Prediction of Number of Suicides in India (Publisher IEEE 2017 ) 
* Suicide Victim Data was analyzed for the Prediction of Number of Suicides in India. 
* Linear regression algorithm was used for prediction. 
* Data of different places was used for prediction. 
* Data set used was of made public by crime branch bureau.

Drawbacks of Existing system: - 

* The existing system has considered only 12 states and given results on basis of them. 
* Union Territories were not taken into consideration during study. 
* Age group between 15-29 was only considered for study.  No proper simulation and results were explained. 
* Only 2011 data has been used for prediction which is less efficient.

**Research Methodology:**

This section represents the research methodology which has been developed for this particular study. In order to understand the behavior and trend in suicide data, there are two logical methods which deals with our research problem in very effective way. Following are the two approaches for our research work:

1. Descriptive and Statistic Approach: This method is used to find out the pattern of suicides with respect to age group, gender, marital status, social status, education along with the professional occupation.
2. Predictive Approach: In this method, data will be used to generate model to predict future causes of suicide by utilizing the information present in the existing data.

Our research aim is to predict causes of suicide in India based on the existing dataset of India’s registered suicide cases which is obtained from NCRB.

Data of the registered suicide cases which is made available for the public by the Indian Government at NRCB website was obtained in order to perform this research.

1. Dataset: In order to investigate, it is important to identify the attributes and characteristics present in a Dataset which contains the information of total suicide in particular state along with other meaningful information which is as follows:
2. State:

This column contains the name of state in India like West Bengal, Andhra Pradesh etc. The total number of unique states which are present in dataset is 35.

1. Year:

This dataset contain information from 2001 - 2012.

1. Gender:

The value in this column is male and female.

1. Age Group:

There are different age group in dataset which are from 0-14 to 60+.

1. Total Number of suicides:

This column contain the sum of total number of suicide in particular state according to its gender, age, and state.

1. Type/Cause:

This column tell us about the reason of attempting suicide like illness, Family Problems, Bankruptcy, unemployment etc.

1. Marital Status:

This column represents the information whether the person who committed suicide was married, unmarried, was a divorcee or a widow.

1. Professional Occupation:

This particular field represents the information whether the victim was a student, employer, house wife or an unemployed person.

1. Educational Level: This field depicts the information regarding the educational background of a victim.
2. Number of records: The total number of records which are present in our dataset is approximately 109200.

Data pre-processing is one of the most important part for improving the accuracy and performance of our model. Data pre-processing is a data mining technique by which we can clean dataset for reducing redundancy and missing values because real world data or raw data is often incomplete, noisy and also contain error. The selection of incorrect data or feature may result in poor result and accuracy for that reason data pre-processing is necessary.

Pre-processing of data involves multiple step by which we can achieve consistent and complete data. The data preprocessing step are given below

1. Data Cleaning:

The first step of pre-processing a data is to clean the data. Find data which contain missing values and remove those data or put any other value like average of that column etc. the goal of data cleaning is to remove missing , noisy and inconsistent data from dataset.

1. Data Integration:

Adding or mixing of data from different databases are put together. After gathering all data conflicts with in the whole data is resolved.

1. Data Transformation:

Data is transformed by using normalization or aggregation method

**Technology Used:**

Machine Learning:-

Machine learning is a study of computer science that provides computers the ability to learn without being explicitly programmed. Machine learning is used to study algorithms that learn from and make predictions on data. Machine learning is related to computational statistics, which also focuses on prediction making. Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The learning process begins with observations or data, examples, direct experience, or instruction, in order to find patterns in data and make better decisions in the future based on the examples provided. The objective is to allow the computers learn automatically without human assistance and adjust actions accordingly.

**Results and Discussion :**

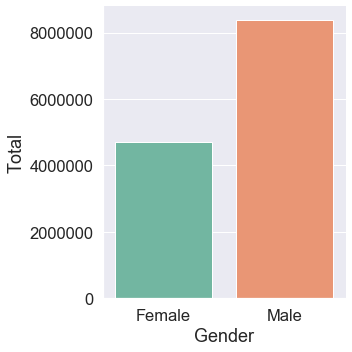
After the pre-processing of data and developing the methodology for this research paper, extracting meaningful information from that is an important part in order to understand the pattern. The intention or the objective for investigating or analyzing data is to comprehend and interpret the findings of the study correctly in order to draw conclusion from the experiment.

Cloe to 800 000 people die due to suicide every year, which is one person every 40 seconds. Many more attempt suicide. Suicide occurs throughout the lifespan and is the second leading cause of death among 15-29 year olds globally.

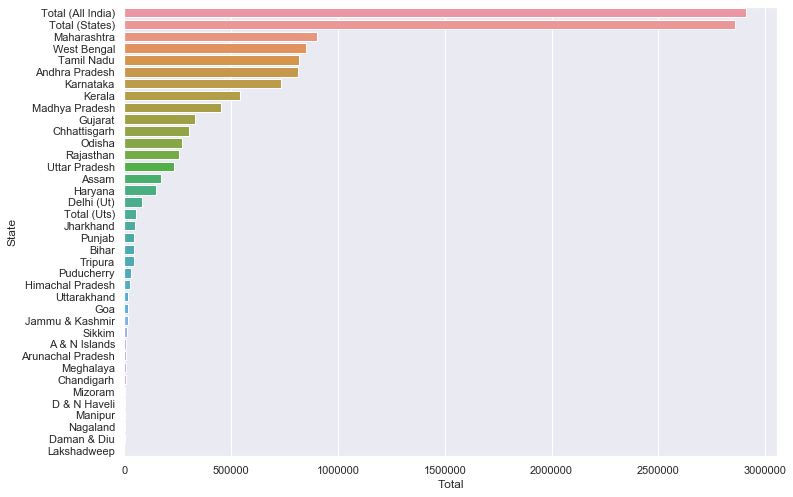
1st step:

Data analysis:

1. Gender analysis:

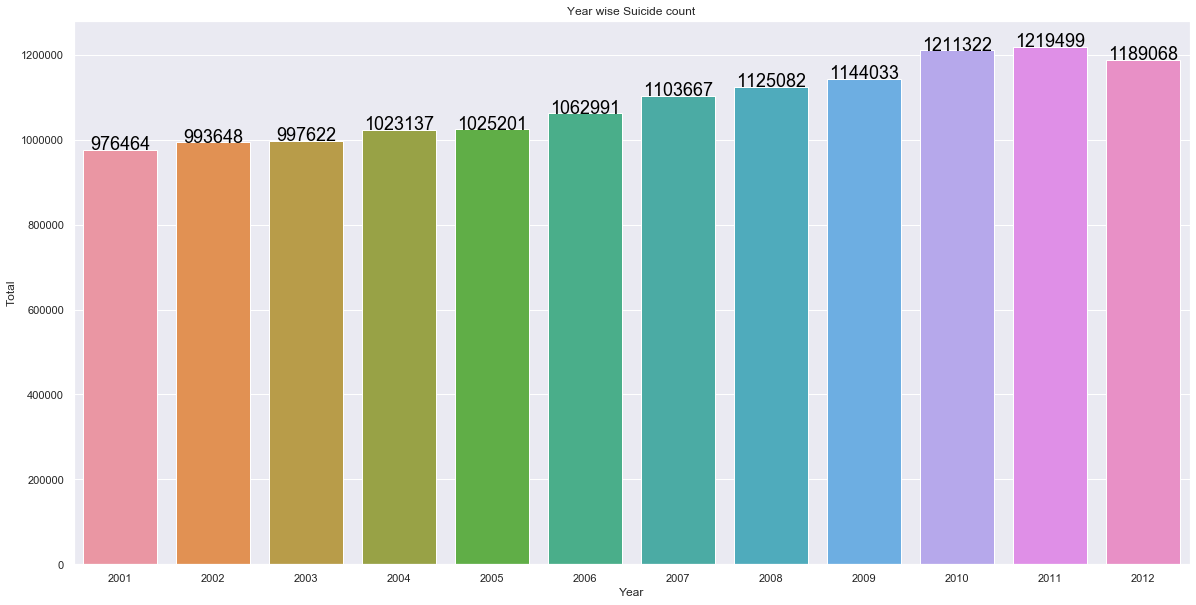


1. State wise analysis:



Here we found Highest no.of suicide cases occur in Maharashtra, west Bengal and Tamil- Nadu.

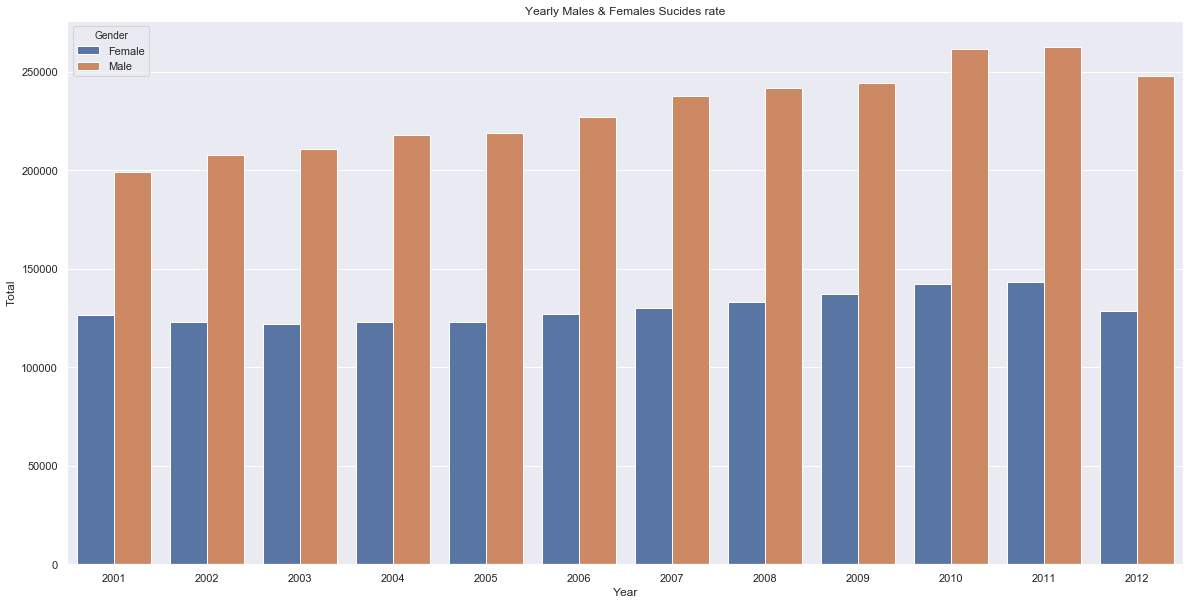
1. Year wise analysis:



From Year 2001 to 2011 suicides rate is increased by 24.76%.

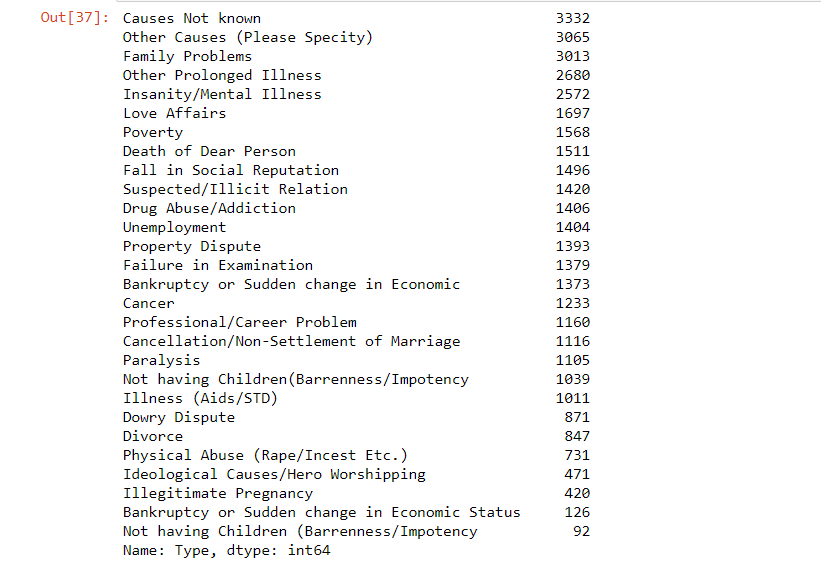
Year 2011 recorded highest suicides(405989).

Yearly Males, Females suicide rates:

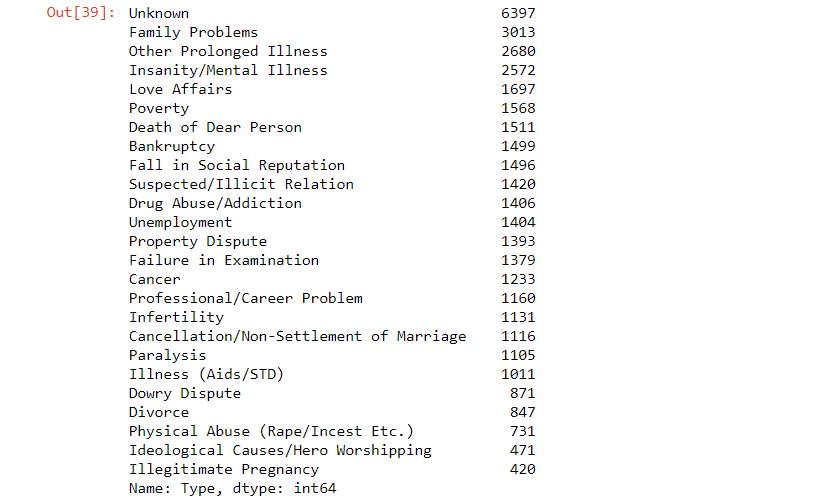


Men appear to be more vulnerable for suicide -- Reasons to be analyzed

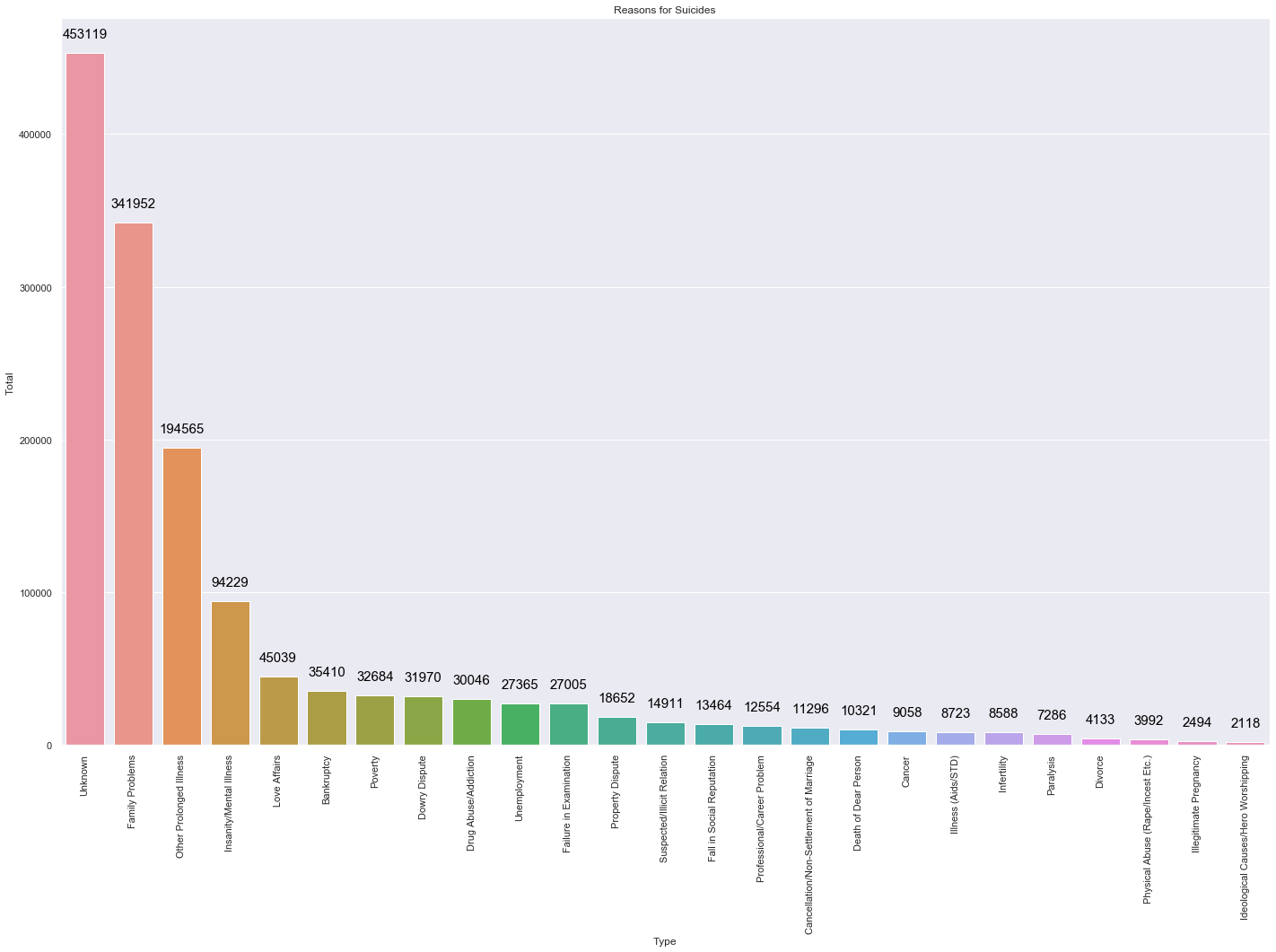
Considering the data where cause of death is specified as separate set for analysis.



Some reasons seem to be repetitive need to unify the reasons to one cause



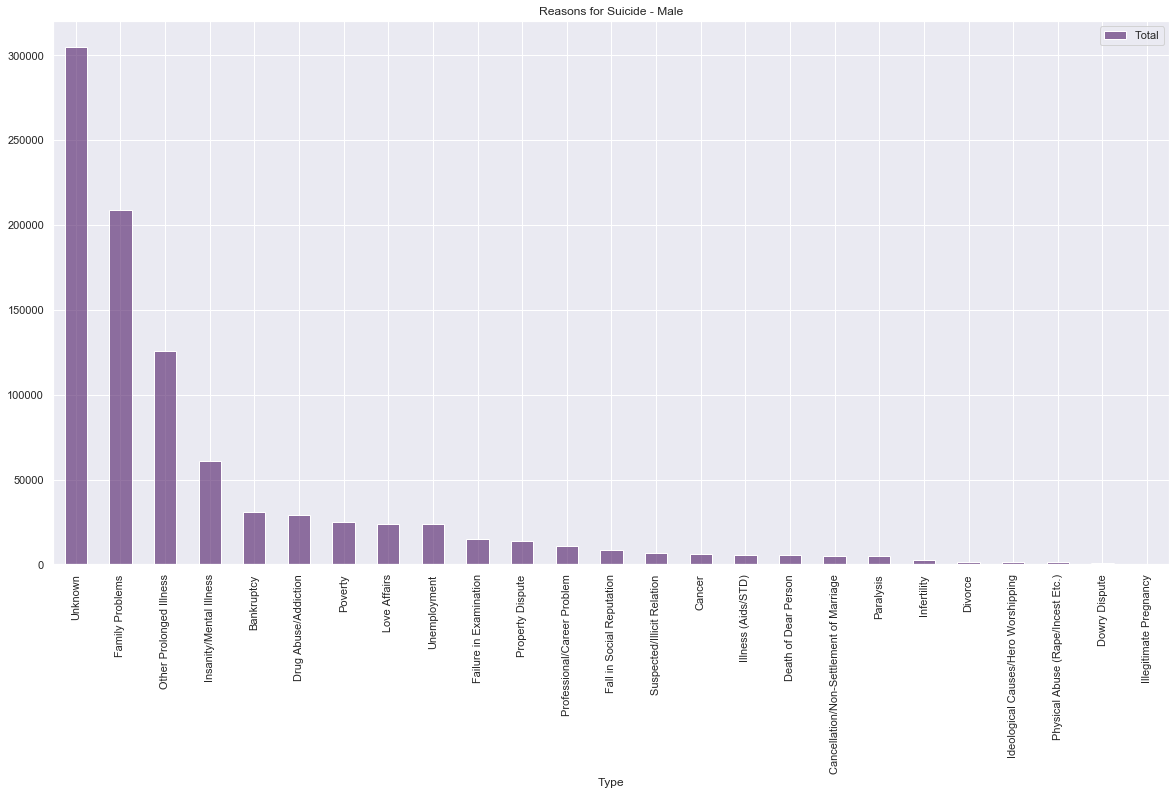
Reasons and Suicides:



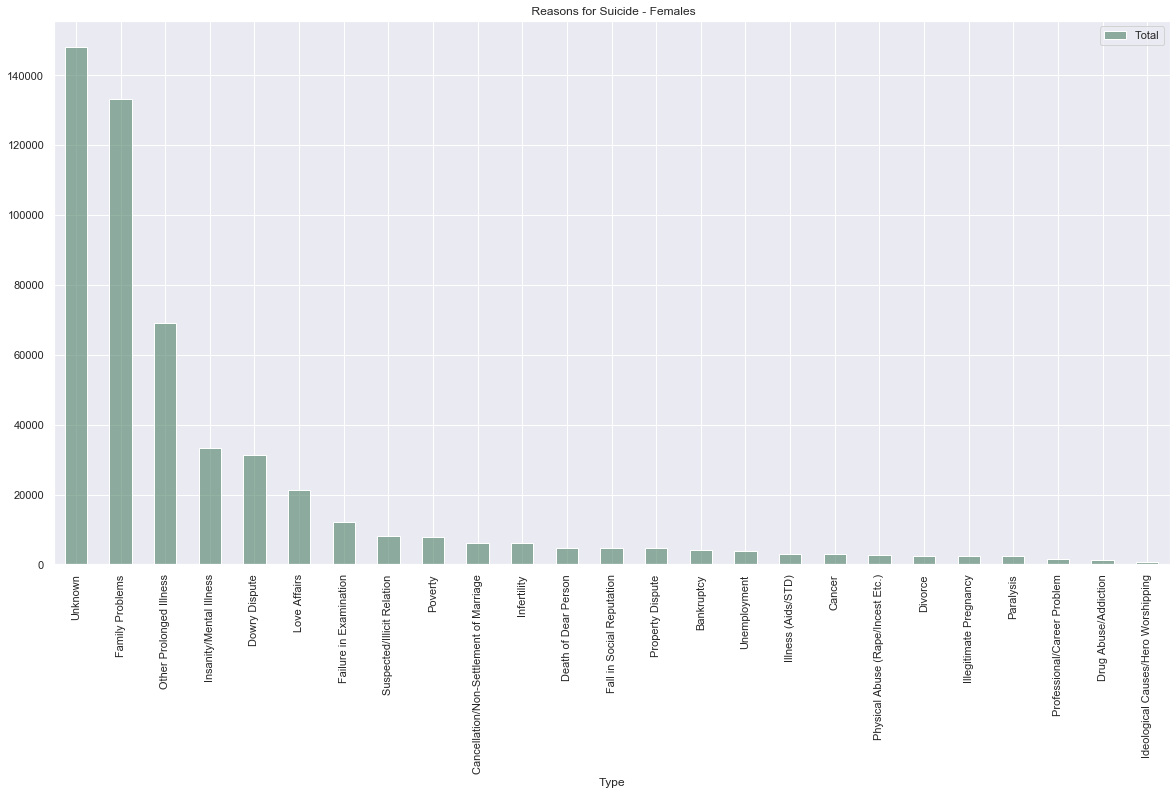
Observation: Family problems, IIIness, Love afairs are in the reasons list. --Are we proud of our family system.

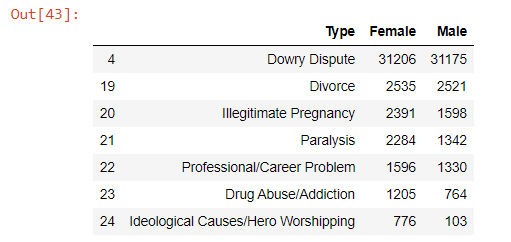
Suicide Reasons and Gender:

1. Reasons for Males suicides:

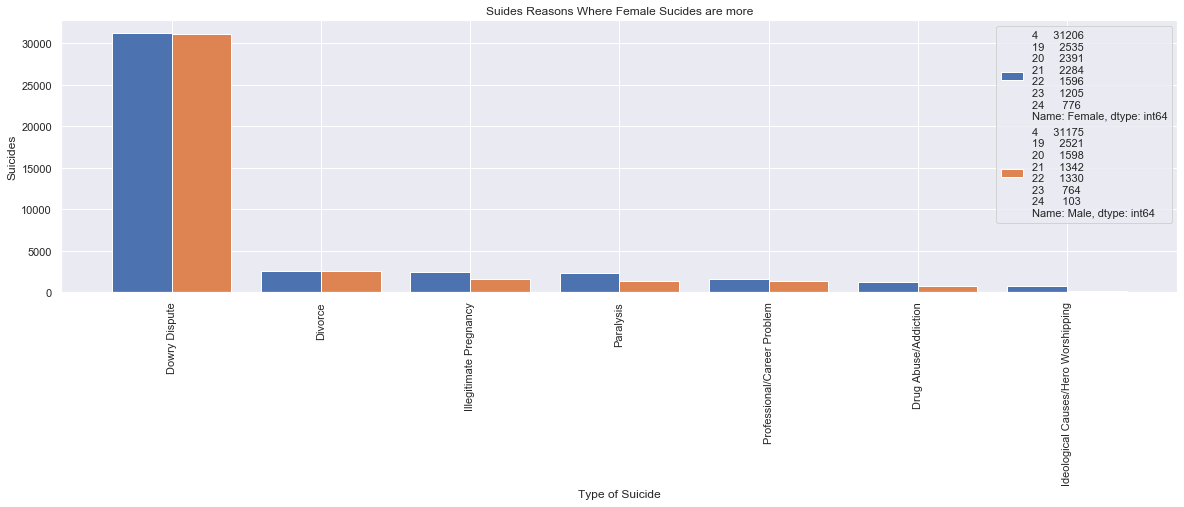


1. Reasons for Females suicides:

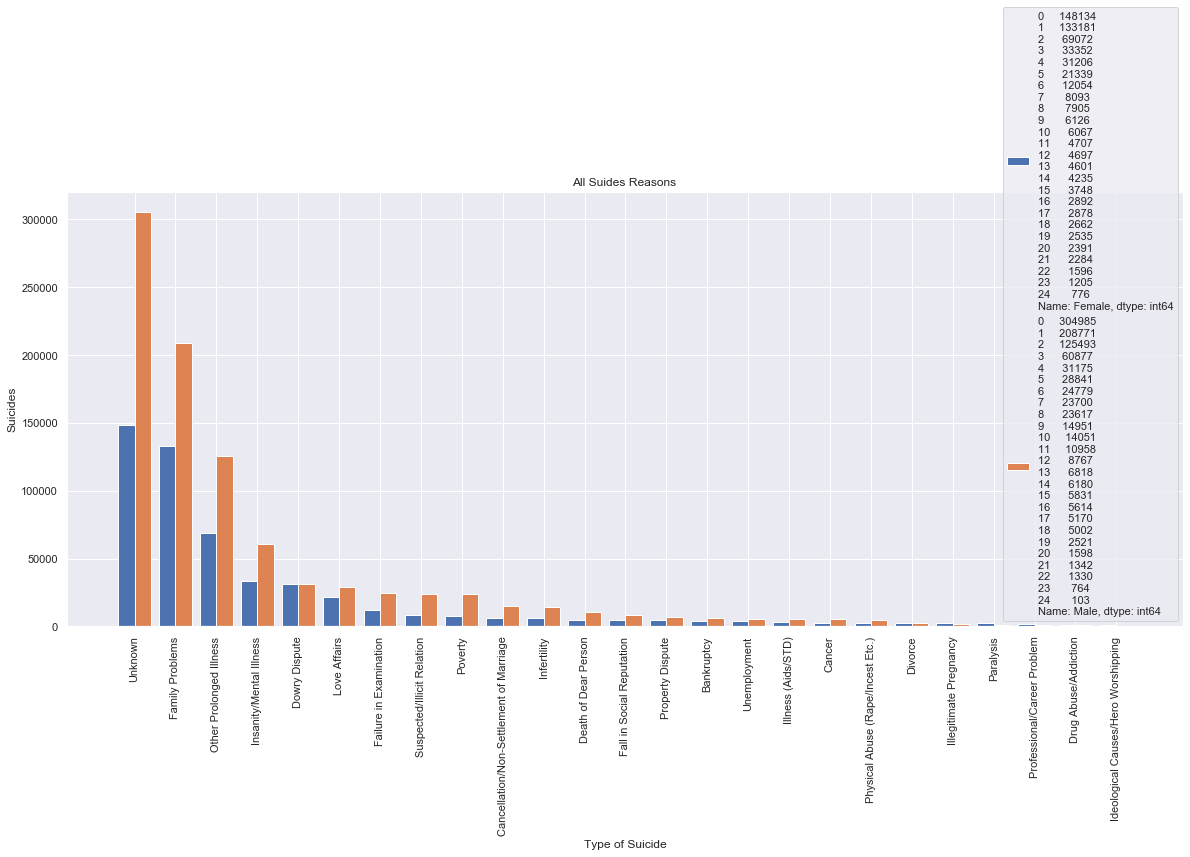




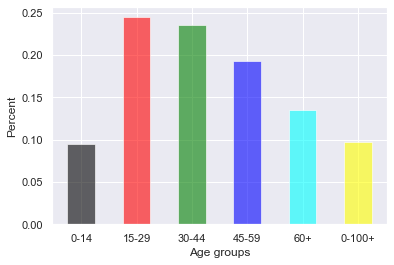
Suicides Reasons where female suicides are more:



1. *Men seem to be more vulnerable to commit suicide.*
2. *"Family problems" is the major reason for all the suicides irrespective of the gender*
3. *Dowry Dispute is second highest reason for women to commit suicide.*
4. *Dowry Dispute , Infertility , Physical abuse are the reasons where Women suicide rate is more than men.*
5. *Love affairs is the reason where suicide rate is alomost same in Men and Women.*

**

*Age Group:*

**

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.

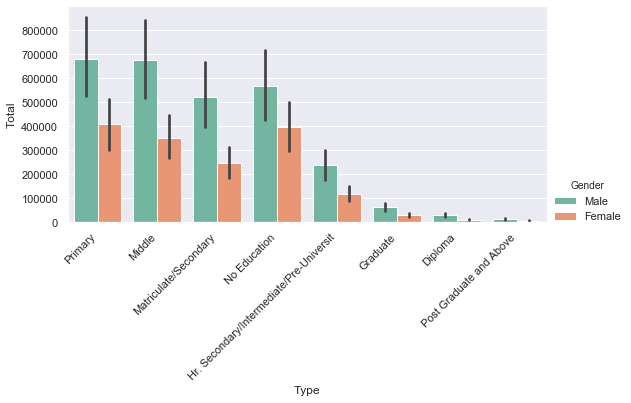
It can be due to several reasons like:

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

Data Preprocessing:



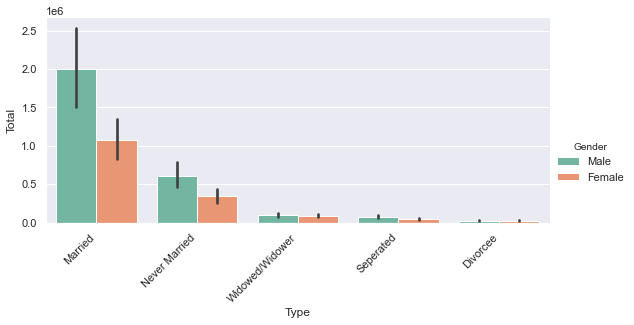
*Education Level:*

**

*It appears that people with low education tend to commit more suicide.*

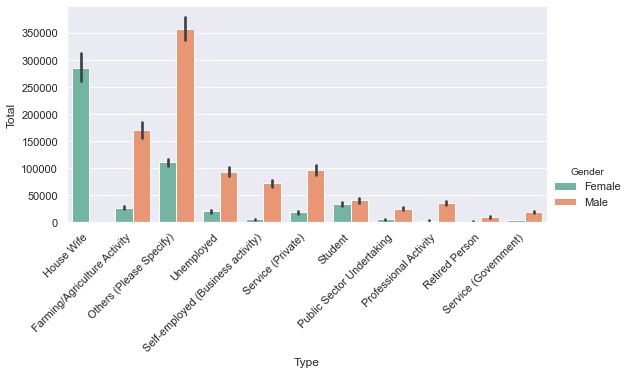
*People with Diploma and Graduate tend to commit least no. Of suicide.*

*Social Status:*

**

*It appears that marries people count for the majority of suicide cases.*

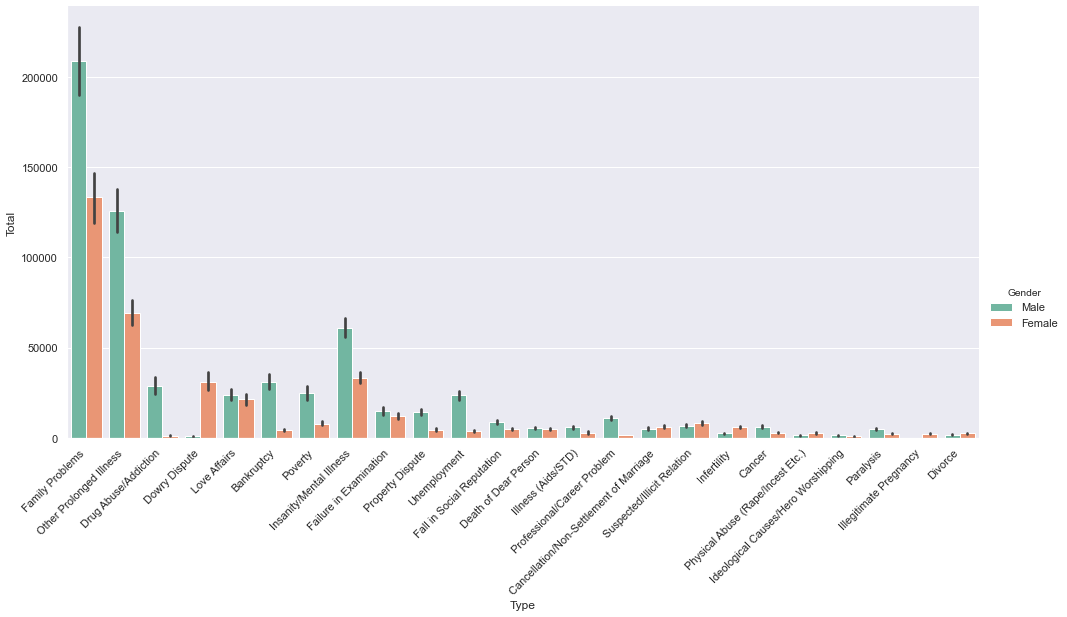
*Profession of People:*

**

Farmers and housewives tend to commit more suicide compared to others.

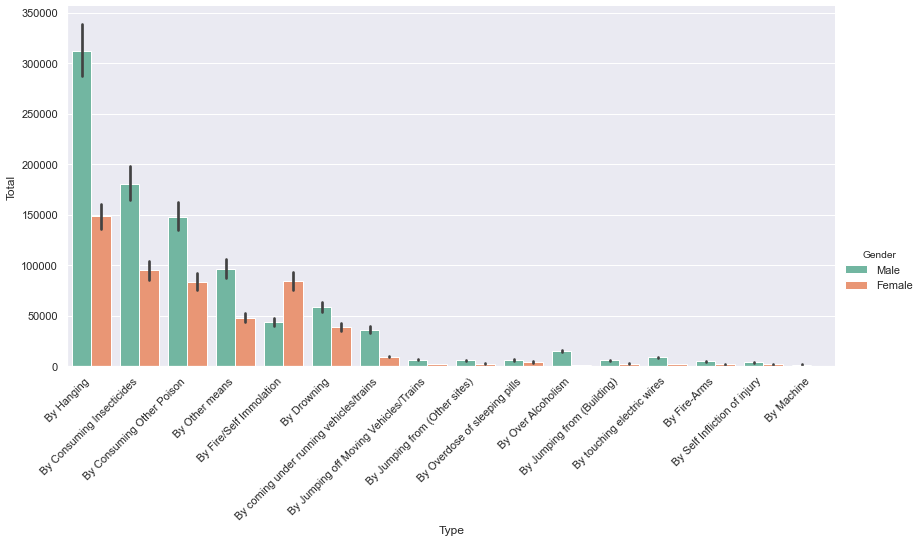
This makes sense because most of the Indian farmers have debt and their life depends on the yield of their crops, if the yield is not good then they will not be able to clear their debt and in the worst case they might commit suicide.Housewives might have issues in their marriage which this might be a reason for such a high number of cases.

Cause category:

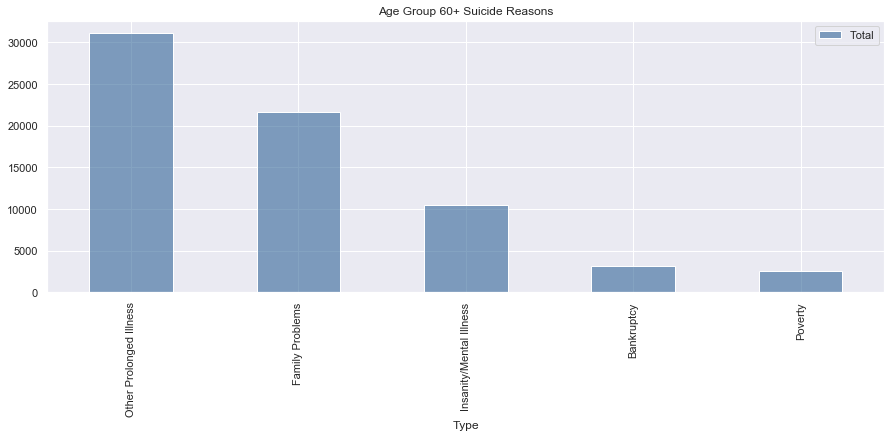


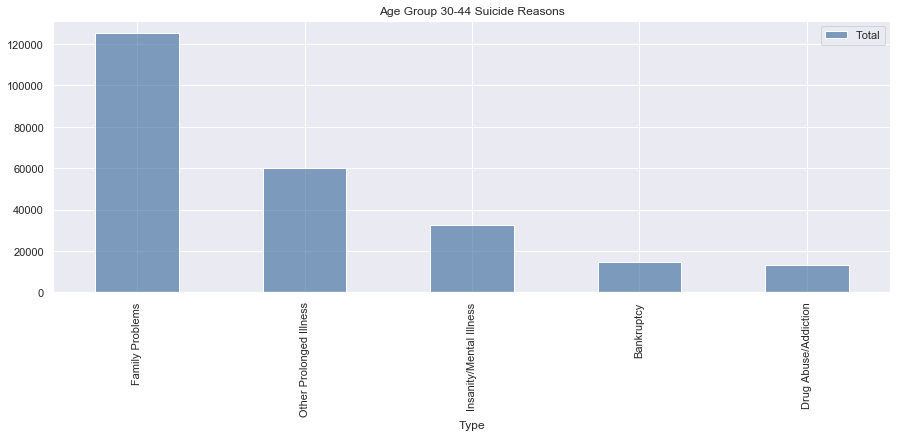
The leading cause of suicide is the Family problem, which is in line with the observation of that married groups are more likely to commit suicide, followed by Prolonged illness and Insanity/Mental Ilness.The only difference between males and females on in the order of types of suicide causes is on "Dowry Dispute", Suspected/Illicit Relation and Cancellation/None-Settlement of Marriage where female suicide has a higher rate.

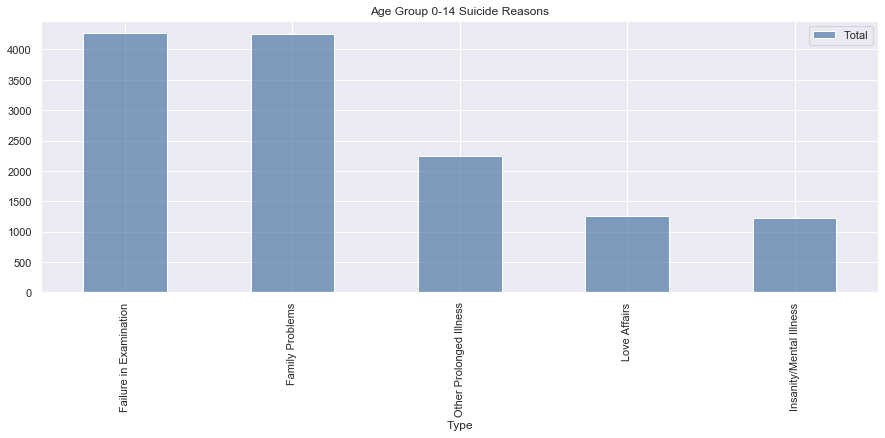
Means of commiting suicide:

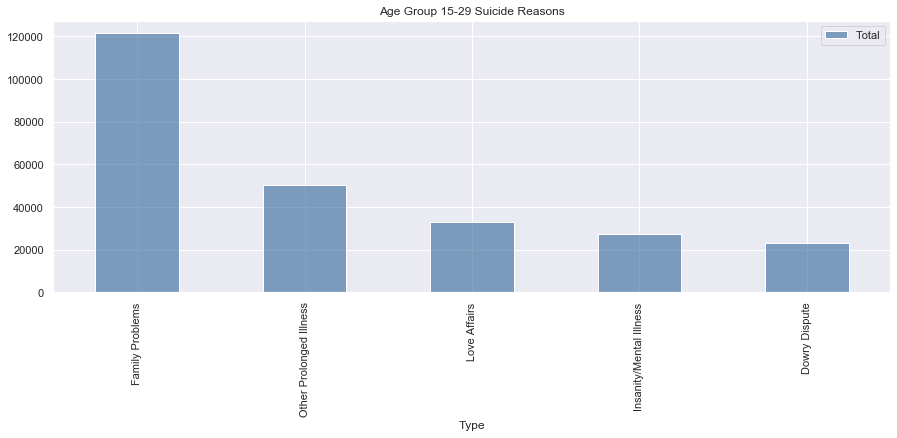


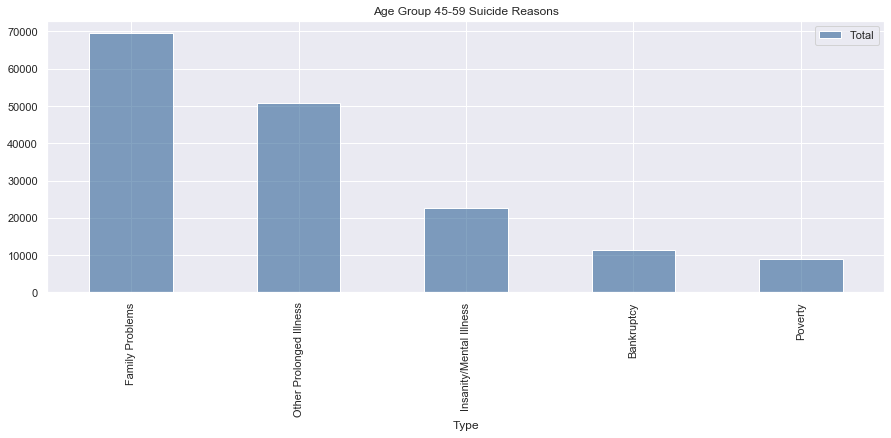
Age wise suicide cause analysis:







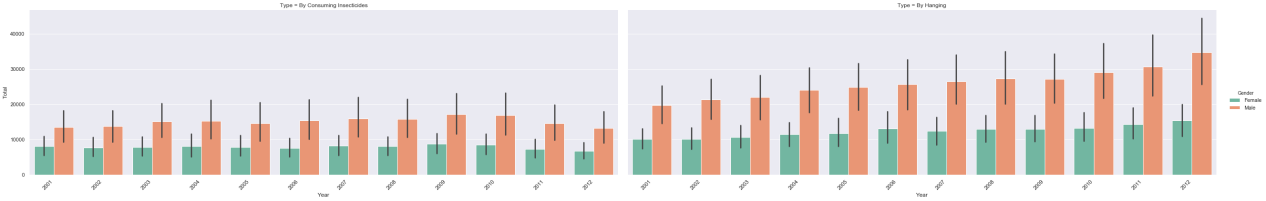
**

**

## **From top 5 Reasons of each age group the following observation were drawn:**

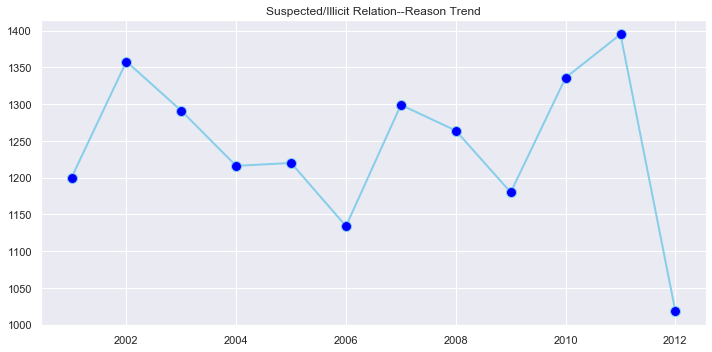
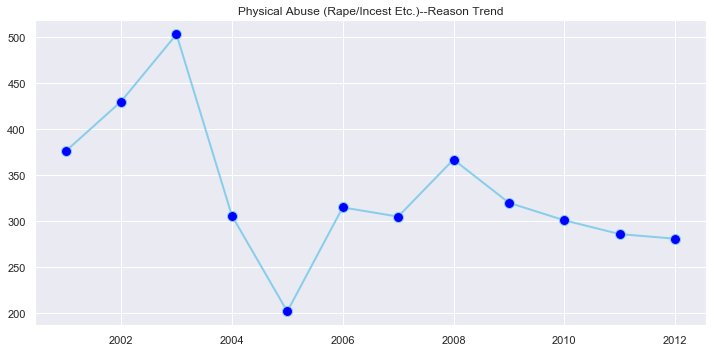
1. Age group 0-14, failure in examination was one of the top reasons
2. Age group 15-29, family problems and love affairs are one of the top reasons
3. Age group 30-44 and 45-59 have Family problems as one the top reasons, followed by reasons liked mental illness and bankruptcy.
4. Prolonged illness is the strongest reason for age group Age 60+.
5. Family Problems is the strongest reason in all the age groups.
6. In general, illness is appearing as the strongest reason irrespective of age group and gender

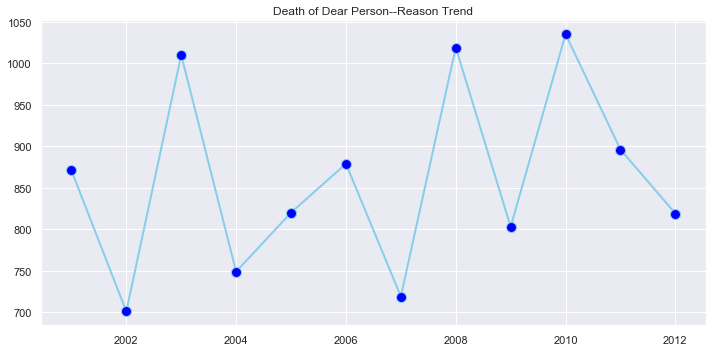
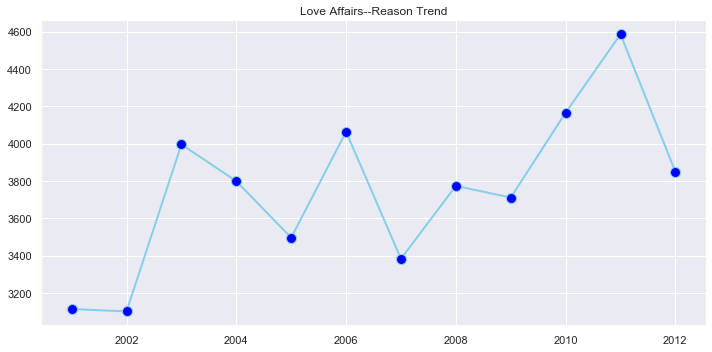
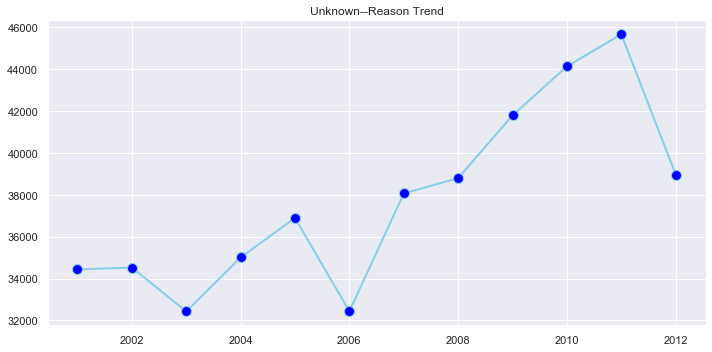
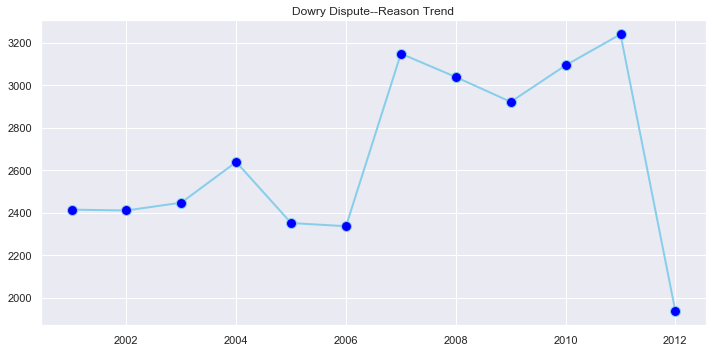
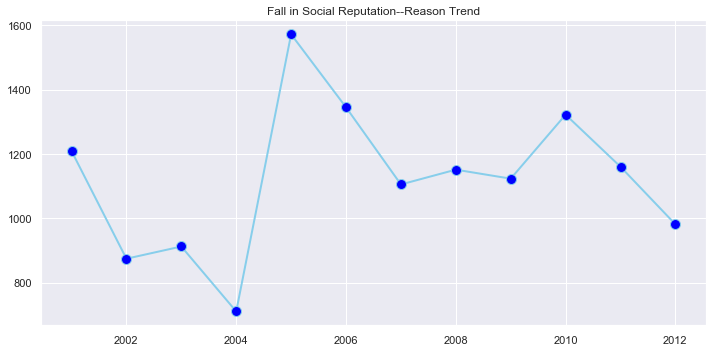
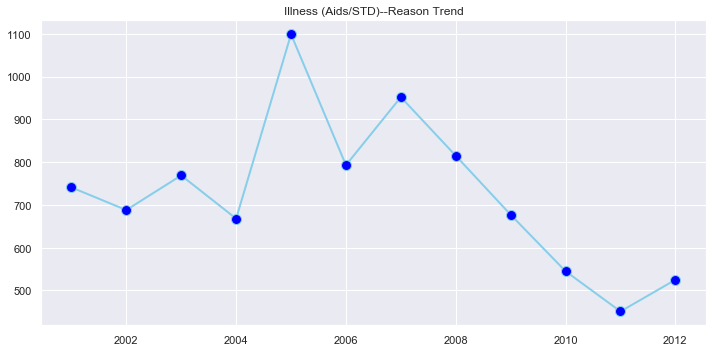
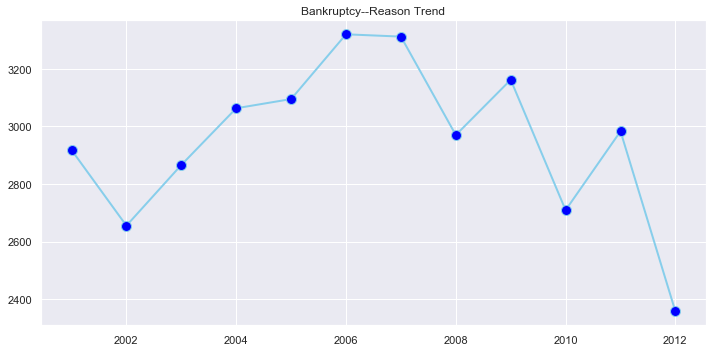
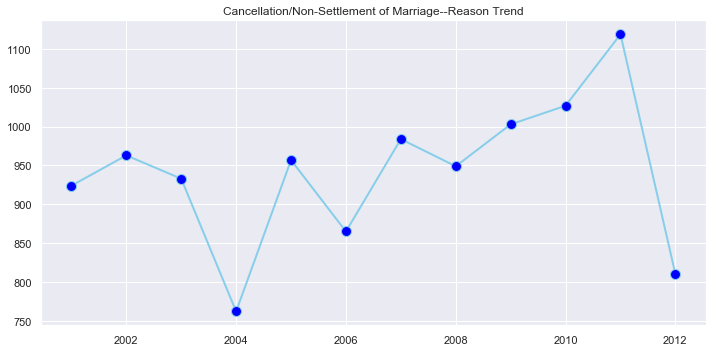
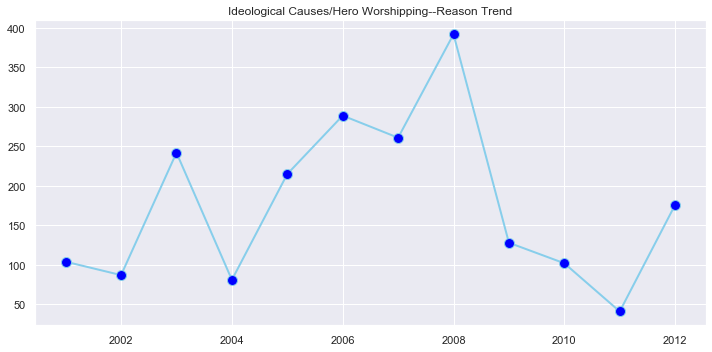
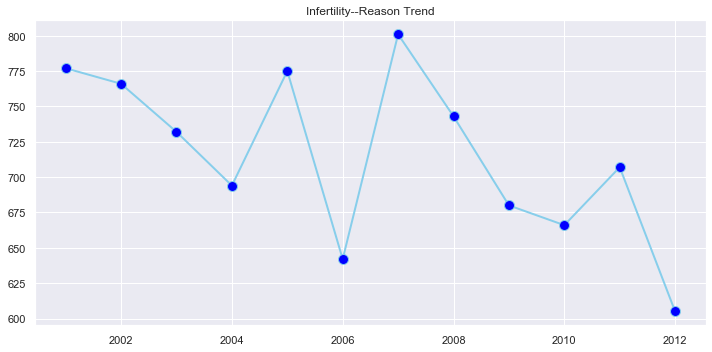
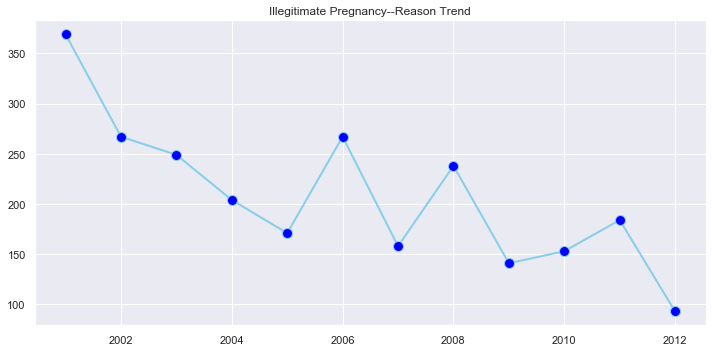
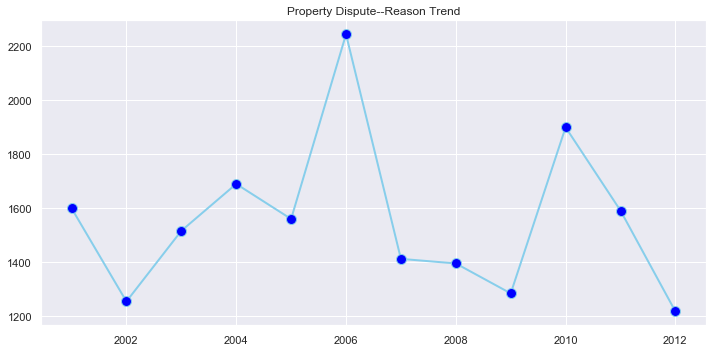
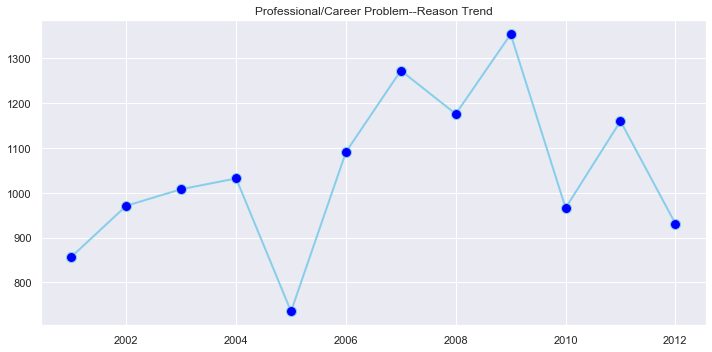
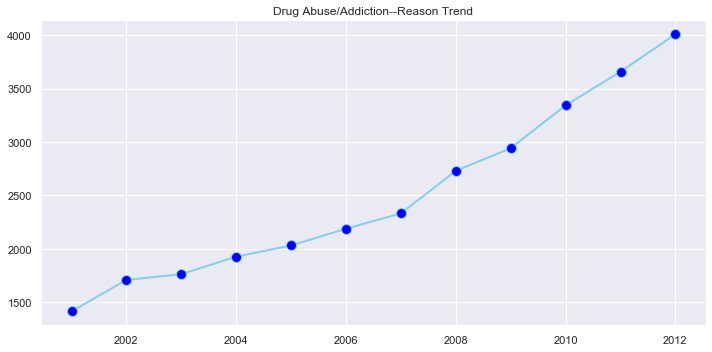
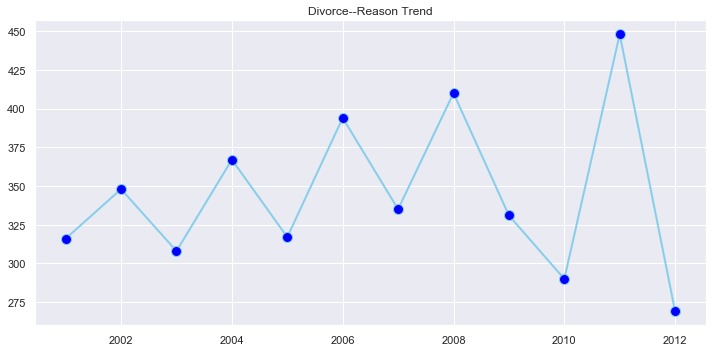
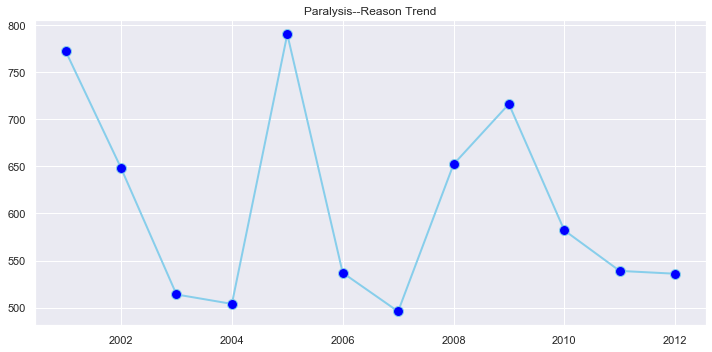
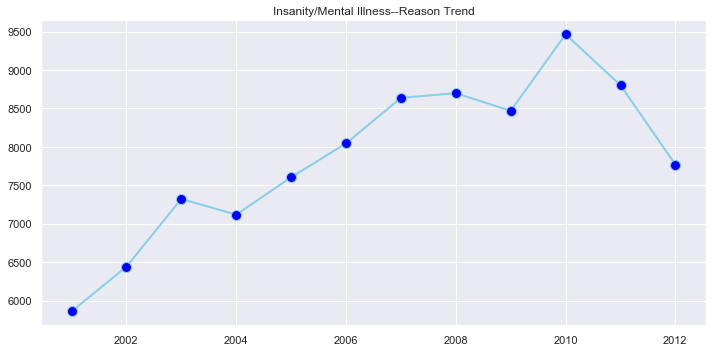
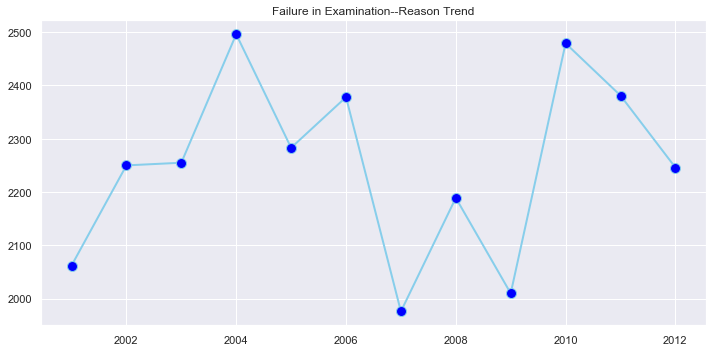
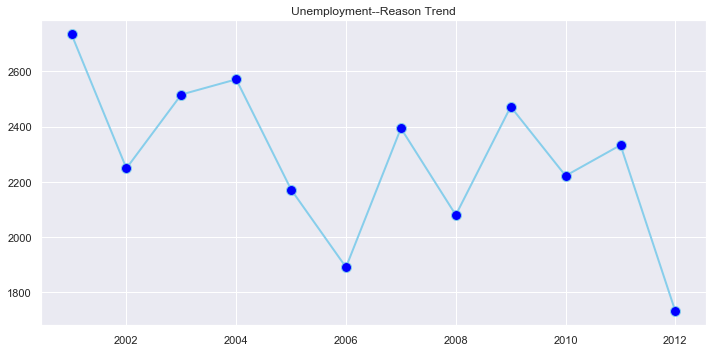
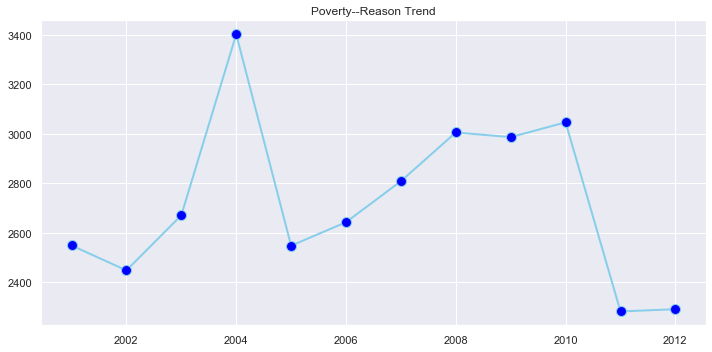
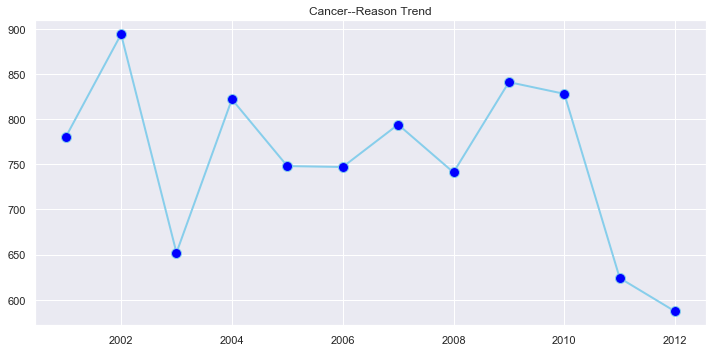
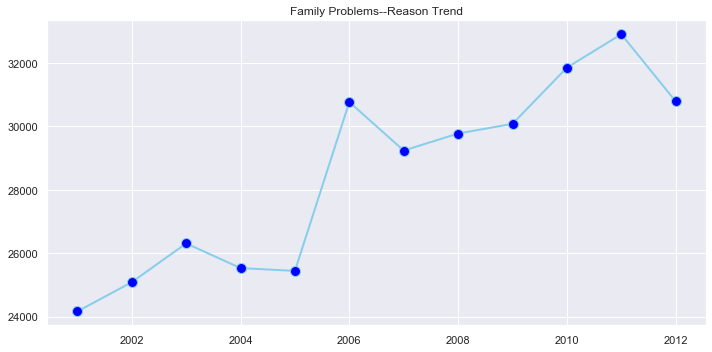
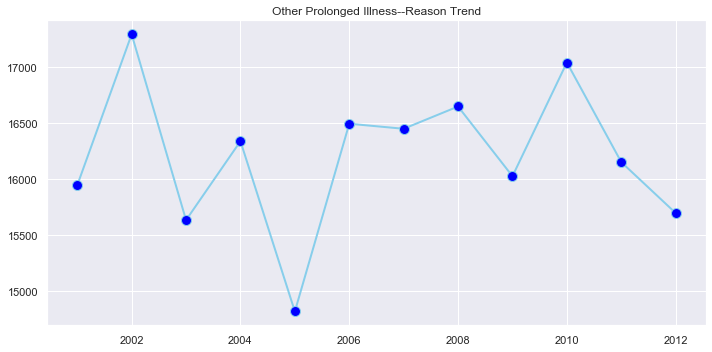
Suicide by hanging is a priority for Suicide Prevention:



Suicide by hanging increased by 56% (from 3.9 to 6.1 per 100,000) among males and by 24% (from 2.1 to 2.6 per 100,000) among females over the study period while incidence of insecticide poisoning decreased by 44% (from 2.7 to 1.5 per 100,000) among males and by 52% (from 1.7 to 0.8 per 100,000) among females.

Yearwise Reasons count - How it is changing?

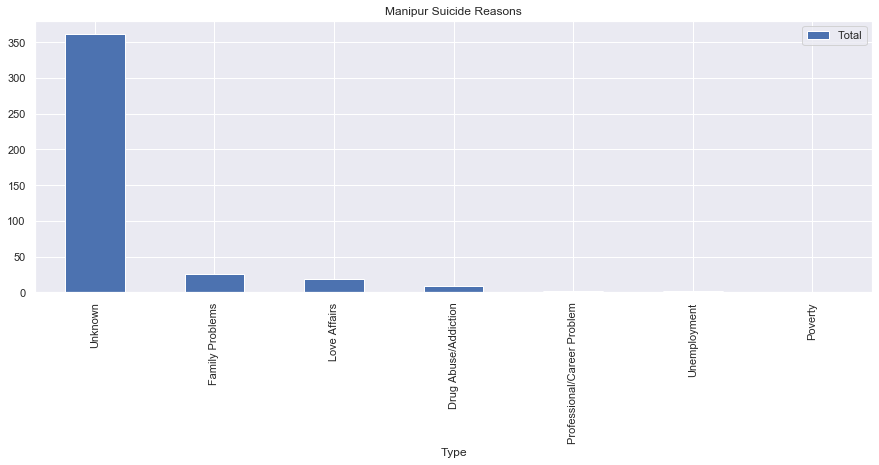
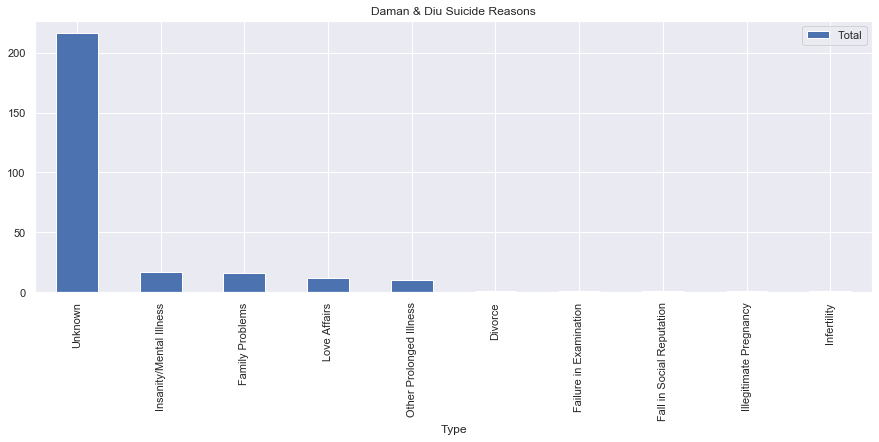
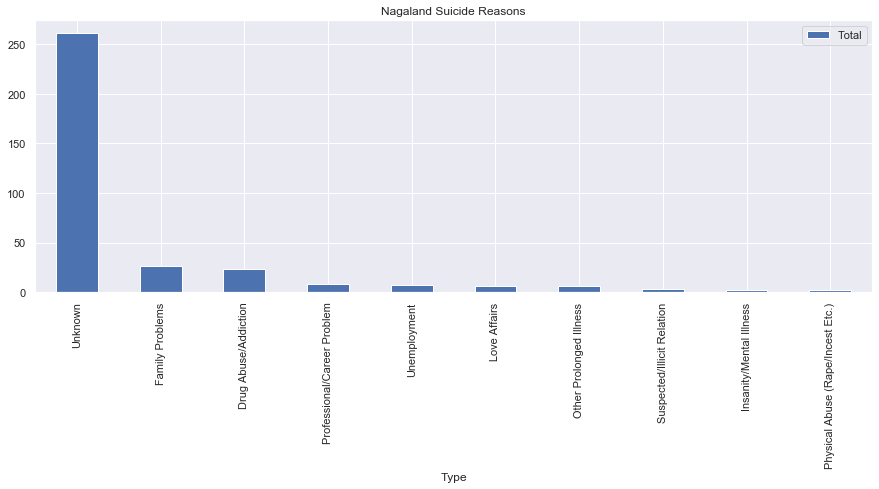
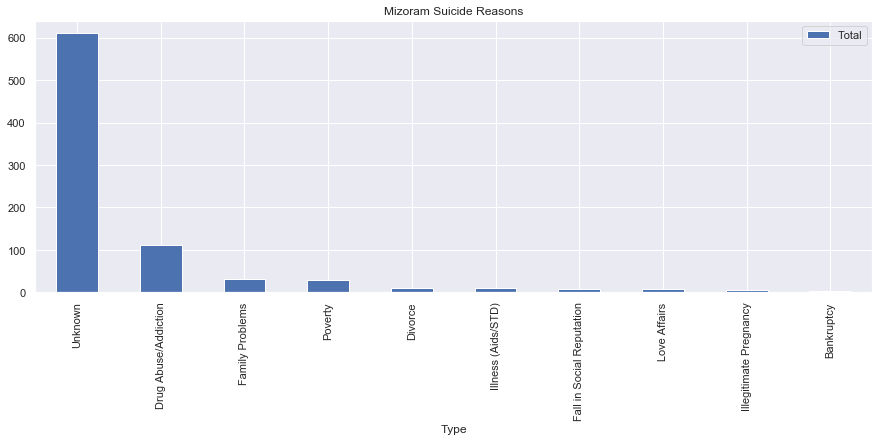
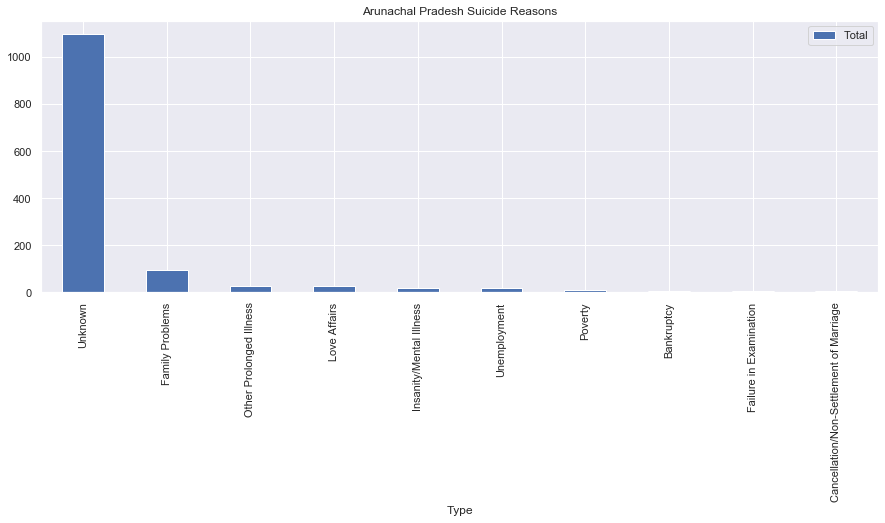
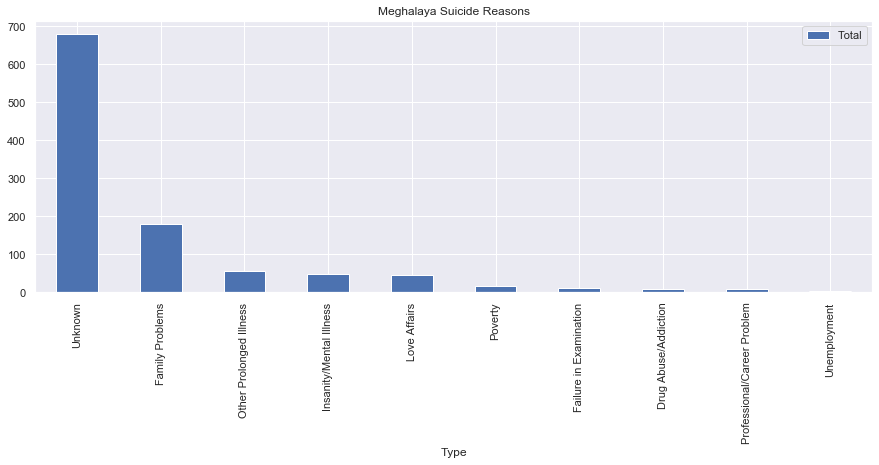
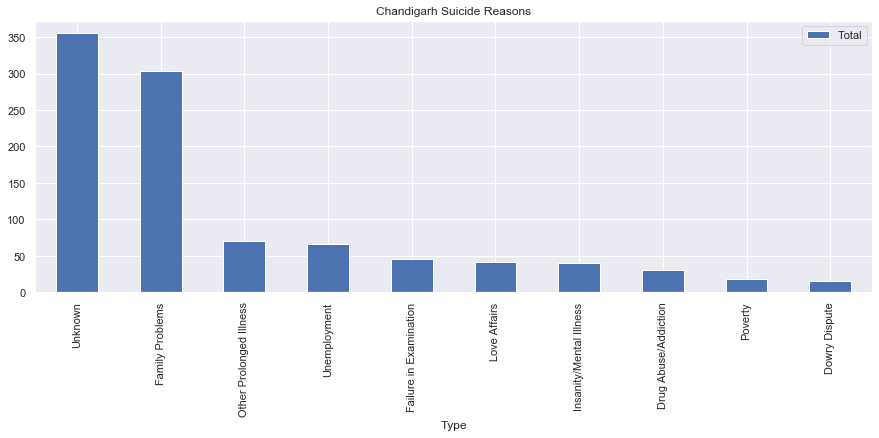
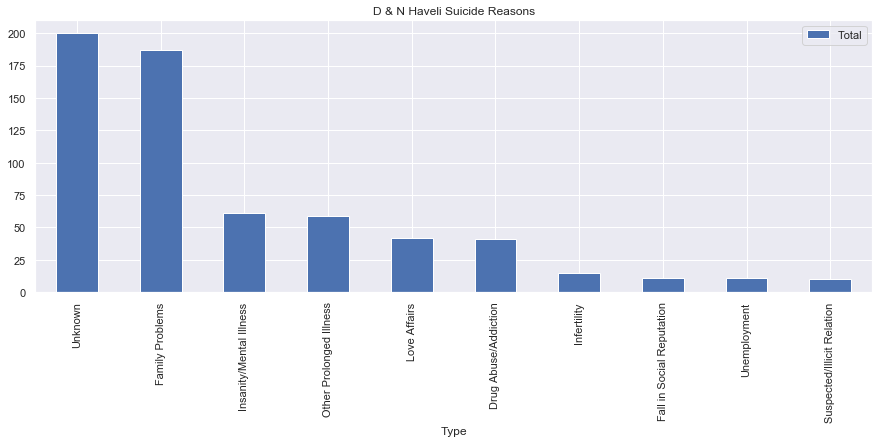
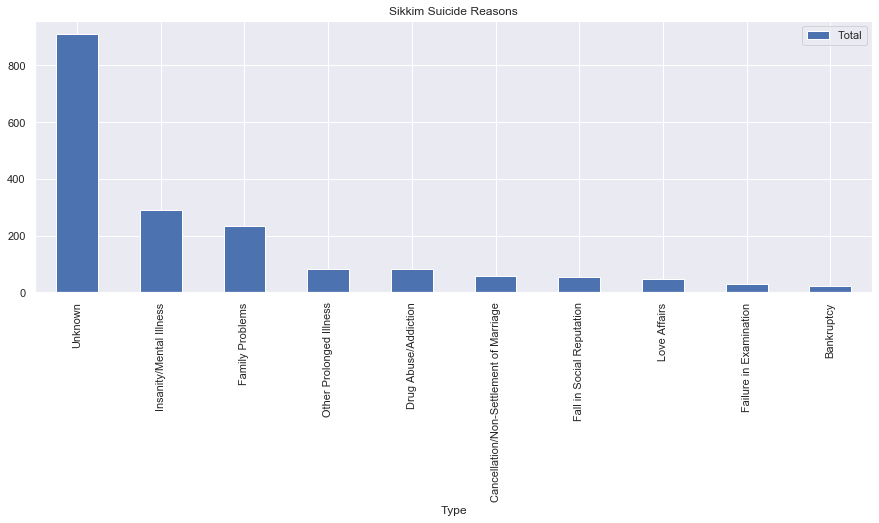
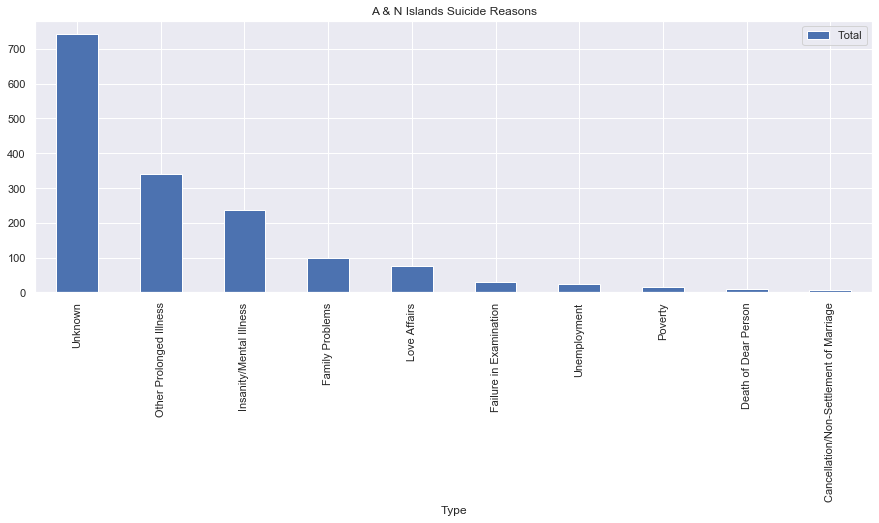
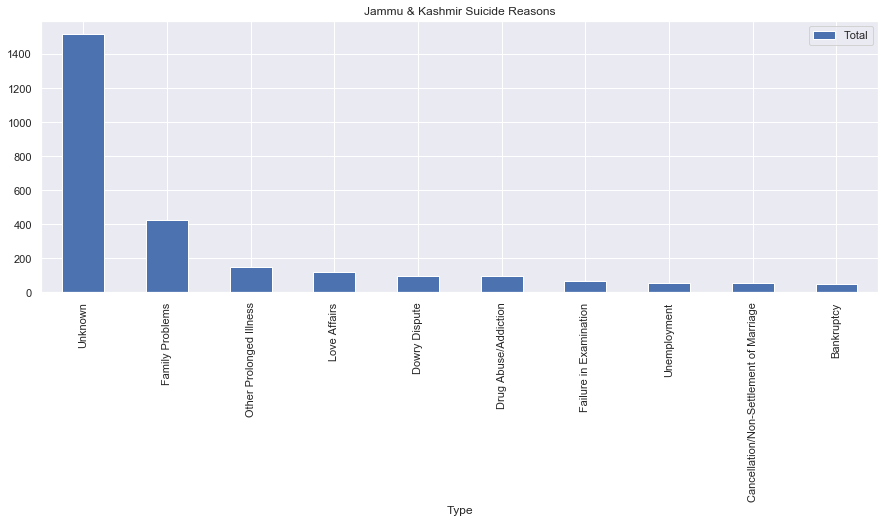
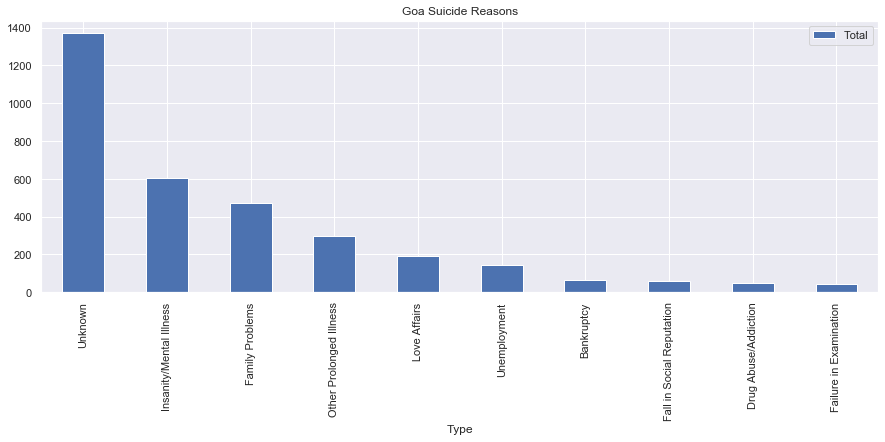
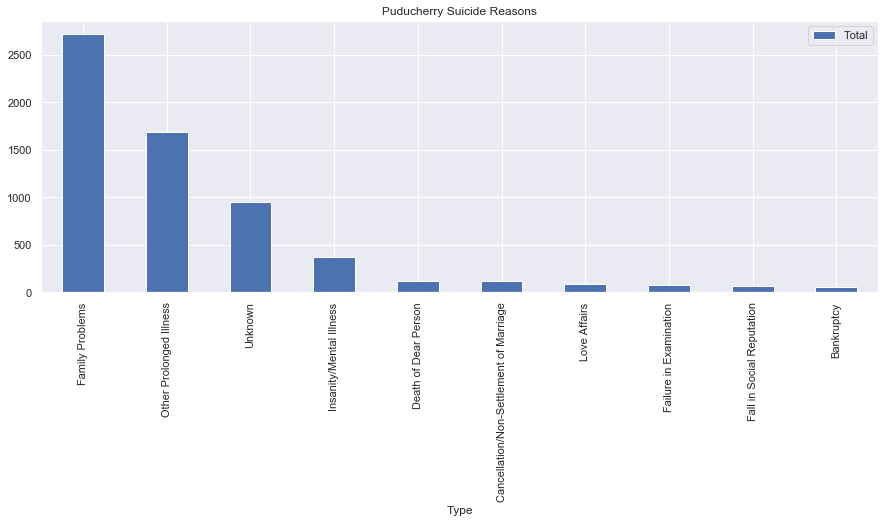
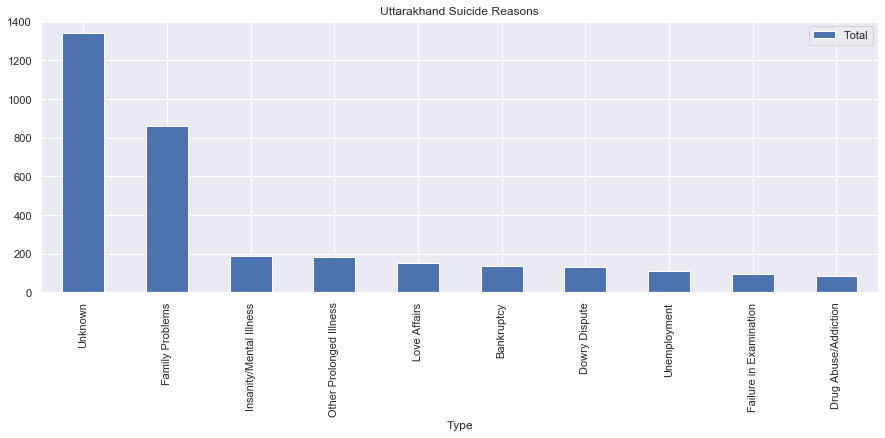
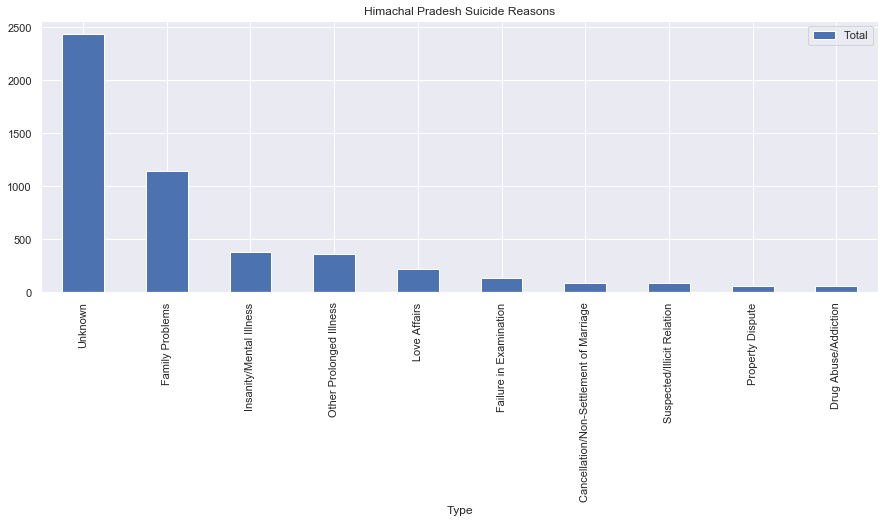
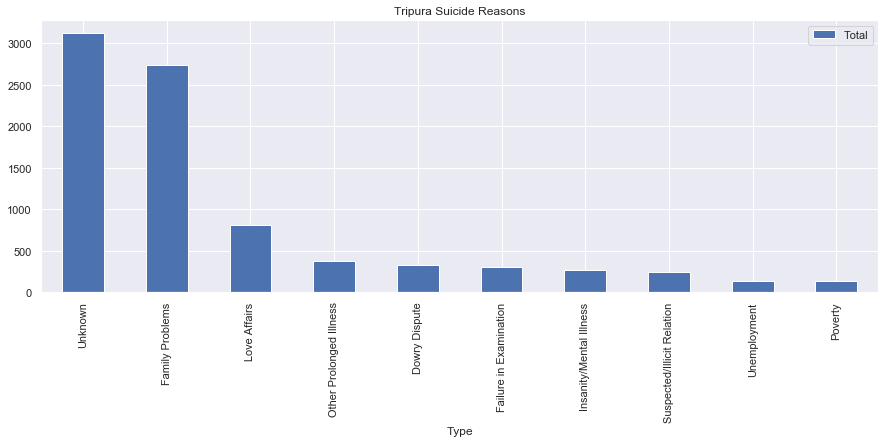
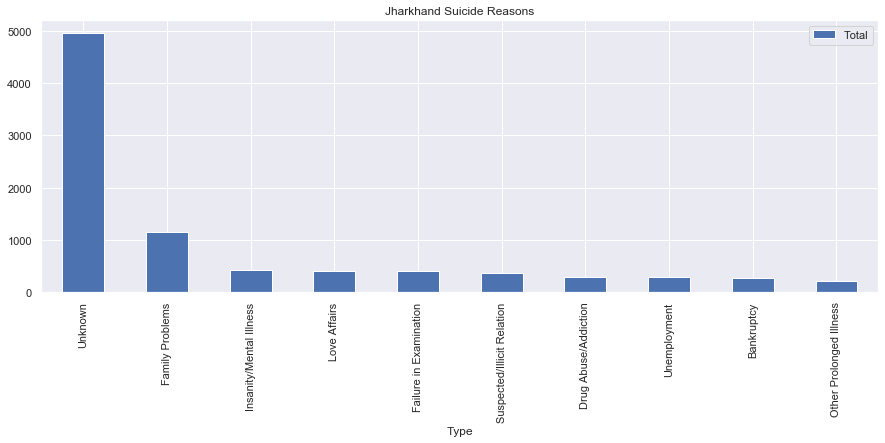
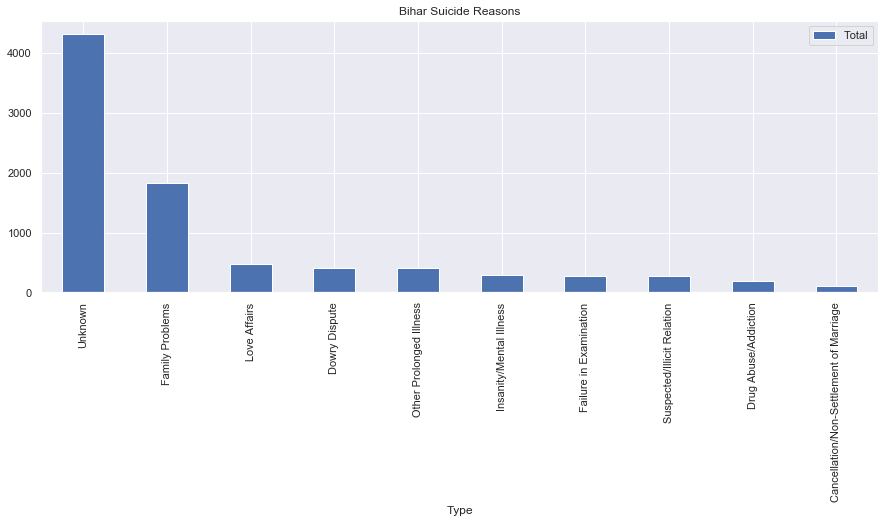
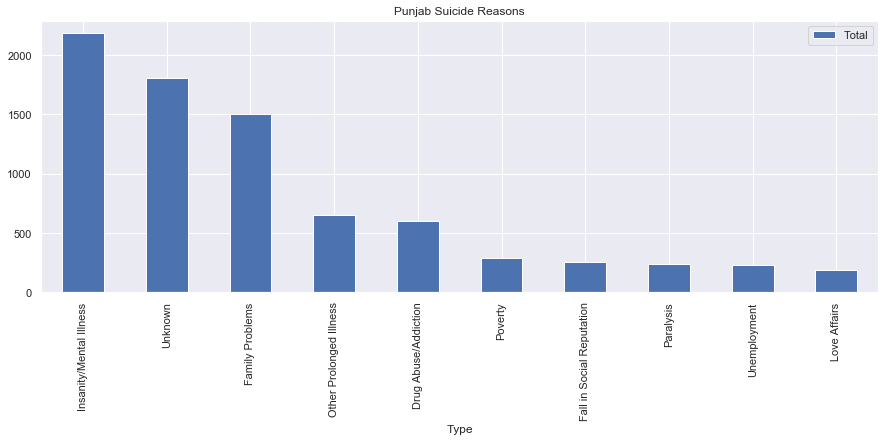
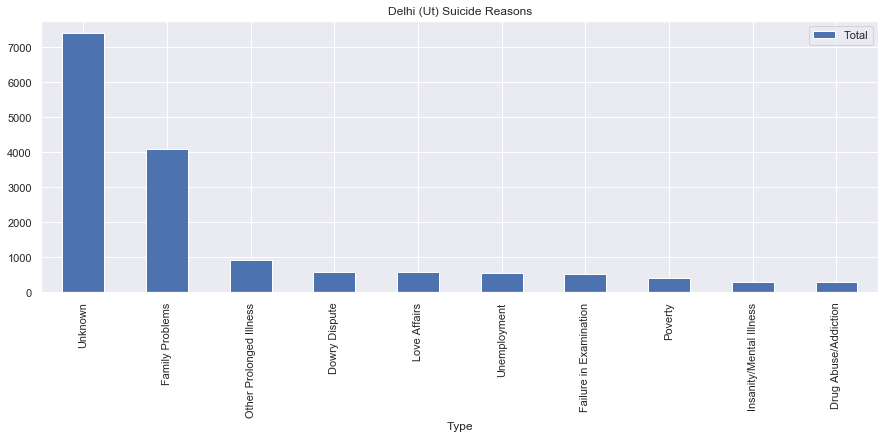
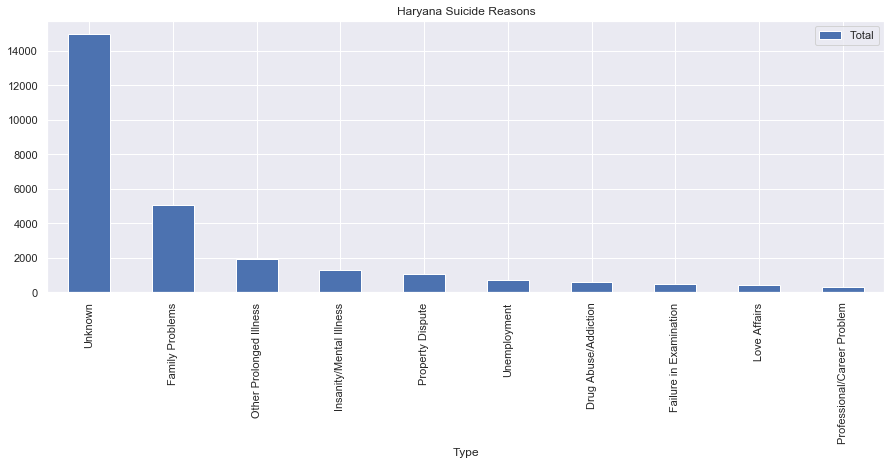
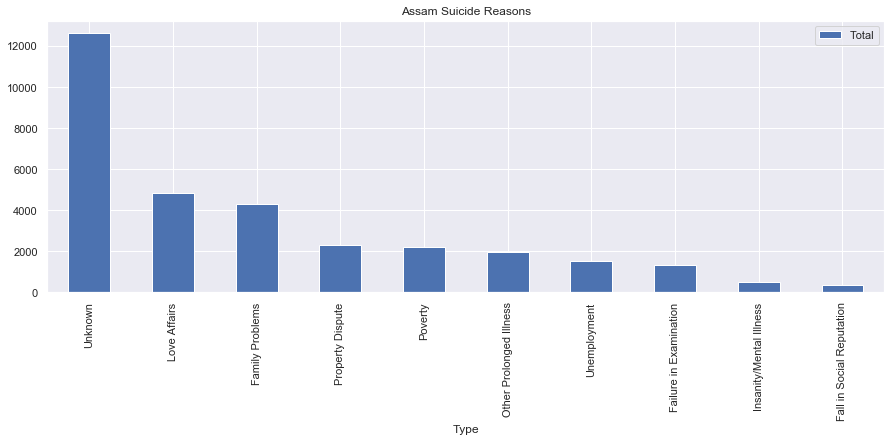
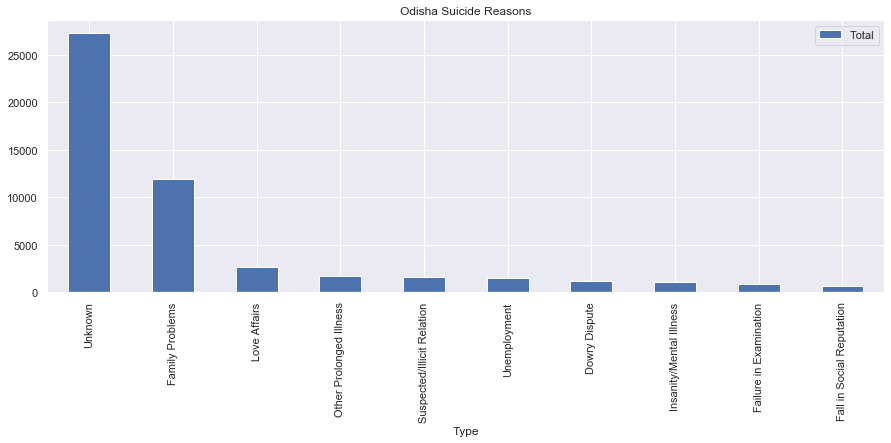
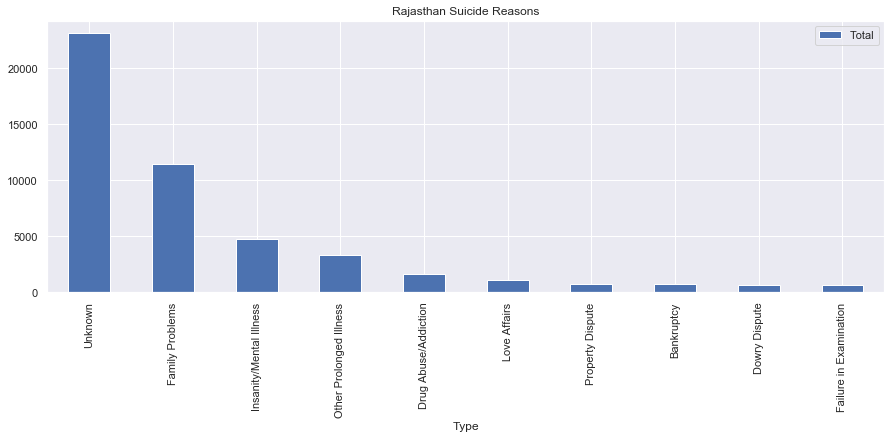
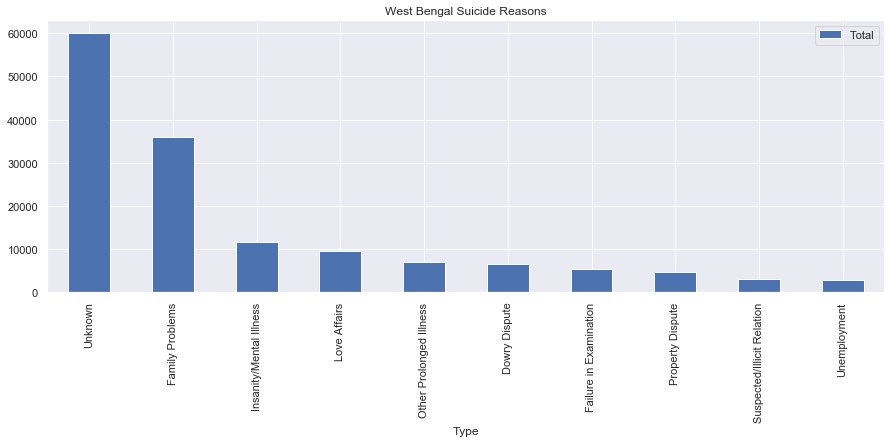
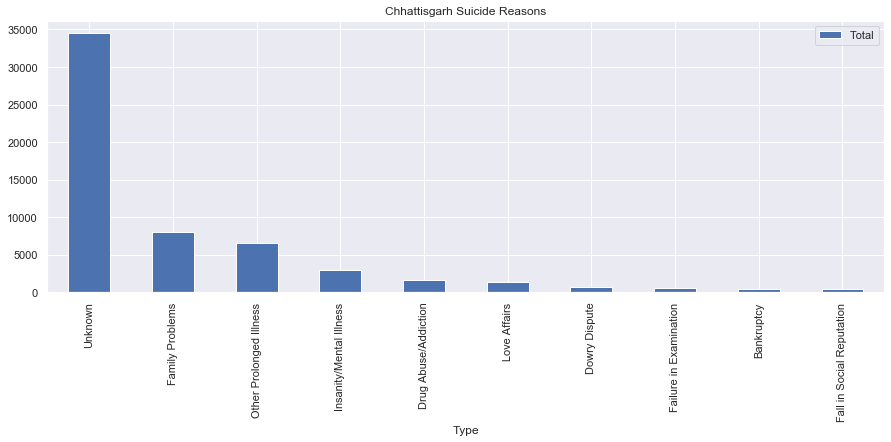
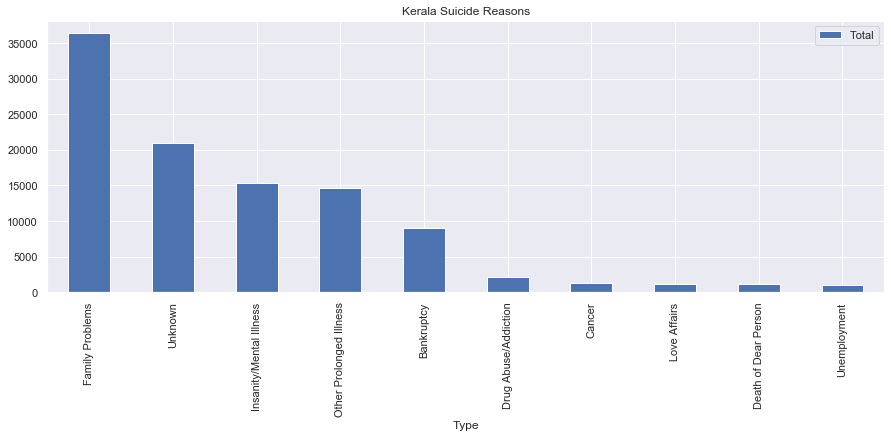
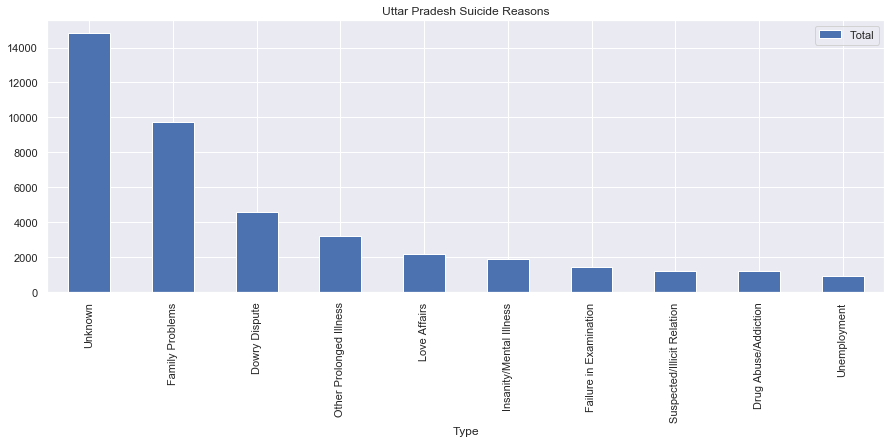
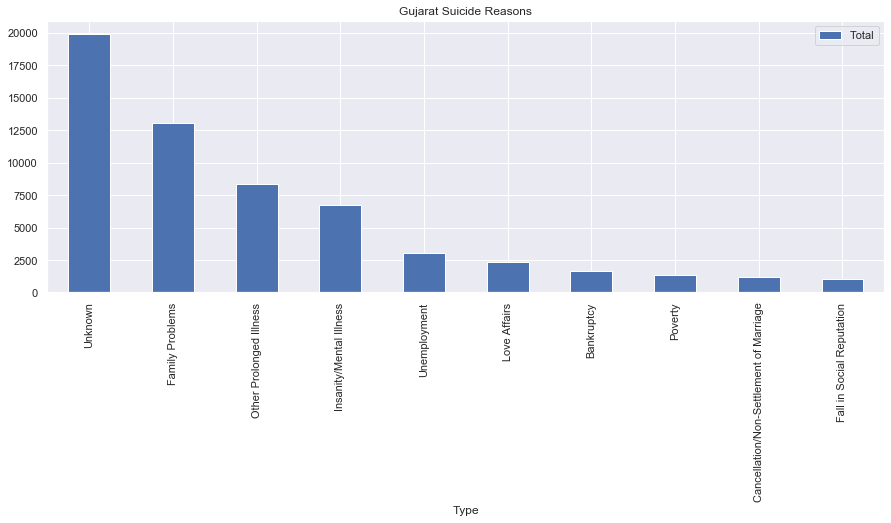
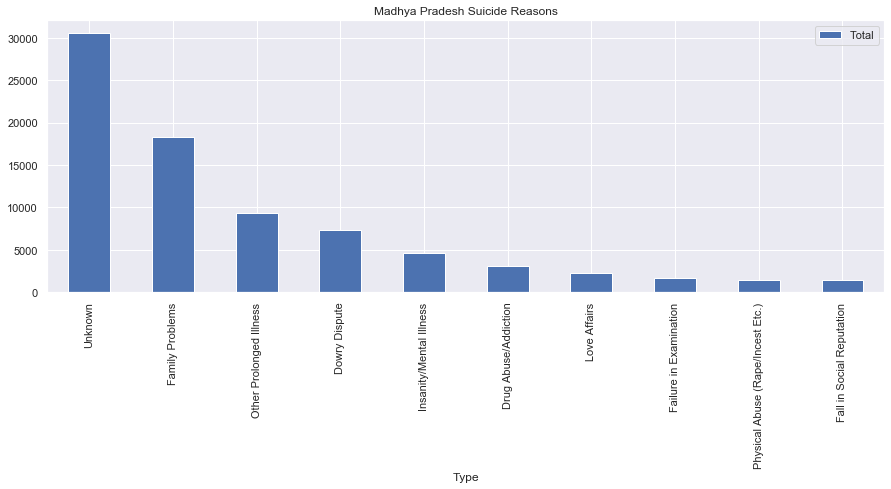
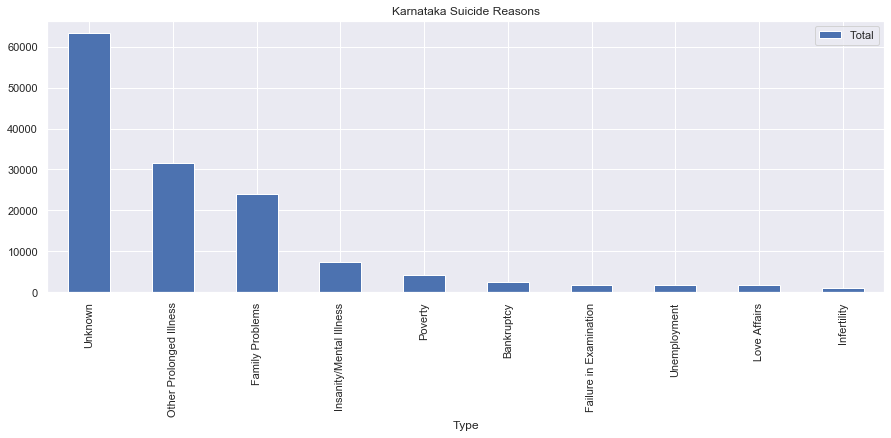
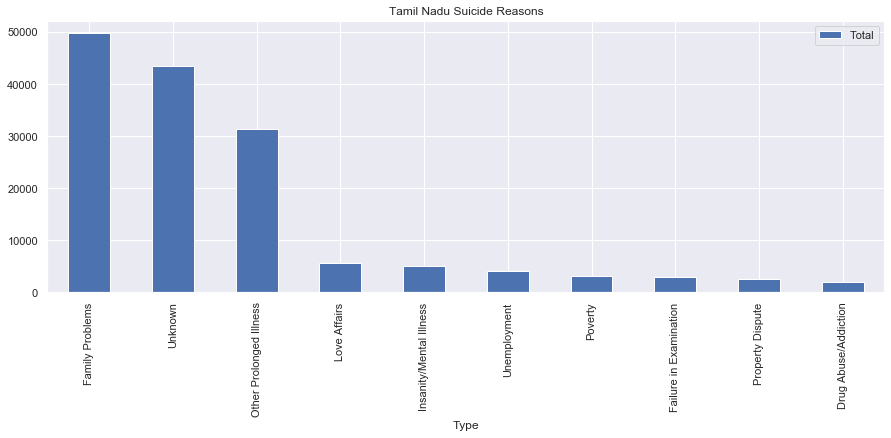
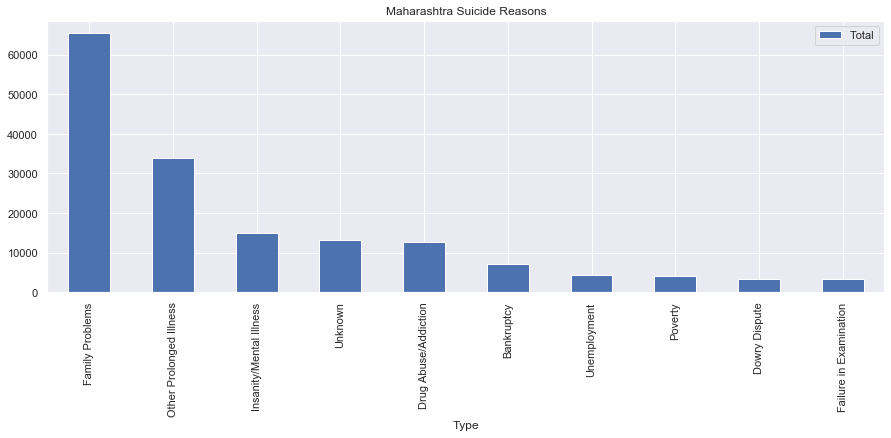
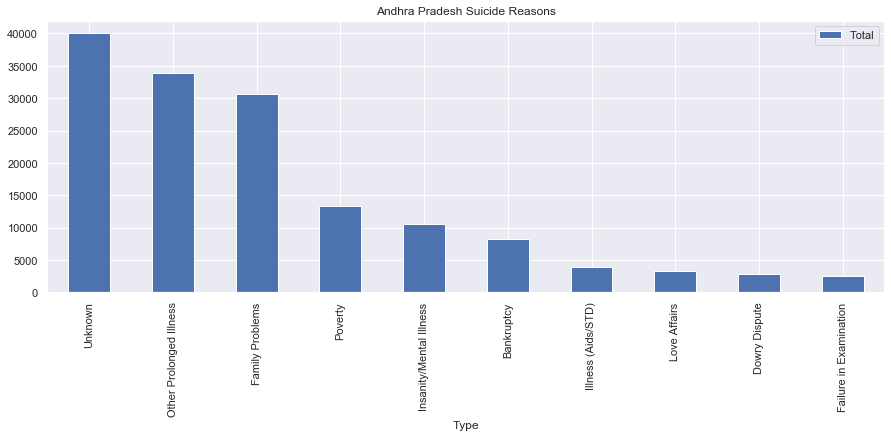
**

**

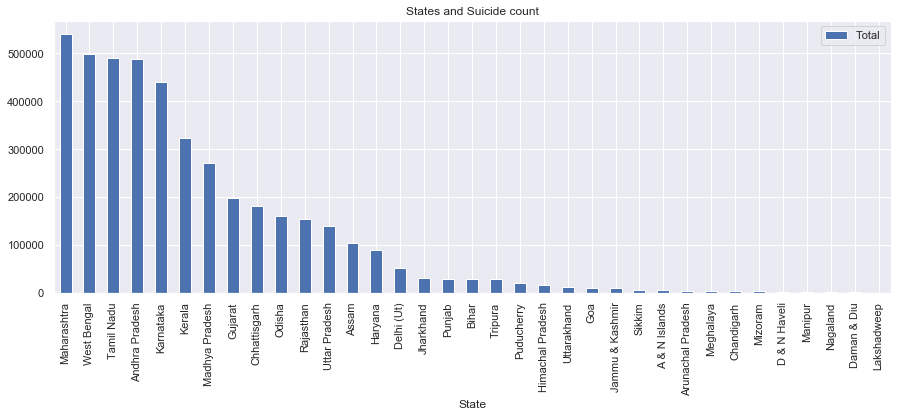
*Observations :*

1. *Drug Abuse/Addiction trend is is linearly increasing -- More than doubled*
2. *Family problems also shows upward trend 3 .Unemployment trend is showing decline - Employment opportunites are growing*
3. *Love affairs show increased trend.*

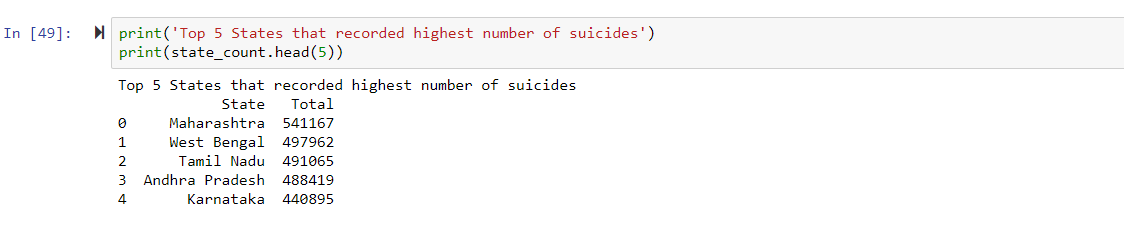
States and suicide reasons:



States and overall suicides count:



Top 5 States that recorded highest number of suicides:



## **What can be done to help reduce suicidal rate in male gender?**

* Male were raised in a manner where they have to act tough and hide their emotions.
* Family members and friends have to show their support no matter if its female or male to share what they going through and encourage them to share their emotions.
* On top of that, Government also should campaign to support citizens to reach out for help from each others and provide therapist to those that are actually in need which can help them come up with solutions that can be dealt with their mental illness.
* It doesn't only have to be some depression kind of illness but also people who are going through stress out can seek out for solution to deal with it.
* As for students who gets bullied at school, should seek help from their parents, in addition they can go see school council to complain. School should also support by taking a strict actions against those students who bully so, in the future students will think twice before bullying anyone.

**Future work:**

In future, the applied model of this research can be used in order to predict the causes independently for the every age group and also to classify the causes according to the male and female separately. Also, this research could have also be utilized to predict the amount of suicide in a timely manner.

**Conclusion:**

The results obtained gives us the clear vision about what type of population is highly affected by this problem. Government should take the preventive measures in bringing down number of suicides in our country by giving more attention on the population that is mostly affected in their respective states. It’s not only the government but even it is job of us also to work hand in hand with the government and help them in reducing suicides from our nation by providing counselling to the population of the respective states who are greatly affected by it.

1. Males were predominantly affected.
2. From Year 2001 to 2011 sucides rate is increased by 24.76% .
3. People who commit suicide are mostly:
   * Married
   * Farmers and housewives
   * Youngsters (15-29 age) and middle age (30-44)
4. Highest no. of suicide cases occur in Maharashtra, West Bengal, and Tamil Nadu.
5. Most favored method of suicide amongst males and females was hanging followed by poisoning.
6. Male dominance was apparent for each method of suicide except for self-immolation.
7. Males were relatively more likely to use hanging and poisoning while females were more likely to prefer drowning and self-immolation as methods of suicide.
8. Among females, significantly younger females resorted to hanging when compared to older females who preferred drowning.

#### These data provide a range of information to identify vulnerable groups, to formulate appropriate suicide prevention strategies. For hanging, limiting fictional portrayals of this method may be useful areas for prevention. Further restrictions on production and sales of highly hazardous pesticides may also help with further reductions in suicide by pesticide poisoning.[¶](https://render.githubusercontent.com/view/ipynb?color_mode=auto&commit=6cbe010153c4821c873905316582a6e480df64a8&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f766964686967616e64686939342f416e616c797369732d6f662d537569636964652d696e2d496e6469612d77686f2d7768792d616e642d686f772d2f366362653031303135336334383231633837333930353331363538326136653438306466363461382f53756963696465253230496e6469612e6970796e62&nwo=vidhigandhi94/Analysis-of-Suicide-in-India-who-why-and-how-&path=Suicide+India.ipynb&repository_id=272565103&repository_type=Repository" \l "These-data-provide-a-range-of-information-to-identify-vulnerable-groups,-to-formulate-appropriate-suicide-prevention-strategies.-For-hanging,-limiting-fictional-portrayals-of-this-method-may-be-useful-areas-for-prevention.-Further-restrictions-on-production-and-sales-of-highly-hazardous-pesticides-may-also-help-with-further-reductions-in-suicide-by-pesticide-poisoning.)

**References:**

1. Suicide in India: a systematic review Article in Shanghai Archives of Psychiatry · April 2014 DOI: 10.3969/j.issn.1002-0829.2014.02.003 · Source: PubMed
2. Prediction of Suicide Causes in India using Machine Learning ,Article · December 2017 , DOI: 10.31645/jisrc/(2017).15.2.01
3. Analysis of Suicides in India ,Vaibhav Khandelwal, Divyanshi Mangal
4. Journal of Network Communications and Emerging Technologies (JNCET), Volume 8, Issue 4, April (2018), ISSN: 2395-5317 ©EverScience Publications, Analysis of Suicide Database to Reduce Number of Suicides in India
5. Chavan B, Singh G, Kaur J, Kochar R. Psychological autopsy of 101 suicide cases from northwest region of India.

*Indian J Psychiatry.* 2008; **50**: 34-38. doi: http://dx.doi. org/10.4103/0019-5545.39757