Factors Affecting Mental Health in Employees and Their Relation to Suicide Rates at a Workplace

Shambhavi Pandey, Mohammad Bilal Khan and Sagar Thakkar, Clark University

ABSTRACT

The number of suicide attempts in the world are increasing and one of the main reasons for this is mental disorders, such as depression and mood disorders. The objective of this project is to predict the likelihood of an employee at workplace to commit suicide based on their behavior. Many factors are considered to analyze the reasons behind why a person would commit suicide and how it can be prevented. The present research mainly focuses on the working population because suicide rates are the highest amongst that population. The logistic regression model created in this study predicts the suicidal tendency among employees with 96% accuracy. This model has its implications in decreasing the suicide rate amongst the young generation by helping them to prevent self-harm.

INTRODUCTION

As per the WHO report regarding suicide rates, it has been found that more than 800,000 people die every year due to suicide, which is one person in every 40 seconds. Suicide is one of the most emerging problems in the world with an increasing trend year by year. Mental Disorders including depression and mood disorders are considered to be the most important risk factor of suicide (Causes of Suicides). In the US, as per the American Foundation of Suicide Foundation, suicide is the tenth leading cause of death, with an average of 123 deaths per day, in which men commit suicide 3.53 times more often than women. As per the report from Bertolote and Fleischmann, 98% of the people who died from suicide had mental disorders. Suicide is an extremely complex issue and the mortality data from the WHO suggests that suicide rates vary considerably between different age groups, different regions, and gender. The rate of suicide is highest in the middle age group as compared to other age groups. Additionally, it is more common among men as compared to women all across the world. As per the Statistica suicide report of 2018, the suicide rates vary from country to country, but the highest rates tend to take place in more diverse countries. In recent years, the suicide rates have increased significantly, but what is more concerning is the unpredictability of suicide tendency. During 2015, a German aircraft crashed into French Alps killing 150 passengers (Kulish & Eddy, 2015). The reason for the crash was quickly assigned to the co-pilot Andreas Lubitz, who was undergoing treatment for the suicidal tendency. According to the therapist and health inspector, Mr. Lubitz was fit to fly an aircraft and no signs of suicidal tendencies or aggression toward others were documented.

In this paper, we try to examine how factors such as age group, country, and gender effect the suicide tendency along with other factors including whether the person is self-employed or not, the mental health history of a family, whether the person has sought treatment for a mental health condition or not, whether the person has worked remotely or face to face in an office, whether their employer has provided health benefits to them, who is undergoing mental disorders, the importance of confidentiality, the difficulties of taking a leave of absence, and how discussing the case of mental health with coworkers can lead to a suicide attempt.

The main focus of this paper is to create a model to predict suicide tendency among the employees at the workplace based on specific demographics, work environment, support or lack thereof, and behavioral factors. By predicting the suicide tendency accurately, we can

prevent suicide from happening, by helping the right person at the right time. If a certain set of characteristics are present in person, steps can be taken by the company by providing guidance and medical treatment in order to save their life. Based on recurring factors, we will be predicting the suicide tendency among employees in order to determine which people are at most risk and how we can provide them with help.

LITERATURE REVIEW

Vast studies have been done on suicide and its causes ranging from genders to human resources. The complex working of a human brain to simple social conditions have been proven to have some other connection with suicide.

Chang et al (2009) used WHO and Taiwanese mortality statistics to explore the correlation between the Asian economic crisis of 1997-1998 and suicide rates. The study found that the male suicide rates in 1998 rose significantly in Japan, Hong Kong, and Korea during this economic recession, while there was no significant increase in female suicide rates. As per the US Bureau of Labor Statistics report from 2011 to 2013, Census of Fatal Occupational Injuries (CFOI) suicide rates increased with each successive age group over a period. The age group between 15 to 49 years had the highest suicide rates as compared to the other age group.

In a study conducted by Retamal C. Pedro & Humphreys Derek in 1998, "Occurrence of suicide and seasonal variation", they established a relationship between season and suicide rates in Chile. The suicide rate in Chile during the month of December was the highest with 10.9% whereas, suicide rates during the summer was only 7%. A similar study conducted by Jonviea Chamberlain of University of Massachusetts Amherst found evidence that in Alaska seasonal effect on suicide is more present in women as compared to men (Chamberlain, 2013). As per the study conducted at the Advanced Critical Care Medical Center, Yokohama City University Medical Center, it was found that there is a positive correlation between suicide rates and family history that include both genetic and non-genetic factors (Nakagawa et al, 2009). In the research, it was concluded that 15% of the suicide cases out of the 460 suicide attempts have a connection to family history.

METHOD

The data selected is from a survey conducted by Open Sourcing Mental Illness in 2014 available on Kaggle.com. Figure 1 depicts the flow diagram of the process. The original data was cleaned to correct any spelling errors and standardize the answers. 39 observation from the original 1259 were also excluded due to multiple missing values. Further, all the yes responses were then converted to 1 and no to 0 for questions related to benefits, remote work, treatment, family history and self-employment. Once the data was cleaned, 11 factors from the initial 27 were selected that were relevant towards the current study. Out of the 11 factors, age and gender were part of the demographics while the remaining 9 factors were part of history and work environment. Table 1 shows the original questions for work environment factors. Once the factors were selected, visualization was used to understand the pattern present in the data as a part of descriptive analysis. This was followed by model creation. The data was partitioned in a 70-30% ratio, where 70% data set was used to train the model followed by validating the model created using the remaining 30% data set.

DATA DESCRIPTION & VARIABLE USED

The survey results originally consisted of 27 factors and 1259 rows. The 27 factors were divided into three categories: work related factors, history, and family background. To better comprehend the results, a combination of predominantly work-related factors

including age, location and a small quantity of history factors were selected. The target or the dependent variable used in the dataset was "Suicide Tendency", which is a string variable in the form of 0 and 1, that has been asked by employees in a survey conducted by OSMI. Among the 27 factors given in the survey questions, it was found that the 11 factors considered in the research are significant in predicting the suicidal tendency among the employees at the workplace.
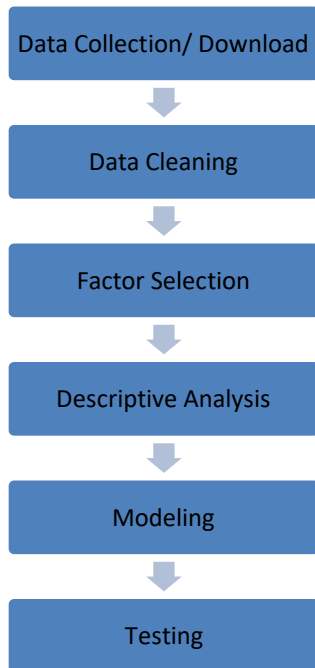
Figure 1 Process

Factor	Survey Question
leave	How easy is it for you to take medical leave for a mental health condition?
coworkers	Would you be willing to discuss a mental health issue with your coworkers?
anonymity	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
benefits	Does your employer provide mental health benefits?
remote_work	Do you work remotely (outside of an office) at least 50% of the time?
treatment	Have you sought treatment for a mental health condition?
family_history	Do you have a family history of mental illness?

self_employed	Are you self-employed?
Seek Help	Does your employer provide resources to learn about mental health Issues and seeking help?

Table 1 Variables Used

There were no missing values found in the dataset on examining it.

Table 2 shows the Kurtosis and Skewness for Age, Leave and Suicidal Tendency. The Kurtosis values for Age, leave and suicidal tendency are 1.9413, -0.6433 and 1.7751, respectively. The Skewness values for Age, leave and suicidal tendency are 0.9528, 0.7431 and 1.9422, respectively. Overfitting with the data was not found to be a significant issue.

	Age	Leave	Suicidal Tendency
Kurtosis	1.9413	-0.6433	1.7751
Skewness	0.9528	0.7431	1.9422

Table 2 Kurtosis & Skewness for Age, Leave, and Suicidal Tendency

The basic statistics details of the target variable and the independent variables chosen was checked. It was found that Age has a mean of 32.05 with Min and Max value of 5 and 72 respectively and the majority of the employees lie in the range of 26-35 age group. The suicidal tendency among the employees is represented based on the factors selected such as self-employment, leave, anonymity, seek help, treatment sought, age, country, state, gender, benefits.

DESCRIPTIVE ANALYSIS

Figure 2 shows the Suicide tendency in the US. On analysis of data set on suicide tendency based on mental health, the suicide tendency among the employees working in the US among various states in the US was found, which was observed to be high in the states such as California which consists of a high number of companies and busy work schedules.

Figure 3 shows the suicidal tendency, self-employment and suicidal rate among various age groups. The analysis found that higher suicidal tendency was reported by people between 26-32 years of age while most of the suicidal tendency among the people are generally in the age group range of 19-45 years.
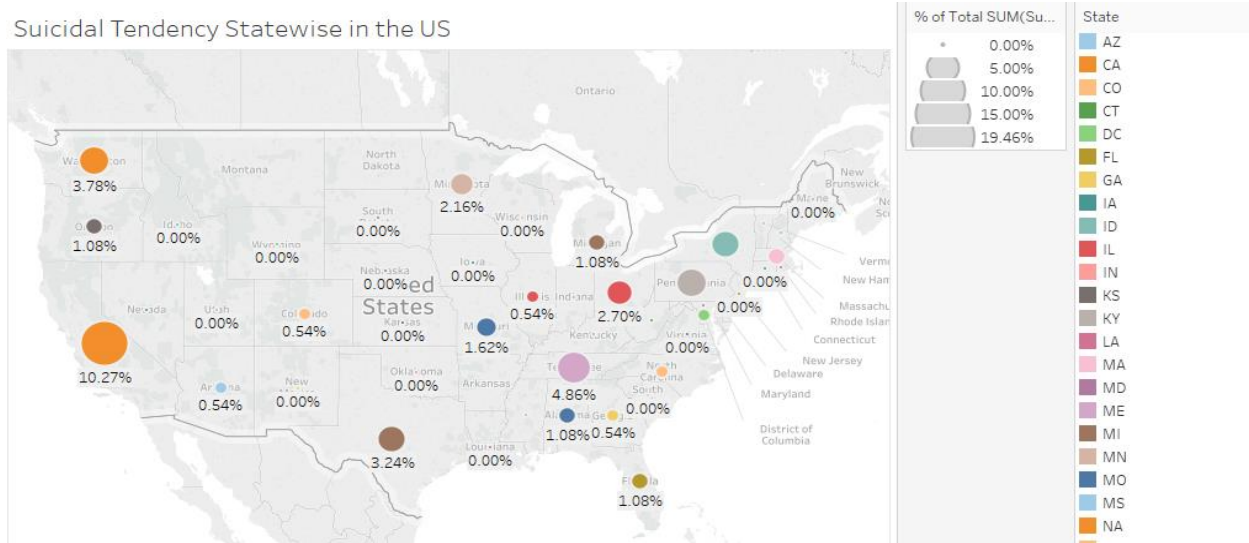
Figure 2 Suicide tendency in US

An increase in the suicidal tendency in the people of the US during the recent time has been observed, with California showing highest suicidal risk among the people. Remote work and the suicidal tendency among the employees were found to be negatively correlated. Similarly, suicidal tendency was also found to be negatively correlated with benefits provided by the employers to the employees. The factors such as benefits, family history, anonymity, and discussion with the coworkers were found to significantly affect the suicidal tendency among the employees. The employees with family history of mental disorders, without health benefits provided by the employers, and uncertainty regarding their anonymity regarding mental health in the workplace tend to have more suicidal tendency. However, it has been analyzed that a lack of discussion regarding mental health with the coworkers increases the suicidal tendency among the employees.

Benefits	Remote Work	
	No	**Yes**
No	148	27
Yes	10	0

Table 3 The variation of Suicide Rates with benefits

The presence of more established companies in states like California also raises questions regarding benefits provided to employees and remote work done by individuals. Table 3 shows the variation of suicide rates with benefits and remote work. There has been found a negative correlation between remote work and suicidal tendency as well as benefits (provision or awareness) and suicidal tendency.

Another representation of the same parameters indicates that the company which does not provide a benefit tend to have more suicide tendency. Further, a fear of disclosure of mental issue information does play an important role in contributing towards suicide tendency. Employees who does not have much confidence in the anonymity policy of the company (or is not aware of it) tends to have more suicide tendency. There is a positive correlation between the employees who have not taken any treatment and the suicide tendency among

5

them. Also, when it comes to the relation between the easiness to get the leave by the employee during the mental illness and suicide rate, it has been found that the more difficult it is for the person to get the leave during the mental health, the more chances are there for the employees to develop a tendency if suicidal thought.
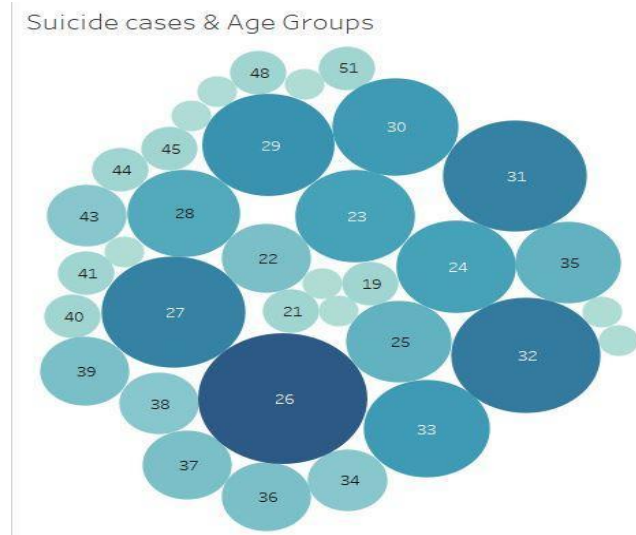
Figure 3 Suicide tendency and self-employment & Suicide rate among various Age Groups

Apart from social conditions, work conditions play a major factor for suicides. For the current research analysis, the effect of being self-employed or working for a company on suicide tendency was considered. Using visualization as a tool, suicide tendency is found to be much higher in people working for a company as compared to the ones who are self-employed. Figure 3 shows the significantly higher rates of suicide attempts among men between age 22 to 43. The visualization is in sync with researches conducted by CDC. In 2014, center for disease control (CDC) conducted a study to evaluate innovative and promising Strategies to Prevent Suicide among Middle-Aged Men (35-64) Even though the data in the current research does not fully sync with CDC's conclusion, a common age group of men in their 30's who are more prone to suicide is still observed. For this reason, age group was chosen as one of the factors contributing factor for suicide tendency.

The aim of the current paper is to examine the relationship between suicide and work. Which work factors/conditions can contribute towards increasing the suicide tendency among its employees? Since majority of the data in this paper is for united states, state with the highest suicide rate will also be examined. When considering impact of work on suicide, the impact of economic recession on suicide will also be considered. Apart from working conditions, attention is also given to social conditions/factors (age, gender, family history) that may be more prone to develop suicide intention.

The methods that will be used to predict the suicidal tendency among the employees based on certain set of sets including both demographics and work related include the decision tree, the artificial neural network and logistic regression. The aim is to come up with the model that predict the suicidal tendency among the employees as accurately as possible.

MODELLING

BASELINE MODEL

First, the baseline model has to be decided to which the model's accuracy will be compared. On analyzing the dataset, it was found that the out of the sample size of 1218, there are 1033, who don't have suicide tendency i.e. equal to 84.81%, while 185 employees out of 1288 i.e. 15.18% of the employees have a suicide tendency due to the factors taken into consideration. The authors aim to develop a model that can predict the suicide tendency among the employees with an accuracy of more than 84.81%. Also, on looking at the suicide tendency among employees' gender wise, it was found that 79.8% of the employees that have a suicide tendency are males, while 20.1% of the employees with suicide tendency are females. From this, it can be concluded that males are more likely to develop suicide tendency as compared to females.

Models

The models for the data sets were created. The decision tree model was the first created. The data set was partitioned in the 70-30 ratio, with 70% of the data used in training the model, while the rest 30% of the data used in validating. The training dataset consists of 852 employees' details and the validation dataset consists of 366 employees' details.

Training data set was used to create a decision tree model to predict the target variable, which is suicidal tendency among the employees and then validated the model using the validation data set. With the training dataset, it was found that the accuracy of model of to be 93.54% with 797 of the employees correctly classified in terms of suicidal tendency, while on validating the model with the validation dataset, the accuracy of the model was found to be 84.68% with 268 of the values as correctly classified and 47 of the values as wrongly classified.

The decision tree model, the first split is done based on the factor coworkers, which represents how often an employee discusses its mental health illness with coworkers and what are the chances that this discussion leads to suicidal tendency among the employees. On looking at the decision tree model, it has been found that out of 852 employees 723 did not develop suicidal tendency based on taken factors, while 129 of the employees develop suicidal tendency. After that, out of the 675 employees, who discussed their mental illness with the coworkers, 626 of the employees did not develop suicidal tendency, while 49 of them developed suicidal tendency. Similarly, out of 177 coworkers, who did not discuss their mental illness with the coworkers 97 of the employees did not develop suicidal thought, but 80 of these employees developed suicidal thought.

With all the steps to be remained same, a logistic regression model was created to train the model with the partition of the dataset in 70-30 ratio, with 70% of the employee details used for training the model, while the rest of 30% used to validate the model.

On validating the model with the validating dataset, the accuracy of the model in predicting the suicidal tendency was found to be 96.44% with 353 of the employee's suicidal tendency correctly classified out of 366, while 13 of them are wrongly classified with misclassification rate of 3.55%.

For creating an artificial neural network, the string variables had to be converted to an integer variable. As the value of variables such as treatment, benefits, family history, and others had yes or no response, Yes was converted to 1 and No to 0, followed by normalizing to make the training process faster and reduce the chances of getting stuck in local optima and then proceeded to create an artificial neural network. Similar to previous models, a 70-30 ratio data partition was created.

Artificial Neural network achieved an accuracy of 96.17% with 352 employee's suicidal tendency correctly classified out of 366, while 14 of them are wrongly classified with misclassification rate of 3.82%. Tabular summary of result for the models is represented in Table 4.

Model	Accuracy	Sensitivity	Specificity
Decision tree	84.68	0.893	0.5
Artificial Neural Network	96.17	0.98	0.862
Logistic regression	96.44	0.98	0.86

Table 4 Accuracy of Models created

Among the three models, the Logistic Regression Model was found to be the preferred model because of the high accuracy and low misclassification rates. To check the consistency of the preferred model, the testing/validations data set ratio was varied (Table 5). Among the three models, Logistic regression experienced smallest variation and was the most consistent model. There are neither issues of overfitting nor missing values in the model.

Model	60-40 Ratio	65-35 Ratio	70-30 Ratio	75-25 Ratio
Decision Tree	88.29%	86.50%	84.68%	85.01%
Logistic regression	95.68%	96.01%	96.44%	96.34%
Artificial Neural Network	95.27%	94.37%	96.17%	94.74%

Table 5 Accuracy of Models with Variation in Data set Ratio

Discussion

The coefficients and P-values of all the factors calculated in the logistic regression can be seen in Table 6. From all the factors considered in the paper, it can be observed that gender and absence of protection of a person's identity, does not have any effect on development of suicidal tendency. It was found that factors that contributed to the model are Self-employment, family history, treatment, remote work, benefits, seek help, anonymity, leave, and coworkers. Factors with p value greater than 0.05 are not significant in this model. Table 6 also shows the beta coefficients and odds ratio for the factors that were used in the model. It is interesting to see that if an individual is self-employed, they have 33.3 times higher odds of developing suicidal tendency as compared to an individual working for a company. Similarly, if an individual work in a company with very strict leave policy, they have 3.2 times higher odds of developing suicidal tendency. It is not surprising to see that presence of mental health in the family history have a positive impact. An individual with family history of mental issue have 141 times higher odds to developing suicidal tendency.

Factors	Coefficients	Odds Ratio	P>|z|
Gender Male	0.71	2.03399126	0.123
Self-employment=yes	-3.505	0.03004677	0
Family history=yes	-4.954	0.00705513	0
Treatment=Yes	4.366	78.7280887	0
Remote work=Yes	4.447	85.3704481	0
Benefits=yes	4.368	78.8857024	0
Seek Help=yes	3.487	32.6877373	0
Anonymity=no	0.165	1.17939312	0.8
Anonymity=yes	3.401	29.9940791	0
Leave	-1.177	0.30820196	0
Coworkers=Yes	4.962	142.879269	0
Constant	-1.369	0.25436119	0.01

Table 6 Coefficients and P-values of Factors in Logistic Regression

The model indicates that when employees are aware of the benefits provided by the company, organization provides a positive environment and a platform for employees to easily ask for help and assure them of the confidentiality of their health significantly reduces the chances of suicidal tendency. If employees are aware that their company have measure to provide support, they are more likely to monitor their mental health and to ask for help without any reservations. Accepting treatment also reduces chances of mental health, as seen in the model, and a company can improve their work environment by promoting their employees to take treatment without any hesitation. Providing a safe support group within the organization can be beneficial (with priority to maintain the anonymity). It is important for an organization to have policies and measures to support their employees but more importantly, it is important for an organization to create a harmonious work environment.

LIMITATION

The survey was conducted focusing selectively on tech companies. Inclusion of other industry can provide a deeper insight into the issue. The expectation of an employee from its organization various with industry and can provide other factors that may have impact on employee's health. The study was limited to the questions asked in the survey conducted by Open Sourcing Mental Illness. A more comprehensive study with other factors can be conducted. Many organizations have health care facility with in the office. A deeper understanding of how the presence of facilities or awareness of the facilities and services offered by these facilities affect the tendency can be a new direction.

CONCLUSION

Today, one of the emerging problems in the world is Suicide. Approximately, one person commit suicide in every 40 second. The suicide tendency is even higher in males as compared to females. In United States alone, Suicide is the tenth leading cause of death. In the recent years, the suicide rate has increased significantly, but what is more concerning is the unpredictability of suicide tendency. From the initial analysis of the data, evidence was observed for higher suicide among middle aged males as compared to females. It was also found that remote work and the suicidal tendency among the employees are negatively correlated as the suicidal tendency among the employees decrease with the increase in the tendency among the people to do work remotely. Similarly, the cases of the suicide rate decrease with the increase in the benefits provided by the employers to the employees. The company which does not provide a benefit, or the employees are not aware of the policies tend to have more suicide tendency. Individual relationship was also established between various work-related factors and suicidal tendency. Since these factors present themselves as a combination in a work place, models were created to see the effect of all the factors on suicidal tendency. Among the models created, logistic regression and artificial neural network had most accuracy, however, logistic regression is the most accurate and most consistent model for predicting suicidal tendency among employees in a workplace.

REFERENCES

Analysis-of-suicide-rates-by-country- and-sex| Kaggle. (n.d.), from *https://www.kaggle.com/liadkeller/analysis-of-suicide-rates-by-country- and-sex/data*

Causes of Suicides | Kaggle. (n.d.), *from https://www.kaggle.com/josephgpinto/causes-of-suicides/data*

Chamberlain, J. (2013). Seasonal Variation of Suicide Rates within Alaska: Associations of Age and Sex.

Chang, S. S., Gunnell, D., Sterne, J. A., Lu, T. H., & Cheng, A. T. (2009). Was the economic crisis 1997–1998 responsible for rising suicide rates in East/Southeast Asia? A time–trend analysis for Japan, Hong Kong, South Korea, Taiwan, Singapore and Thailand. Social science & medicine, 68(7), 1322-1331.

Illness, L. O. (2016, November 03). Mental Health in Tech Survey. Retrieved from *https://www.kaggle.com/osmi/mental-health-in-tech-survey*

Kulish, N., & Eddy, M. (2015, March 31). Co-Pilot Treated for Tendencies Toward Suicide. New York Times, p. A1(L). Retrieved from *http://link.galegroup.com/apps/doc/A407634658/SCIC?u=mlin_c_clarkunv&sid=SCIC&xid= 50a2 a337*

Nakagawa, M., Kawanishi, C., Yamada, T., Iwamoto, Y., Sato, R., Hasegawa, H., ... & Hirayasu, Y. (2009). Characteristics of suicide attempters with family history of suicide attempt: a retrospective chart review. BMC psychiatry, 9(1), 32.

OSMI Mental Health in Tech Survey 2018. (n.d.). Retrieved from *https://osmi.typeform.com/to/xztgPT*

Reginald Harris,"Suicide in the workplace,"Monthly Labor Review,U.S. Bureau of Labor Statistics, December 2016, Retrieved from *https://doi.org/10.21916/mlr.2016.54*.

Retamal C. Pedro, & Humphreys Derek. (1998). Occurrence of suicide and seasonal variation. Revista de Saúde Pública, Vol 32, Iss 5, Pp 408-412 (1998), (5), 408. Retrieved from

http://goddard40.clarku.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=tr u e&db=edsdoj&AN=edsdoj.29906b0464d2437aa3d6f948c7c79cda&site=eds-live

Suicide. (n.d.). Retrieved from https://www.nimh.nih.gov/health/statistics/suicide.shtml

Suicide. (n.d.). Retrieved from https://ourworldindata.org/suicide#economic-recessions Suicide

Suicide Rate by Country 2019. (n.d.), from
http://worldpopulationreview.com/countries/suicide-rate-by-country/

Statistics. (n.d.). Retrieved from *https://afsp.org/about-suicide/suicide-statistics/*

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Shambhavi Pandey
Clark University
Shambhavipandey112@gmail.com
+1 774-243-4084

Mohammad Bilal Khan
Clark University
bilal707786@gmail.com
+1 508-667-7873