

# Analysis of Mental Health and Suicide Rates

Kshitij G Saha

PESIUG20CS222

CSE Department, PES University

kshitijsaha@gmail.com

Vibhav Deepak

PESIUG20CS496

CSE Department, PES University

vibhavdeepak28@gmail.com

Kanumuri Sri Charan

PESIUG20CS192

CSE Department, PES University

royalcharan2002@gmail.com

**Abstract**—In the present day, the world is witnessing an alarming upward trend in suicide rates. This high rate of suicide has a very evident correlation with the mental health of people at large. This paper aims to concretize the correlation of suicide rates with mental health parameters such as country, age, gender, and other parameters pertaining to an individual's mental health including access to mental health infrastructure and related parameters such as medical professionals, etc. We intend to do so by assessing potential trends using exploratory data analysis, meta-analysis, and other mechanisms to further ascertain the correlation between the two factors.

**Index Terms**—Mental health, suicide, Exploratory data analysis, data analytics, global suicide rate, mental health services

## I. INTRODUCTION

In the past decade, mental health has become an issue that has come to the forefront of global attention. The long-drawn Covid-19 pandemic further advertised an integral part that mental health plays in one's self and society at large. However, the elusive and intangible nature of the parameters pertaining to mental health makes it nearly impossible to have a quantifiable metric to map the quality of mental health in general. Access to adequate services and infrastructure pertaining to mental health is still remote in several parts of the globe which is deeply concerning considering the rising need for mental health professionals to tackle the global epidemic of mental illness. The alarming increase in rates of depression and related mental illnesses is a grave concern for health experts worldwide and needs to be dealt with. Mental health services and infrastructure play an irreplaceable role in tackling these rampant problems but they still remain scarce in developing countries with limited resources. Awareness about mental health and related illnesses remains notably low; hence, people continue dealing with anxiety, depression, and other associated problems without direction or guidance.

Unfortunately, mental health problems such as clinical depression, acute anxiety, and related issues often lead to an increased likelihood of suicide. Suicide is a profoundly disheartening global phenomenon that accounts for 1.4% of all deaths worldwide. According to the World Health Organization report of 2021[1], it has caused nearly 800,000 deaths. Suicide is never caused due to a singular reason or process of disease but is instead the outcome of several complex factors intertwined together. These factors include social, psychological, genetic, environmental, and biological characteristics and causes. The high rate of suicide however can be correlated with the

deterioration of mental health at large. Trends in mental health and suicide rates can be analyzed by studying parameters such as sex, age group, and geographical location as well as the provision of health services such as mental hospitals, and mental health professionals such as psychologists, psychiatrists, nurses, and others. Finding the underlying correlation between these factors and their impact on the overall suicide rate around the globe stands to be the aim of this paper. This paper aims to utilise techniques in data analytics including meta-analysis and Exploratory data analysis to study the trends in suicide rates with regard to the factors mentioned above. Analysing this correlation would elucidate the lack in the current mental health infrastructure and the most adversely affected groups in society who are most vulnerable to mental illnesses hence increasing the likelihood of suicide.

## II. REVIEW OF LITERATURE

Xue-Lei Fu et al [2] performed a systematic review and meta-analysis on suicide rates among people with mental illnesses by using the Stata software. They performed Freeman Turkey double arcsine transformation to stabilize variances and then performed a random-effects meta-analysis. The Q tests and I<sup>2</sup> statistics were used to assess heterogeneity across all studies and publication bias was assessed by visual inspection of funnel plots and Begg's Test. The pooled suicide rate among people with serious mental illness was 312.8 per 100 000 person-years (95 percent CI 230.3–406.8 per 100 000 person-years), with substantial heterogeneity (I<sup>2</sup> = 99.4 percent; p < 0.001) (Fig. 2). Visual inspection of the funnel plot symmetry and Begg's Test (p = 0.238) suggested an absence of publication bias (online Supplementary eFig. 1). Meta-regression regression analysis found publication year (p = 0.454), mean age (p = 0.870), or proportion of female (p = 0.687) was not significantly associated with the pooled suicide rate indicating generally favorable results and meaningful conclusions.

Shahid Alvi et al[3] conducted a comparative study of sociocultural factors and immigrant women's mental health by using descriptive, bivariate, and regression analysis. A total of 7 predictor variables (social support, victimization, patriarchal rights, collective efficacy, education, work status, and age) were assessed to test their impact on immigrant and non-immigrant women's mental health. A global mental health dependent variable was constructed using two separate survey questions about different forms of mental health;

namely, depression and perceived control and mastery. Before scale building, a reliability analysis was run on all mental health measures to assess Cronbach's alpha and ensure that the internal consistency of measures would be acceptable.

Peter E. Politser[4], et al evaluated the ability of repeated psychiatric screening tests (the General Health Questionnaire [GHQ]) to predict medical and psychiatric service use in a Health Maintenance Organization (HMO), the Harvard Community Health Plan (HCHP) using exploratory data analysis. In doing the same they used three-dimensional plots, scatterplots, residual scatterplots, and methods of transforming data such as computing absolute values, logarithms, etc.

Alessandro Crociata et al[5] performed an exploratory study on cultural access and mental health by performing regression analysis, and multicollinearity analysis and further tackled with the endogeneity problem by using a generalized method of moments (GMM) methodology. Their analysis concluded that continued access to certain forms of cultural experiences entailed a positive impact on mental health.

### III. METHODOLOGY

#### A. Dataset

The datasets used in this project are taken from kaggle and are available at the following links:

1.<https://www.kaggle.com/datasets/twinkle0705/mental-health-and-suicide-rates>

2.<https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016>

The first dataset contains 4 CSV files titled "Crude Suicide Rates", "Age Standardized Suicide rates", "Facilities", and "Human Resources". The first two files consist of data pertaining to suicide rates in different parts of the world bifurcated on the bases of countries, sex, and age groups. The third CSV file consists of country-specific statistics pertaining to medical facilities such as mental hospitals, health units etc. The last csv file of the first dataset dataset consists of details regarding the number of nurses, psychologists, psychiatrists, and social workers in different countries. The dataset contains 29 columns in total. The second dataset consists of the overview of the suicide rates fro 1985 to 2016 bifurcated in columns based on country, year, age group, suicide number, population, suicide rate per 100k population, HDI and GDP.

#### B. Pre Processing

The CSV file entitled "facilities" was first checked for outliers. It was found that the majority of values for the attributes "day treatment" and "residential facilities" were missing and hence it was concluded that these attributes would not have a major impact on our analysis. These attributes were dropped. Other missing values were replaced by their median values across. Outliers were found using boxplots and the box plots were then normalised. The CSV file entitled "human resources" was also checked for outliers and since the missing values were considerably lower, they were once again replaced by the median values of the respective attributes. Following

this, box plots were plotted once again to check for outliers. Once there were no remaining null values, the data was finally clean and ready for analysis. The other two files present in the dataset had no missing values or outliers and hence were already clean.

#### C. Exploratory Data analysis

To better understand the data and to analyze the underlying trends of the data and correlation between the various attributes, exploratory data analysis was performed on all three sections of the dataset. Figure 1. Shows a color map depicting the suicide rates of different countries around the world which gives a very clear insight on the countries as well as the geographical regions with the highest suicide rates.

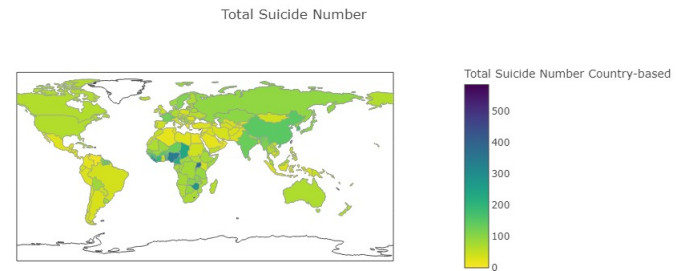


Figure 1

Figure 2 shows the suicide rates of different countries in the form of a bar graph to further visually classify countries based on their suicide rates.

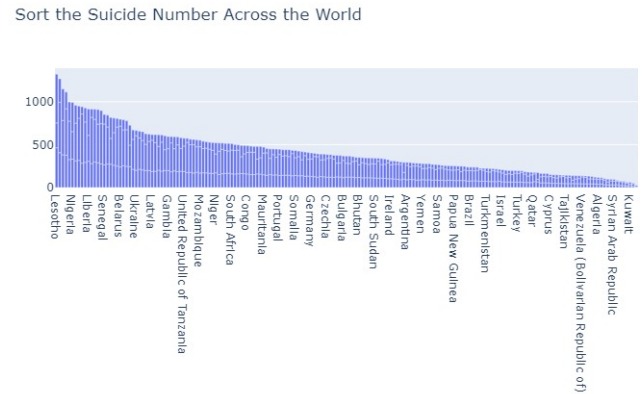


Figure 2

Figure 3 shows the rate of suicide with respect to gender ratio, i.e the suicide rate of men, women and both sexes combined represented as a bar graph. One bar graph is constructed based on the data of crude suicide rates and the other graph is age standardized.

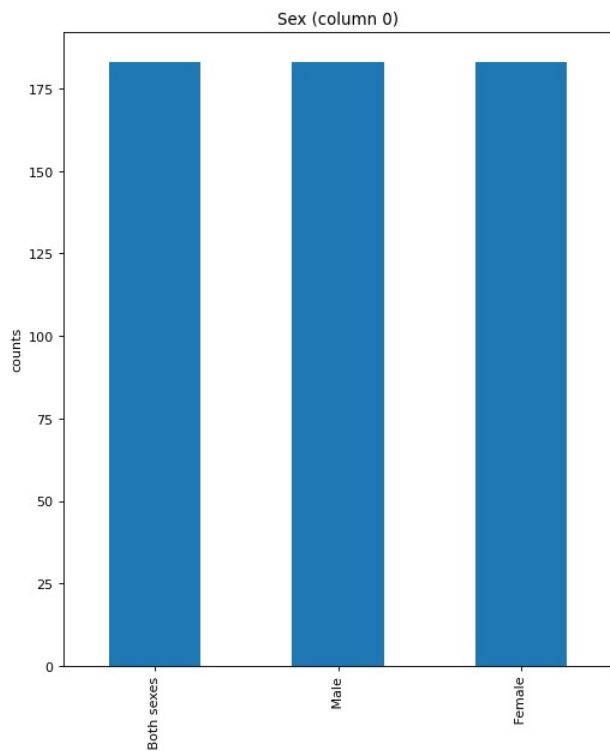


Figure 3

Figure 4 shows the scatter and density plots for crude suicide rates to analyse the pattern of suicide rates among various regions and age groups.

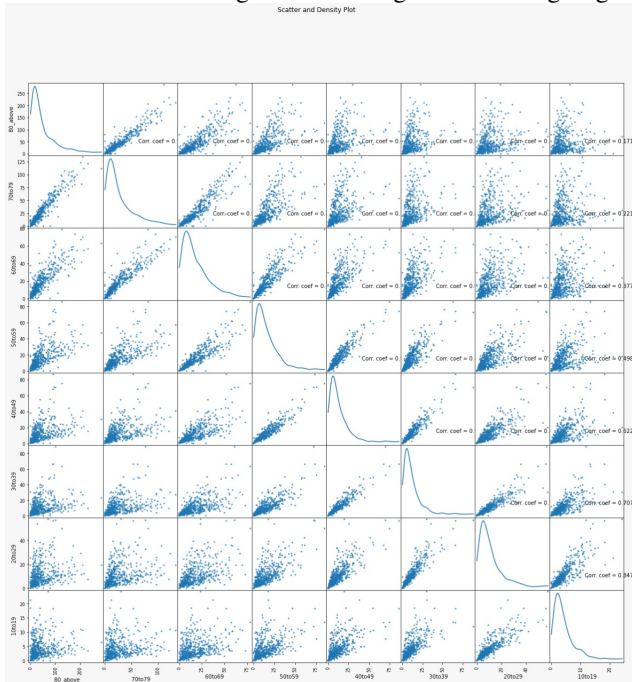


Figure 4

Figure 5 shows the scatter and density plots for age standardized suicide rates to analyse the pattern of suicide rates among various regions over the years.

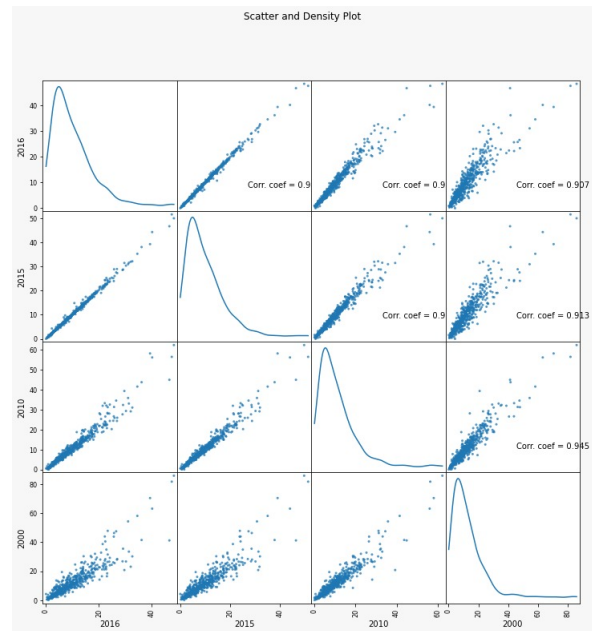


Figure 5

Figure 6 shows the correlation matrix for crude suicide rates and age standardized suicide rates.

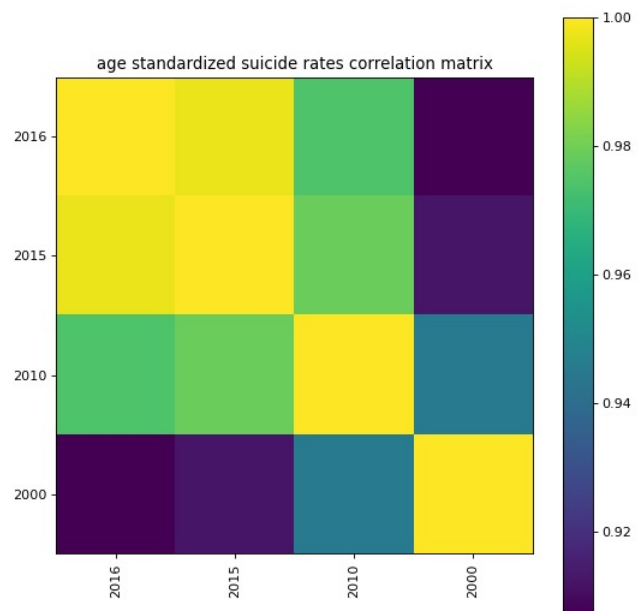


Figure 6.1

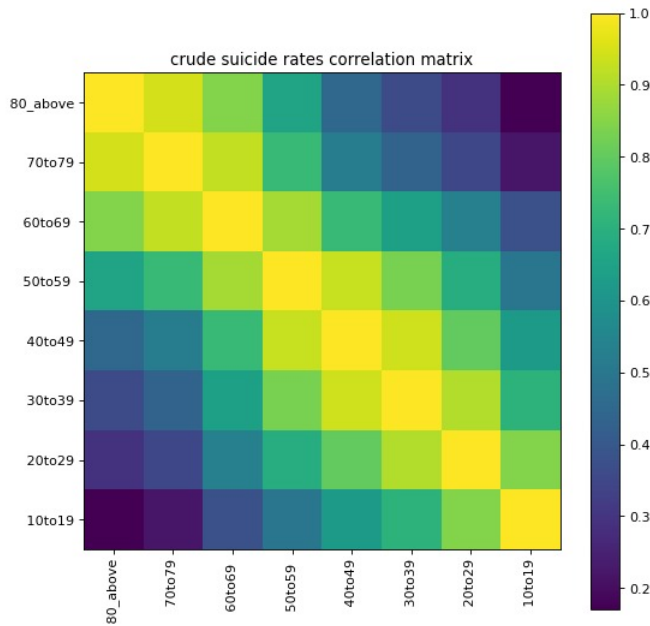


Figure 6.2

Figure 7 is a correlation matrix for the human resources different countries have indicating the correlations that different mental health service members have between each other.

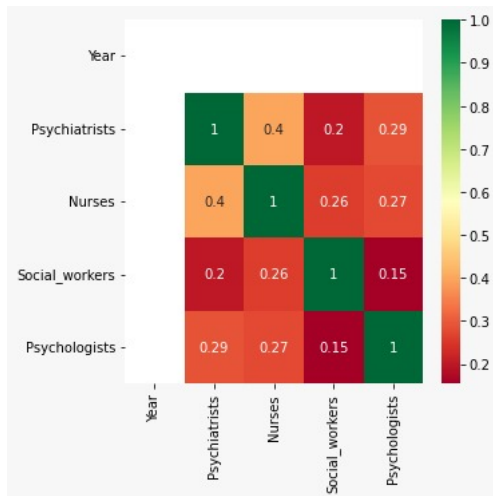


Figure 7

Figure 8 is a correlation matrix for the facilities different countries provide.

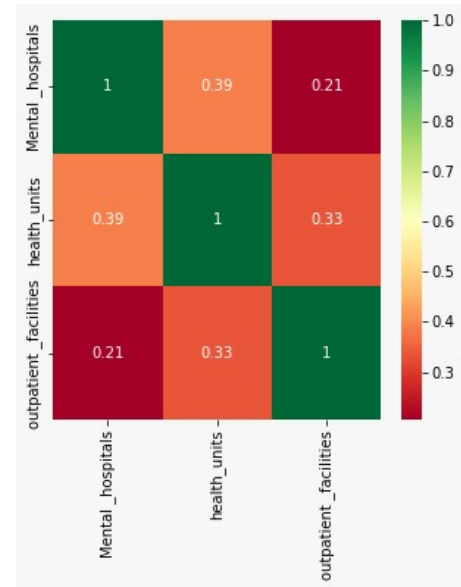


Figure 8

**Aggregate Suicide Rate:** Considering the complex nature of mental health itself and the fact that suicide in itself is caused due to an amalgamation of multiple factors, it is relatively difficult to quantify and measure and to further attempt to use the regular predictive models in the situation. In the case of analyzing trends in suicide rates and their correlation to mental health infrastructure etc, exploratory data analysis proves to be the ideal method to make inferences and conclusions.

Plotting the global suicide rates based on the age-standardized suicide rates dataset starting from the year 2000 to 2016 indicates a fall in the number of suicides over the past decade. This declining trend is witnessed throughout performing analysis on the dataset.

However, the reduction can also be attributed to “hidden suicides” wherein suicides go unreported. Underreporting of suicides might be indicative of the falling suicide rates is several developing countries that don’t have robust investigative mechanisms and this phenomenon leads to a drop in the global sum of suicide rates.

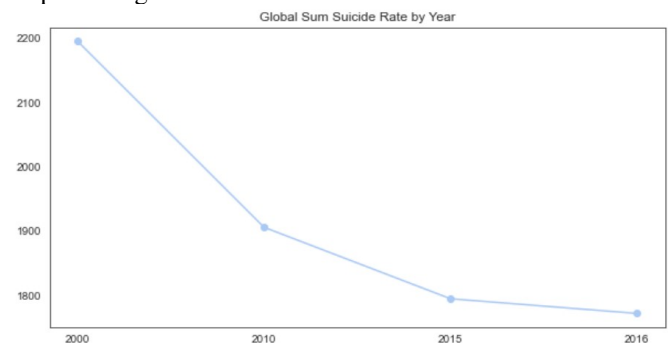


Figure 9

Further, by plotting the suicide rate per population of a hundred thousand globally, we witness the following trend that has been displayed in Figure 10.

Suicide Rate per 100k pop



Figure 10

To further elucidate the global geopolitical suicide rate we plotted the world map based on suicide rates to further observe these trends.

Worldwide Suicide rate in 2016

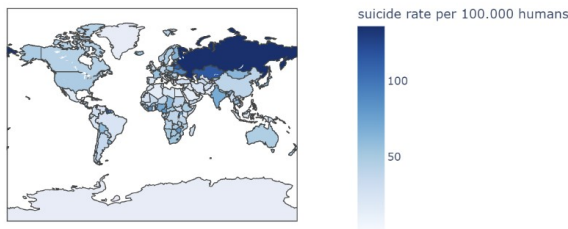
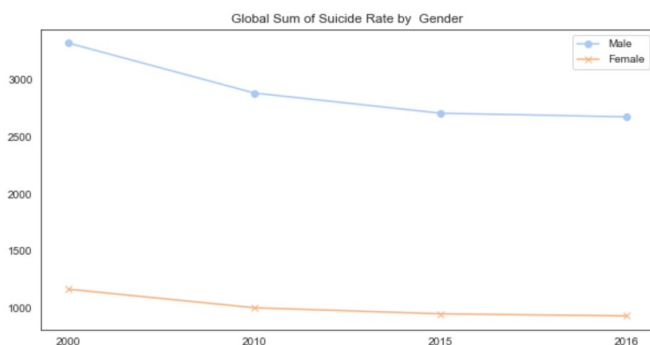


Figure 11

**Suicide Rates on the basis of Gender:** Similarly, plotting the suicide rates bifurcated on the basis of the gender indicated that the rate of suicides among males remained significantly higher than that of females. The cause of this disparity cannot be attributed to a singular factor. This could be due to various reasons including lack of data, lack of proper mental health infrastructure, or due to other psychological differences between males and females.



World Suicide by Gender (1985-2015)

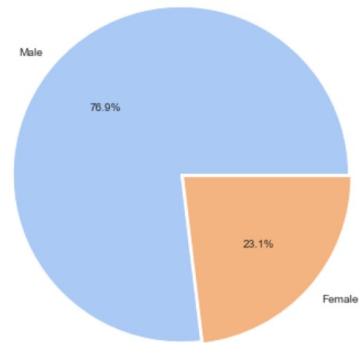
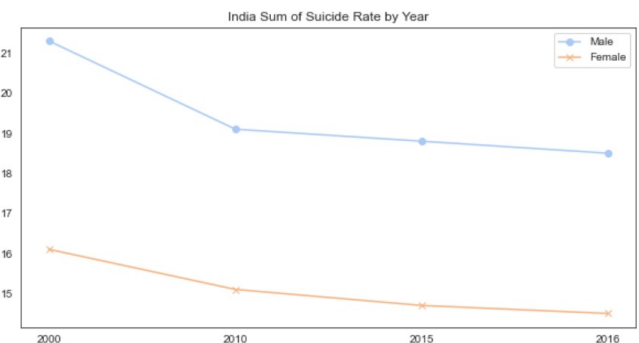


Figure 12

To further understand the trends in suicide rates in India, gender-separated suicide rates were plotted for India. The trends in India seemed to mirror the global trends, i.e male suicide rates were significantly higher in comparison with female suicide rates.



India Suicide Rate per 100k pop



Figure 13

Continuing on the study of male vs female suicide rates, we plotted the top ten countries with the highest mean suicide rates.

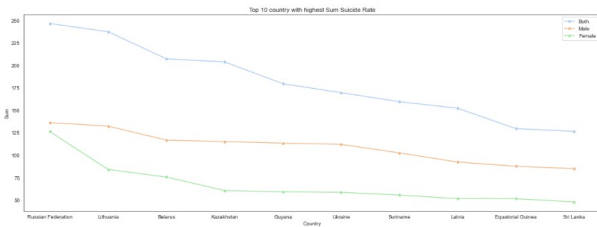


Figure 14

Figure 14 indicates, very interestingly, that several of the countries with the highest suicide rates lie in Northern Asia and Eastern Europe which might indicate geopolitical instability and other factors to drive up the suicide rates. Similar issues faced by these nations neighbouring each other might be the cause of the alarmingly high suicide rates.

Plotting suicide rates per hundred thousand population we obtain the following graph continuing to indicate the huge disparity between male and female suicide rates around the world.

Suicide rate per 100 pop by Gender

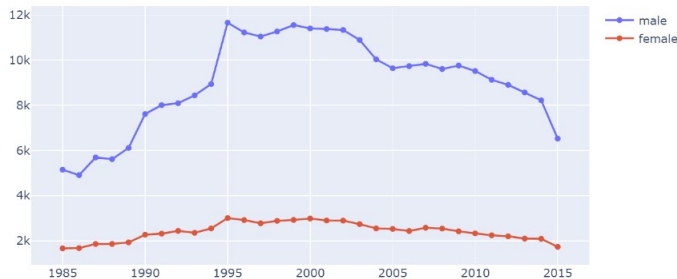


Figure 15

In Figure 16, using the “Age Standardized Suicide Rate” dataset, we plot a vertical bar graph to indicate the countries with the highest suicide rates of both males and females. The common trend again indicates that Asian countries recorded the highest suicide rates in the world which also coincides with the fact that mental health facilities, infrastructure as well as trained professionals are the least in numbers in the Asian continent.

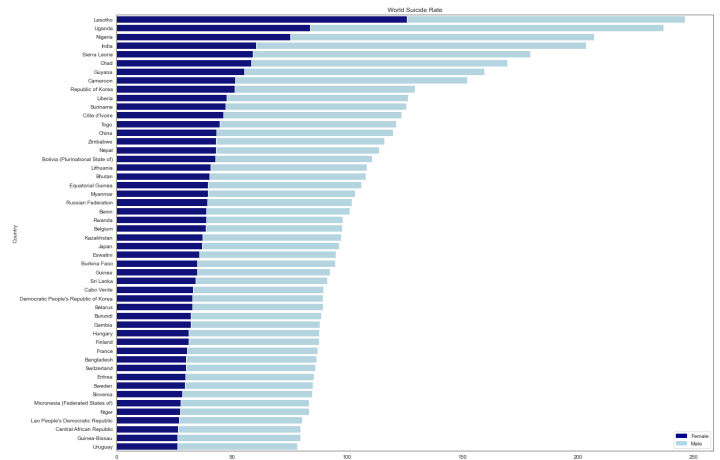
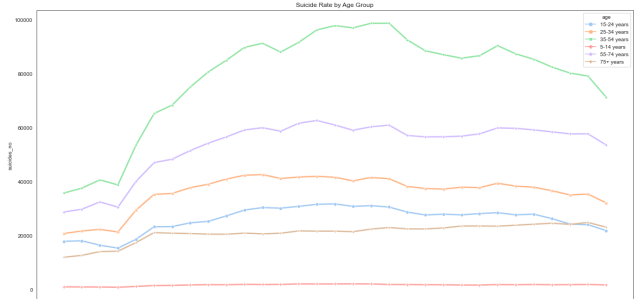


Figure 16

**Suicide rates based on Age, Groups and Generation:** Subsequently, we decide to adopt a different perspective on the problem at hand which is to analyse the trends in suicide rates. We chose the dataset documenting the “Overview of Suicide Rates from 1985 to 2016” and study the age range as well as generational trends in suicides. On plotting the suicide rates based on different age groups, we viewed the following results as displayed in figure 17.



World Suicide by Age (1985-2015)

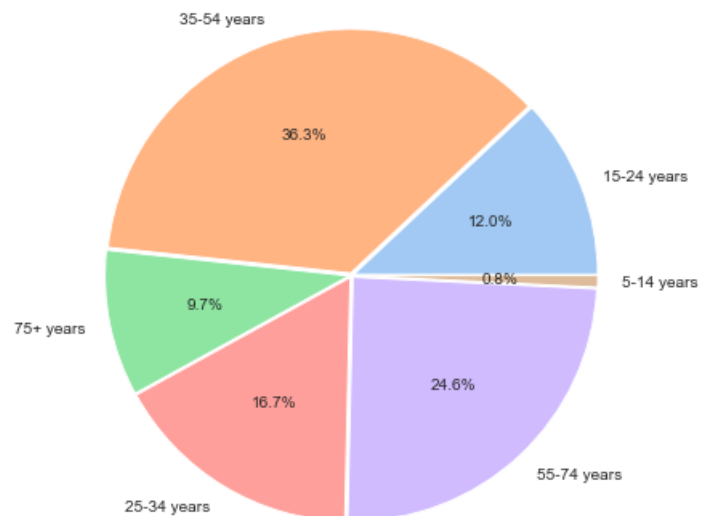




Figure 17

The graph indicated the disproportionately high rate of suicides between the age group of 35-54 years. This stark trend points to the clear indication that suicide remains highest in the working-class middle-aged group due to a multitude of factors including increased responsibilities in the family, the stress in the professional field, and several other attributes. Contrary to what one might believe, the suicide rates in the age group between 25-34 years remain comparatively lower. Another shocking revelation is the fact that the age group with the second highest suicide rates is the age group between 55-74 years. While the reasons for this trend are hard to deduce, this certainly busts several myths about suicide rates and their correlation to age itself. These trends can be further analysed to find out the underlying reasons that can be attributed to these disparities.

Next, we attempt to display the trends in suicides based on different generations i.e boomers, generation X, millennials, generation Z etc. Figure 18 indicates the graph based on generational trends in suicide rates.

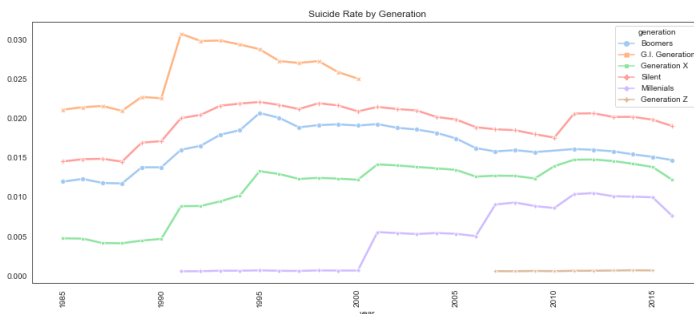


Figure 18

The highest rate of suicide was recorded in the G.I Generation also known as the Greatest Generation. The generation is generally defined as people born from 1901 to 1927. Shaped by the Great Depression and strengthened by the Second World war, this generation was dubbed the “greatest” because they were at the forefront of World War II and were said to have fought the war, not for recognition but because it was the right thing to do. The international destabilization caused due to the war as well as the massive economic turmoil due to the Great Depression were the two biggest causes of the high suicide rates.

Following the G.I generation is the Silent Generation which is generally defined as people born from 1928 to 1945. This generation follows the G.I generation and marks the period when the world was reeling from the aftereffects of the Second World War. This era also marked the Korean War and the Civil Rights movement. The Silent Generation was dubbed so majorly because they were considered to be the “Silent Majority” in most cases.

Millennial suicide rates indicate a climb upward since the year 2000 when they reach the middle-aged demographic and then display another sharp climb after 2005. The suicide rate

continues to increase following 2005, which drops after 2015. The suicide rate in Generation Z shows a steady, low trend however considering the overall trends displayed by other generations, a possible increase in the subsequent years can be a fairly reasonable deduction.

#### D. Modelling

We performed simple linear regression on the ‘age-standardized suicide rate’ dataset to further observe the metrics of the dataset and perform analytics. We melt the dataset and split the data into training and test sets to perform the same. On performing Simple Linear Regression on the dataset, we obtain the following metrics:

**MAE: 5.302658864579461**

**MSE: 52.41057017998232**

**RMSE: 7.239514498913744**

**r<sup>2</sup>: 0.159998886958364**

We observe that the r<sup>2</sup> score isn’t substantially big enough to extract any solid conclusions considering the complex nature of suicides in general. However, we observe the following indications: 1. Being male increases the suicide rate by 8.08 points. 2. For every year we go back from this dataset, the suicide rate goes up by 0.154. 3. The formula for calculating the suicide rate of a male can therefore be expressed as:

$$y = 8.01x - 0.154$$

#### CONCLUSION

After extensive modelling and bridging the gap between accuracy and precision, to predict and forecast suicides across the world in varying living conditions. The simple linear model we employed, forecasted with certainty and had a relatively low RMSE value, which is required for better predictions. Our inferences state clearly that being male is a major factor in determining the prediction of deaths by suicide since it increases the risk of suicide by 8.08 points. As the year’s progress, suicide rates are comparatively rising, and this is alarming. Being born in Gen Z, we must take steps to ensure longer lifespans for humans and reduce the mental stress and problems that come with our day-to-day lives.

#### REFERENCES

- [1] World Health Organization (2021). Suicide data. Retrieved from <https://www.who.int/teams/mental-health-and-substance-use/suicide-data>
- [2] Fu X-L, Qian Y, Jin X-H, Yu H-R, Wu H, Du L, Chen H-L, Shi Y-Q (2021). Suicide rates among people with serious mental illness: a systematic review and meta-analysis. *Psychological Medicine* 1–11. <https://doi.org/10.1017/S0033291721001549>
- [3] Shahid Alvi • Arshia Zaidi • Nawal Ammar • Lisa Culbert 2012 A Comparative and Exploratory Analysis of Socio-cultural Factors and Immigrant Women’s Mental Health within a Canadian Context, Springer Science+Business Media, LLC 2012
- [4] Peter E. Politser, Donald M. Berwick, Jane M. Murphy, Paula A. Goldman, and Milton C. Weinstein Uncovering Psychiatric Test Information With Graphical Techniques of Exploratory Data Analysis
- [5] Alessandro Crociata • Massimiliano Agovino • Pier Luigi Sacco Cultural Access and Mental Health: An Exploratory Study Springer Science+Business Media Dordrecht 2013