Factors affecting depression:an exploratory data analysis

Kirtana Nambiar
Data Analytics Engineering
George Mason University
Virginia, United States
nambiarkirtana@gmail.com

***Abstract*—Despite mental health disorders being a severe disease and affecting the entire globe, it still has a lot of stigma and human rights violation associated with it [1]. Also, CDC (Centers for Disease Control) states that a person's mental health can change overtime with proper care and treatment [3]. In this paper, I have attempted to understand depression and the various factors effecting it through exploratory data analysis using data.world mental health dataset collected from 1990 to 2019 across different countries of the world. Preliminary findings show how age, gender, level of education, employment status, income group of country according to world bank, etc. affect depression. This will help us identify the likelihood of a person developing depression if they fall under the category of high risk. This analysis could help us look for niche areas that drastically affect depression and hence create awareness and act proactively to fight depression and disabilities associated with it.**

I. INTRODUCTION

Mental health is as important as physical health and helps us attain the right balance in our day-to-day activities. If unattended to mental health, it could lead to disability just as any physical issue, rendering us incapable to lead a healthy life. According to WHO (World Health Organization), suicide is the fourth leading cause of death among 15-to-29 year of age. People suffering from mental health disorders die prematurely at least two decades prior.

Around 3.8% of the total world population have depression. This is as alarming as any deadly disease like cancer and if not researched and studied thoroughly can lead to a major problem around the globe [2].

The lack of awareness and stigma associated with it makes it more difficult to prevent and treat a perfectly curable disease. Through analysis using data exploration techniques like histograms, bar plots and by creating regression models to understand the relation between response variable i.e., depression value and predictor variables i.e., age, sex, etc., I have attempted to get understanding of external factors on mental health of an individual.

II. LITERATURE SURVEY

A. Gender differences in mental health [6]:

This paper discusses how gender plays a role in mental health differences and emphasizes how there should be different gender specific treatments and gives some recommendations. A gender-based approach to mental health is important as it helps to identify gender inequalities which promote mental health issues and thereby curate a custom solution to this problem. The paper also talks of social stigma of women being more prone to emotional breakdowns and men seeking alcohol, restricting people from seeking medical help and reinforce such stereotypes. This leads to lower rate of cases being reported and reduces the accurate identification and treatment of such psychological problems.

The author followed an approach of collecting and reviewing 15 years of articles (1991-April 2006) related to gender difference in mental health; to formulate this paper. Some facts mentioned in the paper after thorough review were, in countries like Alexandria and Egypt, depressive symptoms in girls were double than boys. In Oman, however there were no differences. But adolescent boys showed higher depression and anxiety levels than adolescent girls. There was no gender difference in case of anti-depressants medication and the response to it.

Certain factors like women from poor background find themselves without access to healthcare more often than men with same economic status, the health services majorly focus on reproductive functions and neglects the other age group, a lack of female medical

personal sometimes becomes a barrier, etc. further discouraged women from getting treated for mental disorders. In developing countries like India, rural women have accepted higher levels of suffering by gender inequality in terms of fertility, marriage, work norms, domestic violence, and poor psychological health and do not seek treatment.

Men respond to stress by expressing anger or substance abuse as they have been socialized to express and women on the other hand exhibit dysphoria. The paper further states that mental issues are closely associated to gender roles, as each gender interacts with social determinants.

In conclusion, the paper emphasizes the role of gender alone in mental health issues and relates the gender related stress to socially dictated norms. The author further states that women's status and life opportunity are low worldwide and hence their lack of security in terms of income, sense of self-worth, confidence and psychological and physical well-being contributes majorly to women being susceptible to higher mental disorders than men. The author attributes gender inequality as the main factor for mental issues being higher in women as compared to men. Based on this conclusion, the paper also goes on to recommend treatments for mental health disorders to be gender specific and not gender neutral.

Limitations:

However, socio-economic factors like literacy, income, age, disease lineage, etc. should also have been considered along with gender factor, as these factors could also interact with the gender factor in arriving at the conclusion of mental health of a person.

Another limitation is that the author chose a non-systematic narrative approach based on the author's selection of articles, for the basis of this paper. The author's selection could be subjective and hence the results could be skewed.

B. *Education and income: which is more important for mental health?[7]*

This paper states that there is huge gap between our knowledge of social position indicators and socio-economic factors that are health sensitive. Therefore, to understand this relation, both rich and poor countries socio-economic factors and mental health, needs to be studied upon. The paper, therefore, considers Santiago, the capital of Chile as it's sample population, to understand the prevalence of common mental disorders (CMD).

The approach followed was a three-stage clustered design, where data is collected by survey from 1996-1998. 200 household out of 35 boroughs were considered randomly selected with a probability proportional to population size. A larger sample was considered amongst affluent boroughs. Revised Clinical Interview Schedule (CIS-R), a standardized set of questions were used for survey. A person with score greater than 12 or above, were regarded as suffering from CMD. The study also included socio-economic variables like education level I.e., primary, secondary, or higher, monthly per capita income and if there was sudden decrease in income, quality of the house, tenure I.e., whether rented or owned, crowding I.e., number of bedrooms divided by people in the house and lastly occupation status based on Chilean Institute of National Statistics scale to categorize into low, moderate, and high income of the breadwinner.

The association between CMD and each socio-economic variable was examined using logistic regression models, both before and after adjusting simultaneously for sex, age, working status, physical illness, and social support and other factors considered above. They also did outlier analysis by removing households with elderly people alone.

The results after analysis were less educated, lower per capita income, income decrease, low ranked occupation of the main breadwinner, poorer housing quality, and overcrowding all attributed to higher probability of having common mental disorders.

In conclusion, the paper states that in both developed and developing countries, common mental disorders were most prevalent in socially disadvantaged groups. However, other studies in Europe and America, showed no association between income and CMD. The paper narrates that they found an inverse relation between education and CMD.

Limitations:

The sample collected was from middle income nation. However, if same study is applied to other developed countries, the conclusion would not be the same as there are huge socio-economic, cultural differences. Hence, the study should have considered samples representing poor and rich countries.

C. *Sex differences in the effect of education on depression: Resource multiplication or resource substitution? [8]*

The paper states the resource substitution hypothesis where education improves economic well-being for women because socio-economic factors make them depend on education for their well-being. Resource multiplication hypothesis states men benefit from education as they provide higher income and authority.

The paper uses the 1995 survey of Aging, Status, and the Sense of Control (ASOC) and follow up interviews in 1996 and 2000. It is a national telephone probability sample of US households who are randomly selected from pre-screened sample to avoid business or unemployed people in the household. The survey included only English-speaking adults. It has two sub samples, one with age above 60 and the other with all other age group.

The regression model analyzes relation between depression and education and sex. The other models establish a relation between each other variable.

Limitations:

The sample population is selected randomly, and the set of questions used in the interview is not standardized and hence could skew the results.

III. METHODS

We have used dataset from data world of mental health and work bank data as primary sources. Python and R languages are used for statistical analysis and visualization. Multilinear regression model and random forests models have been used to fit the data and understand the relationship between continuous and categorical variables. Skewness and model diagnostic plots have been used further to analyze and transform data.

IV. RESEARCH QUESTIONS:

- 1. How level of education and employment status of a person can influence depression factor?
- 2. Is there a pattern between factors like age, gender of a person and mental disorders?
- 3. Which country has the most mental disorders reported and what could be the factors causing it e.g., Low-income countries have very less treatment facilities

V. EXLORATORY DATA ANALYSIS AND RESULTS

- On analysis using histograms and skewness value, the data for prevalence of depression was found to be right skewed and a log transformation seemed to correct the skewness. Post that the data was checked for missing values and was replaced using a left join with mental health dataset. After the data preprocessing, the data was ready to be visualized using matplotlib.

Below is a graph of prevalence of depression against level of education for both genders.

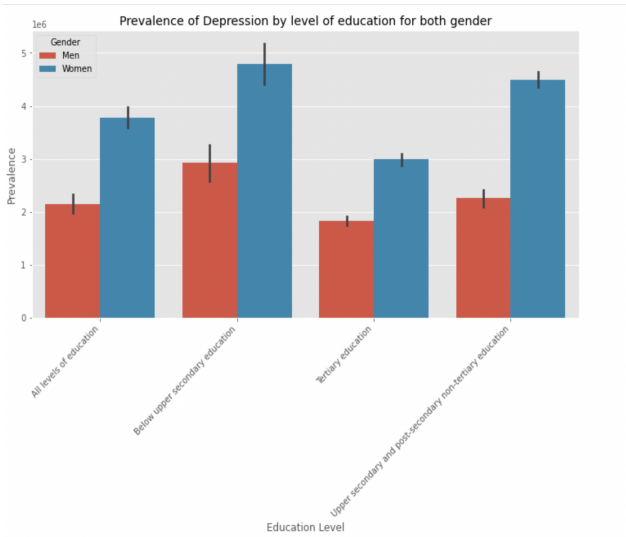

Fig. 1. Prevalence of depression against level of education for both genders.

People with tertiary level of education exhibited the lowest depression percentage.

Tertiary is the highest level of education among the other two i.e. below upper-secondary and upper-secondary. Also, higher the level of education a person acquired, lower values of depression were observed from the graph. This indicates level of education and depression rates are inversely proportional to each other. This could be because

education provides employment, fixed income, economic independence and this in turn boosts self-esteem leading to better mental health. Also, the graph shows women have higher rate of depression than men.

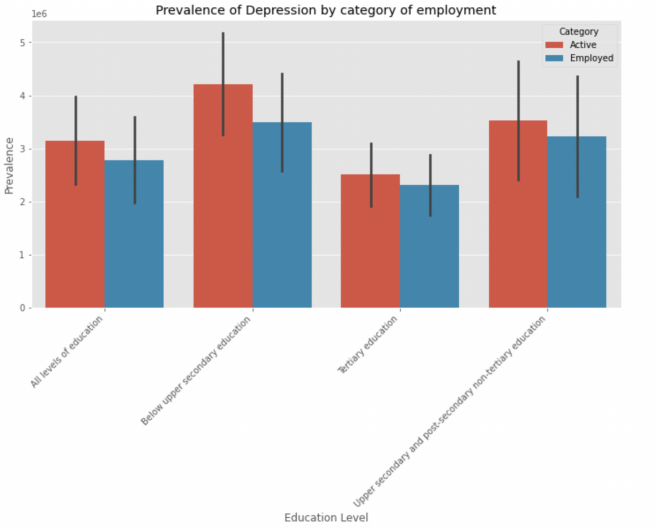

Fig. 2. Prevalence of depression against level of education for employment status.

The above is a graph of prevalence of depression against level of education for both employment status. The graph shows in each level of education, the employed group has the least factor of depression amongst actively seeking employment and employed people in that level. This indicates the employment is one factor in determining depression levels in an individual.

On further fitting a random forest model for feature selection to understand the important features, it was found education level and gender have the most significant influence on depression level. The model was tested by splitting the data into testing and training set and got an accuracy of 86.45%. Below is the tree constructed using random forest.

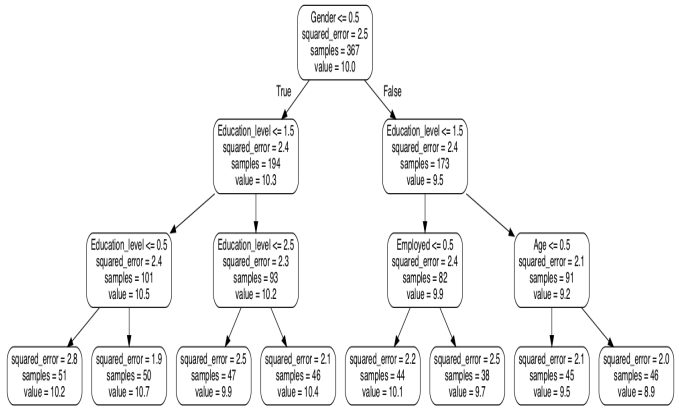

Fig. 3. Decision Tree.

The multi-linear regression model defined that education level is inversely proportional to depression level in a person.

- For answering the second research question, we consolidated data from different age groups across 1990-2019 and preprocessed the data to get in correct format and took log transformation to correct skewness of data.

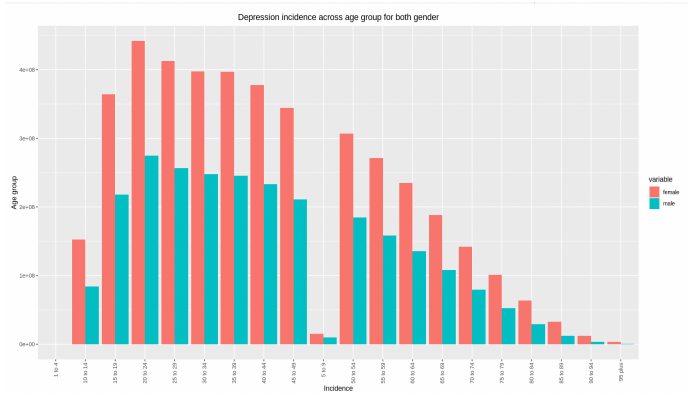

Fig. 4. Incidence of depression across age group for both genders.

The above is a graph of incidence of depression across age group for both genders. From the graph, it can be inferred that females have higher depression rates than male in any age group. Secondly, the age group from 20 to 49 exhibit the highest peak in depression in both men and women. This could be because of several factors; however, the dataset does not include environmental factors like work pressure or financial issues as this age group mostly covers the working group population. After fitting a linear

model on log values, the accuracy of the model improved from 57.71% to 96.81%.

- The third research question was to understand why certain countries have higher depression rates as compared to others and if so does income group have an effect on it. For this we joined data from world bank and mental health data. The below bar graph shows Prevalence of Depression across different income groups

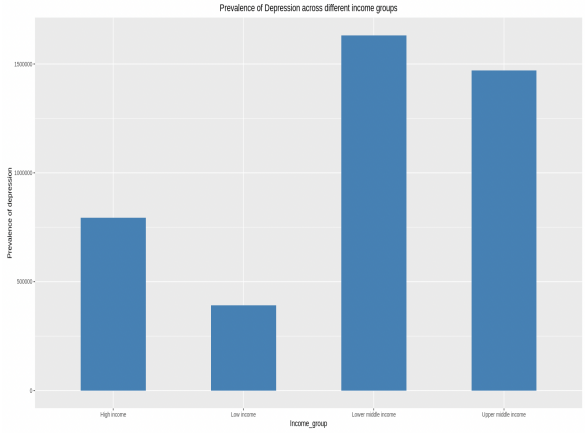

Fig. 5. Prevalence of depression across age group with different income groups.

The lower middle and upper middle group countries exhibited the highest rate of depression while the lowest income exhibited the lowest value for depression. To understand further, we plotted different world regions.

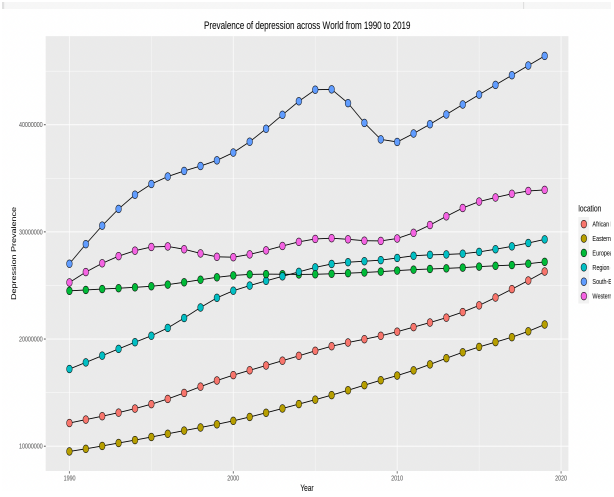

Fig. 6. Prevalence of depression across different world regions.

This showed the South-East Asian regions to have maximum rates of depression which includes nations like India, China, etc. One reason could be these are developing nations and financial stability is yet to attend. However, with this logic the third-world countries should exhibit high depression rates too, which isn't the case. The low rates of depression in low-income countries be because of scarce or no testing facilities, etc.

When plotted country wise against depression value as below:

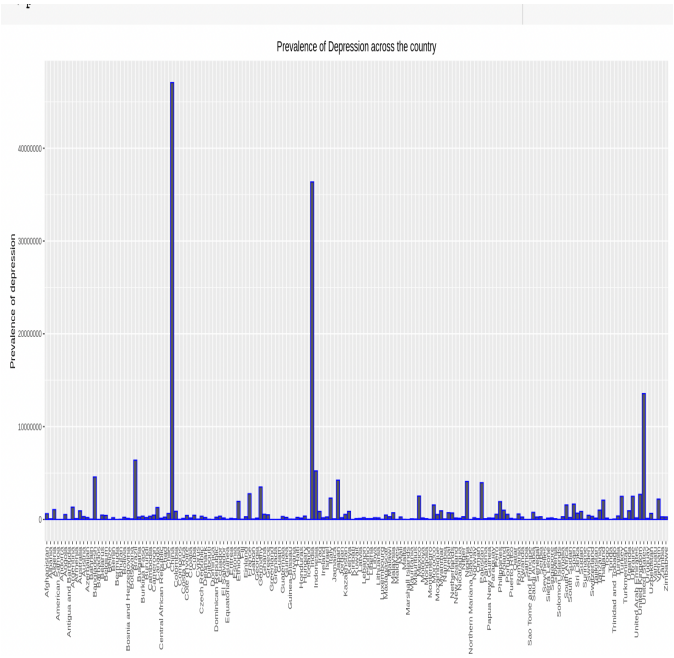

Fig. 7. Prevalence of depression across different countries of the world.

Entity	Income.group	prevalence
<chr>	<chr>	<dbl>
China	Upper middle income	47066531
India	Lower middle income	36364488
United States	High income	13557739

Table. 1. Top three countries with depression

We found China, India, and United States to have highest depression rates which is in accordance with the above two graphs. But as we can see, these countries belong to upper middle, lower middle and high-income groups respectively as per world bank. Hence, it is difficult to arrive at a conclusion why certain countries have low or high depression rates based on income groups alone. It could be several factors like socio-economic factors, work culture, financial condition of the country, education pressure on students, lesser facilities, etc. Hence, this dataset is insufficient to understand depression levels with country or regions of the world.

VI. DISCUSSION, IMPLICATIONS, AND LIMITATIONS

The preliminary data analysis of depression related dataset helped identify certain important factors like female population are at a higher risk of having depression than men. The age group of 20 to 49 are likely to develop depression and if female then the risk is higher. However, in both cases we do not know the reason which could be attributed to external factors which is outside the scope of the dataset. Also, level of education has a huge impact on mental well-being as it brings social and financial security. Higher the level of education, lesser the depression rates and employed population exhibited less rates of depression too. This further needs to be analyzed as to what is responsible for it, could be psychological factors, financial independence, etc.

When it comes to countries or regions of the world, the relation between them and income group and depression isn't very clear. This clearly indicates that there are other factors to be considered and the dataset is insufficient to answer that question.

VII. CONCLUSION AND FUTURE SCOPE

In this study, we tried to understand the impact of external factors on mental health and what a huge crisis "Depression" is, as it continues to spread like a pandemic. The emphasis of this analysis was to create awareness and understand how impactful certain mundane factors like age and gender can be on a person's well-being. Lack of knowledge and stigma associated with depression is deep rooted, making this fight difficult. However, with adequate research and data analysis on data available today, this struggle can be easily overcome.

Certain factors like why female population is more prone or the age group is more susceptible is something that needs to be further analyzed to solve this problem. It could be biological, socio-economic factors like wage gap or inequality, working-class related issues like toxic work culture, financial instability etc. to name a few. Countries with fewer or poor resources could be reporting less and hence lower income group countries are exhibiting lesser depression rates. However, it is very difficult to conclude with certainty without considering external factors. Hence, this study includes an extensive scope to explore with more features to be included in the dataset.

REFERENCES

[1] Dattani, S., H. Ritchie and M. Roser, "Mental Health," Our World in Data, 2018. https://ourworldindata.org/mental-health#depression.

[2] World Health Organization, "Depression," World Health Organization, Sep.13,2021.https://www.who.int/news-room/fact-sheets/detail/depression.

[3] Centers for Disease Control and Prevention, "Learn about mental health," *Centers for Disease Control and Prevention*, 2018. https://www.cdc.gov/mentalhealth/learn/index.htm.

[4] Lobo, "Mental Health Around The World, Mental Illness In Other Countries," Synergy Health Programs, Mar. 31, 2020. https://synergyhealthprograms.com/a-look-at-mental-health-around-the-world/.

[5] "data.world," *data.world*.https://data.world/vizzup/mental-health-depression-disorder-data/workspace (accessed Dec. 04, 2021).

[6] "Gender differences in mental health status and women's mental health in China," Reproductive Health Matters, vol. 3, no. 6, p. 170, Nov. 1995, doi: 10.1016/0968-8080(95)90192-2.

[7] R. Araya, "Education and income: which is more important for mental health?," Journal of Epidemiology & Community Health, vol. 57, no. 7, pp. 501–505, Jul. 2003, doi: 10.1136/jech.57.7.501.

[8] C. E. Ross and J. Mirowsky, "Sex differences in the effect of education on depression: Resource multiplication or resource substitution?," Social Science & Medicine, vol. 63, no. 5, pp. 1400–1413, Sep. 2006, doi: 10.1016/j.socscimed.2006.03.013.

[9] The World Bank, "WDI - The World by Income and Region," *datatopics.worldbank.org*,2020. https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.

[10] Global Health Data Exchange, "GBD Results Tool | GHDx," Healthdata.org, 2019. http://ghdx.healthdata.org/gbd-results-tool.