Health Data Visualization

Chirdroop Sagar
Department of Applied Data Science
San José State University
San José, United States
chidroop.sagar@sjsu.edu

Deepali Zutshi
Department of Applied Data Science
San José State University
San José, United States
deepali.zutshi@sjsu.edu

Himant Goyal
Department of Applied Data Science
San José State University
San José, United States
himant.goyal@sjsu.edu

***Abstract*—The collection and analysis of data in healthcare plays a vital role in correlation of symptoms, forecasting, identification of trends, and improving clinical care and patient outcomes. Apart from the physical health of a person, their mental health is a key factor that impacts a person's life. The aim of this project is to use Python and Tableau to analyze and visualize mental health trends across the world and find factors that have a major impact on people's health, both physical and mental.**

Keywords—Mental Health, Physical Health, Analyze, Visualize, Python, Tableau.

I. Project Background

A major part of providing healthcare to people involves analyzing data and drawing meaningful insights from it to assess the factors that impact people and how the care provided to them can be improved. A person's quality of life is deeply impacted by their physical and mental well-being. While physical health includes factors such as the nutrition, diet, activity level, physical ailments, and sleep cycle, the mental health of a person includes their psychological, social and emotional well-being. It impacts the ability of a person to feel, think, act, and make choices. It may be affected by anything from biological factors to life experiences and even the abuse of alcohol and drugs. With increasing changes in our environment, there has been a steady increase in the number of people experiencing mental illnesses and it is important to understand the factors that are causing this, how these trends have changed over time and what can be done to help people battle mental illness and lead better and happier lives. Mental health disorders are currently not addressed as effectively as they should be by the health system and there is a great imbalance between the needs of the patients and the kind of treatment they are offered. According to Mental Health America, a non-profit community-based organization, 2.5 billion people suffer from mental health disorders world-wide. The most prominent disorders observed are depression, anxiety, and alcohol and drug use. In the United States alone 50 million people suffer from some kind of mental illness of which 42.5 million people have anxiety and 21 million people have depression. A problem this serious must be addressed and analyzed to find solutions. With this project, we aim to analyze mental health data from different sources, draw meaningful conclusions about what affects people's mental health, the kind of disorders they face and what kind of treatment options are available to them. We then make meaningful suggestions that can lead to improvement in treatment and access to mental health resources.

II. Motivation

The outburst of the global pandemic affected the entire world in terms of people's livelihood, their physical and mental health. It also had impacted social, economical and political aspects of their life. The healthcare industry has generated tons of data over the last decades, and this project focuses on drawing inferential insights by performing a quantitative and qualitative analysis of the data. To better understand the agents that play key roles in the mental well-being of people and the cognitive effect they have on people's lives, it is imperative to study and analyze these issues deeper to come up with innovative ways to combat such problems.

III. Objectives

Objective is to study the factors affecting the mental and physical health of people, various disorders and their impact, Wellness and health care programs available to patients, health care facilities and their reach across different geographies. The paper aims to create awareness among individuals about the various disorders, their causes, and ways to improve personal life. Papers cover various types of healthcare facilities and how they are funded by federal and state governments.

IV. Literature Review

As part of the background research for this project, works discussing the factors impacting mental health and their correlation to the physical health of a person were studied and analyzed. The correlation between the mental well-being of a person and their professional working environment is evaluated. The author carries out the study in developing countries and draws important insights. Talking about workplace conditions, the author states that most research in mental health has been based in developed countries and is focused on fair employment conditions. There is substantially more prevalence of Common Mental Diseases (CMD) in high stress working conditions. The author goes on to explore the impact of globalization on more occurrences of CMD and says it has led to exploitation of employees with employees being paid much less in one part of the world compared to the other. Globalization has led to a steep increase in the stress levels of employees in developing nations and is characterized by a demand-control model where there is significant difference between the level of demand from a job and the amount of control an employee has over it. This leads to job-strain and exponentially increases the likelihood of developing CMD. The author concludes that there is a direct correlation between

the mental health of a person and the kind of environment they work in and until serious interventions are not taken to improve these conditions, there will be a steady increase in the number of people impacted by CMD.[1]

Another research paper studies mental health problems and available treatment options for college students. People attending college usually exacerbate their disorders that started earlier in their lives. Among all the disorders, anxiety seems to be the most common with 11.9% students suffering from anxiety, 7%-11% suffering from depression, 6.7% with suicidal ideation, q.6% with a suicide plan, and 0.5% actually making an attempt at suicide. Other than these, eating disorders and drug and alcohol abuse play a key role in the metal well-being of students. As part of the treatment options, evidence based practice (EBP) seems like the best course of treatment. Motivational Interviewing (MI) in combination with EBP has shown good outcomes with patients continuing treatments instead of dropping it completelyt. Insurance plays a key role in receiving good mental care and those who don't have access to it can actually make use of their parents' insurance upto 26 years of age. Colleges receive help from service providers and these are a major source of treatment options to students and must be made available to them to improve patient outcomes.[2]

Another study divided mental health into three categories based on LEP (Limited English Proficiency): 1) speaking a language other than English at home and reporting speaking English "not well or at all" (no English), 2) speaking a language other than English at home and reporting speaking English "well or very well" (bilingual), and 3) speaking only English at home (English only). They conducted their research using the CHIS (California Health Interview Survey). Adults aged 18 to 64 were included (n = 41,984). This research explains 56 % of Asian/PI respondents who could speak English needed mental health assistance using bivariate analyses. On the other hand, they discovered that just 11% of Asian/PIs who could not communicate in English need mental health services. They confirmed that the same population had 85 percent lower likelihood of receiving treatment than those who could not speak English using Multivariate analysis.[5]

In another study by Bildt and Michelsen eximined how different genders' mental health was affected by their working conditions. They discovered that women with occupations such as no education at the expense of the employer, job strain, high alcohol consumption, low occupational pride, and poor mental health. Men with poor mental health were identified as those who worked in jobs with low occupational pride and shift work. Women with poor mental health and non-occupational factors such as physical inactivity, poor social contact quality, insufficient coping strategies, demanding life events, and a high perceived physical load outside of work. It included things like poor coping strategies and smoking for men. In multivariate analyses involving non-occupational and occupational factors, they discovered that occupational factors were more significant for women. According to the findings, 1.6 percent of men in the survey required inpatient treatment due to psychiatric diagnoses. 3.9 percent of women had the same problem. Alcoholism affected 16.7% of men and 7.4% of women. They concluded that the male gender consumed more alcohol than the female gender.[6]

V. Methodology

A. Dataset

This project has been developed using a combination of 3 datasets. The "Mental Health in Tech Data", containing 1200 instances and 27 columns, is a survey dataset that records the factors affecting the mental health of people working in the tech industry. For every response, the survey includes factors such as the size of the company, whether the person has physical ailments, the kind of leave policy offered by the employer, if they had an option to work remotely, etc.

The second dataset "National Mental Health Services Survey" by the U.S. Department of Human and Health Services is a datest documenting the different kinds of treatment facilities in the United States and contains 12275 rows and 383 columns. The survey includes data on the characteristics of the facilities and the kind of demographic they serve.

The third dataset "Mental Health Depression Disorder Data" records the trends of mental health disorders like anxiety, depression, substance abuse, family history, etc in different countries from 1990 to 2007. It includes the percentage of the population affected by different disorders.

B. Data Cleaning & Preprocessing

This project uses three datasets to provide visual insights into the reasons behind mental and physical health and what are the different types of disorders the people suffer from around the world as well as examine the treatment operations for the same in their local region. The first Dataset we used was the survey dataset as mentioned in the above section. This dataset consisted of features like 'Age', 'Gender', 'Country', 'State', 'self_employed', 'family_history', 'comments' etc. a total of 1259 rows and 26 columns. Most of the data types were Object / String and only Age was of numerical type. The data was initially explored in Microsoft excel and sorted some of the values and filtered some of the attributes according to the project needs. The dataset was uploaded using Jupiter Notebooks and we noticed that the dataset had null values in certain columns such as 'self_employed', 'state', 'work_interfer', and 'comments'. With the help of a heat map, we looked further into the missing values of the same features. The 'Comments' column was dropped and not used for our analysis as it had around 86% of data in that column missing. The 'state' column had around 40% of it's data missing, we filtered the data based on countries in order to provide statistics. The 'age' and 'gender' columns were filtered via tableau Calculations in order to the desired age range. The 'work_interfere' column had 20% missing values and with the help of the filters options we were able to filter columns such as 'work_interfere' to the desired values and thus eliminating the missing values.

The second dataset we used was the 'Mental Health Depression Disorder Data' which provides information on the

type of disorders and how their distribution is around the world from the years 1990 to 2007. It consisted of features such as 'Code', 'Year', 'Schizophrenia (%)', 'Bipolar disorder (%)', 'Eating disorders (%)', 'Anxiety disorders (%)', 'Drug use disorders (%)', 'Depression (%)', 'Alcohol use disorders (%)'. Only the 'Code' had 15.14% of missing values. Since the 'code' column was the same as the 'country' column we dropped that column. Plotted a correlation matrix in order to find the linearity between the different features. We noticed that the 'Bipolar_disorder' and 'Eating_disorder' had high correlation of 0.71 and the same was observed between 'Eating_disorder' and 'Anxiety_disorders' which was very helpful for our analyses in Tableau. With the help of the filter options in Tableau, we were able to filter the top 10 countries for comparison between various disorders. Created calculated fields to obtain the count and analyses between various disorders in different countries and further filtered it to obtain only the top countries affected by the disorders. The dataset initially had data for the years 1990 to 2007, we later filtered years accordingly in order to avoid high variance when calculating cumulative years.

The third Dataset we used was the 'National Mental Health Services Survey' which consisted of 12275 rows and 38columns. This dataset describes the various treatment facilities in the United States that offer mental and physical health care options. This dataset had no missing values which were confirmed by plotting a heat map and finding the percentage of missing values. However, we had to perform data transformation as most of the data had categorical features. Some features have numbers in the range 0 to 2 that were converted to yes, no, and other / none with the help of Tableau parameters and filters. The handbook provided by the U.S. Department of Human and Health Services gave us insights on what the range of numbers signified, each number had a unique feature for example 0 was 'no' and 1 was 'yes', and -1 was 'other/None' was filtered with the help of Tableau. We further filtered the types of treatment for various mental and physical health patients with respect to the top 10 states in the United States offering the treatments for that particular.

VI. Discussion

The data analyzed and evaluated in this study points to a direct correlation between the state of mental faculties of a person and the environment around them. The findings echo those of the reviewed literature and back them up with concrete evidence. Analysis of the survey data of the people working in the tech industry reveals that a majority of people suffered from some or the other mental illness during their life with over 50% of them reporting onset between the ages of 25-30. The data also revealed that the work environment affected not only the their mental health but also their physical health. Employer policies regarding the ease of taking leaves, the option to work remotely play a key role in the development of mental illnesses and their ability to contribute at work. Further analysis also revealed that men and women were disproportionately affected, with more cases of mental illness in men than in women. Research also revealed that drug use and eating disorders are very common mental disorders after depression and anxiety with maximum people being affected in the United States, Australia, Finland, Portugal, and Iran. The study of available facilities for mental health treatment reveals that states on the eastern and western coasts tend to have a greater number of facilities available compared to the midwest with California and New York serving the most number of people from the LGBTQ community than any other state. It was also observed that a majority of the facilities offering treatments were pricate non-profit organization followed by private for-profit organizations and at last by public agencies. These facilities offer a variety of treatments ranging from outpatient healthcare to suicide prevention and are funded the most by local agencies, followed by state and then the federal agencies.While we believe that these findings are substantial, a more detailed study using a greater range of factors and conditions will help bring about change and improv on the kind of care offered to patients.

VII. Key Learnings & Future Work

Apart from diet and exercise, working/employment conditions, and culture, the project investigates the various reasons for mental and physical health disorders and the treatment options available.

Any health care system can only be strong when it addresses both aspects of health, physical as well as mental. We discovered through our research that different people in different environments suffer from various mental and physical disorders that require different treatment options.

Employers need to establish awareness/wellness programs, also it is significantly important to address the employees concerns while maintaining privacy. Employers need to foster work-life balance in its culture and policy.

Mental and physical health requires significant investment from the government and corporation, so there are enough facilities available to patients at affordable costs. It requires an integrated approach to address the concerns in the physical and mental health via treatment options, facilities, awareness programs, education, jobs, social welfare schemes etc.

The project can be enhanced by using acknowledged datasets from recognized sources enabling us to achieve accurate results. Furthermore, we can implement more advanced data visualizations and create a stream flow data system that allows us to create real-time visual representations using cloud services.

Acknowledgment

This research would not have been possible without the help and guidance of Prof. Arafat Gabareen Mokhtar. We benefited enormously from his teachings and would like to thank him for the time he spent discussing the subject, techniques and addressing all our questions. We would also like to thank Samhitha Upadhyaya for her support throughout

the development of this project. Finally, we would also like to thank our fellow students at San José State University for their encouragement and support.

References

[1] P. Chopra, “Mental health and the workplace: Issues for developing countries,” *International Journal of Mental Health Systems*, vol. 3, no. 1, p. 4, 2009.

[2] P. Pedrelli, M. Nyer, A. Yeung, C. Zulauf, and T. Wilens, “College students: Mental Health Problems and treatment considerations,” *Academic Psychiatry*, vol. 39, no. 5, pp. 503–511, 2014.

[3] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] “Mental health: Lessons learned in 2020 for 2021 and forward,” *World Bank Blogs*. [Online]. Available: https://blogs.worldbank.org/health/mental-health-lessons-learned-2020-2021-and-forward. [Accessed: 18-May-2022].

[5] Sentell, T., Shumway, M. & Snowden, L. Access to Mental Health Treatment by English Language Proficiency and Race/Ethnicity. *J GEN INTERN MED* 22, 289–293 (2007). https://doi.org/10.1007/s11606-007-0345-7

[6] Bildt, C., & Michélsen, H. (2002). Gender differences in the effects from working conditions on mental health: a 4-year follow-up. International Archives Of Occupational And Environmental Health, 75(4), 252-258. doi: 10.1007/s00420-001-0299-