Introduction

Mental health has become a prevalent public health issue in today's world. Growing awareness through education, media, and advocacy has encouraged more open conversations about mental health. However, despite these advances, misconceptions still persist and there is a need to dispel the stigma surrounding the topic of mental health. Promoting good mental health awareness is key to prevention and early intervention for mentally distressed individuals.

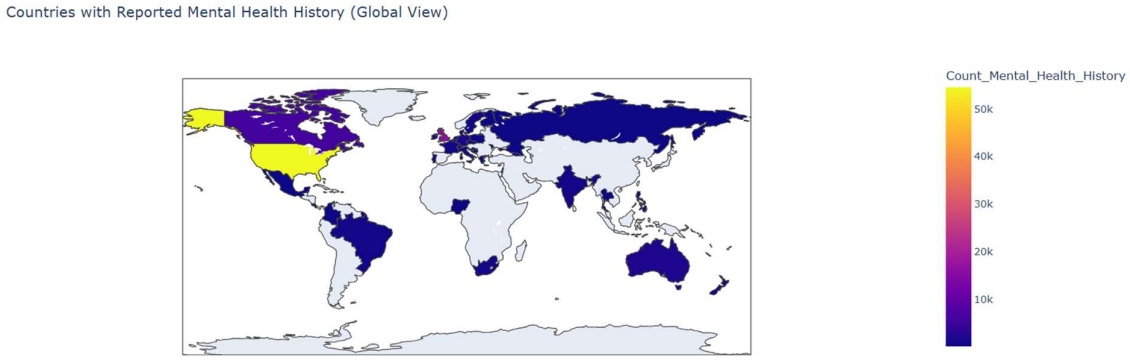
Dataset 1

This dataset contains about 300,000 records with 17 features. The dataset contains data from 2014 to 2016 with categories of features like

1. **Demographics** (Gender, Country)
2. **Work Factors** (Occupation, Self-employed, Work interest)
3. **Social & Psychological Factors** (Days Indoors, Growing Stress, Changes habits, Mood swings, Coping Struggles, Social Weakness, Family history of mental health)
4. **Social support** (Open to Mental health interview, Care options, Treatment)

Plot 1

Firstly, there is a need to establish the prevalence of mental health issues in the world to show that those suffering from poor mental health are not alone.



I chose to use a choropleth plot as it is easy to visualize the number of people with mental health history in different regions of the world.

One limitation of the dataset is that there are quite a few countries with no data regarding mental health, hence these countries appear as colorless on the map, which could result in people mistaking those countries to have no mental health issues.

Dataset 2

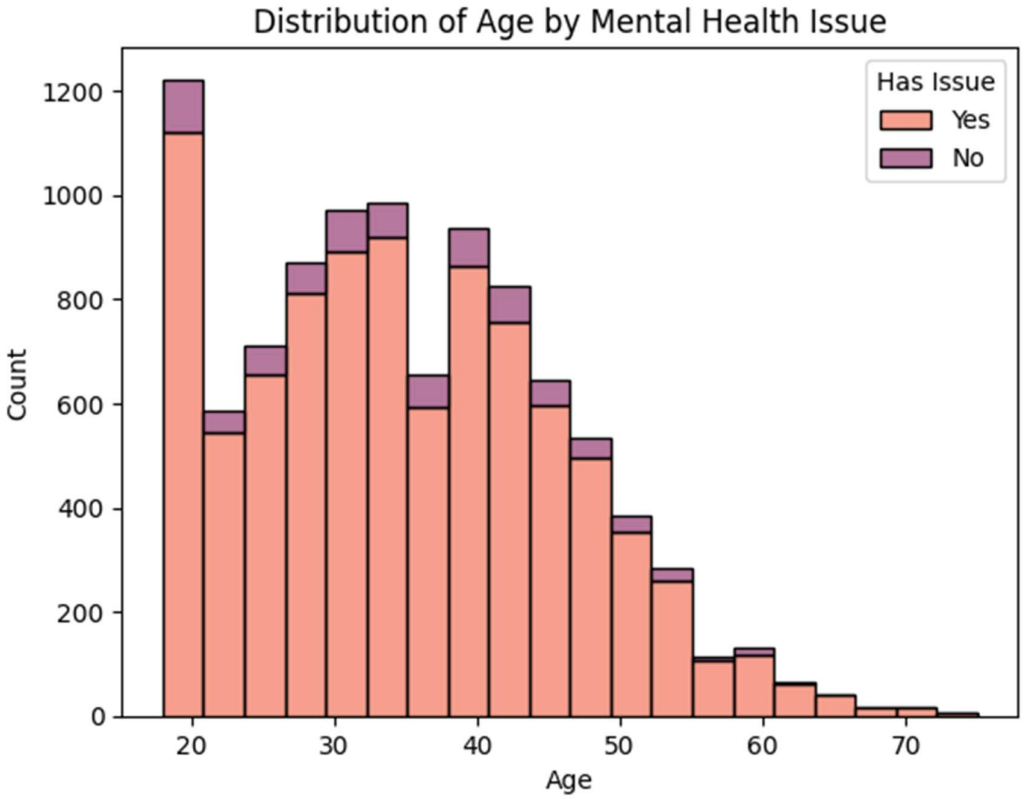
This dataset contains 10,000 records with 50 features and a **binary target label** indicating the presence of a mental health issue. The features of the dataset are split into the following 6 categories:

1. **Demographics** (Age, Gender, Country, Education level, Marital status, Income level)
2. **Work / Academic Factors** (Employment status, Work hours per week, Remote work status, Job satisfaction, Work stress level, Work-life balance, Workplace bullying, Employer mental health support)
3. **Lifestyle & Habits** (Exercise frequency, Sleep duration, Caffeine intake, Alcohol consumption, Smoking habits, Screen time, Social media usage, Hobby time, Diet quality, Financial stress)
4. **Symptoms & Emotional State** (Low mood, Loss of interest, Sleep problems, Fatigue, Appetite changes, Low self-esteem, Difficulty concentrating, Anxiety, Panic attacks, Mood swings, Irritability, Obsessive thoughts, Compulsive behavior, Self-harm, Suicidal thoughts)
5. **History & Background** (Family history of mental illness, Prior diagnosis, Treatment history, On Therapy, On Medication, Trauma history)
6. **Social Support** (Perceived social support, Number of close friends, Feeling understood, Loneliness, Comfort discussing mental health)

This global mental health dataset captures how demographics, work conditions, lifestyle habits, psychological symptoms, background, and social support jointly relate to mental health outcomes.

Plot 2: Who is likely to suffer from poor mental health?

To promote mental health awareness, there is a need to identify the demographic of the population most susceptible to poor mental health, so that future health service interventions can be more timely and tailored to the needs of the people.



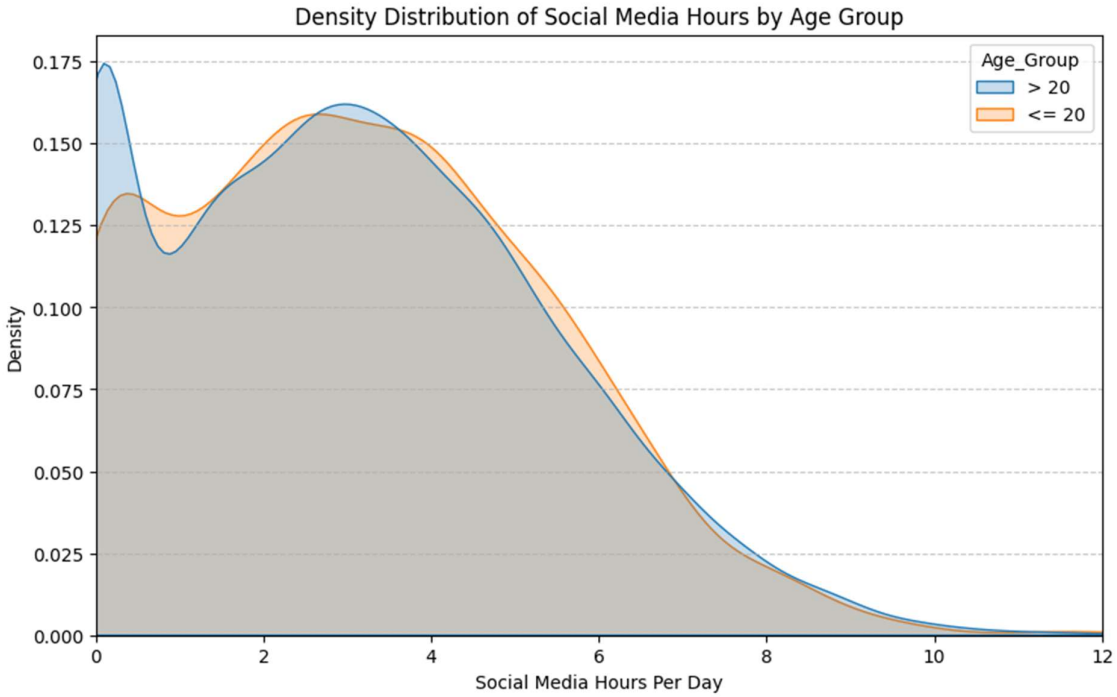
Plot 2 is a stacked histogram with 20 bins showing the distribution of age by mental health issue. I chose to use a histogram so that the distribution of ages can be easily visualized across different bins / ages groups. I also chose to stack the bins so that it is easy to compare the proportions between those who have mental health issues and those who do not **for each age group**.

From plot 2, it is evident that those aged 20 and below are the most likely to suffer from poor mental health since the difference in proportion between the orange bar and the purple bar is the greatest in this bin. Hence one can conclude that special attention needs to be given to those in this age group when constructing mental healthcare policies and services.

However, one limitation of this dataset is that it is imbalanced, with more respondents being aged 20 and below, while there are few elderly respondents aged above 60. This may give the impression that the older one gets, the less likely one is to suffer from poor mental health, which is not true.

Plot 3: Why are these people more likely to suffer from poor mental health?

After discovering that those aged 20 and below are the most likely to suffer from poor mental health, the next step is to uncover the reasons behind it. After performing exploratory data analysis amongst the different variables, social media hours appeared to be a more significant variable in potentially influencing mental health.



Plot 3 is a density plot depicting the distribution of social media hours between those aged 20 and below against those aged above 20. This **density plot was created after filtering out all those respondents who do not have mental health issues**. I chose to use a density plot so that the distributions for both groups of data can be clearly seen in the same figure and comparisons can be easily made. For example, statistics like variance and skewness can be easily inferred from the smooth curves of the density plot.

From plot 3, it can be observed that amongst those who suffer from mental health issues, those aged 20 and below spends more time on social media as compared to those aged above 20. This **may** imply that spending more time on social media increases the probability of deteriorating mental health. Hence targeted interventions can be aimed at reducing the amount of hours people spend on social media.

One limitation of this dataset is that the ages of the respondents ranged from 18 to 75, with no data for people aged 17 and below.

Links

Github repository: https://github.com/Xiang-le/DSA4262

Dataset 1: https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset

Dataset 2: https://www.kaggle.com/datasets/dhrubangtalukdar/global-mental-health-and-lifestyle-survey-dataset