THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

COLLEGE OF INFORMATION SCIENCE & TECHNOLOGY

ANXIETY LEVELS THROUGHOUT THE COVID: A COMPARATIVE STUDY OF AGE GROUPS AND U.S. STATES

QIUHAO ZHU
SPRING 2023

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Information Science & Technology
with honors in Design and Development

Reviewed and approved* by the following:

Kaamran Raahemifar
Teaching Professor of IST
Thesis Supervisor

Edward J. Glantz
Director of Master's Program and Teaching Professor
Honors Advisor

* Electronic approvals are on file.

Abstract

Mental health is a crucial component of overall health and well-being, encompassing emotional, psychological, and social aspects. It significantly influences individuals' thoughts, feelings, and actions, playing a vital role in their capacity to cope with stress, foster relationships, and make decisions. Regrettably, mental health disorders are widespread and can result in severe consequences such as disability, social isolation, and, in extreme cases, suicide.

The COVID-19 pandemic has had a profound impact on mental health, exacerbating pre-existing mental health conditions and giving rise to new ones. This paper aims to explore the current state of knowledge on the impact of the COVID-19 pandemic on mental health, with a particular focus on the ways in which social isolation, economic instability, and increased rates of domestic violence and child abuse have affected mental well-being.

To better understand these issues, the paper will delve into the psychological effects of the pandemic by comparing and analyzing the levels of stress experienced by people of different age groups and from various U.S. states as the Covid pandemic progressed from 2020 to 2021. In doing so, we will examine how the pandemic contributes to feelings of loneliness, depression, and anxiety. Additionally, we will analyze the role of economic instability during the pandemic, which has caused job loss, financial strain, and social instability, all of which contribute to deteriorating mental health.

Contents

Acknowledgment

I would like to express my heartfelt appreciation to Dr. Kaamran Raahemifar for his supervision and assistance throughout the development of this thesis. Dr. Kaamran's extensive knowledge, expertise, and insightful feedback have been invaluable in shaping my research and ensuring its successful completion. Their unwavering support, encouragement, and commitment to my growth as a researcher have inspired me to strive for excellence in my work. I am deeply grateful for the opportunity to learn from Dr. Kaamran and for the guidance they have provided during this important stage of my academic journey.

I would also like to extend my sincere gratitude to Ms. Mehak Juneja for their invaluable assistance with the code used in this thesis. Mehak's expertise and guidance have been instrumental in helping me navigate complex programming challenges, and her support has significantly contributed to the successful completion of this work. I am truly grateful for Mehak's patience, dedication, and willingness to share their knowledge, making this research journey an enriching and rewarding experience.

In addition, I would like to express my profound thanks to my parents for their unwavering support and encouragement throughout my academic journey. Their love, understanding, and belief in my abilities have been a constant source of strength and motivation, empowering me to pursue my goals and overcome any challenges along the way. I am deeply grateful for their sacrifices and commitment to my education and personal growth.

Lastly, I am immensely grateful to Dr. Edward Glantz for his assistance and direction in the development of this thesis. Dr. Glantz's expertise, constructive feedback, and thoughtful guidance have been instrumental in shaping my research and ensuring its successful completion. His encouragement and unwavering support have inspired me to push the boundaries of my knowledge and grow as a researcher. I am truly thankful for the opportunity to learn from Dr. Glantz and for the invaluable insights he has provided throughout this academic journey.

Introduction

Impact of the COVID-19 Pandemic on Mental Health

As the world continues to grapple with the COVID-19 pandemic, it has become increasingly clear that the virus is not only a public health crisis but also a mental health crisis. Numerous studies have shown that the pandemic has had a profound impact on people's mental health, leading to increased levels of stress, anxiety, depression, and other mental health conditions (Bueno-Notivol et al., 2021; Fiorillo et al., 2020; Ho et al., 2020; Salari et al., 2020). The lockdowns, social distancing measures, and economic instability associated with the pandemic have created a perfect storm of stressors, exacerbating pre-existing mental health conditions and leading to new ones in those who were previously mentally healthy (Brooks et al., 2020; Rajkumar, 2020; Wang et al., 2020).

Social Isolation and Mental Health

One of the key ways in which the pandemic has affected mental health is through social isolation. Social distancing measures have forced people to spend more time alone, leading to feelings of loneliness and disconnection from others (Killgore et al., 2020; Palgi et al., 2020). The lack of social support during this time has been particularly challenging for vulnerable populations such as the elderly, those with pre-existing mental health conditions, and individuals with limited access to technology or resources (Galea et al., 2020; Holmes et al., 2020; Torales et al., 2020).

Economic Instability and Mental Health

In addition, the pandemic has created economic instability, leading to job losses, financial insecurity, and increased stress, all of which can have a significant impact on mental health (Pfefferbaum & North, 2020; VanderWeele et al., 2021). Moreover, the pandemic has increased the risk of domestic violence and child abuse, further exacerbating mental health problems (Bradbury-Jones & Isham, 2020; Campbell, 2020). Being confined at home with an abusive partner or parent has led to increased rates of violence and abuse, creating a cycle of trauma and stress for victims. Healthcare workers, who have been at the forefront of the pandemic, have also been exposed to increased levels of stress and trauma, leading to burnout and other mental health conditions (Lai et al., 2020; Shanafelt et al., 2020).

Research Gaps and Future Directions

It is clear that the pandemic has had a profound impact on mental health, and the effects of this crisis are likely to be long-lasting. However, there is still much that is not understood about the mechanisms underlying this impact, and there is a need for further research in this area. In particular, it is important to understand the differential impact of the pandemic on different populations, including people from different age groups.

Literature Review

To address the knowledge gap and contribute to the understanding of the pandemic's varying effects on diverse populations, I conducted a comparative literature review, focusing on studies that track perceived physical and mental health over time. The first article, “The Centers for Disease Control and Prevention's Healthy Days Measures – Population tracking of perceived physical and mental health over time” by Moriarty et al., (2003), serves as the foundation for this review. Two additional articles, “Assessing the Impact of Social Determinants on Health-Related Quality of Life” by Case et al., (2022) and “A Population Health Approach to Mental Health: Lessons From the Global Burden of Disease Study” by Patel et al., (2016), provide further context and support for the examination of population health metrics. This review synthesizes the findings and methodologies employed in these articles, offering a comprehensive overview of the current state of population health research.

Monitoring and assessing population-level physical and mental health is crucial for informing public health policy and interventions. These three articles provide insight into different approaches to tracking population health, including the Centers for Disease Control and Prevention's (CDC) Healthy Days Measures, social determinants of health, and the Global Burden of Disease (GBD) study. This literature review aims to analyze and synthesize the findings and methodologies from these articles, providing a comprehensive understanding of population health research.

Article 1: The Centers for Disease Control and Prevention's Healthy Days Measures – Population tracking of perceived physical and mental health over time.

Moriarty et al., (2003) presented the CDC's Healthy Days Measures as an innovative tool for tracking population-level physical and mental health. The authors argue that these measures, which consist of four core questions assessing self-rated health, recent physically unhealthy days, recent mentally unhealthy days, and recent activity limitation days, provide a simple and effective method for monitoring health trends and evaluating interventions. By analyzing data from the Behavioral Risk Factor Surveillance System (BRFSS), the study demonstrates the strong convergent validity and good test-retest reliability of the Healthy Days Measures. This research provides valuable evidence for the utility of these measures in public health research and practice, as they can be used to identify health disparities and inform policy decisions aimed at improving population health.

Article 2: Health-related quality of life and social determinants of health following COVID-19 infection in a predominantly Latino population

Case et al., (2022) delve into the impact of social determinants of health (SDH) on health-related quality of life (HRQoL) following Covid. The study employs a cross-sectional design and multivariate regression models to examine the associations between SDH, such as education, income, employment, housing, and HRQoL. The results reveal that these social determinants significantly influence HRQoL, highlighting the need for public health efforts to address these factors in order to improve population health outcomes. This research emphasizes the importance of considering social determinants in public health policy development and intervention design, as they play a critical role in shaping individuals' health and well-being.

Article 3: The Lancet Commission on global mental health and sustainable development

The Lancet Commission on global mental health and sustainable development, as discussed by Patel et al. (2016), reflects on the insights gained from the Global Burden of Disease (GBD) study, emphasizing the substantial contribution of mental disorders to the global disease burden. The authors advocate for a population health approach to effectively address the far-reaching consequences of mental disorders worldwide. This approach entails identifying and managing risk factors, promoting mental well-being, and broadening access to evidence-based interventions. The GBD study's findings underscore the significance of embracing this population health strategy for mental health, as it aims to mitigate the global impact of mental disorders and enhance overall population health outcomes.

Synthesis and Discussion

The three articles provide complementary perspectives on population health tracking and assessment. Moriarty et al. (2003) offered the CDC's Healthy Days Measures as a simple, reliable, and valid tool for monitoring population-level physical and mental health. Case et al., (2018) emphasize the importance of social determinants in shaping the health-related quality of life, while Patel et al., (2016) advocate for a population health approach to mental health, using the Global Burden of Disease study as a prime example.

The CDC's Healthy Days Measures, as discussed by Moriarty et al. (2003), provide a valuable starting point for population health tracking, as they capture both physical and mental health dimensions. Case et al., (2018) build on this by demonstrating the significant impact of social determinants on health-related quality of life. This highlights the need for public health policies and interventions to address underlying social factors, such as education, income, employment, and housing, in order to improve overall population health outcomes.

Meanwhile, Patel et al., (2016) emphasize the importance of a population health approach to mental health, which involves addressing risk factors, promoting mental well-being, and enhancing access to evidence-based interventions. This approach aligns with the findings of Moriarty et al., (2003) and Case et al., (2018), as it highlights the need for comprehensive, multi-dimensional strategies to tackle the complex issue of population health.

In conclusion, these three articles collectively underscore the importance of utilizing a variety of tools and approaches to monitor and assess population health. The CDC's Healthy Days Measures offer a solid foundation for tracking physical and mental health, while the consideration of social determinants and the adoption of a population health approach provide valuable insights for developing effective public health policies and interventions. By integrating these perspectives, researchers, policymakers, and practitioners can work together to improve population health outcomes and reduce health disparities across communities.

Building on this understanding, I will provide a more comprehensive examination of the population health approach, delving into the assessment methods, associated factors, and implications for clinical practice and research. In the following sections, we will explore various assessment methods that have been used to evaluate population health and identify the most effective strategies for addressing health disparities.

Assessment Methods

1. Self-report questionnaires: Several studies have used self-report questionnaires to measure the frequency of symptoms related to anxiety and depression during the last 12 days. These screening tools and measurements, such as the Generalized Anxiety Disorder-7 (GAD-7;

Spitzer et al., 2006) and the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001), have demonstrated strong psychometric properties and are widely used in clinical practice and research. They provide a quick and cost-effective method for assessing symptom severity and treatment outcomes (Löwe et al., 2008).

2. Daily symptom diaries: Some researchers have employed daily symptom diaries to track anxiety and depression symptoms. These diaries enable a more detailed and accurate assessment of symptom frequency and severity, as they reduce recall bias and provide real-time data (Wiegle, 2021). However, these methods can be time-consuming and may be subject to participant adherence and consistency (Stone et al., 2003).

Associated Factors

1. Demographic factors: Several studies have examined the relationship between demographic factors and the frequency of anxiety or depression symptoms. Factors such as age, gender, and socioeconomic status have been shown to influence symptom frequency (Alonso et al., 2004; Seedat et al., 2009). For instance, women and individuals with lower socioeconomic status have been reported to experience more frequent symptoms of anxiety and depression (Kessler et al., 2005).
2. Environmental factors: Research has also identified environmental factors that are associated with the frequency of anxiety and depression symptoms. Factors such as life events, social support, and work-related stress have been shown to influence symptom frequency (Hammen, 2005; Paykel, 2003). For example, individuals who experience negative life events or have limited social support are more likely to report frequent anxiety and depression symptoms (Hammen, 2005; Kendler et al., 2004).

3. Comorbidity: Another important factor associated with the frequency of anxiety and depression symptoms is the presence of comorbid mental health disorders. Research indicates that individuals with comorbid disorders, such as substance use disorders or other mood disorders, often report more frequent symptoms of anxiety and depression (Kessler et al., 2008; Lamers et al., 2011). This suggests that addressing comorbid conditions is crucial for effectively managing anxiety and depression symptoms.

Implications for Clinical Practice and Research

1. Treatment planning and monitoring: Assessing the frequency of anxiety and depression symptoms can aid clinicians in treatment planning and monitoring. For instance, by identifying symptom patterns, therapists can tailor interventions to target specific symptoms and monitor treatment progress (Titov et al., 2011). Additionally, frequent symptom assessment allows for early detection of relapse or treatment resistance (Judd et al., 1998).
2. Research: The use of indicators based on the reported frequency of symptoms has facilitated research on various aspects of anxiety and depression. Longitudinal studies that employ daily symptom diaries or self-report questionnaires can provide valuable insights into the course of these disorders, their risk factors, and protective factors (Trull & Ebner-Priemer, 2014).

   Moreover, this approach enables researchers to examine the effectiveness of interventions and identify factors that may predict treatment response or relapse (Boswell et al., 2015).
3. Public health policy: Data on the frequency of anxiety and depression symptoms can inform public health policies and strategies. By identifying populations at risk and

understanding the factors that contribute to symptom frequency, policymakers can develop targeted interventions to address mental health disparities and improve overall mental health outcomes (World Health Organization, 2021).

In summary, the use of indicators based on the reported frequency of anxiety and depression symptoms has significant implications for clinical practice, research, and public health policy. Self-report questionnaires and daily symptom diaries provide valuable assessment tools, while various demographic, environmental, and comorbidity factors are associated with symptom frequency. Understanding these factors and utilizing this approach can improve treatment planning, monitoring, and the development of public health strategies aimed at reducing the burden of anxiety and depression disorders.

Methodology

The objective of the methodology section is to compare and analyze the level of stress experienced by people of different age groups and from various U.S. states as the Covid pandemic progressed from 2020 to 2021. Data Source: The dataset used in this study was obtained from the Centers for Disease Control and Prevention (CDC) website (https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm). The dataset contains information on self-reported symptoms of anxiety or depression based on frequency during the last 7 days.

The research methods employed in this study encompass data cleaning, data filtering, reindexing columns for streamlined data analysis, and constructing a machine learning model. These methods are crucial for guaranteeing that the dataset is accurate, relevant, and appropriately formatted for subsequent analysis.

1. Data Cleaning: The initial step in the research process is data cleaning, which involves removing any inconsistencies, errors, or inaccuracies within the dataset. In this study, the Python programming language and the Pandas library were utilized to perform data cleaning tasks, such as handling missing values, correcting data entry errors, and standardizing data formats.
2. Data Filter: Following data cleaning, the dataset was filtered based on the “state” and “age” column, as it is the most common group among the entire dataset in **Figure 1**. This filtering process ensured that only relevant data was retained for further analysis, enabling a more focused examination of the research question. By narrowing down the dataset in this way, the researchers could concentrate on the most pertinent information and avoid potential distractions or biases from unrelated data.

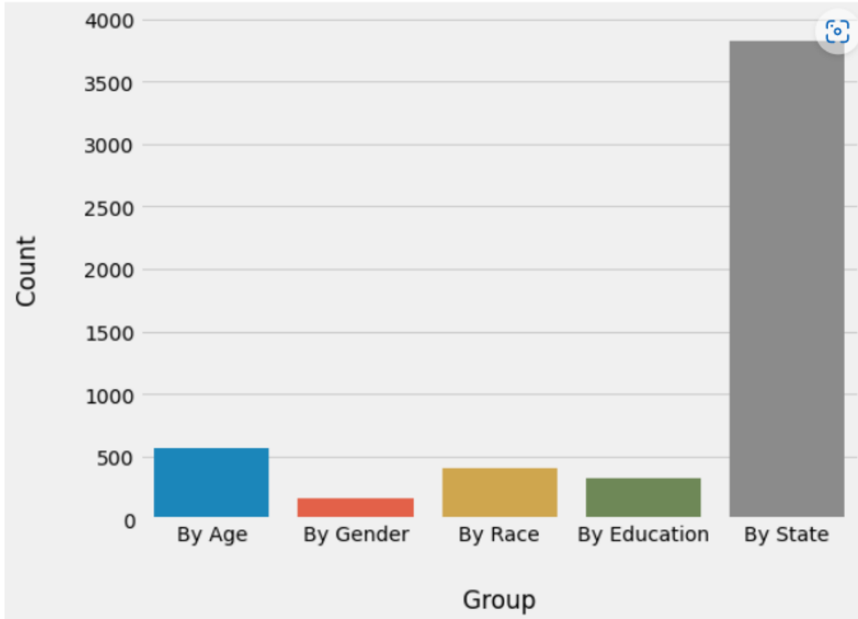

Figure 1: Bar Chart Depicting the Frequency of Observations for Each Group in the Dataset

3. Reindex Columns: After filtering the data, the remaining columns needed to be reindexed to maintain consistency within the dataset. This step is crucial in the methodology section of the thesis, as it demonstrates the attention to detail and the importance of maintaining a well-structured dataset for subsequent analysis. Reindexing columns allows for a more coherent dataset structure, ensuring that the data is easily accessible and interpretable.
4. Date Conversion: The primary objective of this research is to examine the changes in anxiety levels over the years. To achieve this, the first step involves converting the relevant column containing date information in the dataset to a datetime object type. This conversion is essential for time series analysis, as it enables the researchers to accurately track changes in anxiety levels over time and identify any patterns or trends that emerge.

Once the conversion is complete, extract the year from the datetime column and create a new column named "year" to store this information. This new column will be instrumental in separating the data based on the years of interest.

With the "year" column in place, the data is then divided into two distinct dataframes according to the years 2020 and 2021. Separating the dataset in this manner enables a focused comparison of anxiety levels between these two specific years, which is the primary aim of the research.

5. Machine Learning Model: In this study, we aimed to develop predictive models using supervised learning techniques, specifically XGBoost and Random Forest algorithms. The data objects in the dataset were well-defined and well-structured, making it suitable for supervised learning and eliminating the need for additional data categorization.

First, the dataset was divided into training and testing sets following an 80/20 ratio to ensure data validity. The x_test and y_test were printed out for reference. The data was then fitted into a linear regression model, and eXtreme Gradient Boosting (XGBoost) was chosen as an ideal candidate for building the model. XGBoost is a supervised learning algorithm that can be used for both regression and classification tasks (Chen & Guestrin, 2016) and is widely recognized for its high performance and accuracy in various machine learning applications (Mello, 2020).

To develop the predictive models, we employed two machine learning algorithms, XGBoost and Random Forest, and compared their performance using the Root Mean Squared Error (RMSE) metric. The necessary libraries, including XGBoost and Scikit-learn, were imported. An XGBoost regressor (xgb_r) was initialized with the objective set to 'reg:linear', the number of estimators (n_estimators) set to 10, and a random seed value of 123. The regressor was then fitted with the training data (x_train and y_train). Using the fitted model, predictions were made for the target variable in the test data (x_test) and stored in the pred variable. The RMSE between the predicted values (pred) and the true target values in the test data (y_test) was calculated and printed for reference.

Next, we utilized the Random Forest algorithm from Scikit-learn's ensemble module. A RandomForestRegressor was initialized with 100 estimators and a random state of 0. This regressor was then fitted with the training data (x_train and y_train). Similar to the XGBoost model, predictions were made for the target variable in the test data (x_test) using the fitted Random Forest model and stored in the rF_pred variable. The RMSE between the predicted values (rF_pred) and the true target values in the test data (y_test) was calculated and printed for comparison with the XGBoost model.

Lastly, a scatter plot was created to visualize the predicted values against the true target values. A new column was added to the x_train dataframe with a range from 1 to the length of the dataframe. The plot was scaled logarithmically on both axes, and the line of equality was drawn to compare the predictions with the true values. The plot's x-axis was labeled ‘True Values’, and the y-axis was labeled ‘Predictions’ as shown in **Figure 2**.

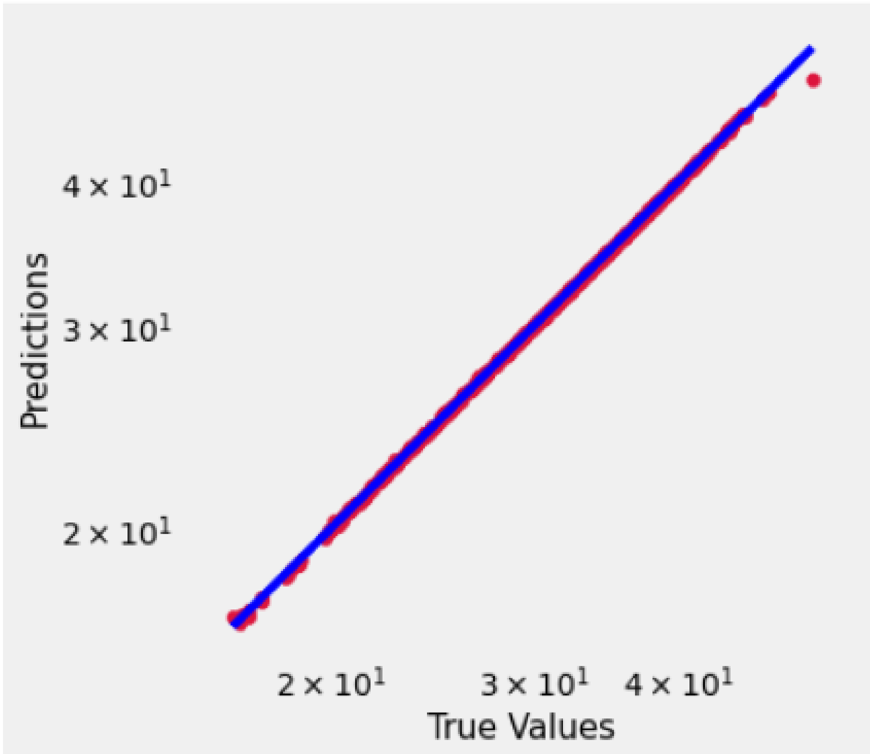

Figure 2: Scatter Plot Comparing True Values and Predictions on a Log-Log Scale, with a Blue Line Representing the Ideal 1:1 Relationship

Results and Discussion

In this section, we present the findings of our analysis, highlighting the trends and patterns observed in the data. Our primary focus is on the levels of anxiety during the years 2020 and 2021, as captured by the dataset. We will discuss these findings in detail, examining the implications and potential factors that may have contributed to the observed changes in anxiety levels.

Data Visualization and analysis

To further analyze and visualize the data, a plot is created with the X-axis representing the "Time Period Start Date" and the Y-axis representing the "Value", which corresponds to the level of anxiety. This visual representation helps identify any significant changes or patterns in anxiety levels during the years 2020 and 2021. The plot provides a clear illustration of the data, allowing for a more comprehensive understanding of the trends and fluctuations in anxiety levels over time.

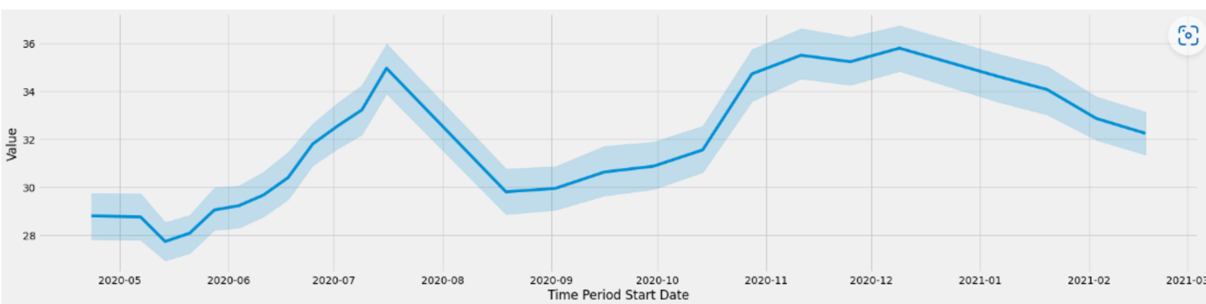

Figure 3: Line Plot illustrating the Trend of 'Value' (level of anxiety and depression level) Over Time, with 'Time Period Start Date' on the X-axis for the Selected State

As we can see in **Figure 3**, there is a peak in the value (depression level) between 2020-06 and 2020-08. The level remained high during 2020-11 to 2021-01.

Furthermore, I conducted data analytics to gain deeper insights into the dataset. In Graph 2, we observe an even distribution of the number of people across different periods, which suggests that the dataset is reliable for further analysis.

An evenly distributed dataset is crucial for obtaining reliable results in research, as it ensures that the data points are not concentrated in a specific period or skewed towards certain values. This equal distribution of data points allows for a more accurate representation of the population, thus enabling researchers to draw meaningful conclusions based on the analysis (D'Agostino & Pearson, 1973).

As demonstrated in **Figure 4**, the number of people in each period is distributed evenly, which is an indication of the dataset's reliability. This even distribution can be attributed to the data collection process, where the researchers ensured that they gathered data from different periods in a balanced manner. Consequently, this balanced dataset helps to minimize the risk of bias and provides a more accurate reflection of the changes in anxiety levels over time.

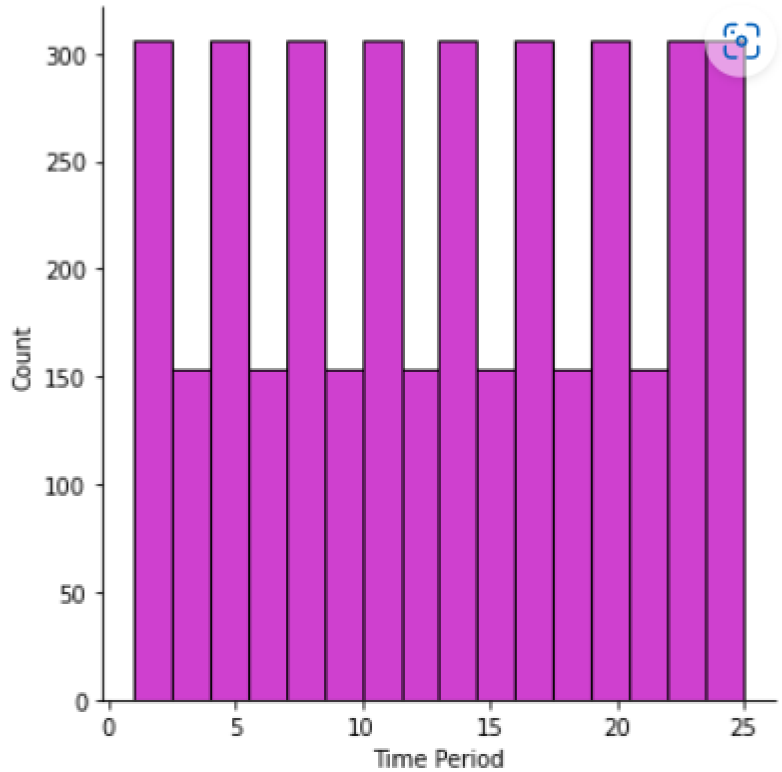

Figure 4: Histogram Depicting the Distribution of 'Time Period' in the Dataset, with Frequency on the Y-axis

Figure 5 illustrates the relationship between Value and Time Period using a linear regression model. Specifically, it highlights how the Value, which represents depression levels, has changed over time since the onset of the COVID-19 pandemic.

Upon closer examination, it becomes evident that there is a noticeable upward trend in depression levels as time progresses. This suggests that the mental health impact of the pandemic has been significant and far-reaching. Factors contributing to this increase may include prolonged periods of social isolation (Jeffers et al., 2022), economic instability (Parker, K., Minkin, R., & Bennett, J. 2020), and the ongoing uncertainty surrounding the crisis (Wang et al., 2022).

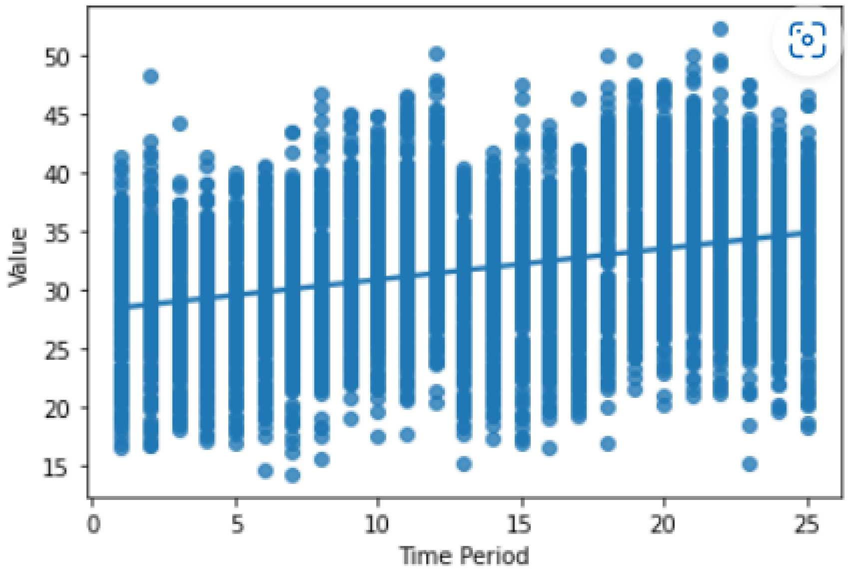

Figure 5: Scatter Plot with Regression Line Displaying the Relationship Between 'Time Period' and 'Value' (level of anxiety and depression level) in the Dataset

Figure 6 extracts three age groups "18 - 29", "40 - 49" and "70 - 79" from the age column and presents an analysis of depression levels among various age groups as time progresses. It is evident from the data that young adults (aged 18-29) experience higher depression levels compared to both middle-aged individuals and elderly people (aged 70-79).

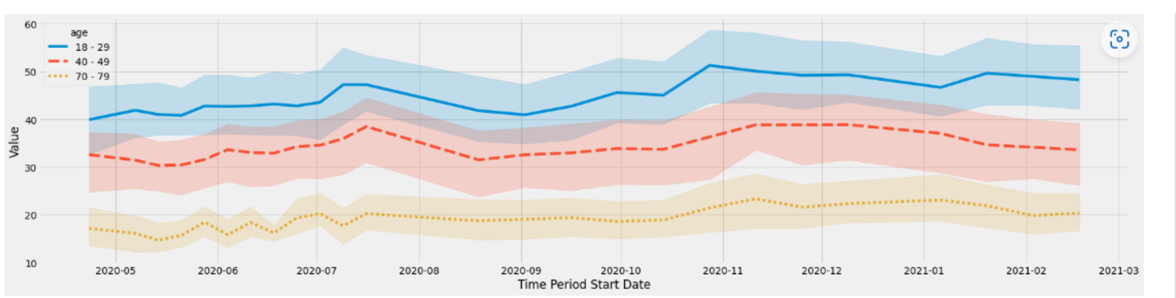

Figure 6: Line Plot Depicting the Trend of 'Value' Over Time by 'Time Period Start Date' for Age Groups '18 - 29', '40 - 49', and '70 - 79'

This trend could be attributed to multiple factors, such as the substantial impact of the pandemic on the lives and routines of younger individuals (Germani et al., 2020). Young adults may face heightened stress due to disruptions in their education or career opportunities, along with increased social isolation resulting from pandemic-related restrictions (Loades et al., 2020). In contrast, middle-aged and older adults might have more stable life circumstances, allowing them to better cope with the challenges posed by the pandemic (Birditt et al., 2021).

Figure 7 below extracts data on three states – "California," "New York," and "Wyoming" – from the States column and presents an analysis of anxiety levels as time progresses. It is evident that residents in California and New York experienced higher levels of anxiety than residents in Wyoming during the COVID-19 pandemic (Jia et al., 2021). Specifically, residents in California exhibited the highest anxiety levels among all three states. As the graph illustrates, anxiety levels peaked during early to mid-July 2020 in all three states.

Several factors contributed to the increased anxiety in California and New York, such as higher population density, greater economic impact, and the rapid spread of the virus in these states (Saeed et al., 2022). The anxiety level in New York also surged in early February 2021, in contrast to the plummeting anxiety levels in Wyoming during the same period. This spike in New York could be attributed to new virus variants, increased case numbers, and uncertainty surrounding vaccine distribution (Saeed et al., 2022).

The depression level in Wyoming, on the other hand, plummeted in early February 2021. This decrease may be linked to a combination of factors, including successful containment of the virus, lower population density, and effective vaccine distribution (Jackson et al., 2021). Additionally, the relatively rural nature of Wyoming may have provided residents with more opportunities to maintain social connections and engage in outdoor activities, which are known to have a positive impact on mental health (Jackson et al., 2021).

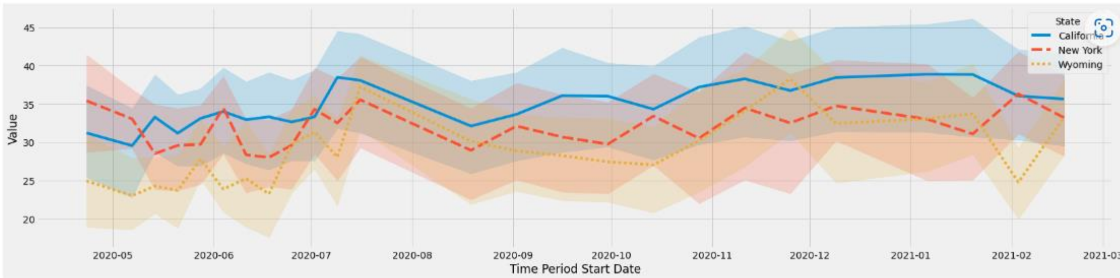

Figure 7: Line Plot illustrating the Trend of 'Value' Over Time by 'Time Period Start Date' for States 'California', 'New York', and 'Wyoming'

XGBoost model discussion

Below is the running result of the XGBoost model. Based on the previously described XGBoost model training process, we obtained a Root Mean Squared Error (RMSE) of 0.138691. This value represents the average difference between the predicted values and the true target values in the test data. A lower RMSE indicates a better fit of the model, implying that the model's predictions are closer to the actual values.

As demonstrated in **Figure 8**, an RMSE of 0.138691 suggests that the XGBoost model has a relatively good fit, as the error is quite small. However, to fully assess the performance of the model, we also need to compare this RMSE value with other models' performance, such as the Random Forest model.

```
from sklearn.ensemble import RandomForestRegressor

regressor = RandomForestRegressor(n_estimators = 100, random_state = 0)
regressor.fit(x_train, y_train)
rF_pred = regressor.predict(x_test)
rf_error=np.sqrt(mean_squared_error(y_test, rF_pred))
print("RMSE: % f" %(rf_error))
```

```
C:\Users\Owner\AppData\Local\Temp/ipykernel_11220/1547343061.py:4: DataConversionWarning: A column-vector y was passed when a 1
d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
  regressor.fit(x_train, y_train)

RMSE:  0.138691
```

Figure 8: Random Forest Regressor Model Performance, with the Root Mean Squared Error (RMSE) Value Calculated from the Model's Predictions on the Test Set

In the methodology section, **Figure 2** demonstrates that the majority of the points align closely with the prediction line, indicating that the model's predictions are generally accurate. However, there are a few data points that deviate from the line of best fit. These outliers could potentially highlight areas where the model's performance is less accurate or might indicate the presence of noise or unique characteristics in the data that the model is unable to capture effectively. Despite these deviations, the overall distribution of points in the plot suggests that the model performs well in predicting the target variable for most instances.

Discussion

In general, the increase in depression levels between June and August 2020 in the United States can be attributed to various factors, such as the continued impact of the COVID-19 pandemic, economic instability, social unrest, and political uncertainty. According to a study conducted by Ettman et al. (2020), the COVID-19 pandemic and associated lockdown measures have had a significant impact on mental health in the United States, with depression rates increasing since the start of the pandemic.

Additionally, a study by Twenge and Joiner (2020) found that the pandemic has had a disproportionate impact on younger adults, who have reported higher levels of depression and anxiety. They suggest that this may be due to the disruption of important life milestones, such as graduation and job opportunities, and the lack of social interaction caused by social distancing measures.

Furthermore, the period between June and August 2020 was marked by significant social and political unrest in the United States, with protests against racial injustice and police brutality occurring in various cities. According to a study by Fitzpatrick et al. (2020), exposure to civil unrest has been linked to increased rates of depression and anxiety.

Overall, the increase in depression levels between June and August 2020 in the United States can be attributed to a complex set of factors related to the ongoing COVID-19 pandemic, economic instability, social and political unrest, and social isolation.

Another peak in depression levels in the United States during December 2020 and January 2021 can be attributed to various factors related to the ongoing COVID-19 pandemic, including the winter surge in cases, economic instability, and social isolation.

The winter surge in COVID-19 cases in the United States during this time was the most severe of the pandemic, with daily case numbers and hospitalizations reaching record highs (CDC, 2021). This surge was accompanied by an increase in deaths and a strain on the healthcare system, which may have contributed to increased anxiety and depression among the population.

Additionally, the economic instability caused by the pandemic, including job losses and financial insecurity, may have contributed to increased depression rates during this time. The COVID-19 pandemic has had a significant impact on the economy, with many businesses closing and individuals losing their jobs or experiencing reduced income (Bureau of Labor Statistics, 2021).

Finally, social isolation and loneliness have been linked to increased depression rates during the pandemic. The winter months are traditionally a time of increased social interaction and gatherings, but the pandemic made it difficult for people to celebrate holidays and spend time with loved ones. This isolation, coupled with the uncertainty and stress of the pandemic, may have contributed to the peak in depression rates during this time.

Conclusion

In this paper, we analyzed the impact of the COVID-19 pandemic on mental health in the United States during the early stages from 2020 to 2021. Our findings demonstrate a significant surge in overall depression rates during June and July of 2020, likely attributable to social instability, including the "Black Lives Matter" protests that swept the nation (Fitzpatrick et al., 2020). Additionally, another spike in depression levels occurred in January 2021, which may be linked to record highs in daily case numbers and hospitalizations (CDC, 2021).

Upon examining age groups, our analysis revealed that individuals aged 18-29 experienced higher levels of anxiety and depression compared to other age groups. This could be attributed to the unique challenges faced by younger adults during the pandemic, such as navigating remote learning and entering an uncertain job market (Twenge and Joiner, 2020).

When considering regional differences within the U.S., our study showed that residents of California experienced higher levels of depression than those in New York and Wyoming. This disparity may be due to California's higher population density and the rapid spread of the virus, which contributed to heightened anxiety and stress (Saeed et al., 2022). In contrast, Wyoming exhibited lower levels of anxiety and fewer COVID-19 cases, likely due to its lower population density and the abundance of outdoor activities available, which can contribute to improved mental health (Jackson et al., 2021).

In conclusion, our analysis highlights the multifaceted nature of the pandemic's impact on mental health in the United States. Our findings demonstrate significant surges in depression rates during specific periods, and our XGBoost model, with a relatively good fit (RMSE = 0.138691), provides valuable insights into the various factors affecting mental health. However, further investigation and comparison with other models, such as the Random Forest model, are necessary to enhance our understanding. By examining age groups, regional differences, and other demographic factors, we can better inform future public health strategies and interventions to support the population during times of crisis.

Future works

1. Longitudinal studies: Investigate the long-term effects of the pandemic on mental health, extending the analysis beyond the early stages (2020-2021) to explore how mental health indicators evolve over time and how they may be influenced by ongoing social, economic, and public health factors.
2. Demographic analysis: Conduct a more in-depth examination of the specific factors contributing to anxiety and depression within different age groups, ethnicities, and socioeconomic backgrounds to identify targeted intervention strategies for vulnerable populations.
3. Regional policy impact: Assess the effectiveness of different state-level policies and interventions aimed at mitigating the pandemic's mental health consequences, including access to mental health resources, telehealth services, and community support programs.
4. Comparison with other countries: Expand the scope of the analysis by comparing the United States' mental health impact during the pandemic with other countries to identify potential best practices and lessons learned in managing mental health challenges on a global scale.
5. Resilience factors: Explore the factors that contribute to resilience and improved mental health during crises, such as the role of social support, outdoor activities, and coping strategies, to develop evidence-based recommendations for fostering resilience in future public health emergencies.

6. Attention mechanisms in neural networks can be utilized to identify and focus on relevant features when making predictions. By incorporating attention mechanisms in deep learning models, researchers can gain more insight into which factors are most influential in predicting mental health outcomes during the pandemic (Ive et al., 2018).
7. Post-pandemic mental health: Investigate the long-term psychological consequences of the pandemic, including potential increases in post-traumatic stress disorder (PTSD), grief, and other mental health disorders, as well as the ongoing demand for mental health services and support.

By pursuing these avenues of future research, we can further our understanding of the pandemic's mental health impact and inform the development of targeted interventions and policies to better support individuals and communities during times of crisis.

References

Alonso, J., Angermeyer, M. C., Bernert, S., Bruffaerts, R., Brugha, T. S., Bryson, H., de Girolamo, G., Graaf, R., Demyttenaere, K., Gasquet, I., Haro, J. M., Katz, S. J., Kessler, R. C., Kovess, V., Lépine, J. P., Ormel, J., Polidori, G., Russo, L. J., Vilagut, G., Almansa, J., … ESEMeD/MHEDEA 2000 Investigators, European Study of the Epidemiology of Mental Disorders (ESEMeD) Project (2004). Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta psychiatrica Scandinavica. Supplementum*, (420), 21–27. https://doi.org/10.1111/j.1600-0047.2004.00327.x

Birditt, K. S., Turkelson, A., Fingerman, K. L., Polenick, C. A., & Oya, A. (2021). Age Differences in Stress, Life Changes, and Social Ties During the COVID-19 Pandemic: Implications for Psychological Well-Being. *The Gerontologist*, *61*(2), 205–216. https://doi.org/10.1093/geront/gnaa204

Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychotherapy research : journal of the Society for Psychotherapy Research*, *25*(1), 6–19. https://doi.org/10.1080/10503307.2013.817696

Bradbury-Jones, C., & Isham, L. (2020). The pandemic paradox: The consequences of COVID-19 on domestic violence. *Journal of clinical nursing*, *29*(13-14), 2047–2049. https://doi.org/10.1111/jocn.15296

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet (London, England)*, *395*(10227), 912–920. https://doi.org/10.1016/S0140-6736(20)30460-8

Bueno-Notivol, J., Gracia-García, P., Olaya, B., Lasheras, I., López-Antón, R., & Santabárbara, J. (2021). Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. *International journal of clinical and health psychology : IJCHP*, *21*(1), 100196. https://doi.org/10.1016/j.ijchp.2020.07.007

Bureau of Labor Statistics. (2021). Labor Force Statistics from the Current Population Survey. Retrieved from https://www.bls.gov/cps/

Campbell A. M. (2020). An increasing risk of family violence during the Covid-19 pandemic: Strengthening community collaborations to save lives. *Forensic Science International. Reports*, *2*, 100089. https://doi.org/10.1016/j.fsir.2020.100089

Case, K. R., Wang, C. P., Hosek, M. G., Lill, S. F., Howell, A. B., Taylor, B. S., Bridges, J., MacCarthy, D. J., Winkler, P., & Tsevat, J. (2022). Health-related quality of life and social determinants of health following COVID-19 infection in a predominantly Latino population. *Journal of patient-reported outcomes*, *6*(1), 72. https://doi.org/10.1186/s41687-022-00473-8

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

D'Agostino, R., & Pearson, E. S. (1973). Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{}$ b_1. *Biometrika*, *60*(3), 613–622. https://doi.org/10.2307/2335012

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ambulatory assessment: An innovative and promising approach for clinical psychology. *European Psychologist*, *14*(2), 109–119. https://doi.org/10.1027/1016-9040.14.2.109

Ettman, C. K., Abdalla, S. M., Cohen, G. H., Sampson, L., Vivier, P. M., & Galea, S. (2020). Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. *JAMA network open*, *3*(9), e2019686. https://doi.org/10.1001/jamanetworkopen.2020.19686

Fiorillo, A., & Gorwood, P. (2020). The consequences of the COVID-19 pandemic on mental health and implications for clinical practice. *European psychiatry : the journal of the Association of European Psychiatrists*, *63*(1), e32. https://doi.org/10.1192/j.eurpsy.2020.35

Fitzpatrick, K. M., Drawve, G., & Harris, C. (2020). Facing new fears during the COVID-19 pandemic: The State of America's mental health. *Journal of anxiety disorders*, *75*, 102291. https://doi.org/10.1016/j.janxdis.2020.102291

Galea, S., Merchant, R. M., & Lurie, N. (2020). The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA internal medicine, 180(6), 817-818.

Germani, A., Buratta, L., Delvecchio, E., & Mazzeschi, C. (2020). Emerging adults and COVID-19: The role of individualism-collectivism on perceived risks and psychological maladjustment. *International journal of environmental research and public health*, *17*(10), 3497.

Hammen C. (2005). Stress and depression. *Annual review of clinical psychology*, *1*, 293–319. https://doi.org/10.1146/annurev.clinpsy.1.102803.143938

Holmes, E. A., O'Connor, R. C., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Cohen Silver, R., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A. K., Shafran, R., Sweeney, A., Worthman, C. M., … Bullmore, E. (2020). Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The lancet. Psychiatry*, *7*(6), 547–560. https://doi.org/10.1016/S2215-0366(20)30168-1

Ive, J., Gkotsis, G., Dutta, R., Stewart, R., & Velupillai, S. (2018). Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic* (pp. 69-77).

Jackson, S. B., Stevenson, K. T., Larson, L. R., Peterson, M. N., & Seekamp, E. (2021). Outdoor Activity Participation Improves Adolescents' Mental Health and Well-Being during the COVID-19 Pandemic. *International journal of environmental research and public health*, *18*(5), 2506. https://doi.org/10.3390/ijerph18052506

Jeffers, A., Meehan, A. A., Barker, J., Asher, A., Montgomery, M. P., Bautista, G., ... & Marcus, R. (2022). Impact of Social Isolation during the COVID-19 Pandemic on Mental Health, Substance Use, and Homelessness: Qualitative Interviews with Behavioral Health Providers. *International Journal of Environmental Research and Public Health*, *19*(19), 12120.

Jia, H., Guerin, R. J., Barile, J. P., Okun, A. H., McKnight-Eily, L., Blumberg, S. J., Njai, R., & Thompson, W. W. (2021). National and State Trends in Anxiety and Depression Severity Scores Among Adults During the COVID-19 Pandemic - United States, 2020-

2021. *MMWR. Morbidity and mortality weekly report*, *70*(40), 1427–1432. https://doi.org/10.15585/mmwr.mm7040e3

Judd, L. L., Akiskal, H. S., Maser, J. D., Zeller, P. J., Endicott, J., Coryell, W., Paulus, M. P., Kunovac, J. L., Leon, A. C., Mueller, T. I., Rice, J. A., & Keller, M. B. (1998). A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. *Archives of general psychiatry*, *55*(8), 694–700. https://doi.org/10.1001/archpsyc.55.8.694

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *The American journal of psychiatry*, *161*(4), 631–636. https://doi.org/10.1176/appi.ajp.161.4.631

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., Ormel, J., Petukhova, M., McLaughlin, K. A., Green, J. G., Russo, L. J., Stein, D. J., Zaslavsky, A. M., Aguilar-Gaxiola, S., Alonso, J., Andrade, L., Benjet, C., de Girolamo, G., de Graaf, R., Demyttenaere, K., Fayyad, J., Haro, J. M., Hu, C.y, Karam, A., Lee, S., … Ustün, T. B. (2011). Development of lifetime comorbidity in the World Health Organization world mental health surveys. *Archives of general psychiatry*, *68*(1), 90–100. https://doi.org/10.1001/archgenpsychiatry.2010.180

Killgore, W. D. S., Taylor, E. C., Cloonan, S. A., & Dailey, N. S. (2020). Psychological resilience during the COVID-19 lockdown. *Psychiatry research*, *291*, 113216. https://doi.org/10.1016/j.psychres.2020.113216

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lai, J., Ma, S., Wang, Y., Cai, Z., Hu, J., Wei, N., Wu, J., Du, H., Chen, T., Li, R., Tan, H., Kang, L., Yao, L., Huang, M., Wang, H., Wang, G., Liu, Z., & Hu, S. (2020). Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. *JAMA network open*, *3*(3), e203976. https://doi.org/10.1001/jamanetworkopen.2020.3976

Lamers, F., van Oppen, P., Comijs, H. C., Smit, J. H., Spinhoven, P., van Balkom, A. J., Nolen, W. A., Zitman, F. G., Beekman, A. T., & Penninx, B. W. (2011). Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *The Journal of clinical psychiatry*, *72*(3), 341–348. https://doi.org/10.4088/JCP.10m06176blu

Loades, M. E., Chatburn, E., Higson-Sweeney, N., Reynolds, S., Shafran, R., Brigden, A., Linney, C., McManus, M. N., Borwick, C., & Crawley, E. (2020). Rapid Systematic Review: The Impact of Social Isolation and Loneliness on the Mental Health of Children and Adolescents in the Context of COVID-19. *Journal of the American Academy of Child and Adolescent Psychiatry*, *59*(11), 1218–1239.e3. https://doi.org/10.1016/j.jaac.2020.05.009

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical care*, *46*(3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093

Mello, A. (2020). XGBoost: Theory and practice. *Medium https://towardsdatascience. com/xgboost-theory-and-practice-fb8912930ad6*.

Moriarty, D. G., Zack, M. M., & Kobau, R. (2003). The Centers for Disease Control and Prevention's Healthy Days Measures - population tracking of perceived physical and mental health over time. *Health and quality of life outcomes*, *1*, 37. https://doi.org/10.1186/1477-7525-1-37

Parker, K., Minkin, R., & Bennett, J. (2020). Economic fallout from COVID-19 continues to hit lower-income Americans the hardest. Pew Research Center. Retrieved from https://www.pewresearch.org/social-trends/2020/09/24/economic-fallout-from-covid-19-continues-to-hit-lower-income-americans-the-hardest/

Palgi, Y., Shrira, A., Ring, L., Bodner, E., Avidor, S., Bergman, Y., Cohen-Fridel, S., Keisari, S., & Hoffman, Y. (2020). The loneliness pandemic: Loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. *Journal of affective disorders*, *275*, 109–111. https://doi.org/10.1016/j.jad.2020.06.036

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. M., Huang, Y., Jordans, M. J. D., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., … UnÜtzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *Lancet (London, England)*, *392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

Paykel, E. S. (2003). Life events and affective disorders. *Acta Psychiatrica Scandinavica*, *108*, 61-66. Retrieved from: https://doi.org/10.1034/j.1600-0447.108.s418.13.x

Pfefferbaum, B., & North, C. S. (2020). Mental Health and the Covid-19 Pandemic. *The New England journal of medicine*, *383*(6), 510–512. https://doi.org/10.1056/NEJMp2008017

Rajkumar R. P. (2020). COVID-19 and mental health: A review of the existing literature. *Asian journal of psychiatry*, *52*, 102066. https://doi.org/10.1016/j.ajp.2020.102066

Salari, N., Hosseinian-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., Rasoulpoor, S., & Khaledi-Paveh, B. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. *Globalization and health*, *16*(1), 57. https://doi.org/10.1186/s12992-020-00589-w

Saeed, H., Eslami, A., Nassif, N. T., Simpson, A. M., & Lal, S. (2022). Anxiety Linked to COVID-19: A Systematic Review Comparing Anxiety Rates in Different Populations. *International journal of environmental research and public health*, *19*(4), 2189. https://doi.org/10.3390/ijerph19042189

Seedat, S., Scott, K. M., Angermeyer, M. C., Berglund, P., Bromet, E. J., Brugha, T. S., Demyttenaere, K., de Girolamo, G., Haro, J. M., Jin, R., Karam, E. G., Kovess-Masfety, V., Levinson, D., Medina Mora, M. E., Ono, Y., Ormel, J., Pennell, B. E., Posada-Villa, J., Sampson, N. A., Williams, D., … Kessler, R. C. (2009). Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. *Archives of general psychiatry*, *66*(7), 785–795. https://doi.org/10.1001/archgenpsychiatry.2009.36

Shanafelt, T., Ripp, J., & Trockel, M. (2020). Understanding and Addressing Sources of Anxiety Among Health Care Professionals During the COVID-19 Pandemic. *JAMA*, *323*(21), 2133–2134. https://doi.org/10.1001/jama.2020.5893

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled clinical trials*, *24*(2), 182–199. https://doi.org/10.1016/s0197-2456(02)00320-3

Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive behaviour therapy*, *40*(2), 126–136. https://doi.org/10.1080/16506073.2010.550059

Torales, J., O'Higgins, M., Castaldelli-Maia, J. M., & Ventriglio, A. (2020). The outbreak of COVID-19 coronavirus and its impact on global mental health. *The International journal of social psychiatry*, *66*(4), 317–320. https://doi.org/10.1177/0020764020915212

Twenge, J. M., & Joiner, T. E. (2020). Mental distress among U.S. adults during the COVID-19 pandemic. *Journal of clinical psychology*, *76*(12), 2170–2182. https://doi.org/10.1002/jclp.23064

VanderWeele, T. J., Fulks, J., Plake, J. F., & Lee, M. T. (2021). National Well-Being Measures Before and During the COVID-19 Pandemic in Online Samples. *Journal of general internal medicine*, *36*(1), 248–250. https://doi.org/10.1007/s11606-020-06274-3

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., Ho, C. S., & Ho, R. C. (2020). Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in

China. *International journal of environmental research and public health*, *17*(5), 1729. https://doi.org/10.3390/ijerph17051729

Wang, F., Pan, F., Tang, Y. Y., & Huang, J. H. (2022). Editorial: Uncertainty Induced Emotional Disorders During the COVID-19. *Frontiers in psychology*, *13*, 943966. https://doi.org/10.3389/fpsyg.2022.943966

Wiegele, P. N., Kabar, I., Kerschke, L., Froemmel, C., Hüsing-Kabar, A., Schmidt, H., Vorona, E., Vollenberg, R., & Tepasse, P. R. (2021). Symptom Diary-Based Analysis of Disease Course among Patients with Mild Coronavirus Disease, Germany, 2020. Emerging infectious diseases, 27(5), 1353–1361. https://doi.org/10.3201/eid2705.204507

World Health Organization. (2021). Comprehensive mental health action plan 2013–2030. World Health Organization. https://apps.who.int/iris/handle/10665/345301. License: CC BY-NC-SA 3.0 IGO

Appendix A: Data dictionary for each column in the table

Indicator: Symptoms of Anxiety Disorder or Depressive Disorder

Description: The dataset focuses on symptoms of anxiety disorder or depressive disorder in participants.

Group: Separating Participants into Categories

Description: Participants are divided into different categories, including states, age groups, gender, etc.

State: Participant's State

Description: The state in the United States where the participant is from.

Subgroup: Participant's Specific Category

Description: The specific category each participant falls under within the broader categories.

Phase: Data Collection Phase

Description: The stage of data collection during which the participant's information was gathered.

Time_period: Time Period Number

Description: The numeric identifier for the period in which the data was collected.

Time_period_label: Time Period Duration

Description: The duration of time elapsed within the specific data collection phase.

Value: Symptom Severity Value

Description: In the dataset, the numerical value assigned to depressive or anxiety disorder symptoms represents the severity of symptoms experienced by individuals diagnosed with these conditions. However, it's essential to note that depression and anxiety are complex conditions that can present differently in different individuals, and a single numerical value cannot capture the full range of experiences. For the sake of simplicity in this thesis, depression and anxiety levels will be treated as equivalent.

Low CI: Lowest Confidence Interval Value

Description: The lowest possible value that the true population parameter could take.

High CI: Highest Confidence Interval Value

Description: The highest possible value that the true population parameter could take.

Confidence Interval: Range of Values with Confidence Level

Description: A range of values that is likely to contain the true value of a population parameter (such as the mean or the proportion) with a certain level of confidence or probability.

Academic Vita

Qiuhao Zhu
20010426alex@gmail.com

EDUCATION

The Pennsylvania State University, Schreyer Honors College University Park, PA

B.S. Information Sciences and Technology **Graduation Date**: May 2023

Achievements: Dean's List 9/9 Semester.

Relevant Courses: Java Programming, Python Pandas, Project Management.

PROFESSIONAL EXPERIENCE

Penn State Office of Research Information Systems University Park, PA

Intern May 2022 – Aug 2022

- Created a data dictionary by joining different tables and crossing reference primary keys and foreign keys.
- Defined data objects in a relational database and built more than ten Entity Relational Diagrams to master SQL skills.
- Added a detailed description of the system governing financial research and compliance to ensure faculty at Penn State adhere to federal regulations and become more knowledgeable in stakeholder analysis.

University Professional Continuing Education Association Remote

Intern May 2021 – Nov 2021

- Analyzed and transformed institutional performance data with Excel and SPSS improving student retention rates by 10% and advancing the university's mission.
- Employed an empirical tool in Excel to analyze clients' online assets to isolate institutional strengths and weaknesses.
- Developed a creative method for benchmarking student-university interaction to gauge university timeliness.

PROJECT EXPERIENCE

Java Programming Class University Park, PA

Project Lead Aug 2022 – Dec 2022

- Built an API system using JavaScript and HTML to allow users to access information about the book in a library system.
- Implemented RESTful API by enabling customers to make/edit reservations in the system using Node.js, and Express.js.
- Employed rigorous testing and validation procedures with a peer, simulating real-world user interactions to optimize the system's functionality, performance, and user satisfaction.

LEADERSHIP & PROFESSIONAL DEVELOPMENT EXPERIENCES

Nittany Data Labs University Park, PA

Vice President Aug 2022 – Present

- Hosting Twitter API workshops for lowerclassmen to help improve their coding skills.
- Sharing academic and career resources to lowerclassmen in Discord to provide additional learning materials to them.
- Organizing and participating in Smoothie Crawl to visualize and extract the key information from datasets.

SKILLS & INTERESTS

Technical Skills: Java, API, JavaScript, Node.js, Express.js, CSS, React, Python, AWS, ML, R, SQL.