EXPLORING MENTAL HEALTH TREATMENT IN THE TECH WORKPLACE: A DATA-DRIVEN ANALYSIS

ISYE7406 — Data Mining & Statistical Learning Final Report

Group 22

Yael Benarroch; 609; ybenarroch3@gatech.edu

Sukriti Raut; 889; sraut9@gatech.edu

Kann Y Tan; 142; ktan71@gatech.edu

Shruti Rajagopal; 901; srajagopal37@gatech.edu

April 16th, 2023

TABLE OF CONTENTS

1. **ABSTRACT** …..3
2. **INTRODUCTION**..3
3. **DATA**…..4
   - *i.* ***Data Source***…..4
   - **ii.** ***Dataset***…..5
   - *iii.* ***Data Preprocessing***…..6
4. **EXPLORATORY DATA ANALYSIS**..7
   - **i.** ***Understanding respondent's & their company's general characteristics***..................8
   - *ii.* ***Understanding relationship between various predictors & treatment***......................9
5. **METHODOLOGY**..13
6. **RESULTS**..14
7. **FINDINGS & DISCUSSION**..20
8. **LESSONS LEARNED**..21
9. **CONCLUSION**..21
10. **APPENDIX**..22
11. **REFERENCES**..24

1. ABSTRACT

In this report, we analyze and present the results of our project that aims to identify the factors associated with an employee seeking treatment for mental health conditions in the workplace, with a focus on technology companies. In our analysis, we utilized various data mining and statistical learning methods, to analyze a survey dataset. The models built to address our research questions showed how numerous factors can play a part in influencing an employee's decision to seek treatment for mental health conditions such as age, gender, work environment, and access to mental health benefits. By utilizing our developed models, we were able to predict which employees are more likely to seek treatment based on demographic and work-related factors, with a moderately high accuracy. Our study adds to the growing body of literature addressing mental health in the workplace and highlights the importance of addressing the stigma surrounding mental health issues. The results of our study could inform policies and practices in the tech workplace and provide guidance for future research in this field.

2. INTRODUCTION

There is no question that sound mental health is as significant as good physical health for overall well-being. It is especially important to feel mentally healthy and fit to be able to perform effectively at the workplace. Among working-age Americans, depression interferes with the ability to complete physical job tasks about 20% of the time and reduces cognitive performance about 35% of the time[1]. It can also be costly for employers and businesses if they don't make mental health a priority. A recent statistic by WHO states that an estimated 12 billion working days are lost globally every year to depression and anxiety at a cost of US$ 1 trillion per year in lost productivity[2]. Therefore, the workplace where we spend one third of our lives can also provide a safe environment and resources to prevent mental health risks and support workers with mental health conditions. The increasing awareness of mental health in the workplace has led to efforts by employers to support employees in seeking treatment. Despite this, many employees still do not seek help for mental health conditions for various reasons. In the U.S., only 57% of employees who report moderate depression and 40% of those who report severe depression receive treatment to control depression symptoms[3].

With this research project, our main goals were to identify the prevalence of mental health conditions at the workplace and the resulting negative consequences, the association between several factors with mental health treatment, and finally to predict the likelihood of employees seeking treatment based on certain characteristics. We formulated the following four questions to address our research goals:

1. What factors are associated with an employee seeking treatment for mental health conditions?
2. Can we predict which employees are more likely to seek treatment for mental health conditions based on their demographic information and work-related factors?
3. Does a company that support/provide mental health benefit lead to more employees seeking treatment for mental health conditions?
4. Which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions?

In this report, we employed various classification and ensemble methods, such as correlation analysis, logistic regression, random forests, boosting, among others, to answer these questions.

3. DATA

i. Data Source

The publicly available dataset utilized for this research project was obtained from Kaggle[4]. It comes from a 2014 survey compiled by the Open Sourcing Mental Illness (OSMI) project. The survey includes responses from over 1200 employees working in the technology and non-technology industries. The survey data covers a broad range of areas using 26 variables, which we grouped into four categories as shown in the *Dataset* section below. The grouping of variables was created to simplify our exploratory data analysis efforts and to create a more targeted approach in our analysis of each research question. The purpose of the dataset is to be used for research, analysis, and other projects related to mental health in the workplace.

ii. Dataset

Category	Variables	Survey Questions
Demographic	Age	
	Gender	
	Country	
	state	If you live in the United States, which state or territory do you live in?
	self_employed	Are you self-employed?
	family_history	Do you have a family history of mental illness?
Company Characteristics	no_employees	How many employees does your company or organization have?
	remote_work	Do you work remotely (outside of an office) at least 50% of the time?
	tech_company	Is your employer primarily a tech company/organization?
Company's Mental Health Services	benefits	Does your employer provide mental health benefits?
	care_options	Do you know the options for mental health care your employer provides?
	wellness_program	Has your employer ever discussed mental health as part of an employee wellness program?
	seek_help	Does your employer provide resources to learn more about mental health issues and how to seek help?
	anonymity	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
	leave	How easy is it for you to take medical leave for a mental health condition?
Employee's Perception and Attitudes	work_interfere	If you have a mental health condition, do you feel that it interferes with your work?
	mental_health_consequence	Do you think that discussing a mental health issue with your employer would have negative consequences?
	phys_health_consequence	Do you think that discussing a physical health issue with your employer would have negative consequences?
	coworkers	Would you be willing to discuss a mental health issue with your coworkers?
	supervisor	Would you be willing to discuss a mental health issue with your direct supervisor(s)?
	mental_health_interview	Would you bring up a mental health issue with a potential employer in an interview?
	phys_health_interview	Would you bring up a physical health issue with a potential employer in an interview?
	mental_vs_physical	Do you feel that your employer takes mental health as seriously as physical health?
Response variable	obs_consequence	Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
Response variable	treatment	Have you sought treatment for a mental health condition?

Table 1 - All Variables & Descriptions

Responses to the survey questions were Yes/No (and occasionally Maybe/Don't Know), thus data type for almost all variables was categorical. Only the *age variable* was continuous which was re-coded into age groupings as categorical values. A quick preview of a few rows of the dataset looks like below. You may read more about the survey and OSMI project by following this link: **https://osmhhelp.org/research.html**

	Example.1	Example.2	Example.3	Example.4	Example.5
Age	37	44	32	31	31
Gender	Female	M	Male	Male	Male
Country	United States	United States	Canada	United Kingdom	United States
state	IL	IN	NA	NA	TX
self_employed	NA	NA	NA	NA	NA
family_history	No	No	No	Yes	No
treatment	Yes	No	No	Yes	No
work_interfere	Often	Rarely	Rarely	Often	Never
no_employees	6–25	More than 1000	6–25	26–100	100–500
remote_work	No	No	No	No	Yes
tech_company	Yes	No	Yes	Yes	Yes
benefits	Yes	Don't know	No	No	Yes
care_options	Not sure	No	No	Yes	No
wellness_program	No	Don't know	No	No	Don't know
seek_help	Yes	Don't know	No	No	Don't know
anonymity	Yes	Don't know	Don't know	No	Don't know
leave	Somewhat easy	Don't know	Somewhat difficult	Somewhat difficult	Don't know
mental_health_consequence	No	Maybe	No	Yes	No
phys_health_consequence	No	No	No	Yes	No
coworkers	Some of them	No	Yes	Some of them	Some of them
supervisor	Yes	No	Yes	No	Yes
mental_health_interview	No	No	Yes	Maybe	Yes
phys_health_interview	Maybe	No	Yes	Maybe	Yes
mental_vs_physical	Yes	Don't know	No	No	Don't know
obs_consequence	No	No	No	Yes	No

Figure 1 - Dataset Preview

iii. Data Preprocessing

There were various character fields that required cleaning for formatting inconsistencies and typing errors. For example, *Gender* had the most variability, which was recoded as “Male”, “Female”, and “Other”. Similarly, a new variable *AgeGroup* was created and defined as “YoungAdult” (*Age* <= 30), “MiddleAgedAdult” (31 <= *Age* <= 45), and “OldAdults” (*Age* > 45).

Some variables in the dataset contained a significant number of missing values as shown below.

Column	Missing_Values
anonymity	819
mental_vs_physical	576
leave	563
state	515
seek_help	363
self_employed	314
care_options	314
work_interfere	264
wellness_program	188
Age	7
Gender	0
Country	0
family_history	0
treatment	0
no_employees	0
remote_work	0
tech_company	0
benefits	0
mental_health_consequence	0
phys_health_consequence	0
coworkers	0
supervisor	0
mental_health_interview	0
phys_health_interview	0
obs_consequence	0

Figure 2 - Variables With Missing Values

To keep the dataset size intact and not lose any critical information, we used the MICE *R* package to impute the missing values rather than remove such records completely[5]. This package generates imputations for each column with missing values based on predictions using other variables. Different methods of imputation were utilized based on column's data type:

- "logreg" for factor level exactly 2
- "polyreg" for factor level greater than 2
- "norm" for the continuous data such as 'age'

We then assessed the distribution of the imputed columns against the original data to confirm the plausibility of these new values. We found the distribution of the imputed values to be similar to the distribution of the records with no missing values. For example, the comparative distribution for responses to whether employees feel that their mental health issues interfere with their work *work_interfere* looks similar as shown below where predicted are the imputed values.

```
actual
    Never     Often    Rarely Sometimes
      213       144       173       465
predicted
    Never     Often    Rarely Sometimes
      351       160       209       539
```

Figure 3 - Data Imputation Results

4. EXPLORATORY DATA ANALYSIS

The two response variables used in this analysis are: *treatment*, which indicates whether the employee has sought treatment for a mental health condition, and *obs_consequence*, which indicates whether an employee has heard or observed any negative consequence for coworkers with mental health conditions in their workplace. The EDA is divided into two sections below:

i. Understanding respondent's & their company's general characteristics

We performed exploratory analysis on the distribution of the various predictors to understand our population better, which is crucial for making any assumptions. The key findings are presented based on the four categories of variables. Charts representing these findings are included in the Appendix.

Demographic

- Most respondents are male and young and middle-aged adults (20-45 yrs old).
- Most are employed with a company (i.e., not self-employed).
- Most employees do not have a family history of mental illness.

*Note that geographic data such as state and country were sparsely populated and not reliable, therefore analysis by geography is not a focus for this study. *

Company Characteristics

- Most employees worked for companies with small employee sizes 6-25 and 25-100 and big employee size >1000.
- Most work at a tech company.
- Most work remotely more than 50% of the time.

Company's Mental Health Services

- About 40% of the employees said that their employers provided mental health benefits whereas the other 60% (more than half) said they did not or were unaware.
- Most employees think that their anonymity will be protected if they choose to take advantage of mental health or substance abuse treatment resources.
- Based on distributions of variables *seek_help*, *care_options*, and *wellness_programs* which deal with promoting mental health care and providing extra resources on mental health issues, most employees responded negatively.
- Most respondents said they found it easy or somewhat easy to take medical leave for a mental health issue.

Employee's Perception and Attitudes

- Most respondents are willing to discuss mental health issues with at least some of their coworkers and supervisors.
- Most respondents said they would not mention mental health issues in a job interview; however, most would maybe bring up a physical health issue.
- Most respondents have also said that they feel their mental health condition sometimes or often interferes with their work.

ii. Understating the relationship between various predictors & treatment

Exploring the relationship between the predictors and our target variable treatment, we tried to understand the prevalence of mental health conditions at the workplace and identify any existence of class biases. They would be important for making fair inferences from the findings. Since most predicting and dependent variables were categorical, it made sense to visualize the relationships using bar plots.

The graphs below show the proportions rather than the counts calculated from contingency tables. In these plots, dark red corresponds to the proportions of employees who did not seek mental health treatment and green corresponds to the proportions of employees who did seek mental health treatment.

From these plots, we see that the proportion of the treatment vs. non-treatment group is similar within each level or factor but vary a lot when compared to another factor. For e.g., half of the remote workers have sought treatment for mental health whereas half of them did not. Similarly, half of the non-remote workers sought treatment and half of them did not. However, the total proportion of employees who work remotely is significantly less than who commute. This seems to be the trend for most of the variables in all four broad categories. This tells us that class (treatment vs. non-treatment) bias might not be an issue with this data set. It also provides some evidence that several of these variables do have a relationship with the outcome variable and might possess some significant predictive power, which we will examine further through the modelling process.

Demographic:

Figure 4 - Exploratory Analysis for Demographic Predictors & Treatment

*Note that geographic data such as state and country were sparsely populated and not reliable, therefore analysis by geography is not a focus for this study. *

Company's Characteristics:

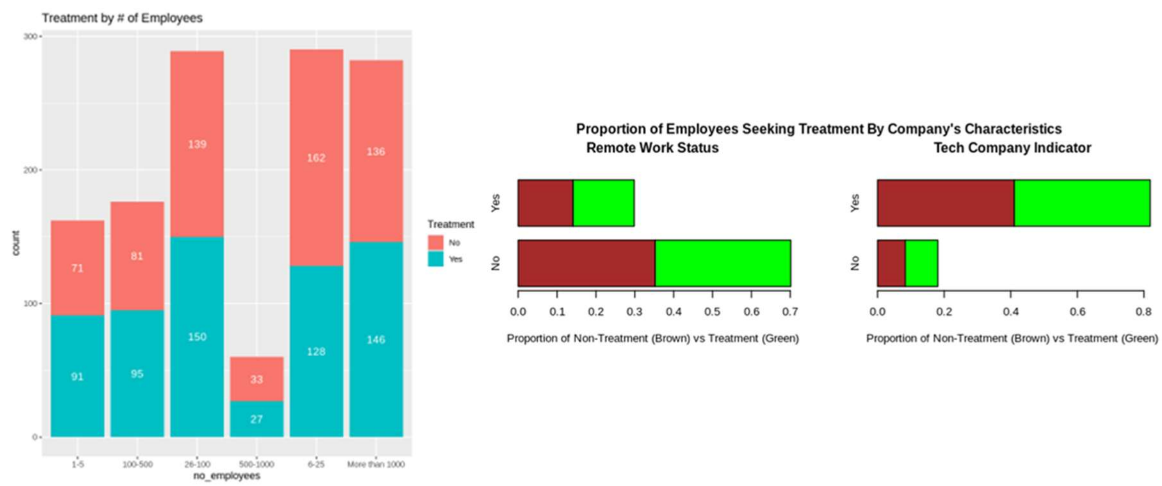


Figure 5 - Exploratory Analysis for Company Characteristics & Treatment

Company's Mental Health Services:

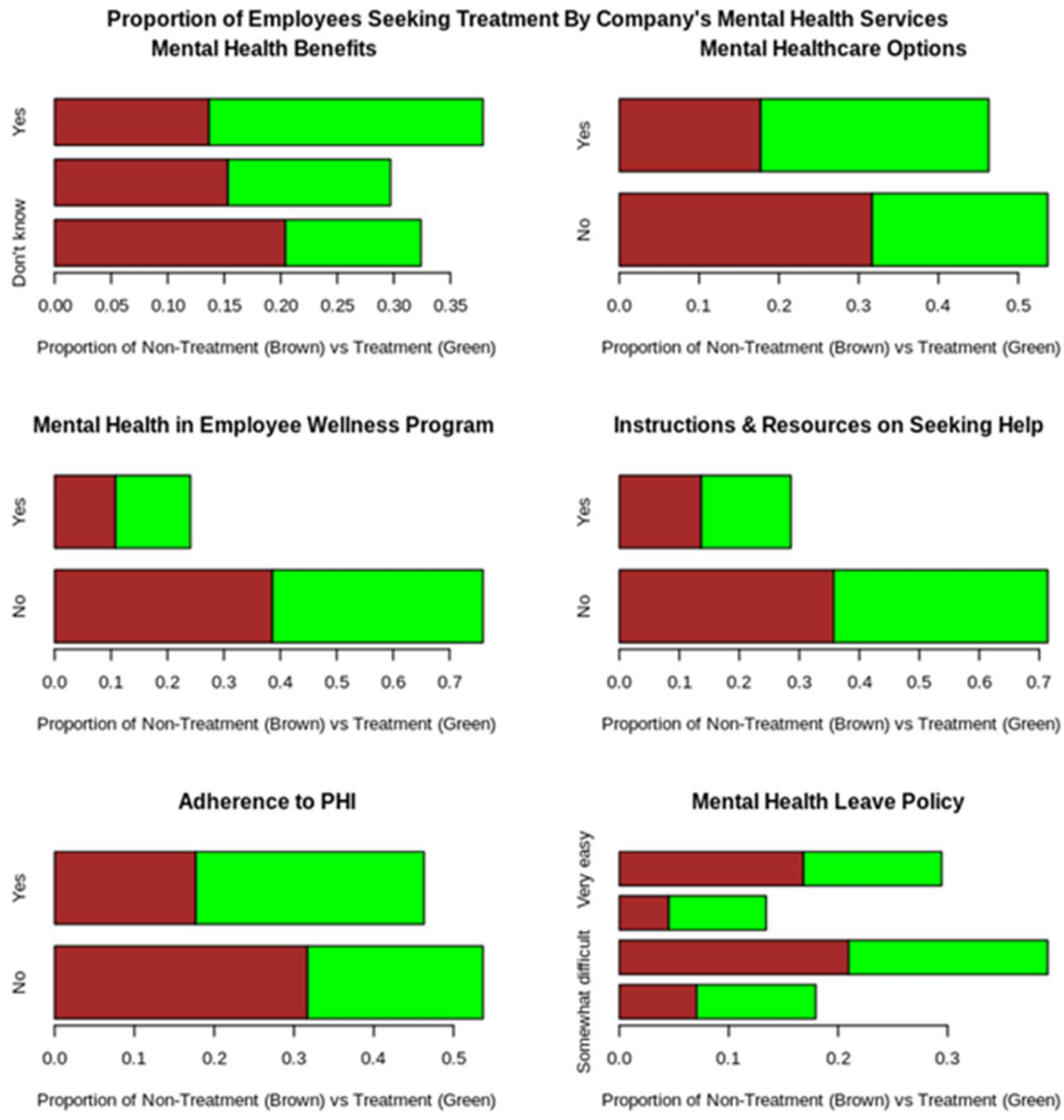
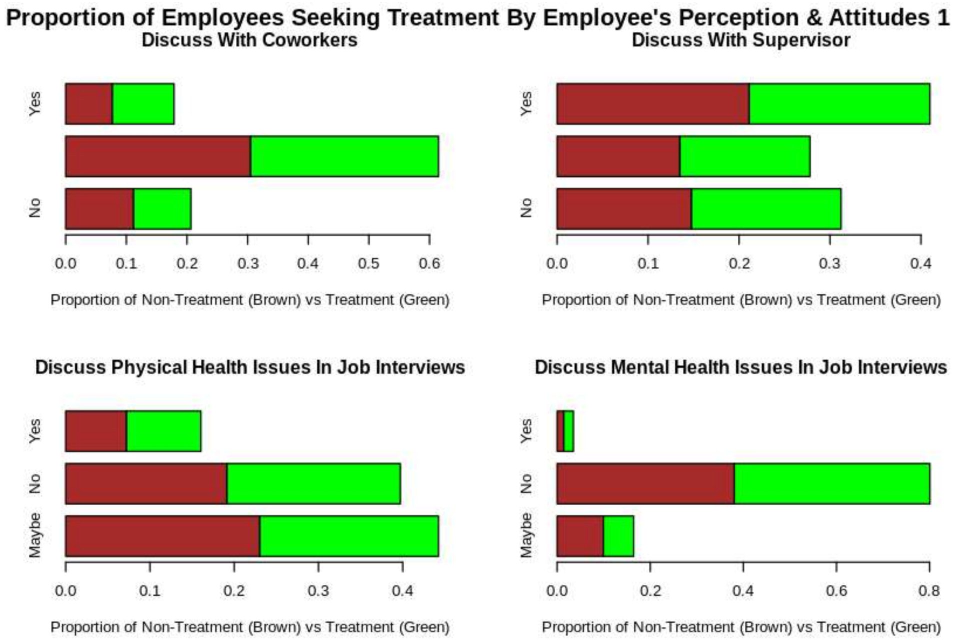


Figure 6 - Exploratory Analysis for Company Mental Health Services & Treatment

Employee's Perception and Attitudes:



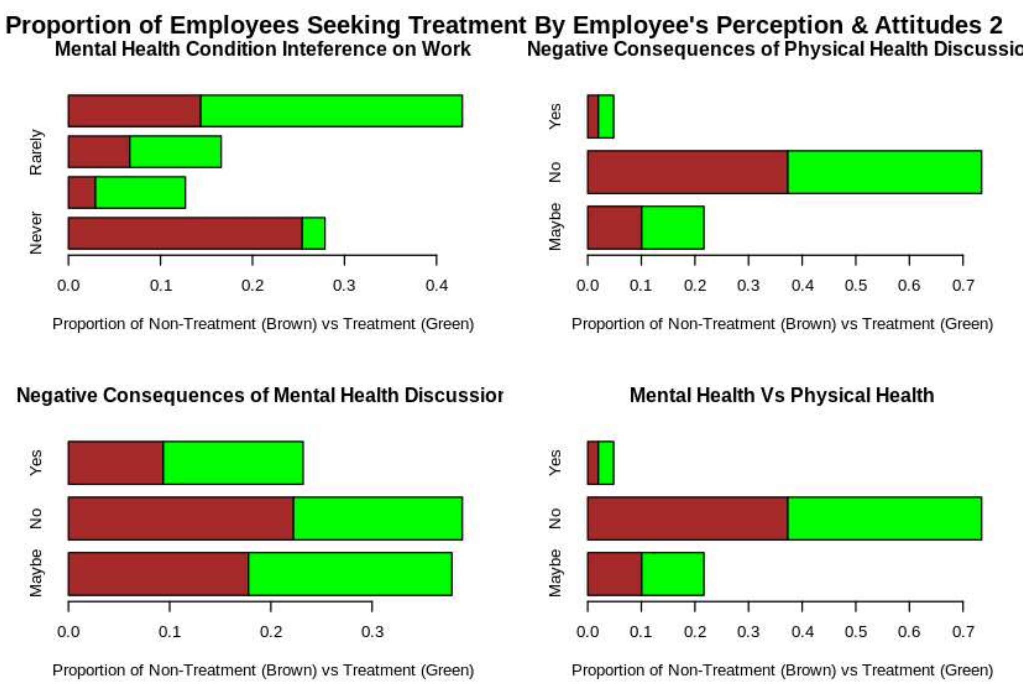


Figure 7 - Exploratory Analysis for Employee Perceptions & Treatment

5. METHODOLOGY

We split the dataset into train (80%) and test (20%) sets, and used the following methods for each of the four research questions:

Research Questions	Response Variable (Y)	Method(s)	Tuning Methods/Parameters	Evaluation Metrics
What factors are associated with an employee seeking treatment for mental health conditions?	*treatment*	1. Correlation Analysis		Chi-squared tests, Cramer's V
		2. Logistic Regression	10-Fold Cross Validation	Log Likelihood, Accuracy Rates
		3. Decision Tree		Accuracy Rates
Can we predict which employees are more likely to seek treatment for mental health conditions based on their demographic information and work-related factors?	*treatment*	1. Gradient Boosting	10-Fold Cross-Validation: Number of trees, Tree Depth, Learning Rate (shrinkage), Terminal Node Size	Relative Influence Factors, Accuracy Rates
		2. KNN	Number of neighbors (K)	AUC, Accuracy Rates
Does a company that support/provide mental health benefit leads to more employee seeking treatment for mental health conditions?	*treatment*	1. Random Forest		Accuracy Rates
		2. Logistic Regression	10-Fold Cross Validation	Accuracy Rates
Which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions?	*obs_consequence*	1. Recursive Feature Elimination (RFE) with Random Forest	10-Fold Cross Validation	CV Accuracy, Feature Importance
		2. Logistic Regression	No. Of Predictors selected with RFE	Accuracy Rates

Table 2 – Methodology

6. RESULTS

Research Questions	Response Variable (Y)	Models/Methods	Testing Accuracy Rate
What factors are associated with an employee seeking treatment for mental health conditions?	*treatment*	1. Correlation Analysis	
		2. Logistic Regression	77.10%
		3. Decision Tree	75.40%
Can we predict which employees are more likely to seek treatment for mental health conditions based on their demographic information and work-related factors?	*treatment*	1. Gradient Boosting	79.20%
		2. KNN	69.30%
Does a company that support/provide mental health benefit leads to more employee seeking treatment for mental health conditions?	*treatment*	1. Random Forest	64.95%
		2. Logistic Regression	63.47%
Which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions?	*obs_consequence*	1. Recursive Feature Elimination (RFE) with Random Forest	
		2. Logistic Regression	85.60%

Table 3 - Analysis Results

1.What factors are associated with an employee seeking treatment for mental health conditions?

Correlation Analysis:

Using the chi-squared test, we examined the relationship between 23 variables and employees seeking mental health treatment. Several factors were significantly associated with treatment-seeking, including *family_history*, *work_interfere*, and *mental_health_consequence*. The highest Cramer's V values, indicating the strength of association between variables, were observed for *work_interfere*, *family_history*, and *state*.

Logistic Regression:

Our analysis revealed that males are less likely to seek treatment for mental health conditions compared to females. Self-employed individuals, employees with a family history of mental health conditions, and those who experience work interference due to mental health issues are significantly more likely to seek treatment when compared to their opposite grouping population in each variable. Additionally, mental health benefits and easier leave policies are associated with a higher likelihood of seeking treatment. Employees who perceive no mental health consequences at work are less likely to seek treatment, while those willing to discuss mental health with all coworkers are more likely to seek treatment.

Decision Tree:

Work interference emerged as the most important factor associated with an employee seeking treatment for mental health conditions. State and family history of mental health conditions also played significant roles in the decision-making process. Employees experiencing work interference often or residing in a state not listed in the decision tree model were more likely to seek treatment.

	variable	chi_squared	p_value	df	cramers_v
X-squared	Gender	51.08	8.108e-12	2	0.2014
X-squared1	Country	69.8	0.01706	47	0.2355
X-squared2	state	96.67	8.052e-06	44	0.2771
X-squared3	self_employed	91.7	1.01e-21	1	0.2699
X-squared4	family_history	178.3	1.158e-40	1	0.3763
X-squared5	work_interfere	350.2	1.34e-75	3	0.5274
X-squared6	no_employees	8.765	0.1188	5	0.08344
X-squared7	remote_work	0.7996	0.3712	1	0.0252
X-squared8	tech_company	1.093	0.2958	1	0.02946
X-squared9	benefits	64.84	8.327e-15	2	0.2269
X-squared10	care_options	53.21	2.993e-13	1	0.2056
X-squared11	wellness_program	3.027	0.08187	1	0.04904
X-squared12	seek_help	0.4464	0.5041	1	0.01883
X-squared13	anonymity	19.83	8.48e-06	1	0.1255
X-squared14	leave	37.98	2.858e-08	3	0.1737
X-squared15	mental_health_consequence	22.33	1.418e-05	2	0.1332
X-squared16	phys_health_consequence	3.371	0.1854	2	0.05174
X-squared17	coworkers	6.001	0.04975	2	0.06904
X-squared18	supervisor	1.726	0.422	2	0.03702
X-squared19	mental_health_interview	12.69	0.001755	2	0.1004
X-squared20	phys_health_interview	3.4	0.1827	2	0.05196
X-squared21	mental_vs_physical	4.473	0.03444	1	0.0596
X-squared22	obs_consequence	30.14	4.021e-08	1	0.1547
X-squared23	age_group	5.794	0.05519	2	0.06784

Table: Association Analysis Results

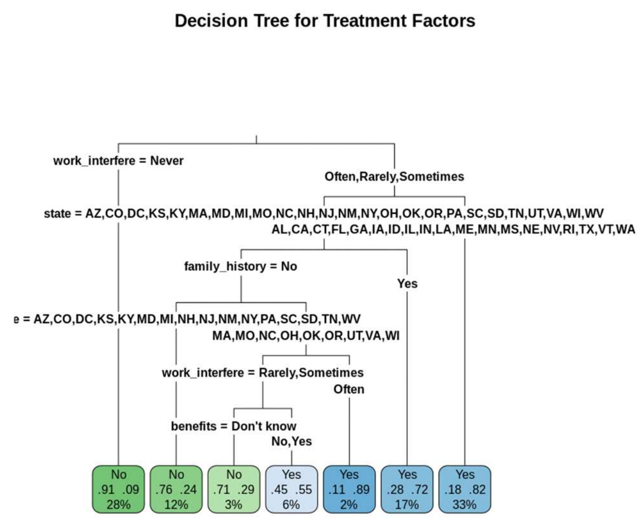


Figure 8 - Correlation Analysis & Decision Tree Results

2. Can we predict which employees are more likely to seek treatment for mental health conditions based on their demographic information and work-related factors?

Gradient Boosting:

Based on 10-fold cross-validation, the best selected model used 60 trees and tree depth of 3 at learning rate 0.1 with 5 as the minimum number of observations in terminal nodes. Based on relative influence factors of the best selected model, the top five demographic & work-related variables that were important in predicting the likelihood of employee seeking mental health treatment were *family_history*, *work_interfere*, *benefits*, *age*, and *gender*. Ensemble methods like boosting can have a strong predictive power (train and test accuracy ~ 80%) as it trains weaker models to create aggregated models to improve prediction and can handle non-linear data better. It is, however, less interpretable.

	n.trees	interaction.depth	shrinkage	n.minobsinnode
	<dbl>	<dbl>	<dbl>	<dbl>
92	60	3	0.1	5

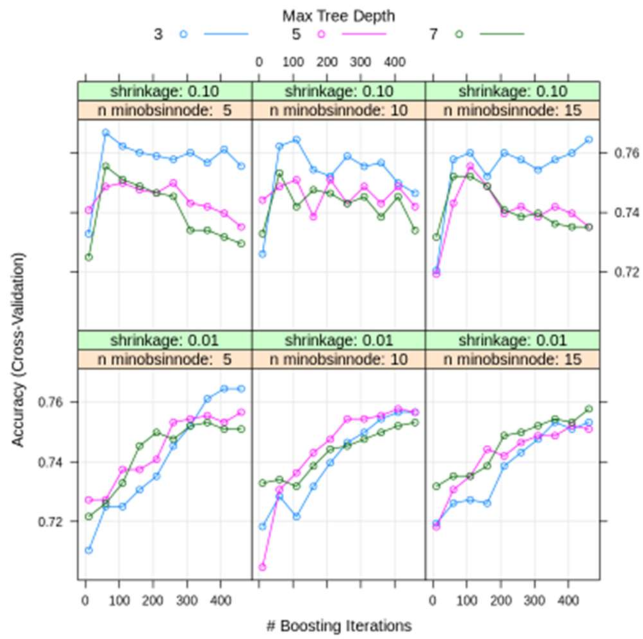


Figure 9 – Gradient Boosting Results

KNN:

The K Nearest Neighbors (KNN) model fitted with hyper-tuned parameters using cross-validation selected 23 as the optimal number of trees. Based on high number of K neighbors, moderate AUCs of 0.51 & 0.42 for train and test datasets, and <74% accuracy rates, KNN did not seem like a strong prediction model for this dataset. The low performance of the model could be due to the high number of dimensions and binary nature of the variables.

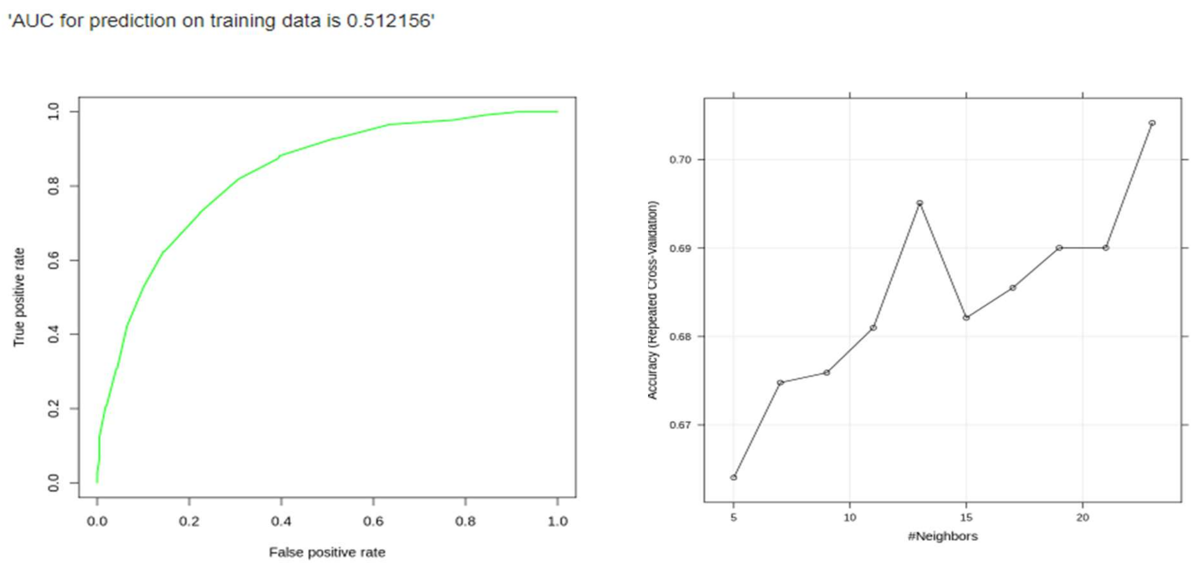


Figure 10 - KNN Results & Prediction Accuracy

3. Does a company that supports/provides mental health benefits lead to more employees seeking treatment for mental health conditions?

Logistic Regression:

Fitting the model with only variables related to mental health benefits provided by the company, the output indicated that employees in companies with mental health benefits are more likely to seek treatment compared to those without such benefits. Similarly, employees working in companies that have more open discussion among coworkers without negative consequences, provide resources to learn more about mental health issues, and provide medical leave related to mental health are more likely to seek treatment compared to their baseline groups. Anonymity protection is also important in a work environment for employees if they opt to seek mental health or substance abuse treatment.

Random Forest:

Random Forest fitted with only variables in scope like the logistic model above, we found that the important factors that decide if employees are more likely to seek treatments are *benefits*, *care_options*, *anonymity*, *leave* and *coworkers*. The findings of the random forest model are consistent with the findings of the logistic regression model. A company that is supportive of mental illness tends to have employees that are more willing to seek treatment. Ways a company can create a positive mental illness environment includes providing mental health care benefits, anonymity, mental health medical leave and discussion amongst coworkers.

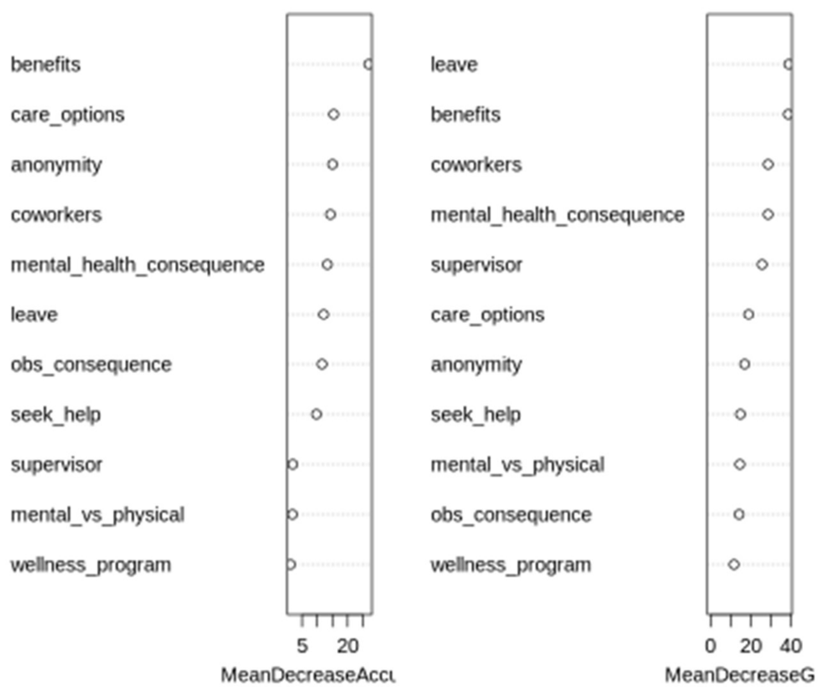


Figure 11 - Random Forest Feature Importance by Accuracy (left) & Node Impurity (right)

4. Which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions?

Recursive Feature Elimination (RFE) with Random Forest:

We employed the RFE algorithm with Random Forest to identify the company characteristics with the most predictive power. Using this method, the optimal number of features used to explain the response variable *obs_consequence* was found to be two. Based on calculated importance, *tech_company* and *no_employees* were found to be the most important features in predicting *obs_consequence.* However, we also found that the cross-validated accuracy did not change when adding a third predictor. Hence, the next important variable, *remote_work,* was also included in the next part of the analysis for better predictive performance.

Logistic Regression:

Our analysis revealed that employees in a tech company are less likely to see negative workplace consequences for those with mental health conditions compared to non-tech companies. Mid-sized companies with 6 to 100 employees are less likely to see negative workplace consequences for those with mental health conditions compared to very small companies with 5 or less employees. Not many coefficients were found to be statistically significant in the regression model. However, the overall logit model does have strong predictive power with 85% train and test accuracy. The model was found to be statistically significant and passed a Chi-Square goodness of fit test.

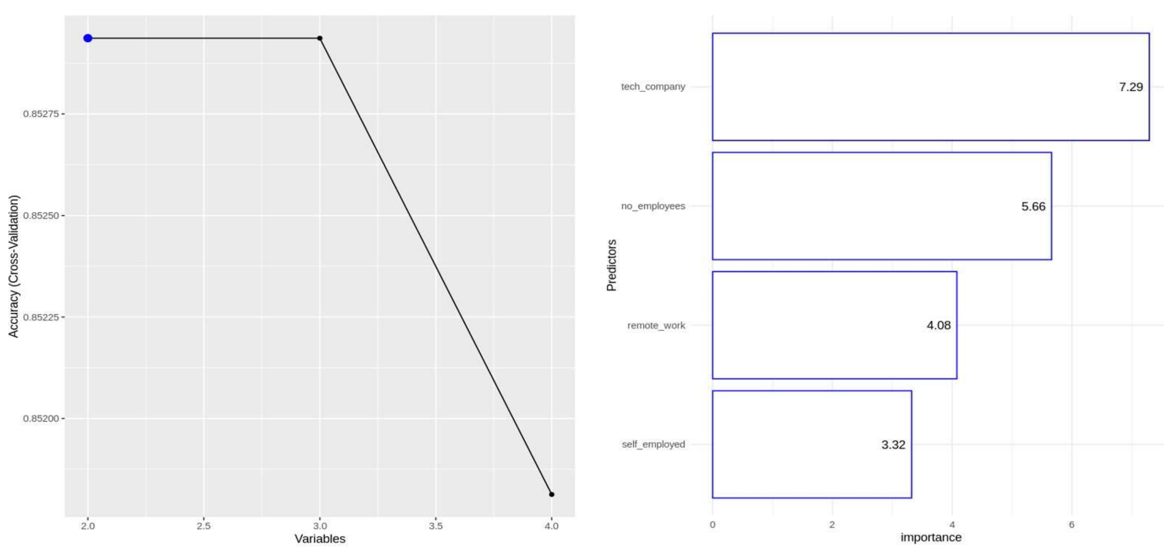


Figure 12 - RFE Cross-Validation Accuracy & Feature Importance

7. FINDINGS AND DISCUSSION

Most of the models identified that employees are more willing to seek mental health treatment if the issue interferes with their work and if they have a family history of mental health issues. Conversely, employees are less likely to seek the treatment if they identify as male. Our society still suffers from the issues of toxic masculinity, thus not creating a supportive environment for our male counterparts to express their emotions. Likewise, employees are more willing to seek care for mental health if the employers provide mental health benefits, protect anonymity, and prioritize medical leave for mental health conditions. Mental health issues are still a taboo in certain cultures and societies; thus, employees might feel safe if their identity is protected. If the awareness for mental health care is institutionalized or is properly advocated by the leadership, employees feel more confident to have open discussion with coworkers, supervisors, and even prospective employers without worrying about negative consequences.

Based on the model outputs, tech companies seem to be more conscious of mental health issues since their employees are less likely to see negative workplace consequences for those with mental health conditions compared to non-tech companies. Finally, employees of mid-sized companies (6-100 employees) are less likely to see negative workplace consequences compared to large or small-sized companies. This finding sounds logical since small-sized companies might lack the appropriate resources to promote such health services and large-sized companies might be prone to overlook such aspects of employee wellbeing against their numerous competing priorities.

Regarding model performances, the full logistic model with 77% testing accuracy rate seems to be performing the best in answering the first research question about factors that are strongly associated with an employee seeking treatment for mental health conditions. Boosting with 79% testing accuracy rate has a better prediction performance than KNN. Both subset models, logistic and random forest, including variables related to mental health benefits only are performing moderately with testing accuracy rates of 64%. To answer which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions, logistic regression seems to be performing better with the testing accuracy rate of about 85%.

8. LESSONS LEARNED

- In-depth exploratory data analysis is a vital step in the data mining process. It is not only important to assess the relationship between the response and predictors, but also the distribution of the predictors themselves prior to any model building.
- The incorporation of cross-validation and/or stochasticity (i.e., in Random Forests) is important in assessing the robustness of the models built. This gives us a better gauge of model accuracy against new data and helps offset bias from the training data.

9. CONCLUSION

In summary, we used various data mining techniques to understand the frequency of mental health issues among working professionals, identify key characteristics associated with employee's willingness to seek mental health treatment and ability to predict based on such characteristics, and analyzed the types of companies that are more likely to see negative consequences for seeking mental health care. We hope these results can contribute to the creation and enforcement of more effective workplace practices to support employees' mental health. Businesses could boost productivity, prevent loss time, improve employee relations and engagement, and gain a competitive advantage in recruiting if mental health is promoted in the workplace. It might not be ethical to use this kind of predictive tool by employers to predict such employees who might need such care. However, employers and insurance companies could make mental health an integral part of their wellness programs and insurance plans. We hope for a future where the annual mental health check-up could be a covered service like the annual physical health visits.

Future considerations for this study include collecting additional data and a more thorough variable selection process to improve model performance and reduce overfitting and exploring other machine learning algorithms for better predictive performance. Expansions of the analysis could include examining other cultural and regional factors on employees seeking treatment and observed negative consequences, as well as longitudinal studies to track changes in behavior over time.

10. APPENDIX

Exploratory Data Analysis: Demographics

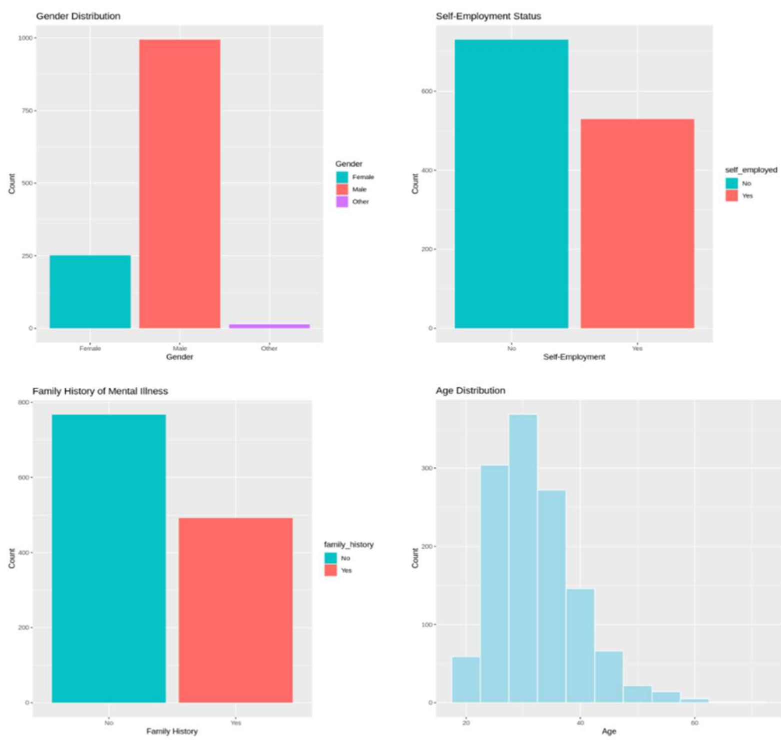


Figure 13 - Exploratory Analysis for Demographic Predictors

Exploratory Data Analysis: Company Characteristics

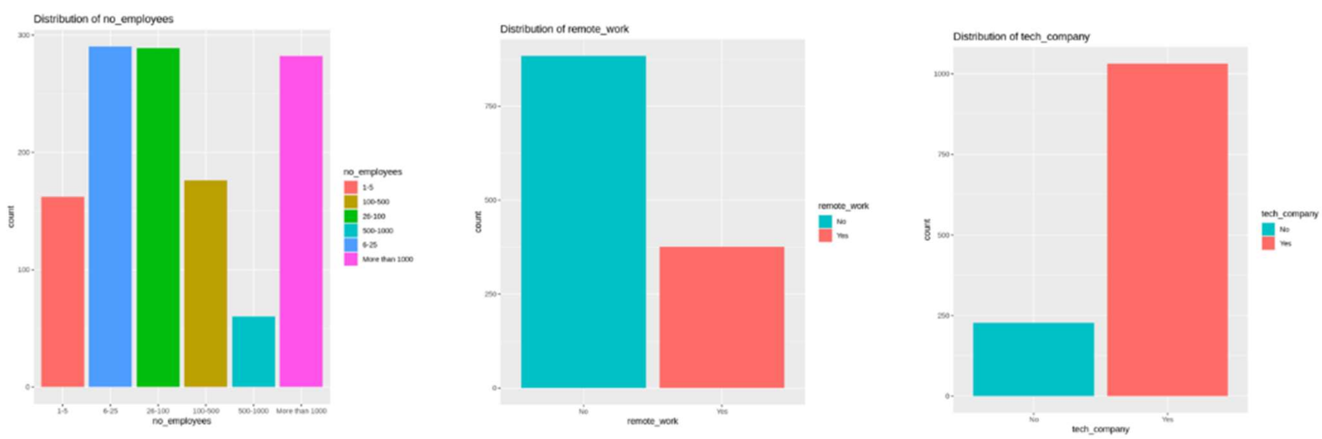


Figure 14 - Exploratory Analysis for Company Characteristics

Exploratory Data Analysis: Company's Mental Health Services

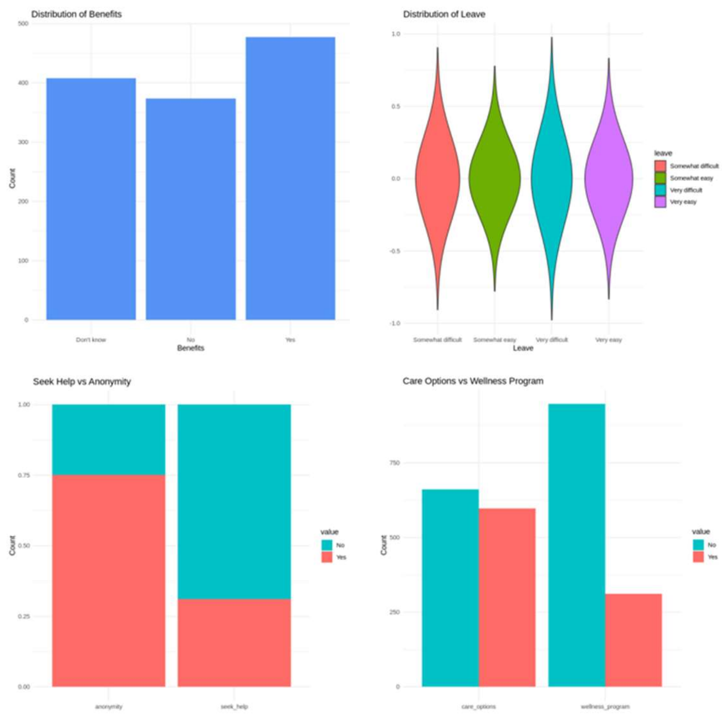


Figure 15 - Exploratory Analysis for Company Mental Health Services

11. REFERENCES

1. Lerner, Debra, and Rachel Mosher Henke. "What Does Research Tell Us about Depression, Job Performance, and Work Productivity?" Journal of Occupational & Environmental Medicine, vol. 50, no. 4, 2008, pp. 401–410., https://doi.org/10.1097/jom.0b013e31816bae50.

2. "Mental Health at Work." *World Health Organization*, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work.

3. Dewa, Carolyn S, et al. "The Association of Treatment of Depressive Episodes and Work Productivity." *The Canadian Journal of Psychiatry*, vol. 56, no. 12, 2011, pp. 743–750., https://doi.org/10.1177/070674371105601206.

4. Open Sourcing Mental Illness, LTD. "Mental Health in Tech Survey." *Kaggle*, 3 Nov. 2016, https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey.

5. "Handling Missing Data with Mice Package; a Simple Approach." *DataScience+*, https://datascienceplus.com/handling-missing-data-with-mice-package-a-simple-approach/.