Mental Health in Tech

ISYE 7406 – Data Mining and Statistical Learning

Group 22: Yael Benarroch, Sukriti Raut, Kann Tan, Shruti Rajagopal

Introduction

The stigma around mental health has been a pervasive issue in the workplace. Employers are now acknowledging the importance of mental health and are taking steps to ensure that their employees feel comfortable seeking treatment. However, despite the increased awareness, many employees still do not seek treatment for their mental health conditions.

This project aims to analyze data from a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace to identify factors associated with an employee seeking treatment for mental health conditions.

Data

- Data Source: Kaggle.
  - https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey
  - Survey Conducted By Open Sourcing Mental Illness (OSMI) project
  - Year Conducted: 2014

- 1,259 observations; 23 variables

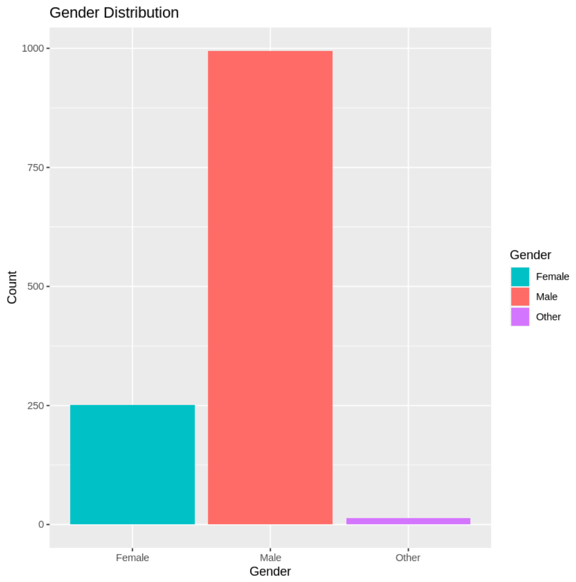
- Data Type: Categorical. Responses to survey questions were Yes/No (and occasionally Maybe/Don't Know)

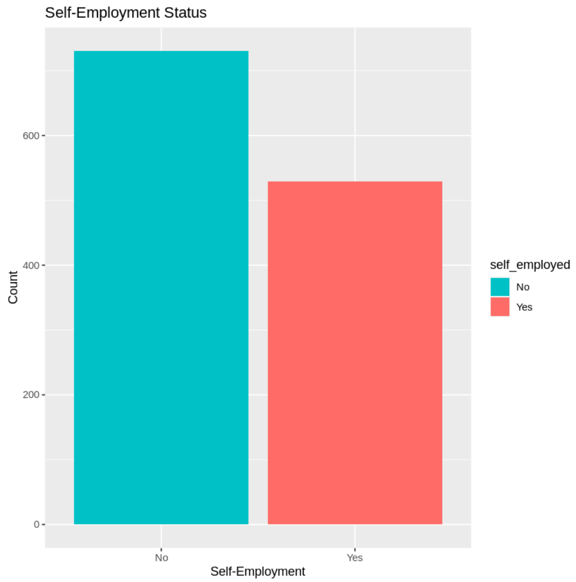
Data Preprocessing

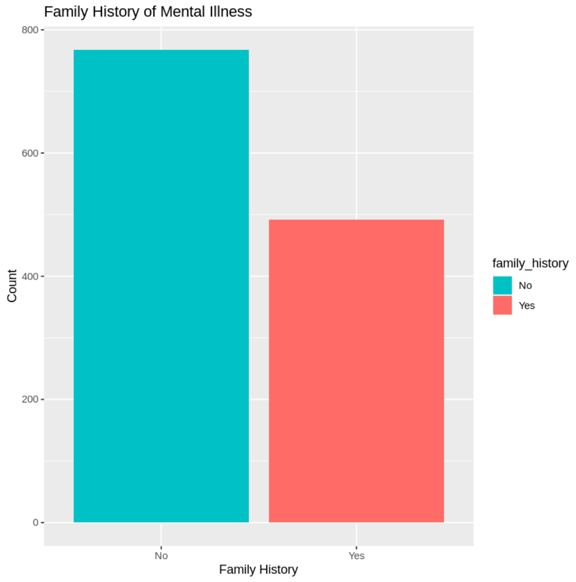
- Character fields had many formatting errors/typos which were cleaned.
  - *Gender* had the most variability which was recoded as "male", "female" and "other"
  - New variable *AgeGroup* created as
    - YoungAdult ($Age < 30$),
    - MiddleAgedAdult ($31 > Age > 45$),
    - OldAdult ($Age > 45$)
- To keep the dataset size, we decided to impute missing values instead of removing the records
- Missing values were imputed using Mice package. 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'

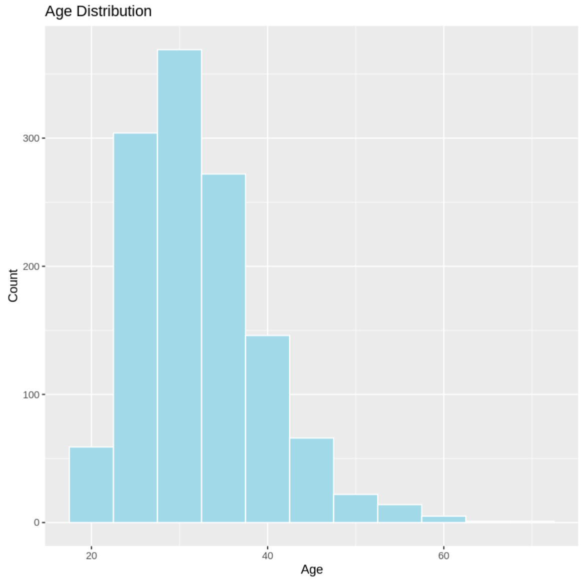
Column	Missing_Values
anonymity	819
mental_vs_physical	576
leave	563
state	515
seek_help	363
self_employed	314
care_options	314
work_interfere	264
wellness_program	188
Age	7
Gender	0
Country	0
family_history	0
treatment	0
no_employees	0
remote_work	0
tech_company	0
benefits	0
mental_health_consequence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

Exploratory Data Analysis – Employees' Demographics





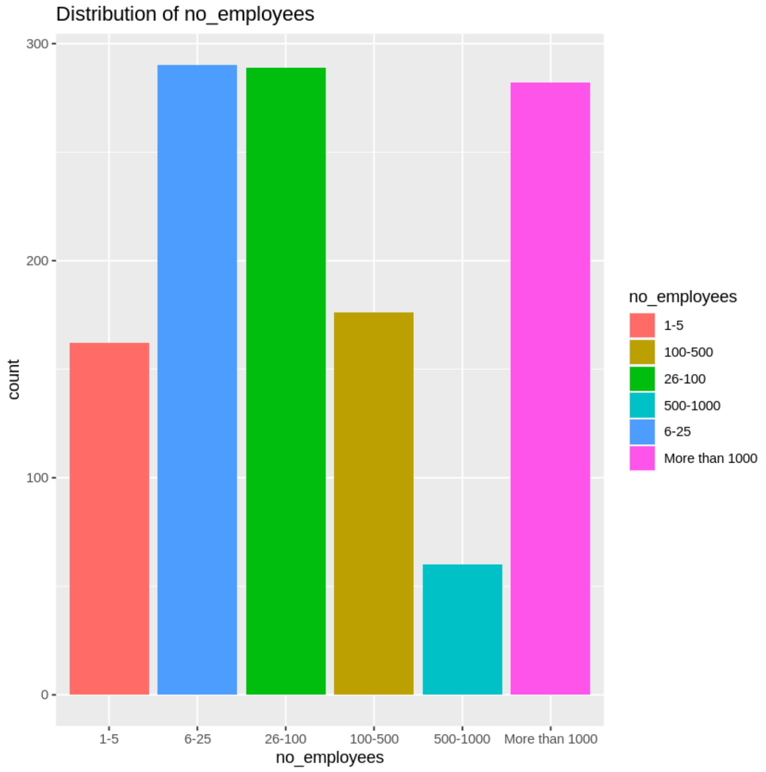


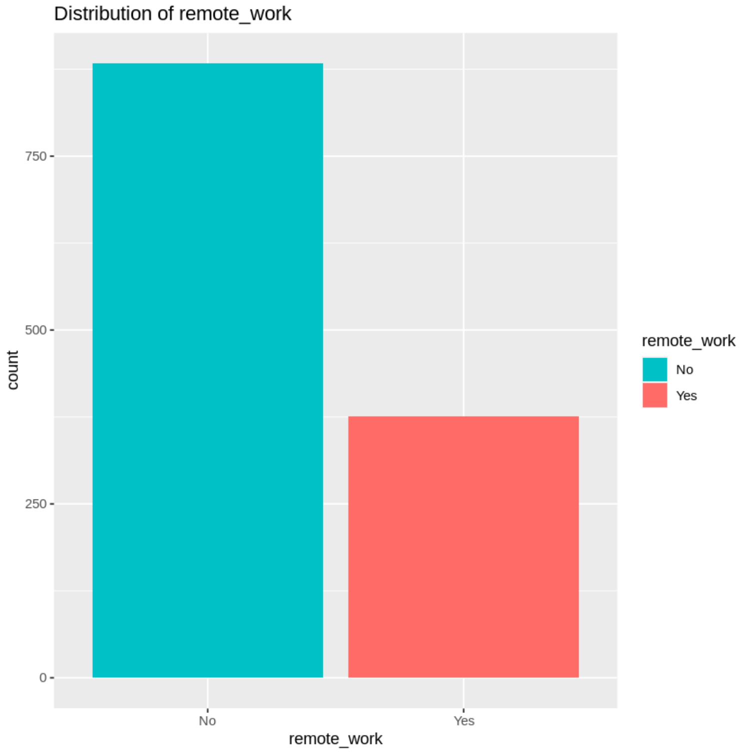


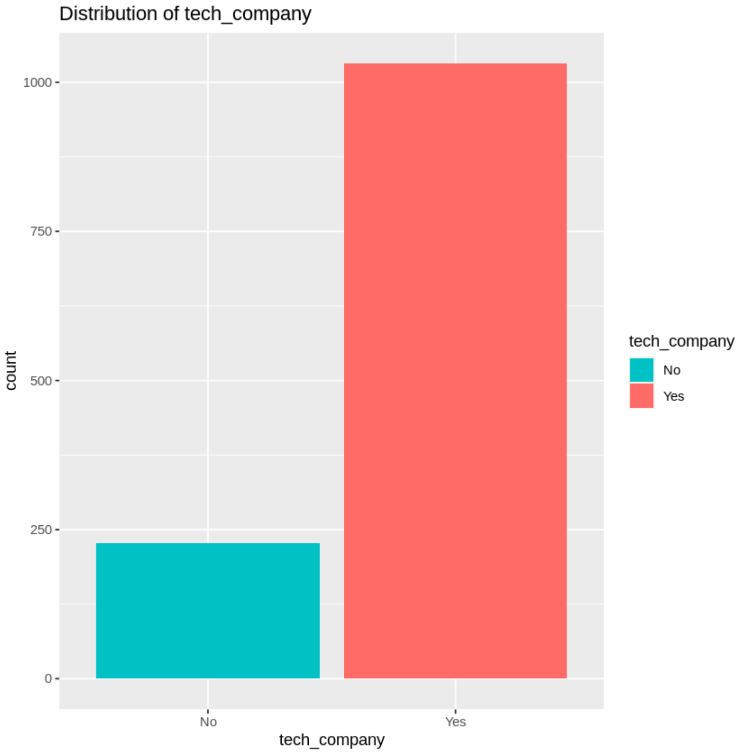
- Variables: age, gender, family_history, self-employed, state, country
- Most respondents are male and younger than 45. Most are employed with a company (i.e. not self employed) and do not have a family history of mental illness.

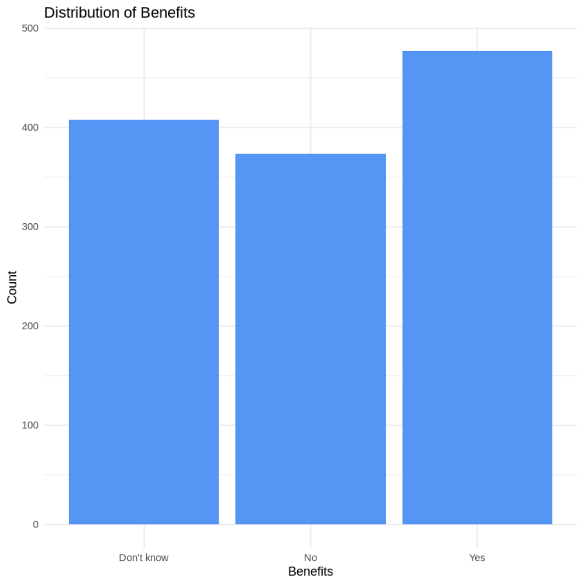
Exploratory Data Analysis – Company Characteristics

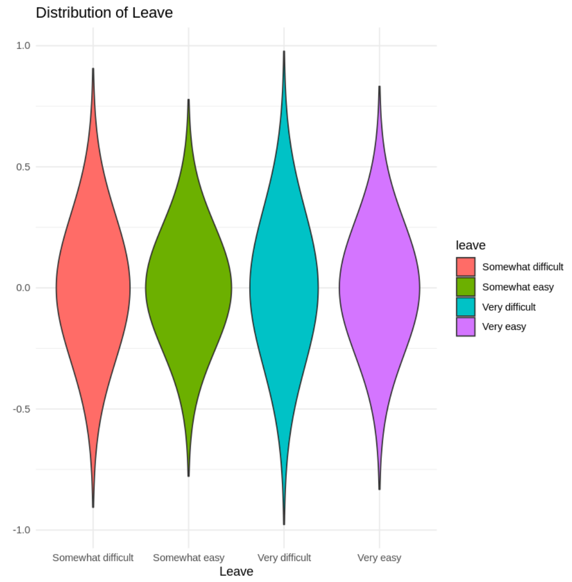
- Variables: no_employees, remote_work, tech_company
- There is significant variety in the distribution of company size. Most respondents work for companies with fewer than 100 employees. Most work at a tech company and do not work remotely more than 50% of the time.

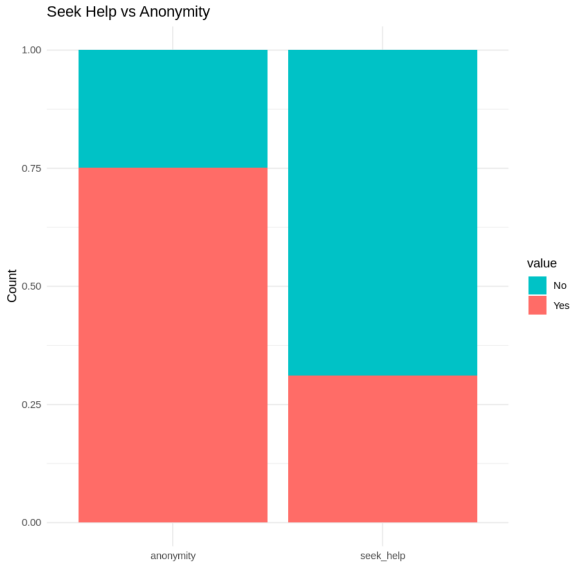


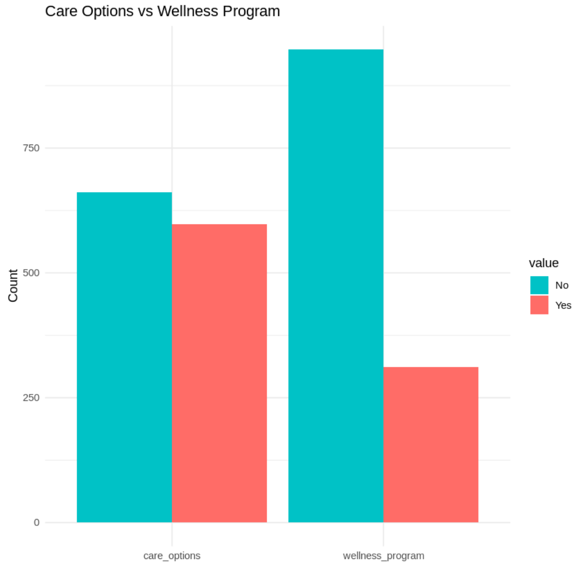












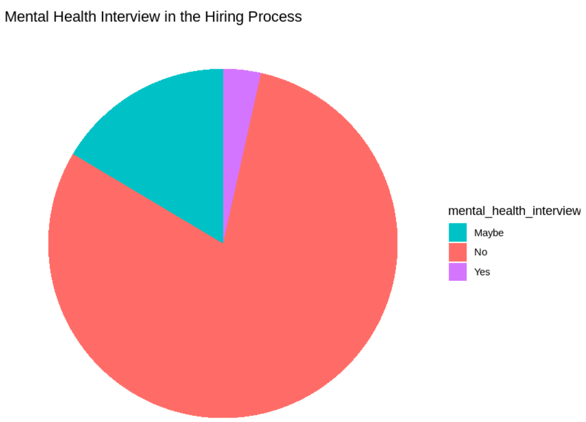
Exploratory Data Analysis – Company's Mental Health Services

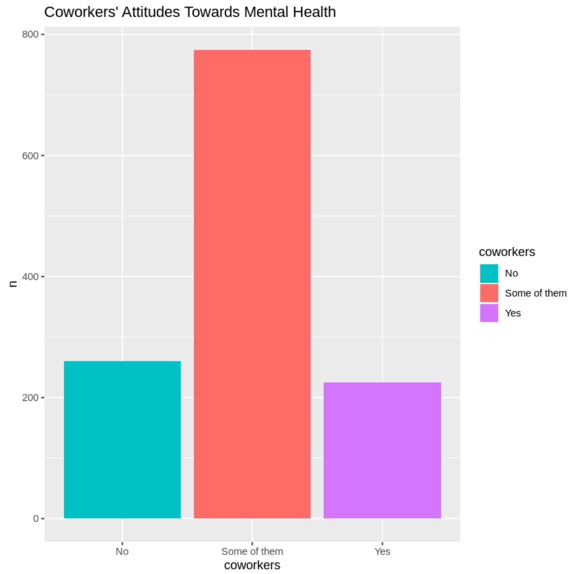
- Variables: benefits, care_options, wellness_program, seek_help, anonymity, leave
- Most respondents are not aware of the mental health care options provided by their employers. Most have also never discussed mental health as part of an employee wellness program with their employers.
- Most employers of the respondents do not provide resources on seeking help for mental health issues. However, most respondents said they found it easy or somewhat easy to take medical leave for a mental health issue.

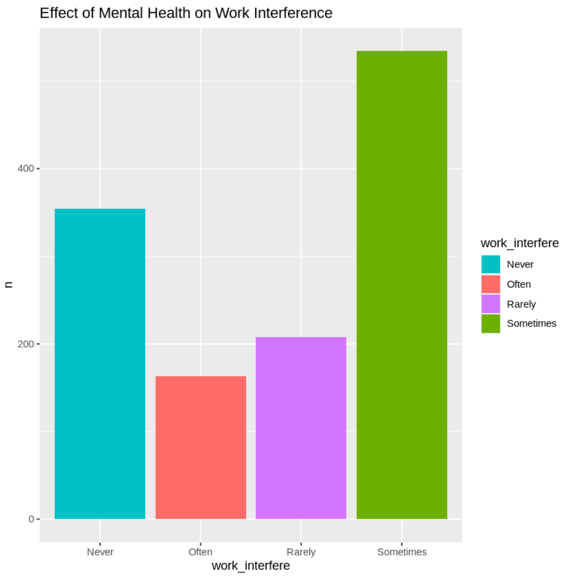
Exploratory Data Analysis – Employees' Perception & Attitudes

Variables: coworkers, supervisor, mental_health_interview, phys_health_interview, work_interfere, mental_health_consequence, phys_health_consequence, mental_vs_physical

- 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.

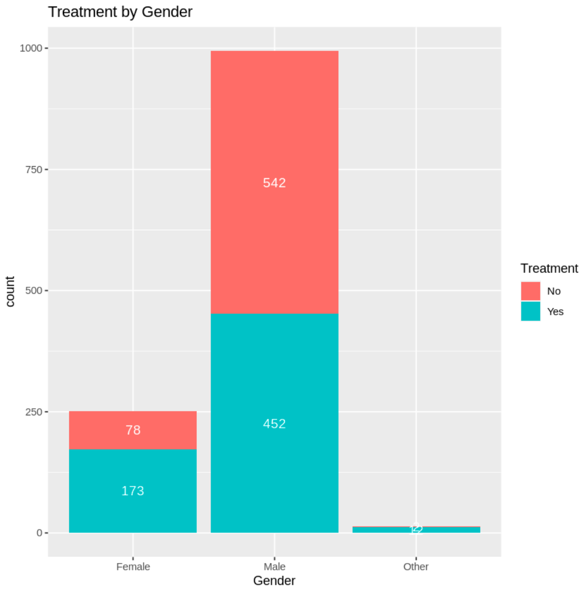


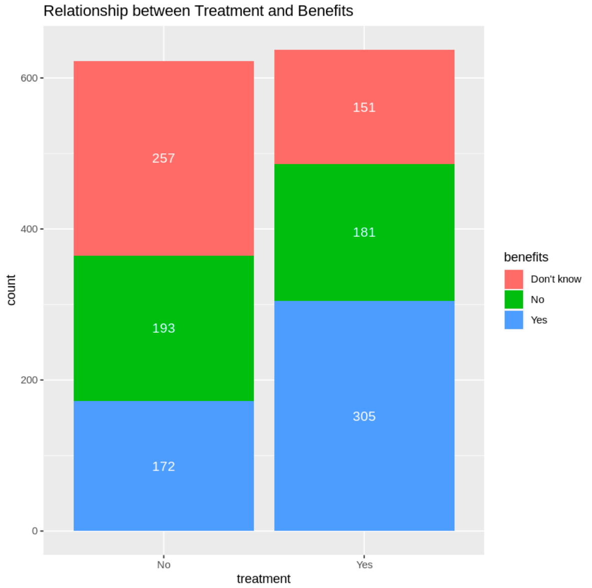




Exploratory Data Analysis – Response Var: *treatment*

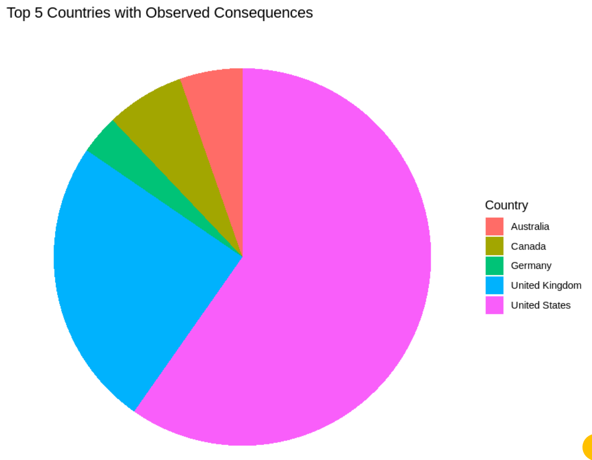
- A larger proportion of female respondents have sought treatment for mental health conditions compared to males.
- Overall, a little over 50% of total respondents said they have sought treatment for mental health conditions.

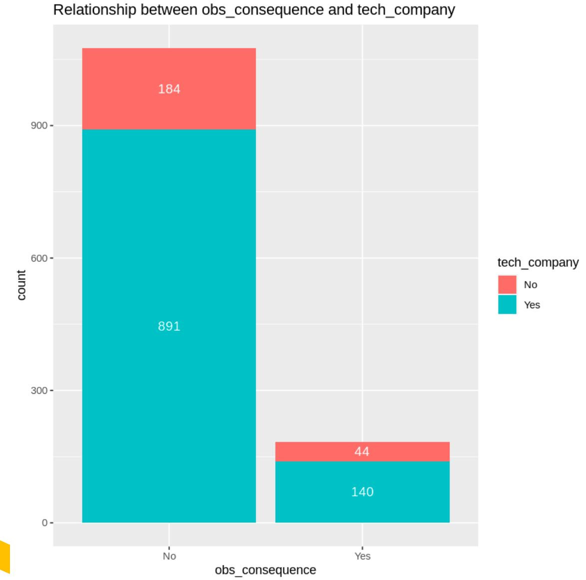




Exploratory Data Analysis – Response Var: *obs_ consequence*

- The US and the UK were the top 2 countries (and CA and WA were the top 2 US states) where respondents said they have heard of or observed negative consequences for coworkers with mental health conditions.
- Of those who said they have heard of or observed negative consequences, most worked at a tech company.





Research Questions & Methodology

Research Questions	Models	Response Variable (Y)
What factors are associated with an employee seeking treatment for mental health conditions?	Logistic, Correlation Analysis, Decision Tree	Treatment
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, KNN	Treatment
Does a company that support/provide mental health benefit leads to more employee seeking treatment for mental health conditions?	Random forest, Logistic Regression	Treatment
Which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions?	Random forest, Logistic Regression	Obs_consequence

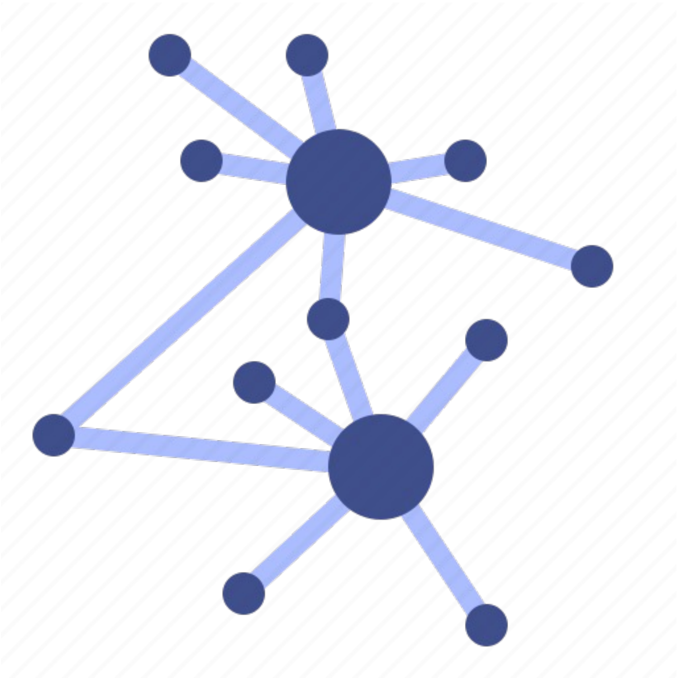
80/20 split on dataset for Train/Test

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

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

Method: Correlation Analysis

- This study used chi-squared test to examine the relationship between 23 variables and employees seeking mental health treatment.
- Results show that several factors were significantly associated with an employee seeking treatment for mental health conditions, such as family history, work interfere, and mental health consequence.
- The highest Cramer's V values, which indicate the strength of association between two variables, were observed in work interfere, family history, and state



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

Method: Logistic Regression

- Males are less likely to seek treatment for mental health conditions compared to females.
- Self-employed individuals are more likely to seek treatment.
- Employees with a family history of mental health conditions are more likely to seek treatment.
- Employees who experience work interference due to mental health issues are significantly more likely to seek treatment.
- 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.
- Employees who are willing to discuss mental health with all coworkers are more likely to seek treatment.

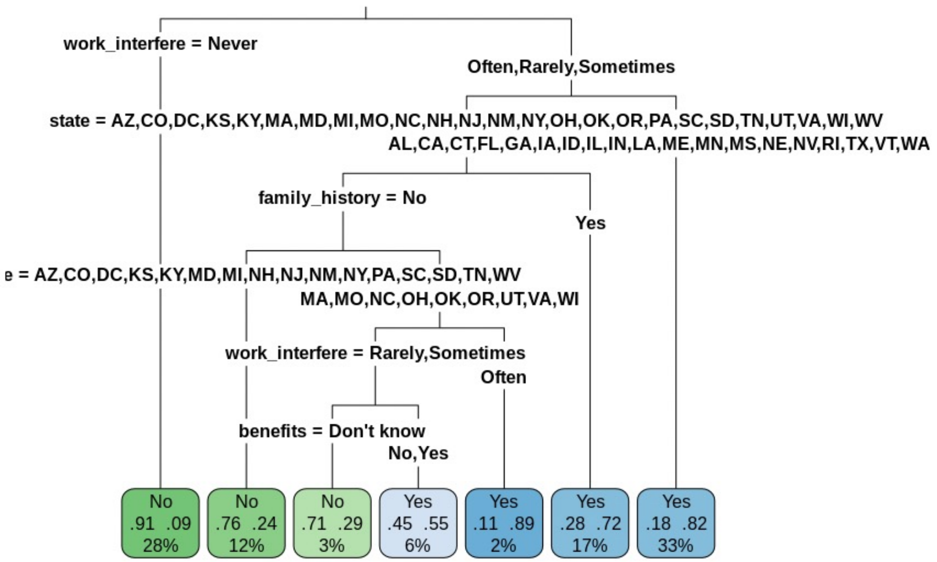
Model predictive accuracy = 77.1%

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

Method: Decision Tree

- Work interference is the most important factor associated with an employee seeking treatment for mental health conditions.
- State and family history of mental health conditions are also significant factors in the decision-making process.
- If an employee experiences work interference often, they are more likely to seek treatment.
- If an employee is in a state not listed in the decision tree model, they are more likely to seek treatment.

Decision Tree for Treatment Factors



Model predictive accuracy = 75.4%

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

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

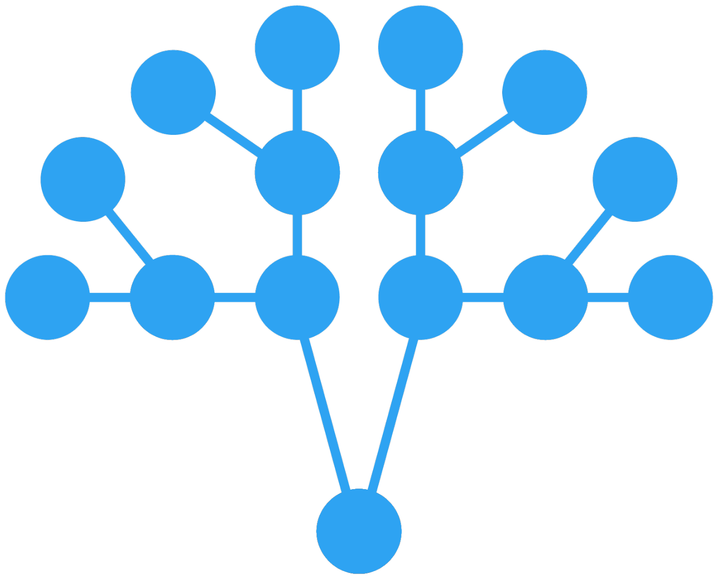
Method: Gradient Boosting

- Cross-validation was utilized to tune the hyperparameters
- The best selected model used: number of trees =60, learning rate (shrinkage)=0.1, tree depth=3, min number of obs in terminal nodes=5
- Based on relative influence of the model, these were top 5 demographic & work-related variables that were important in predicting the likelihood of employee seeking mental health treatment:

  family history, work_ interfere, benefits, age, gender, obs_ consequence

- Based on high accuracy rates, we could potentially use Gradient Boosting to predict which employees are more likely to seek treatment for mental health conditions

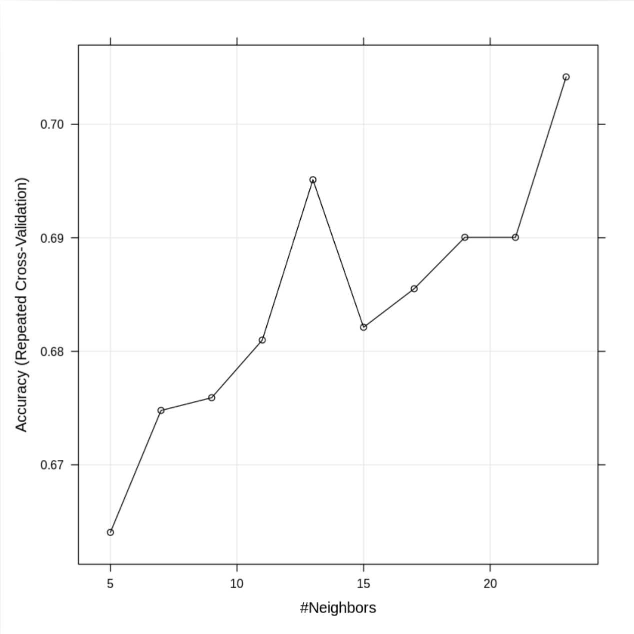
  Train Accuracy: 79.5 %
  Test Accuracy: 79.2 %

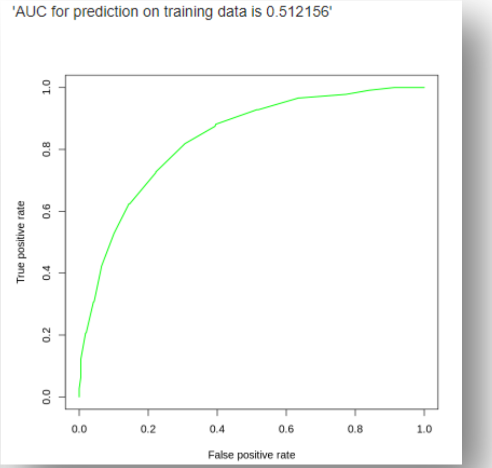


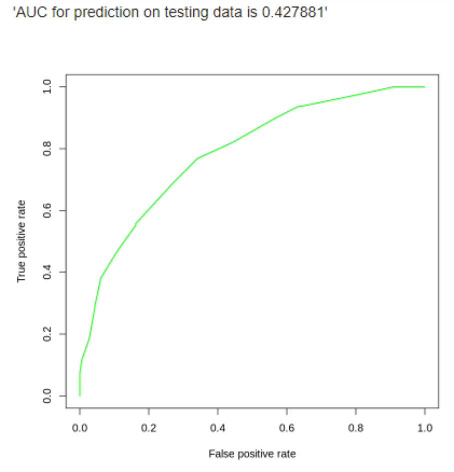
Can we predict which employees are more likely to seek treatment for mental health conditions based on their demographic information and work-related factors?

Method: KNN

- Cross-validation was utilized to tune the hyperparameters
- Based on the highest accuracy rate, optimal number of selected k=23
- Based on high number of k neighbors and moderate AUC, train, & test accuracy rates, KNN does not seem like a strong prediction model for this dataset.







Area Under the Curve: 0.51 & 0.42
Train accuracy: 73.9%
Test accuracy: 69.3%

Does a company that support/provided mental health benefit leads to more employee seeking treatment for mental health conditions?

Does a company that support/provided mental health benefit leads to more employee seeking treatment for mental health conditions?

Method: Logistic Regression

- Employees with companies with mental health benefits are more likely to seek treatment.
- Employees that can take medical leave for mental health has more likely to seek treatment.
- Employees that can discuss mental health issue with their employer without negative consequences are more likely to seek treatment.

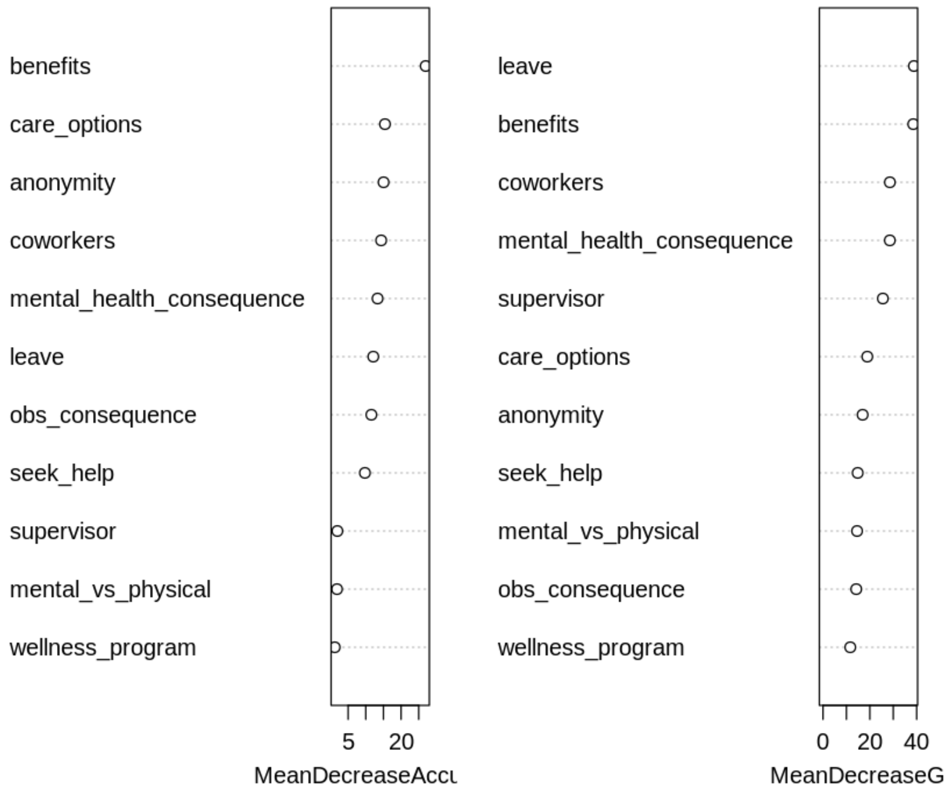
Model predictive accuracy = 64.95%

- Employees in a company where coworker are willing to discuss mental health issue are more likely to seek treatment .
- Employees are significantly more likely to seek treatment when employer provide resources to learn more about mental health issues and how employees can seek help.
- Employees with anonymity protected if they chose to take advantage of mental health or substance abuse treatment are more likely to seek treatment.

Does a company that support/provided mental health benefit leads to more employee seeking treatment for mental health conditions?

Method: Random Forest

- The model was only fitted with variables that are related to mental health benefits of the company.
- Based off calculated importance, the following features were found to be the most important in decrease in node impurity for predicting treatment are leave, benefits and coworkers.
- Based off calculated importance, the following features were found to be the most important in decrease in accuracy for predicting treatment are benefits, care_options and anonymity.

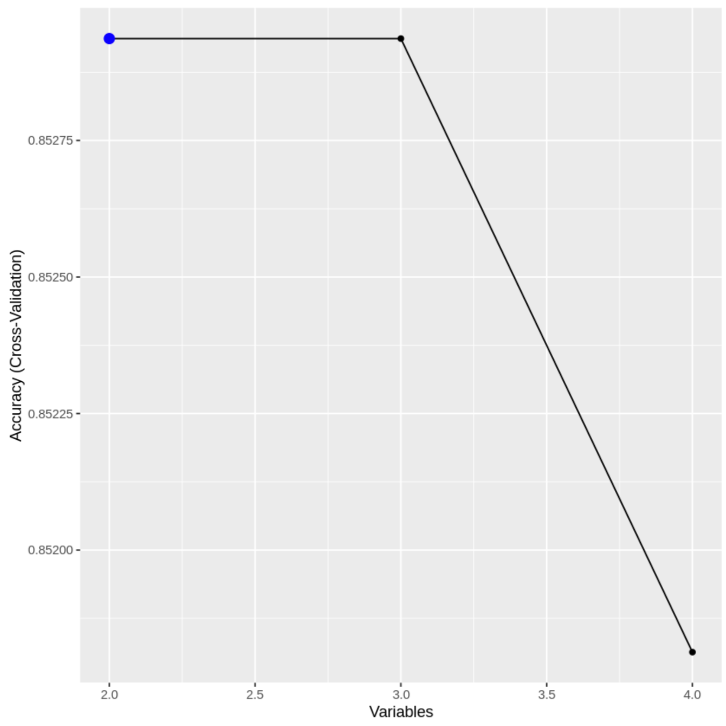


Model predictive accuracy of training data = 85.63%
Model predictive accuracy of testing data = 63.47%

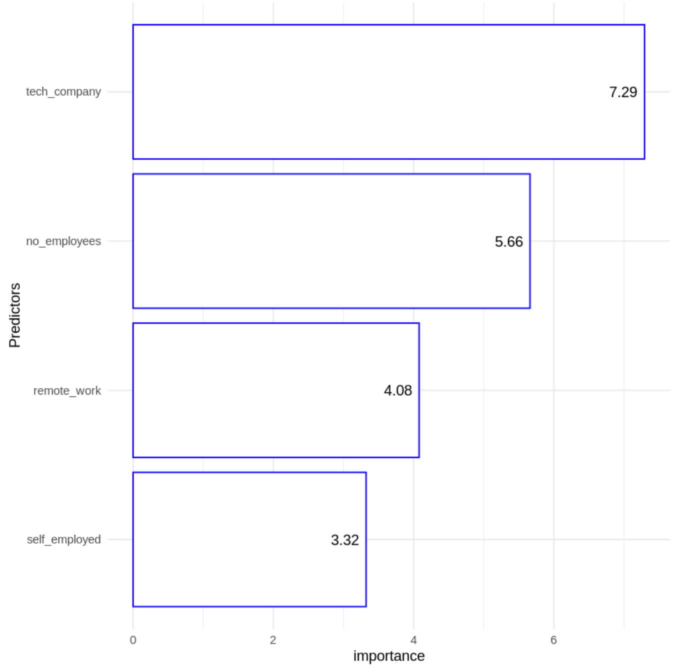
Which type of companies, if any, are more likely to see negative consequences for employees with mental health conditions?

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

Method: Random Forest RFE



- Using a Recursive Feature Elimination algorithm with Random Forest, the optimal number of features used to explain the response variable *obs_ consequence* was found to be 2.
- Based off calculated importance, the following features were found to be the most important in predicting *obs_ consequence*: *tech_ company* and *no_ employees*
- Since the cross-validated accuracy did not change when adding a third predictor, the next important variable, *remote_ work,* was included in the next part of the analysis for better predictive performance.



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

Method: Logistic Regression

- Not many coefficients were found to be statistically significant
- The overall Logit model, however, was found to be statistically significant and passed a Chi-square goodness of fit test.
- Findings:
  - Employees in a tech company are less likely to see workplace consequences for those with mental health conditions compared to non-tech companies.
  - Companies with 6 to 100 employees are less likely to see workplace consequences for those with mental health conditions compared to companies with 5 or less employees. Companies sized more than 100 employees was not found to be statistically significant.
  - Remote work was not found to be statistically significant in predicting observed workplace consequences.

Train Accuracy: 85.2%

Test Accuracy: 85.6%

Future Considerations

- Collect additional data to improve model predictions and reduce overfitting
- Explore alternative machine learning algorithms for better predictive performance
- Consider incorporating feature engineering techniques to enhance model interpretability
- Examine the impact of cultural and regional factors on mental health treatment-seeking behavior
- Investigate the effectiveness of mental health interventions within the workplace
- Conduct longitudinal studies to track changes in employee mental health and treatment-seeking patterns over time
- Assess the influence of company leadership and management styles on mental health outcomes
- Utilize natural language processing techniques to analyze qualitative data, such as employee feedback and surveys

Conclusion

- Explored relationships between factors and employees seeking mental health treatment
- Identified key factors and demographic characteristics influencing treatment-seeking behavior
- Investigated the impact of company-provided mental health benefits on employee well-being
- Analyzed types of companies with fewer negative consequences for employees with mental health conditions
- Highlighted the importance of supportive policies for a positive and inclusive work environment

References

- "Handling Missing Data with Mice Package; a Simple Approach." *DataScience+*, https://datascienceplus.com/handling-missing-data-with-mice-package-a-simple-approach/
- "Fixing the R Warning - Glm.fit: Algorithm Did Not Converge." *ProgrammingR*, https://www.programmingr.com/fixing-the-r-warning-message-glm-fit-algorithm-did-not-converge/