**Springboard Capstone Milestone Report: Mental Health in the Workplace**

Elizabeth Matthews

Introduction to the Problem

According to the National Institute of Mental Health, almost 18% of all adult Americans were diagnosed with mental illness in 2015. The direct and indirect costs of mental health issues to both individuals and their employers can be high. The most frequently cited ways that poor mental health among a company’s employees can negatively impact the company’s bottom line include: lost employee productivity, poor performance in the workplace, high rates of worker absenteeism or employee turnover, accidents on the job, as well as increasing costs related to insurance premiums and medications (Harvard Health, 2010).

Recent research has found that when companies actively promote mental health wellness in the workplace, that there are significant financial benefits as well as improved employee moral (Harvard Health, 2010). As part of its efforts to better understand how to address work place mental health concerns, an employer could look to data science and the development of statistical models that can accurately predict the risk of mental health issues. Developing such a model could enable employers to work proactively to promote mental health well-being in their companies, which could improve their workplace environment and help reduce costs associated with mental health illness.

The proposed project aims to investigate factors associated with mental health issues among workers in the technology sector. Previous analyses Open Source Mental Illness 2014 data set, have not found that tech workers have a higher rate of mental health issues than those in other sectors. That said, given the high prevalence of mental health issues among adults, all employers must be prepared to address this issue.

Data Set & Variables

Data for this project will be obtained from the Open Source Mental Illness Project—The 2014 Mental Health in Tech Survey includes 1259 observations and 27 variables including: demographic information of the respondents, workplace information and attitudes regarding mental health care and consequences of seeking mental health care. In addition to these variables, there is also an additional comments section including qualitative information on respondents’ experiences with mental health issues in the workplace.

The original data set can be obtained on Kaggle @ <https://www.kaggle.com/osmi/mental-health-in-tech-survey>

Limitations

There are a number of limitations in the 2014 OSMI data set. There is no single variable assessing whether the respondent currently is experiencing a mental health disorder, or the classification of the disorder. The two closest variables include the “seek help” and “treatment” variables. Both attempt to assess whether the individual has sought help or treatment for a mental health condition. This may however leave out a large portion of individuals who have mental health concerns, but are not willing, or able to obtain treatment. It also does not provide a time frame for the individual’s treatment and whether treatment was sought recently or during the respondent’s tenure with the current company.

In addition, the demographic information is limited to age, gender, workplace type (tech/not tech), whether there is a family history of mental health issues , the location of the respondent and the size of the respondent’s employer. It would be helpful to have other demographic information such as marital status.

Data Wrangling

Some cleaning of the raw data was necessary. Although I do not plan to use all 27 variables in my analyses, I chose to clean the entire data set. Most of the issues surrounded fixing spacing issues in the variable levels, correcting mistakes in the gender column, addressing missing values and creating a new column for state region. In addition, a new column was created with a binary 1, 0 level to indicate whether or not the respondent made a comment in the survey. The full documentation of the code used in the data manipulation can be found here: https://github.com/belljar26/MentalHealthCapstone/blob/master/DataWranglingCapstone.Rmd

Preliminary Exploration & Initial Findings

Some descriptive and inferential statistics were utilized to better to explore the data set and to provide useful information for developing a model to predict mental health disorders among workers.

*Descriptive Analyses:*

The mean age of the respondent is 32. 78% of respondents are men, 19 % female and 3 % identify as “other”. 376 respondents reported that they work remotely. 50.56% of the respondents have indicated that they have sought treatment for a mental health disorder. Interestingly, 62% either don’t know (if they do) or don’t have access to mental health benefits their place of work.

*Inferential Statistics*

Chi Square Tests & Odds Ratio

As the majority of the variables in this data set are category and there are few continuous variables, Chi Square tests (and Odds Ratios where Chi Square tests were statistically significant) were conducted on several different variables to better understand the behavioral health views and treatment seeking behavior of the respondents.

One of the first areas explored was the relationship between working in a technology company and seeking treatment for a mental health disorder. This relationship was not found to be significant (p=.29). The association between remote work (defined here, as working remotely > 50% of the time) and seeking treatment was examined, but also found to be not statistically significant (p=.38). Thus, working in the technology field and/or working remotely do not seem to be associated with greater seeking of mental health treatment.

In the initial exploratory analyses, some statistically significant relationships were found. As expected, family history of a mental health disorder does appear to be associated with seeking treatment (p<.01). Gender also was found to be statistically significant with treatment seeking (p<.01) with women being more likely to seek treatment. Interestingly, perceptions and observations of consequences for disclosing mental health disorders and treatment seeking were statistically significant (p <.01).

It also seems that company size may have an impact on employees’ willingness to disclose mental health disorders/need for treatment as a Chi Square test showed a statistically significant result (p=.0018), for these variables. There also seems to be a statistically significant relationship (p= .02) between country [where the respondent is located] and beliefs among respondents about consequences for disclosing mental health disorders. Understanding the relationships between company culture and respondents’ fear of disclosure of mental health issues may provide some additional understanding of the factors that influence treatment seeking.

*Qualitative Comments:*

This data set also includes some qualitative comments made by respondents. 163 respondents provided comments about their experiences with mental health issues in the workplace. Leaving a comment was statistically significant with seeking treatment (p< .0002).

Plans & Approaches for Analyses

There are several variables that demonstrate a statistical significance with seeking treatment for a mental health disorders. A GLM model will be developed to elucidate which factors are most predictive of mental health disorders (using the seeking treatment variable). Promising variables to explore in building this model include: family history, gender, “observed consequence [variable]” and age. A regression and classification tree (CART) using rpart may also be done on this data set.

Sentiment analysis may also be done on the qualitative comments to determine the presence of any important themes that might help to further our understanding of mental health issues and work.