Mental Health in Tech

Andrea Fox

Summer 2021

https://github.com/anfox86/Mental-Health-in-Tech

Introduction

For this project I found a dataset that I thought would be interested in which consisted of survey answers from 2014 regarding mental health in the tech industry. It feels like lately the tech industry has really focused on employee mental health and have been more understanding with leave, treatment, and sharing to other coworkers or supervisors about a mental health condition. However, this survey was completed in 2014 when mental health was not widely talked about.

I wanted to come at this project with the mind of a potential client. I am acting as a consultant for another tech company interested in investing in employee mental health but wanted to see how the rest of the industry approached the topic. By doing this it allowed my mindset to focused on how do I help my client instead of simply looking for random questions? This is typically how I work daily anyways, so it has made this project flow better for me than the first project I worked on.

Business Problem/Hypothesis

My dataset came from Kaggle and contains 27 different variables which can be found in table 1 in the Appendix. Since I am coming at this project as a consultant the biggest question I am trying to answer is does the tech industry take mental health conditions seriously? Do they provide their employees with the resources they need to stay mentally healthy? By looking at the existing resources and standards, I can help my client make a more informed decision on whether they are on par with the

rest of the industry or if they need to make some improvements to provide a better mental health environment for their employees. So why is this important to my client? “Experts suggest that stress and anxiety in particular cost companies $30 billion in lost workdays each year.” (Davis, 2018) My client does not want to lose money especially at the amounts that are quoted in this article, but that is not the only reason to invest in employee mental health. There have been studies that “1 in 4 have mental health problems,” but really everyone has mental health just like they have physical health, and both need to be prioritized. (Lopukhina, 2020) It has finally become “ok” to talk about mental health and as well as seek treatment for a mental health condition, which is what my client wants to focus on.

Clean-up

After downloading the dataset from Kaggle, I then uploaded a copy into my Jupyter notebook so I could start my cleaning. One thing I did differently here is I made a list of things I knew needed to be fixed up before I could do a visual and technical analysis. This helped keep me on track and made the process faster and smoother than my previous projects. After I read in the data, I then printed a list of columns so I could see if there were any that needed to be removed. I ended up removing timestamp and comments. Timestamp was not useful to me since all the responses were from 2014, and if I wanted to look at timestamp, I would need more than just one years’ worth of survey results. Comments was a freetext field, so it would be incredibly hard to use it for analysis plus it had several null values. The next step was to look at all the unique responses for gender and try to narrow them down to a handful that would make analysis easier. There were 49 unique responses for gender which I narrowed down to Female, Male, Trans-female, Genderqueer, Non-binary, Other, Gender-fluid, and Androgynous. Looking at all of the unique responses many were misspellings of the world male or female. There were also some responses like cis female that I could update to just female as well as some variations of trans-female/woman that I updated to Trans-female. Other was for the responses where it was not clear on gender such as nah or all or enby.

The next clean-up process was to create an age category column that would be easier to analyze than all the unique ages. I first wanted to see what the minimum and maximum ages which came out to -1726 and 99999999999. Since those ages were outliers, I then had to look at all the unique values within the Age column. I was then able to narrow down what outliers I needed to remove and dropped anything greater than 72 and anything less than 18. From there I created the following age groups: 10-30, 30-40, 50-60, and 60-80. I had to mess with this piece of code to get what I wanted as the first time I tried I changed everything to null, then I tried something else that kept throwing an error about not having more bins than labels or vice versa. Thankfully, I was able to figure it out and created the Age Groups column which is at the end of the dataset. The next thing I wanted to create was number of employees grouped by column, but when I looked at the unique values it was already grouped well without my interference, so I left that column alone.

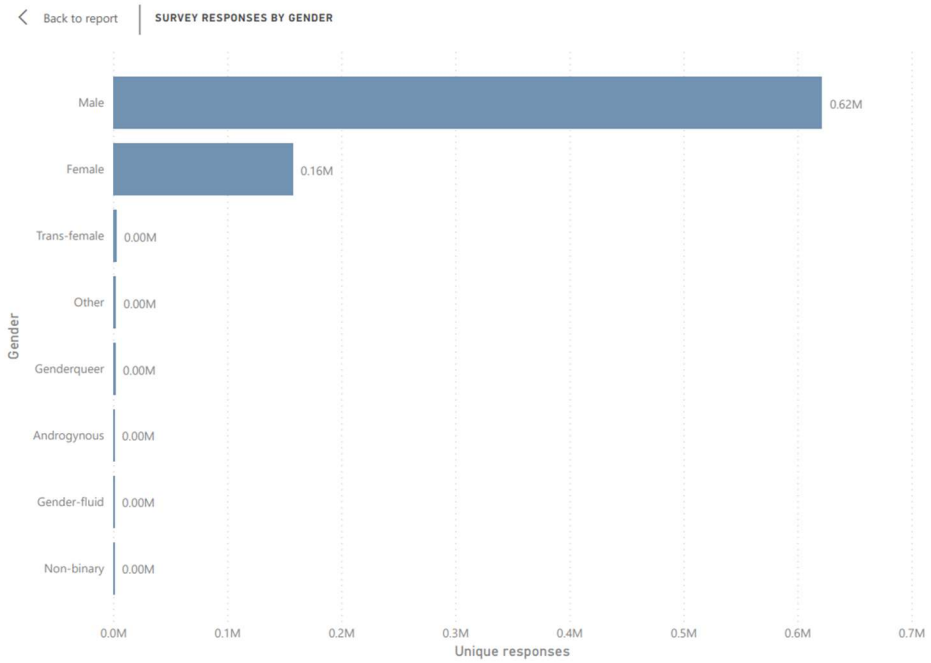
The last thing I needed to look at was the null values. For the column state, I filled in the null values with not applicable since those aligned to countries that did not have states. For self-employed I looked at the value counts for each response and since the 'No' response had a much higher number than 'Yes' , I changed the null values to 'No.' The last column that had null values was work_interfere. For that I looked at the value counts and did not see any that had a high enough number to substitute the null values, so I changed the null values to unknown. I then printed out the dataframe to get a final look and saved it off to a csv.

Visual Analysis

For the visual analysis, I loaded my newly cleaned csv file into PowerBi and created three dashboards. My first dashboard is named Overview, and looks at responses grouped by country, state, gender, age groups, number of employees, and self-employed. This is where I encountered my first surprise. I have included Figure 1, which looks at survey responses by gender. My assumption was that

the responses of females would be much higher than the others, but the male response made all the others look miniscule.

Figure 1 – Survey Responses by Gender

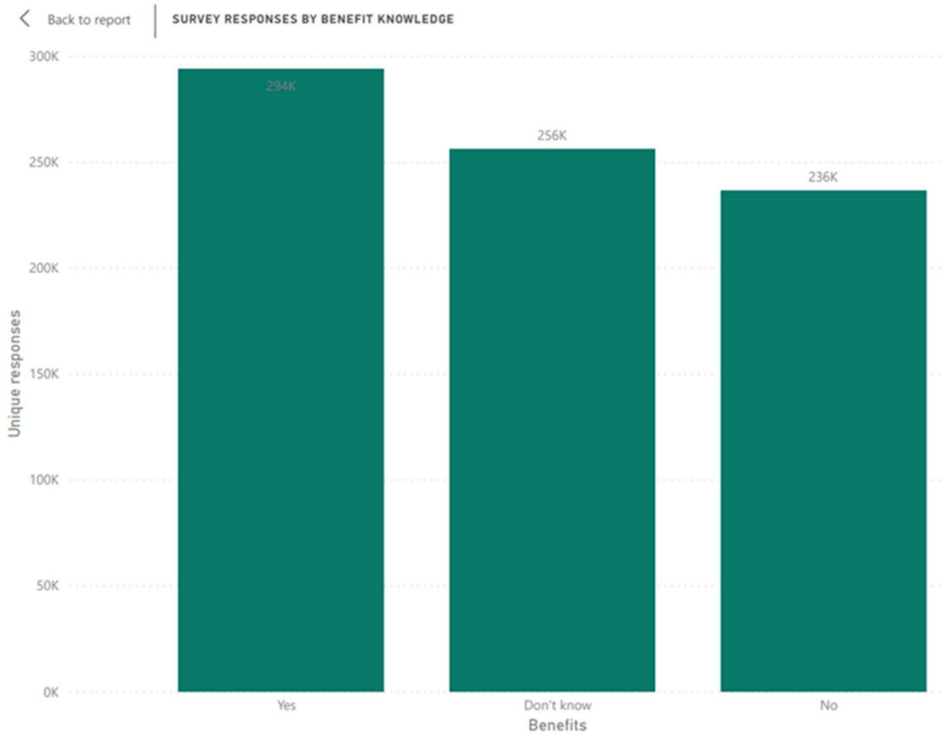


In my experience, females are usually the ones open to talking about mental health as well as answer surveys so finding males having a huge response rate was surprising. However, then I start to think of the lack of females in the tech industry and this does start to make sense. There are more females joining the tech industry work environment now but reading several articles along with working at a tech company myself most employees are male.

The second dashboard is named Benefits and included my visuals regarding employee knowledge of mental health benefits and programs available. Here it looked like there were quite a few employees who knew about existing benefits regarding mental health, but they were not confident in remaining anonymous if seeking help or whether leave was available. Employees also responded saying

that no there are not care options or wellness programs that are currently available at their place of employment.

Figure 2 – Survey Responses by Benefit Knowledge

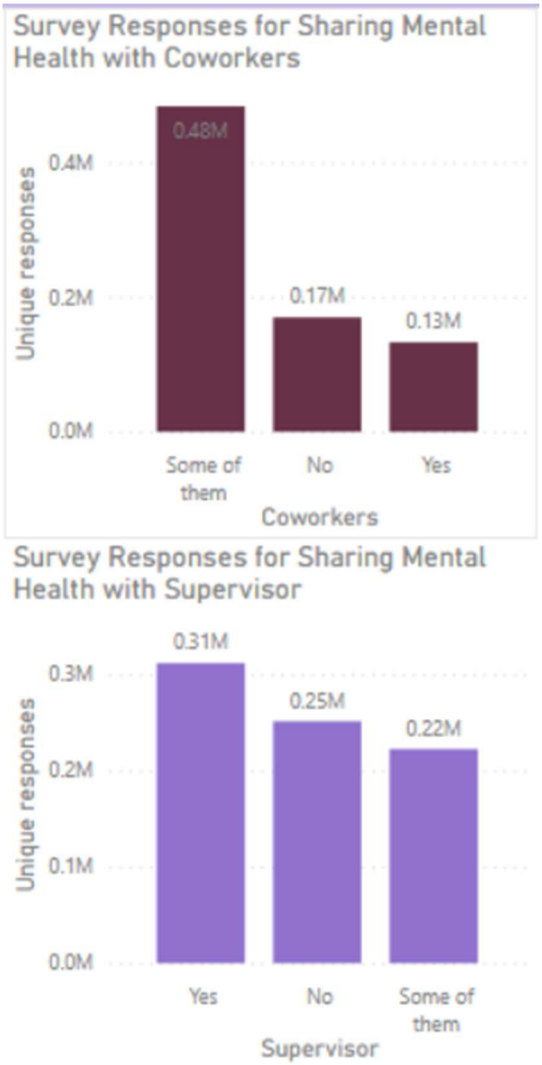


For my client, this would be somewhere I would make suggestions that not only do we need to make sure all benefits are easily accessible as well as published and marketed, but we need to offer wellness programs as well as multiple care options. This gives employees a variety of options to choose for their mental health.

The third dashboard I named In Workplace. This dashboard consisted of survey responses for consequences if employee talked about mental or physical health, whether respondent would bring up mental health in an interview, and lastly does the employer take mental health as seriously as it does physical health. The one I noticed the most in a good way was survey responses for observing other coworkers having consequences for sharing their mental health concerns. There were 0.66 million in the

no category to the 0.13 million in the yes category. This is also something my client needs to think about is making sure that no one has negative consequences for sharing their mental health concerns, as that can easily break an employee's trust. I also found it interesting that more respondents felt comfortable sharing their mental health concerns with their supervisor while the highest response for sharing with coworkers was 'some of them.' It makes sense since your relationship with your direct supervisor should consist of confidentiality as well as have a certain level of trust.

Figure 3 – Survey Responses for Sharing Mental Health with Coworkers and Survey Responses for Sharing Mental Health with Supervisor



The biggest takeaway from this dashboard with the visualizations is that my client will need to foster a welcoming environment where discussing mental health issues can be done so without fear of retaliation.

Technical Analysis

For my technical analysis I have just started. I have created a new Jupyter notebook and read in the cleaned csv. The first thing I wanted to look at was if there was a relationship between a respondent having a family history of mental issues and whether those same respondents had also sought treatment. I wanted to see if there was a potential hereditary issue that needed to be investigated. For this I made sure the responses were yes's and no's then updated them to 1's and 0's. I then looked at correlation between family_history and treatment. After consulting with my resource, my correlation has a moderate, positive relationship between the two variables. (Ratner, 2009) Now that it looks like there is a possibility of a hereditary issue, I plan to investigate further, but not entirely sure how I want to proceed. That will be my next steps is deciding how I want to answer my other questions.

Conclusion

I still have some investigation to do to answer all of the questions I listed in my proposal. I also need to decide how I want to proceed as no modeling technique has stuck out as being the best way to move forward. Either way, my client has somewhere to start with revamping their benefits for mental health which is a start. I hope to provide a more detailed answer in the coming week.

10 Questions

1. Does it matter location when it comes to knowledge of benefits?
2. Why have 8 different types of genders?

3. Do you plan to investigate the different answers with negative consequences between mental and physical health?
4. Will your client keep everything completely anonymous or confidential?
5. What types of wellness programs do you think your client will incorporate into their existing benefits?
6. How does respondent confidence change by size of employer?
7. Will you look at the different states as well as countries for more specifics?
8. How will your client handle the uncertain answers when it comes to sharing mental health concerns in the workplace?
9. If a respondent had observed negative consequences to someone else sharing their mental health how would your client react?
10. Why does your client want to know about other tech companies mental health benefits?

References

Davis, M. (2018, October 8). *The Conversation on Mental Health is Changing the Tech Industry and Supporting Employee Wellbeing*. Retrieved from Thriveworks: https://thriveworks.com/blog/mental-health-tech-industry-employee-wellbeing/

Lopukhina, D. (2020, February 20). *Are We TAckling Mental Health in Tech Companies the Right Way?* Retrieved from Thriveglobal: https://thriveglobal.com/stories/are-we-tackling-mental-health-in-tech-companies-the-right-way/

Open Sourcing Mental Illness, L. (2016, November 3). *Mental Health in Tech Survey*. Retrieved from kaggle: https://www.kaggle.com/osmi/mental-health-in-tech-survey

Ratner, B. (2009, May 18). *The Correlation Coefficient: Its values range between +1/-1, or do they?* Retrieved from Springer Link: https://link.springer.com/article/10.1057/jt.2009.5

Appendix

Table 1 – Variables from Dataset (Open Sourcing Mental Illness, 2016)

Variable	Description
Timestamp	Time survey was submitted
Age	Age of respondent
Gender	Gender of respondent
Country	Country respondent resides
State	State of respondent, if applicable
Self_employed	Is the respondent self employed?
Family_history	Does the respondent have family history of mental illness?
Treatment	Has respondent sought treatment for mental health condition
Work_interfere	If respondent has a mental condition, does it interfere with work?
No_employees	Number of employees at respondents company
Remote_work	Does company respondent works at allow at least 50% work from home?
Tech_company	Is employer primarily a tech company?
Benefits	Does employer provide mental health benefits?
Care_options	Does the respondent know what care options are available for mental health?
Wellness_program	Has employer discussed mental health as part of the employee wellness program?
Seek_help	Does employer provide resources to learn more and seek help?
Anonymity	Is anonymity protected if respondent chooses to use mental health or substance abuse treatment?
Leave	Is it easy to take leave for a mental health condition?
Mental_health_consequence	Do you think talking about a mental health issue with your employer would have negative consequences?
Phys_health_consequence	Do you think talking about a physical health issue with your employer would have negative consequences?
Coworkers	Would respondent be willing to discuss a mental health issue with a coworker?
Supervisor	Would respondent be willing to discuss a mental health issue with supervisor?

Mental_health_interview	Would respondent share about a mental health condition to a potential employer in an interview?
Phys_health_interview	Would respondent share about a physical health condition to a potential employer in an interview?
Mental_vs_physical	Does respondent feel that employer takes mental health as seriously as physical health?
Obs_consequence	Has respondent ever seen any negative consequences happen to a coworker who talked about their own mental health condition at work?
Comments	Any comments respondent felt needed

Figure 1 – Overview Dashboard

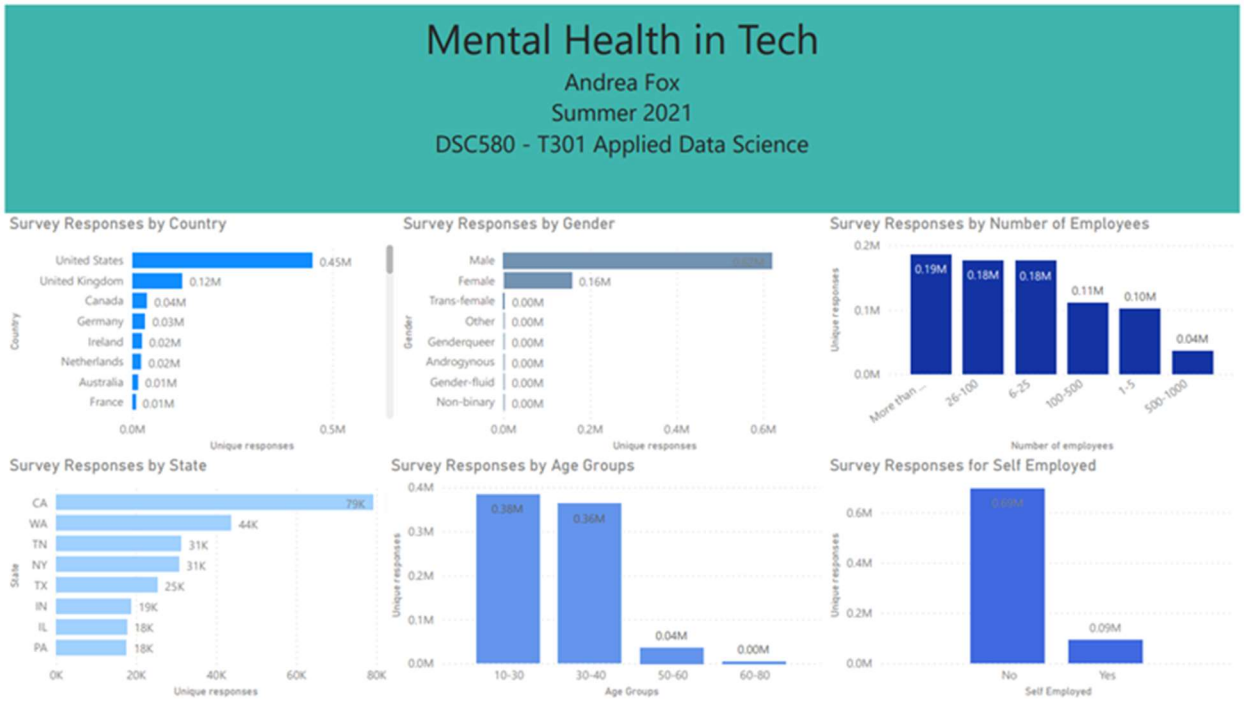


Figure 2 – Benefits Dashboard

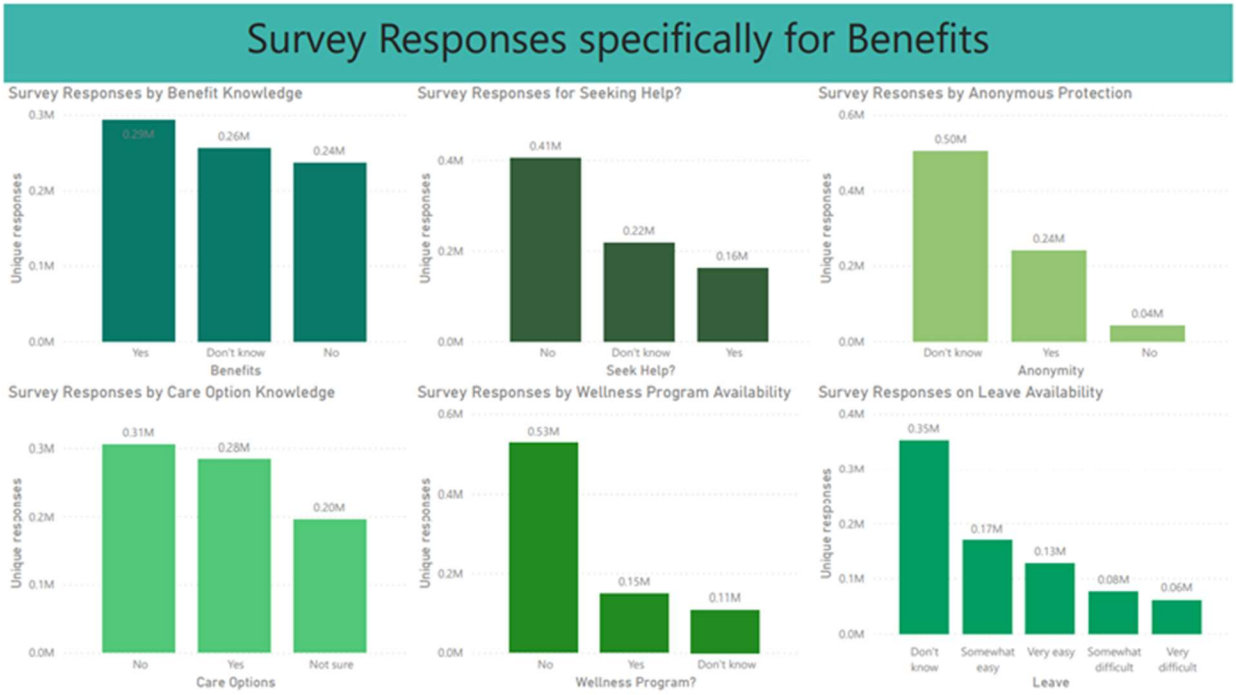


Figure 3 – In Workplace Dashboard