Laura Abood, Kelly Gu, Edison Huang, Alisha Qian, Carlie Stewart, Hana Wang

Dataset: https://www.kaggle.com/datasets/osmi/mental-health-in-tech-2016

As computer science and data science students about to enter the tech field, understanding the mental health landscape within our industry is crucial. This dataset on mental health in tech provides an opportunity to explore trends and patterns related to the well-being among those already in the industry. By analyzing self-reported mental health statuses and workplace conditions, we can first address any common patterns in the findings and then assess whether there are any improvements that can be made for the future of the industry to both improve employee mental health and expand upon existing resources related to mental health.

This dataset will help us identify practices that contribute to better mental health outcomes and uncover stigma surrounding mental health issues. By gaining these insights, we can advocate for healthier work environments and foster a more inclusive culture in our future workplaces. Ultimately, this knowledge equips us to contribute positively to the tech industry, prioritizing mental well-being alongside innovation and productivity.

This survey data offers insights into the attitude towards mental health in the tech workplace, exploring the frequency of mental health disorders among tech workers. The data is compiled by volunteers driven by raising awareness for mental health and supporting those who have mental health disorders. The data was collected from a survey and compiled into a CSV file that includes demographic statistics like the individual age, gender, country/state of residence, as well as other questions regarding the work conditions, insights into the employer, and the stigma surrounding mental health in a work environment.

While the dataset offers significant insights, there is still the possibility of challenges arising as we are working with this data. A key concern to be aware of is the data quality and completeness, as this data is self-reported, being prone to errors. We would have to figure out how to deal with possible missing values or any inconsistencies while maintaining the integrity of our analysis. Another potential challenge is the generalizability of the findings. The data obtained may not represent all sectors within the tech industry, which puts a limit on any findings which could affect the applicability of conclusions drawn.