Final Project

Mariana MacDonald

2022-05-15

1) Identify a topic or a problem that you want to research. Provide an introduction that explains the problem statement or topic you are addressing. Why would someone be interested in this? How is it a data science problem?

Mental health influences on an adult's career.

I believe that, based on a brief research, majority of population with depression are women. I assume that this adds to the fact that women are not treated the same way as men are in most of the companies. Businesses are affected by untreated depressed people. Depression and other mental disorders are very subjective and hard to identify. People often say they are sad, depressed but not necessarily have a mental health issue. Data can help improve the treatment, the trial and error for medication identifying a pattern, the early detection of the symptoms, which can reduce the impact for that person. Companies should consider depression as a condition and maybe even, consider as part of an inclusion program. Knowing how to assist, companies can provide the support, the access to health insurance and benefit from great employees that just need some treatment.

2) Draft 5-10 Research questions that focus on the problem statement/topic.

- What gender tend to be more depressive?
- Which age group has more depression?
- What is their marital status?
- What is the work situation? employed? unemployed?
- How many % in the world suffer from depression
- What are mental health disorder types?

3) Provide a concise explanation of how you plan to address this problem statement.

My plan is to research for data that can prove my predictions that majority of the population with depression are middle age women, who make less money than men (depending on them) and suffer with their career path.

4) Discuss how your proposed approach will address (fully or partially) this problem.

I am sure there are lots of studies out there with the same purpose, but my idea is to help women with depression to grow in companies. Provide awareness to the companies that there is discrimination and they should include these employees like any other disorder/disease and not discriminate.

I will analyze different data sources in hope of an useful outcome.

5) Do some digging and find at least 3 datasets that you can use to address the issue. (There is not a required number of fields or rows for these datasets)

- Original source where the data was obtained is cited and, if possible, hyperlinked.
- Source data is thoroughly explained (i.e. what was the original purpose of the data, when was it collected, how many variables did the original have, explain any peculiarities of the source data such as how missing values are recorded, or how data was imputed, etc.).

DATA 1: https://data.world/vizzup/mental-health-depression-disorder-data (data.world, n.d.)
DATA 2: https://www.kaggle.com/datasets/nilimajauhari/glassdoor-analyze-gender-pay-gap (kaggle, n.d.a)
DATA 3: https://www.kaggle.com/datasets/arashnic/the-depression-dataset (kaggle, n.d.b)

6) Identify the packages that are needed for your project.

I may need more or less packages than described here, it will depend on my future analysis, but for now, I believe I will need: ggplot2, readxl, plyr, Dplyr, magrittr, lm.beta, carData, Hmisc

7) What types of plots and tables will help you to illustrate the findings to your research questions?

Comparison of gender wage gap Comparison of gender with depression Histograms

8) What do you not know how to do right now that you need to learn to answer your research questions?

Logistic regression and machine learning.

data.world, amitd -. n.d. “Mental Health Depression Disorder Data.”
kaggle, multiple contributors -. n.d.a. “Glassdoor- Analyze Gender Pay Gap.”
———. n.d.b. “The Depression Dataset.”