Kevin Rabidou

ANA 502 Second Year Data Practicum

Initial Data Preparation

The data being studied is focusing on Mental Health in the Tech related industry and was sourced from Kaggle with the dataset owner being the Open Sourcing Mental Illness, LTD. All the data collected comes from between early 2014 to late 2015. The dataset started with 27 variables and 1259 entries and after the initial data cleaning and preparation process there are 24 variables and 1233 entries. The variables were timestamp, state, and comments as they were not pertinent to any analysis. The comments will be used potentially in later analysis when it comes to more specific questions but when looking to understand the data and the distribution it is not needed. The state column is not needed currently as I am not looking specifically at the United States (if looking just at US I will include states for a potential more geographic approach).

Of the remaining variables, 1 is an integer type (Age) and the other 23 are object types. The summary statistics that are going to be provided focus on the distribution of responses and looking at the percentage of answers received under each variable. Many of the variables ate binary with yes or no responses and another group incorporate a third response depending on the question type and the participant not knowing the information.

Definitions of each variable

Age: Age listed

Gender: Gender/Sex listed

Country: Country of origin

self\_employed: Is the participant self employed?

family\_history: Is there a mental health history in the participants family?

treatment: Has medical treatment been used or gone after for mental health conditions?

remote\_work: Do you work at least 50% remotely?

tech\_company: Is the company worked for primarily a Tech group?

obs\_consequence: Have negative remarks/actions been seen/heard surrounding mental health conditions in the workplace?

benefits: Are mental health benefits being provided by the employer?

care\_options: Does the participant know the mental health options that are available to them?

wellness\_program: Has there been a discussion on including mental health as part of an employee wellness program?

seek\_help: Are resources available to learn how to seek help on mental health issues?

anonymity: Are you able to stay anonymous if you use any resources for mental health treatment?

mental\_health\_consequence: In your opinion, will discussing mental health issues will have any negative consequences with you employer?

phys\_health\_consequence: In your opinion, will discussing physical health issues will have any negative consequences with you employer?

coworkers: Are there coworkers that you would be comfortable to talk about mental health issues with?

supervisor: Are there supervisors/managers that you would be comfortable to talk about mental health issues with?

mental\_health\_interview: Would you bring up mental health issues during an interview?

phys\_health\_interview: Would you bring up physical health issues during an interview?

mental\_vs\_physical: Is mental health seen as serious as physical health with the emplpoyer?

work\_interfere: Does their mental health condition impact or interfere the work that they do?

no\_employees: How many employees are there in the company that they work for?

leave: How easy or difficult is it to take leave sue to a mental health condition?

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Mean | Median | Standard Deviation |
| Age | 32 | 31 | 7 |

|  |  |  |
| --- | --- | --- |
| Gender | Count | Percentage |
| Male | 973 | 78.9% |
| Female | 242 | 19.6% |
| Other | 18 | 1.5% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | United States | United Kingdom | Canada | Germany | 42 other countries |
| Count (%) | 735 (59.6) | 182 (14.8) | 68 (5.5) | 45 (3.6) | 203 (16.4) |

|  |  |  |
| --- | --- | --- |
| Variable | Yes (%) | No (%) |
| self\_employed | 142 (11.5) | 1091 (88.5) |
| family\_history | 482 (39.1) | 751 (60.9) |
| treatment | 623 (50.5) | 610 (49.5) |
| remote\_work | 365 (29.6) | 868 (70.4) |
| tech\_company | 1009 (81.8) | 224 (18.2) |
| obs\_consequence | 179 (14.5) | 1054 (85.5) |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Yes (%) | No (%) | Don’t know/Not sure/ Maybe/Some of them (%) |
| benefits | 466 (37.7) | 307 (24.9) | 400 (32.4) |
| care\_options | 434 (35.2) | 492 (39.9) | 307 (24.9) |
| wellness\_program | 227 (18.4) | 824 (66.8) | 182 (14.8) |
| seek\_help | 246 (19.9) | 632 (51.3) | 355 (28.8) |
| anonymity | 370 (30.3) | 61 (5.0) | 802 (65.0) |
| mental\_health\_consequence | 286 (23.2) | 476 (38.6) | 471 (38.2) |
| phys\_health\_consequence | 56 (4.5) | 905 (73.4) | 272 (22.1) |
| coworkers | 214 (17.3) | 256 (20.8) | 763 (61.9) |
| supervisor | 501 (40.6) | 385 (31.2) | 347 (28.2) |
| mental\_health\_interview | 38 (3.1) | 994 (80.6) | 201 (16.3) |
| phys\_health\_interview | 196 (15.9) | 490 (39.7) | 547 (44.4) |
| mental\_vs\_physical | 336 (27.2) | 331 (26.9) | 566 (45.9) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Don’t know (%) | Never (%) | Rarely (%) | Sometimes (%) | Often (%) |
| work\_interfere | 262 (21.2) | 206 (16.7) | 170 (13.8) | 457 (37.1) | 138 (11.2) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | 1-5 (%) | 6-25 (%) | 26-100 (%) | 100-500 (%) | 500-1000 (%) | More than 1000 (%) |
| no\_employees | 156 (12.6) | 283 (23.0) | 283 (23.0) | 172 (13.9) | 59 (4.8) | 280 (22.7) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Don’t know (%) | Very difficult (%) | Somewhat difficult (%) | Somewhat easy (%) | Very easy  (%) |
| leave | 552 (44.8) | 96 (7.8) | 121 (9.8) | 262 (21.2) | 202 (16.4) |