**Mental Health in the Tech Industry**

[**https://www.kaggle.com/datasets/anth7310/mental-health-in-the-tech-industry**](https://www.kaggle.com/datasets/anth7310/mental-health-in-the-tech-industry)

Analizar estructura de la base datos

Cargar la base de datos a sql

Crear una imagen que contenga las tablas y sus relaciones

Modelo E-R

Conectar la base de datos con Python y trabajar allí (tratar con queries siempre que se pueda) (usar import sqlite3 as sql)

A hacer:

1. Mostrar cada tabla
2. Averiguar cantidad total de personas encuestadas (ver si no hay ids repetidos)
3. Mostrar cómo responde la gente, darle sentido a lo que significa (ej. ¿el 1 qué significa?, ¿el -1?). En el caso de que respondan con texto, categorizar como un grupo
4. Eliminar datos no válidos
5. Mostrar gráfico de torta con porcentaje de respuestas numéricas y respuestas texto
6. Diagrama de barras de años del survey
7. Histograma de edades
8. Diagrama de barras de género (considerar sólo tres categorías: Male, Female y Nonbinary, el resto tratar de meterlos en estos grupos y si no se puede no considerarlos)
9. ¿Cuál es la pregunta que más se hizo?
10. Mostrar tabla completa de datos (sin primaries keys), unificar
11. Mostrar tabla de encuestados con menos de 30 años

Dejé abajo una tabla con las siguientes questionids para tenerlas a mano.

Tener cuidado en los siguientes puntos: no todas las personas respondieron todas las preguntas. Para hacer el análisis utilizar las personas que sí hayan respondido ambas preguntas (excluido el punto 19).

1. Relación entre cantidad de empleados (questionid = 8) y cuidados de la salud mental en el trabajo (se pueden usar las questionid = 10, questionid = 15, questionid = 23)
2. Relación entre edad (questionid = 1) y self-employed (questionid = 5)
3. Considerar las personas que responden que sí a la questionid = 20 y ver qué responden en la questionid = 12, questionid = 7, question = 72 y questionid = 33
4. Relación entre edad (questionid = 1) y si tuvo empleadores anteriormente (questionid = 22)
5. Encontrar a la persona más joven que cumpla las siguientes condiciones:
   * vive en Estados Unidos (questionid = 3)
   * trabaja remoto (questionid = 118)
6. Considerar la cantidad de respuestas positivas a la questionid = 34 (¿alguna vez fue diagnosticada la persona con una enfermedad mental?) y evaluar si hay congruencia con la questionid = 33 (¿actualmente la persona tiene una enfermedad mental?) y la questionid = 32 (¿ha tenido en el pasado la persona una enfermedad mental?). ¿Debería pasar que las respuestas positivas de la primera pregunta sean igual a la suma de las respuestas positivas de las otras dos?
7. Agrupar las personas por cantidad de preguntas que respondieron y considerar las que tienen más de 66.
8. Quedarse con las personas que más se involucran con la salud mental en el trabajo (mediante puntaje), el puntaje se obtiene de la siguiente forma:
   * 1 punto si se responde Sí a la questionid = 12
   * 1 punto si se responde Sí a la questionid = 14
   * 1 punto si se responde Sí a la questionid = 21
   * 1 punto si se responde Sí a la questionid = 62
   * 2 puntos si se responde NO a la questionid = 108
   * 2 puntos si se responde NO a la questionid = 109
   * -2 puntos si se responde lo contrario a los anteriores puntos en al menos 3 preguntas

|  |  |
| --- | --- |
| question | questionid |
| What is your age? | 1 |
| What country do you live in? | 3 |
| Have you ever sought treatment for a mental health disorder from a mental health professional? | 7 |
| How many employees does your company or organization have? | 8 |
| Does your employer provide mental health benefits as part of healthcare coverage? | 10 |
| Would you bring up a mental health issue with a potential employer in an interview? | 12 |
| Do you know the options for mental health care available under your employer-provided health coverage? | 14 |
| Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)? | 15 |
| Do you have medical coverage (private insurance or state-provided) that includes treatment of mental health disorders? | 20 |
| Do you know local or online resources to seek help for a mental health issue? | 21 |
| Do you have previous employers? | 22 |
| Have your previous employers provided mental health benefits? | 23 |
| Do you currently have a mental health disorder? | 33 |
| Have you ever had a coworker discuss their or another coworker's mental health with you? | 62 |
| Did you ever discuss your mental health with a previous coworker(s)? | 72 |
| Do you think that discussing a mental health disorder with previous employers would have negative consequences? | 108 |
| Do you think that discussing a physical health issue with previous employers would have negative consequences? | 109 |
| Do you work remotely? | 118 |

Ojo = questionid es una columna

No hay una sóla forma de solucionar las preguntas, mientras se argumente correctamente estará bien

¿Cómo conectar al dataset?

!wget https://github.com/LautaroOchotorena/Mental-Health-in-the-Tech-Industry-en-curso-/raw/main/mental\_health.sqlite

#Importamos la librería

import sqlite3 as sql

con = sql.connect("mental\_health.sqlite")

cursor = con.cursor()

#Pedimos lo que queremos

query = """

select \* from Answer

"""

#Ejecuta la query

cursor.execute(query)

#Muestra todos los resultados

cursor.fetchall()

#También existe fetchmany(5), muestra 5 elementos y fetchone() muestra uno

answer = pd.read\_sql\_query(query,con)

anser.head()