**PEC 3**

## **Ejercicio 2**

**Limpieza de los datos.**

**2.1 Selección de los datos de interés a analizar ¿Cuáles son los campos más relevantes para responder al problema?**

Las variables con las que nos vamos a quedar para dar solución al problema son:

1. Age: edad
2. Gender: género
3. Country: país
4. surself\_employed: auto-empleado. ¿Es autónomo (auto-empleado)?
5. family\_history: historia familiar. ¿Tiene antecedentes de enfermedad mental en la familia?
6. treatment: tratamiento. ¿Ha sido tratado por una enfermedad mental?
7. work\_interfere: ¿Si tiene una enfermedad mental, siente que interfiere con su trabajo?
8. no\_employees: ¿número de empleados. ¿Cuántos empleados tiene su compañía o organización?
9. remote\_work: ¿realiza teletrabajo (fuera de la oficina) al menos el 50% del tiempo?
10. tech\_company: ¿su empleador primaro es una organización o empresa de tecnología?
11. benefits: ¿su empleador provee beneficios de salud mental?
12. care\_options: ¿conoce las opciones de cuidado mental de la compañía médica que el empleador provee?
13. wellness\_program: ¿Su empleador ha mencionado alguna vez que tiene un programa de bienestar mental para sus empleados?
14. seek\_help: ¿Su empleador proporciona recursos para saber más sobre aspectos de salud mental y cómo encontrar ayuda?
15. anonymity: ¿Está protegida su privacidad si elige acogerse a ventajas de salud mental o recursos de tratamiento de abusos de substancias?
16. leave: ¿Le sería fácil, acogerse a una baja por situación de salud mental?
17. mental\_health\_consequence: ¿Cree que hablar de un aspecto de salud mental con su empleador, tendría consecuencias negativas?
18. phys\_health\_consequence: ¿Cree que hablar de un aspecto de salud física con su empleador, tendría consecuencias negativas?
19. coworkers: ¿Estaría dispuesto a hablar con sus compañeros de un aspecto de salud mental?
20. supervisor: ¿Estaría dispuesto a hablar con sus supervisores de un aspecto de salud mental?
21. mental\_health\_interview: ¿Mencionaría un aspecto de salud mental con un potencial empleador en una entrevista?
22. phys\_health\_interview: ¿Mencionaría un aspecto de salud física con un potencial empleador en una entrevista?
23. mental\_vs\_physical: ¿Siente que su empleador se toma la salud mental como un aspecto importante de la salud?
24. obs\_consequence: ¿Ha oido u observado consecuencias negativas para sus compañeros de trabajo que se encuentren en situación de enfermedad mental en su puesto de trabajo?

**2.2 ¿Los datos contienen ceros o elementos vacíos? ¿Y valores extremos? ¿Cómo gestionarías cada uno de estos casos?**

-Guión de resolución de la cuestión

Diferencia entre el valor cero que proviene de un hecho medible del cual se pueden realizar operaciones matematicas. Ejm vacios correspondiente a la variable número de hijo este vacio se rellena con Cero.

Vacio se produce cuando se carece de observación. Añadimos NA

Los errores atípicos que tenemos provienen o tienen su origen en errores de muestras luego la forma de tratarlos no es otra que la eliminación de estos.

En el análisis de la variable cuantitativa Age nos encontramos con valores extremos o valores atípicos. Estos son:

-1726, -29, -1, 5, 8, 11, 329, 99999999999

Nota: En el texto debemos incluir las fases de un proyecto analítico.

1. Se trata de encontrar la cuestión que deseamos resolver

2. Consiste en la recogida y almacenamiento de los datos. Conocer de dónde se han extraído los datos y el formato de almacenamiento.

3. Limpieza de datos. Los datos son preparados para el análisis. Para ello es muy posible que se produzca eliminaciones, transformaciones, etc.

4.En esta etapa se produce el estudio de los datos y un aprendizaje de forma automática.

5.Aqui nos encontramos con el estudio de establecer la forma visual más eficiente para la representación de los datos.

6.Resolvemos la cuestión que se planeó en la primera fase del proyecto.