## 



Individual Delivery

23/12/2020 to 13/01/2021 ← The start date to end date

**─**

María del Mar Delgado Domínguez

Data Science

The Bridge

# General Vision

Actualmente, vivimos en una década marcada por los términos diversidad, inclusión e igualdad que están presentes en todos los aspectos de vida. En un sector como el empresarial, no se ha quedado atrás y apuesta por la realización de estrategias que apoyan a diversos colectivos que aún son discriminados por la sociedad.

Cada vez más las empresas compiten por diferenciarse del resto apostando por la creación de un lugar de trabajo más diverso como motor para favorecer el bienestar de los empleados y que esto a su vez impulse la productividad y la innovación empresarial. Uno de los colectivos más representados es el colectivo LGTBI, ya que desde varios años atrás las legislaciones europeas han ido regulando la no discriminación por orientación sexual o de género en el ámbito laboral, obligando a las empresas a tomar medidas al respecto.

Promueven ambientes diversos e inclusivos y lucen en el mes del orgullo la imagen de su marca con los colores de la bandera del orgullo. No obstante, el 31,2% de los ciudadanos europeos ocultan su identidad en el trabajo y solo un 16% lo expresa abiertamente.

Es por ello que el proyecto que se presenta a continuación analiza la situación de las personas LGTBI en el ámbito laboral. El conjunto de datos es extraído de una encuesta realizada a ciudadanos europeos por la Agencia de los Derechos Fundamentales de la Unión Europea, FRA. El cuestionario está compuesto por una amplia gama de temas, relacionados con experiencias sobre discriminación, acoso o violencia, conciencia de derechos, apertura sobre ser LGBTI, experiencias positivas y negativas en el trabajo, educación, condiciones socioeconómicas, etc. Cada temática contiene diversas preguntas y las respuestas pueden ser abiertas o cerradas.

# Goals

El principal objetivo de este proyecto es comprobar si la hipótesis presentada *“Las personas LGTBI eligen no revelar su orientación o identidad sexual en el trabajo a pesar de las estrategias de diversidad e inclusión que existen actualmente en el sector empresarial “,* se confirma a través del análisis y visualización de los datos.

A lo largo del trabajo he ido utilizando diversas librerías como Pandas, Numpy, Seaborn, Matplotlib y plotly para la visualización de los datos. He completado todos los pasos requeridos en la *Option C*, además de realizar algún que otro paso de las otras opciones, por lo que considero que puedo optar por la *Option B*.

# Specifications

## Software

El software que he utilizado para desarrollar el proyecto ha sido a través del sistema operativo Windows 10, no obstante, también puede desarrollarse en otros sistemas operativos siempre y cuando pueda accederse a alguna aplicación que pueda leer el lenguaje utilizado. Este proyecto de análisis de datos ha sido desarrollado a través del lenguaje de programación Python, por lo que para poder visualizar el proyecto es necesario instalarse Python o utilizar algunas de las plataformas online para poder leer dicho lenguaje.

El programa por el que he ejecutado Python es Visual Studio Code, pero cualquier otro que pueda leer y ejecutar Python serian válidos. La versión que he utilizado de Python es la 3.9.

## Hardware

En cuanto al hardware, el tipo de sistema operativo es de 64 bits, con un procesador Intel CORE i5 y una memoria RAM de 8GB. Para poder ejecutar este programa sin problemas considero que lo más importante es tener una memoria RAM de mínimo 8GB. En cuanto al tipo de procesador y sistema operativo puede variar según el software instalado.

## Requirements

El conjunto de datos se ha obtenido de la [página oficial](https://fra.europa.eu/en/tools) de la Agencia de los Derechos Fundamentales de la Unión Europea, FRA, la cual permite exportar una colección de datos que tienen almacenados sobre diversos ámbitos relacionados con los derechos humanos. Por lo que, para poder acceder a su página, se necesita conexión a internet.

En este caso, los datos se han exportado al ordenador y guardado en la carpeta data, después para poder visualizarlos en Python es necesario hacerlo con un programa como Visual Studio Code, en mi caso, el cual permita leer el lenguaje Python.

# Steps

## Research the context

Investigando sobre el tema del proyecto he podido descubrir que son muchas empresas las que apuestan por estrategias específicas para combatir la discriminación LGTBI en el trabajo a través de actividades, talleres, campañas de concienciación, etc, todo ello recogido en manuales de buenas prácticas.

Aun así, hay que tener en cuenta que solo una mínima parte de las personas LGTBI revelan su identidad en el ámbito laboral, es por ello que surge la cuestión sobre si dichas estrategias que son llevadas por grandes multinacionales están dando sus frutos o si por el contrario aún queda mucho camino por recorrer para que se hagan efectivas.

## Get Data

El conjunto de datos ha sido exportado desde la página oficial de la Agencia de los Derechos Fundamentales de la Unión Europea, FRA, el cual está compuesto por un cuestionario realizado por las instituciones europeas hacia la comunidad LGTBI.

En este caso al querer analizar diversas temáticas para obtener los resultados, se ha exportado cada temática en un dataset en formato XLS.

## Data Wrangling

Para poder analizar los dataset, primeramente, se han pasado a formato CSV para después importarlos como DataFrames de Python. Primeramente, se han importado todas las datas conjuntas directamente y posteriormente se han dividido para analizar cada temática individualmente.

Los dataset estaban formados por cuatro columnas, en primer lugar *“CountryCode”* formada por todos los países de la Unión Europea, incluido Reino Unido ya que el cuestionario se realizó en 2019, y los países candidatos del Norte de Macedonia y Serbia. Otra de las columnas es *“question\_label”* donde se incluían las preguntas de cada temática, también está la columna *“answer”* con las respuestas a cada pregunta, que generalmente eran cerradas y distintas según la pregunta. Y la columna *“percentage”* con el porcentaje de las respuestas contestadas según cada país. El resto de columnas fueron borradas porque no aportan ninguna información relevante.

Para acceder a analizar los datos, la columna *“percentage”* ha sido transformada a formato “integer” o “int”, es decir, numérico, ya que estaba en formato *“object”* y no se podía trabajar con ella numéricamente. También se han eliminado las filas “UE-28” de la columna *“CountryCode”,* ya que recogía información de todos los países de la Unión Europea. Otras de las acciones realizadas ha sido filtrar a través de la columna *“question\_label”* cuáles eran las preguntas más interesantes para poder llegar a la hipótesis.

Principalmente, el trabajo ha sido sobre este filtrado para extraer la información relevante sobre la hipótesis y visualizar aquellas preguntas o conjunto de preguntas que aportaban las conclusiones necesarias.

También se han realizado agrupaciones para poder visualizar por países o por respuestas totalizadas. Igualmente, también se ha comprobado si existe correlación entre dos de las preguntas filtradas creando un nuevo dataset y representando la matrix correlación.

## Data Mining / Clean Data

Para poder analizar correctamente los datos, de la base principal se eliminaron los valores Nan, ya que eran muy pocos respecto al conjunto de datos y formaban parte de una pregunta que por el momento no se utilizaría para sacar conclusiones.

Lo más importante es el cambio de “object” a “int” en la columna “percentage”, comentado anteriormente, para poder trabajar con los datos y la eliminación de las columnas que no aportaban información relevante. Además de las agrupaciones necesarias realizadas para extraer información.

## Others...

En cuanto a la visualización de los datos se han utilizado diferentes librerías y gráficos para representar las conclusiones del proyecto, comentadas anteriormente.

A continuación, las preguntas requeridas en la *“Option B”* :

a. ¿Fue posible demostrar la hipótesis? ¿Por qué?  
b. ¿Qué puede concluir sobre su estudio de los datos?  
c. ¿Qué cambiaría si tuviera que hacer otro proyecto de EDA?  
d. ¿Qué aprende haciendo este proyecto?

1. La hipótesis se ha podido demostrar con solo una pregunta ya que la mayoría de las personas LGTBI ocultan su identidad en el trabajo de manera parcial o completa. Además, gracias al análisis de otras preguntas se ha podido ver que la mayoría de los compañeros/as de trabajo no apoyan, protegen o promueven los derechos de la comunidad LGTBI y esto se ha podido correlacionar siendo una de las causas de porque no se muestran abiertamente en el trabajo.
2. Como conclusión destacar que hay muchas personas que no expresan su orientacion o identidad sexual abiertamente en el ámbito laboral y a pesar de que el lugar de trabajo no es uno de los sitios donde más evitan mostrarse tal y como es, un 80% de las personas no muestra abiertamente su identidad por miedo a ser discriminado por ello.
3. Cambiaria la forma en la que comencé haciéndolo, ya que al principio no tenia muy clara la hipótesis y fui buscando conclusiones sin saber que buscaba, eso hizo que me retrasara mucho en el trabajo. Sin duda antes de comenzar a analizar los datos hay que tener claro lo que se pretende buscar y cuales son los objetivos del proyecto.
4. Con este proyecto he aprendido a hacer mi primer proyecto EDA y poder extraer información relavnte a partir de los datos. Además, al hacerlo de manera individual y sin un seguimiento continuo por parte de los profesores me he enfrentado a más retos de los que imaginaba y he aprendido de cada uno de ellos.