SEMANA 1

El objetivo principal de esta semana es iniciar las bases de lo que será nuestro proyecto y portafolio.  
   
Los objetivos específicos son:

* Crear repositorio de GitHub.
* Escoger un problema que nos interese.
* Encontrar un dataset relacionado a la problemática escogida (sugerencia Kaggle).
* Subir datos a github y realizar una primera carga de datos en jupyter notebook.
* Por último, realizar preguntas de interés sobre el dataset para posterior análisis.

En la parte de las preguntas la idea es hacer una lluvia de ideas o “brainstorming” para imaginarnos como puede verse nuestro futuro proyecto. Te recomiendo utilizar mapas mentales o dibujos para plantear tus ideas.

# Proyecto 1. Análisis exploratorio de ofertas de empleo del ámbito data.

## Objetivo

Conocimiento del estado actual del mercado de trabajo en relación a diferentes puestos del ámbito del Big Data, tratando de dar respuestas a los siguientes puntos o preguntas:

* Evolución anual del número total de empleos de Data (2020-2022) por tipología de perfil.
* Relación entre la cantidad de ofertas de empleo y el nivel de experiencia.
* Ofertas de empleo por tipo de perfil (data analyst, Data Science, Data engineer)
* Ofertas por tipo de empleo (PT, FT, CT)
* Relación entre salario y experiencia requerida.
* Relación entre salario y tamaño de la empresa
* Relación entre nivel de experiencia y modalidad de empleo (remoto, híbrido, presencial)
* Localización geográfica de las ofertas
* Relación entre el tamaño de la empresa y salario.
* Relación entre el tamaño de la empresa y la cantidad de ofertas de empleo.
* Relación entre tamaño de la empresa y el requerimiento de la empresa.

## Dataset:

Se trabajará con un archivo CVS procedente de la web <https://ai-jobs.net/salaries/download/> que incluye información de las ofertas publicadas en dicha web. Se tomará como referencia para tener una aproximación a la realidad del mercado de trabajo.

### Exclusiones del dataset

Solo se incluyen las ofertas de empleo publicadas en dicha web. No aparecen ofertas de trabajo en España ni de otras plataformas.

### Estructura del data set:

El dataset está compuesto por las siguientes 11 columnas y sus respectivos atributos:

work\_year: The year the salary was paid.

experience\_level: The experience level in the job during the year with the following possible values:

EN Entry-level / Junior

MI: Mid-level / Intermediate

SE: Senior-level / Expert

EX: Executive-level / Director

Employment\_type: The type of employement for the role:

PT: Part-time

FT: Full-time

CT: Contract

FL: Freelance

job\_title: The role worked in during the year.

Salary: The total gross salary amount paid.

Salary\_currency:

salary\_in\_usd: The salary in USD (FX rate divided by avg. USD rate of respective year via data from BIS).

employee\_residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.

remote\_ratio: The overall amount of work done remotely, possible values are as follows:

0: No remote work (less than 20%)

50: Partially remote

100: Fully remote (more than 80%)

company\_location: The country of the employer's main office or contracting branch as an ISO 3166 country code.

company\_size: The average number of people that worked for the company during the year:

S: less than 50 employees (small)

M:50 to 250 employees (medium)

L: more than 250 employees (large)