# Práctica Final Asignatura Herramientas del Científico de Datos

## Entrega: 09/06/2024

La práctica final de la asignatura consiste en resolver una serie de ejercicios de Python. Se debe adjuntar el script en formato .ipynb y debe ser íntegramente ejecutable.

Se recomienda comentar el código explicando que se ha hecho en cada apartado y porque se ha recurrido a dicha solución.

La práctica está formada por 11 apartados cuya puntuación es de un punto cada uno. La práctica forma el 100% de la nota total, la mitad del total corresponde al ejercicio de Python y la otra mitad para R.

Los datasets son funds.csv y funding\_rounds.csv

Se han obtenidos dos datasets relacionados con la adquisición y fundación de \*startups\* a nivel global, el objetivo general, es realizar un análisis exploratorio de datos sobre los elementos comunes de ambos datasets. Estos son los objetivos a realizar:

1. Convierte los archivos funds.csv y funding\_rounds.csv a pandas dataframes.

2. Busca un elemento común de ambos dataframes y conviértelos en uno solo.

3. Elimina las columnas 'id', 'source\_url', 'source\_description', 'updated\_at'.

4. Muestra el número de nulos que tiene cada columna del dataframe y qué porcentaje supone dentro de cada columna el número de nulos.

5. Modifica todos los valores de 'raised\_currency\_dode' nulos a 'USD'.

6. Elimina el resto de nulos del dataset.

7. Muestra el porcentaje de los diferentes valores de 'raised\_currency\_dode'.

8. Crea un sub-conjunto sin el valor USD y visualiza un diagrama de barras.

9. Crea un nuevo sub-conjunto filtrando del dataset todas las filas por debajo de 10M y visualiza el histograma.

10. Con el conjunto del apartado anterior. Investiga sobre la función contains() <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.str.contains.html>

filtra el dataset por las fechas de creación de las startup en 2000

11. Sobre el conjunto filtrado por fecha visualiza la cuantía obtenida explicada por el tipo de moneda.