

# PEC 1 (20% nota final)

## Presentación

En esta Prueba de Evaluación Continuada se trabajan los conceptos generales de cuál es el ciclo de vida de los datos, y se identifican y conocen sus características. También se trabajan los conceptos esenciales de Web Scraping.

## Competencias

En esta PEC se desarrollan las siguientes competencias del Máster de Data Science:

- Capacidad de analizar un problema en el nivel de abstracción adecuado a cada situación y aplicar las habilidades y conocimientos adquiridos para abordarlo y resolverlo.
- Capacidad para aplicar las técnicas específicas de tratamiento, almacenamiento y administración de datos

## Objetivos

Los objetivos concretos de esta Prueba de Evaluación Continua son:

- Conocer el ciclo de vida de los datos y los principales tipos de datos.
- Aprender a aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios o multidisciplinares.
- Desarrollar las habilidades de aprendizaje que permitan continuar estudiando de una manera que tendrá que ser en gran medida autodirigida o autónoma.
- Desarrollar la capacidad de busca, gestión y uso de información y recursos en el ámbito de la ciencia de datos.
- Entender la utilidad, la legalidad y algunas características de web scraping.

## Descripción de la PEC a realizar

### Ejercicio 1 [70%]

Después de leer el recurso “Calvo, M., Pérez, D., Subirats, L. (2019). Introducción al ciclo de vida de los datos.” contesta las siguientes preguntas con tus propias palabras:

- 1 ¿Qué perfil profesional relacionado con la ciencia de datos te gustaría ser? ¿Y cuál te gustaría menos ser? Razona la respuesta (máximo 200 palabras).
- 2 Lista los diferentes factores que influyen en la calidad de los datos y pon un ejemplo diferente al que se explica en los materiales (máximo 300 palabras).
- 3 ¿En qué tipos de base de datos no es necesario conocer a priori los datos que se quieren almacenar? Pon tres ejemplos de tecnologías que utilicen estas bases de datos (máximo 100 palabras).
- 4 Pon ejemplos visuales con imágenes de las 7 visualizaciones que permiten un nivel más alto de abstracción (adjuntar las 7 imágenes).

### Ejercicio 2 [30%]

Después de leer el recurso “Subirats, L., Calvo, M. (2019). Web Scraping.”, capítulos 1 y 7. Contesta las siguientes preguntas con tus propias palabras:

- 1 ¿Por qué crees que es necesario hacer web scraping)? (máximo 100 palabras)
- 2 ¿Por qué es importante analizar el contenido del archivo *robots.txt*? ¿Qué riesgo corremos si no lo hacemos? (máximo 100 palabras).
- 3 Explica como evitarías saturar el servidor con peticiones web (máximo 100 palabras).

## Recursos

Los siguientes recursos son de utilidad por la realización de la PEC:

### Básicos

- Calvo, M., Pérez, D., Subirats, L. (2019). Introducción al ciclo de vida de los datos. Editorial UOC.
- Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC.

### Complementarios

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 1. Introduction to Web Scraping.
- Minguillón, J. (2016). Fundamentals of Data Science. Editorial UOC.

## Criterios de valoración

La ponderación de los ejercicios es la siguiente:

- Ejercicio 1: 70%
- Ejercicio 2: 30%

Se evaluará la precisión de los ejemplos, así como el respeto al número de palabras máximo establecido para cada pregunta. La idoneidad y claridad de las respuestas también será evaluada.

## Formato y fecha de entrega

Hay que entregar un único documento Word, Open Office o PDF (**preferiblemente este último**) con las respuestas a las preguntas.

Este documento se tiene que entregar en el espacio de Entrega y Registro de AC del aula antes de las **23:59** del día **8 de marzo**. No se aceptarán entregas fuera de plazo.