

# PEC 1 (10% 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 Continuada 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 “Fundamentos de Data Science” contesta las siguientes preguntas:

1. Explique los tres tipos de datos y proponga un ejemplo para cada uno sin utilizar los ejemplos ya planteados en el recurso (máximo 200 palabras).
2. Con la ayuda de un ejemplo real, explique cada una de las fases del ciclo de vida de los datos necesarias para la resolución de un cierto problema (máximo 250 palabras).
3. En la fase de captura del ciclo de vida de los datos, ¿estos siempre pueden ser creados?. Explique los dos principales mecanismos para la captura de datos y desarrolle un ejemplo para cada uno de los mecanismos (máximo 200 palabras).

### Ejercicio 2 [30%]

Después de leer el Capítulo 1 “Introduction to Web Scraping” del recurso “Web Scraping with Python” contesta las siguientes preguntas:

1. ¿Cuándo es legal utilizar web scraping y cuando no?. Explica con un ejemplo para cada caso (máximo 100 palabras).
2. ¿Qué son los robots.txt? (máximo 50 palabras).
3. ¿Cual es la ventaja de establecer un *user agent*? (máximo 50 palabras).
4. ¿En qué consiste el *Throttling download*? (máximo 50 palabras).

## Recursos

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

### Básicos

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

## Criterios de valoración

La ponderación de los ejercicios es la siguiente:

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

## 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 octubre**. No se aceptarán entregas fuera de plazo.