

**Máster universitario de Ciencia de Datos**

**Prueba de Evaluación Continua 1 – PEC1**

**Trabajo sobre los conceptos generales del ciclo de vida de los datos y *Web Scraping***

Autor:

Mario Ubierna San Mamés

|  |
| --- |
| Índice de Contenido |

[Índice de Contenido 3](#_Toc65150183)

[Índice de tablas 4](#_Toc65150184)

[Índice de ilustraciones 5](#_Toc65150185)

[1. Introducción 6](#_Toc65150186)

[1.1. Presentación 6](#_Toc65150187)

[1.2. Objetivos 6](#_Toc65150188)

[2. Enunciado 7](#_Toc65150189)

[2.1. Ejercicio 1 7](#_Toc65150190)

[2.2. Ejercicio 2 9](#_Toc65150191)

[3. Bibliografía 10](#_Toc65150192)

|  |
| --- |
| Índice de tablas |

**No se encuentran elementos de tabla de ilustraciones.**

|  |
| --- |
| Índice de ilustraciones |

**No se encuentran elementos de tabla de ilustraciones.**

|  |
| --- |
| Introducción |

## 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*.

## 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 multidisciplinarios.
* 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*.

|  |
| --- |
| Enunciado |

## Ejercicio 1

*Después de leer el recurso “Calvo, M., Pérez, D., Subirats, L. (2019). Introducción al ciclo de vida de los datos.” Contesta a 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 tu respuesta (máximo 200 palabras).*

Dada a mi poca experiencia profesional no sé si lo que me gusta hoy va a ser lo mismo que lo que me gustará mañana, pero a día de hoy el rol que más encaja conmigo es el de *data scientist*. Soy una persona inquita de mente y por lo que he leído tanto en los recursos de esta asignatura como en otras, el *data scientist* es un filósofo (curioso) [1], pero con datos, por lo que trata de buscar respuestas a preguntas que puedan solucionar o mejorar el futuro a medio/largo plazo, además hace uso de las tecnologías, matemáticas, lógica de negocio y comunicación, las cuales son las áreas del conocimiento que más me gustan.

Por otro lado, lo que sí tengo claro es que no quiero el día de mañana hacer tareas que sean más mecánicas, es decir, que no se haga uso de esa curiosidad o querer saber, por lo que roles como por ejemplo: arquitecto de datos, administrador de base de datos, entre otros no son los roles a los que me quiero dedicar.

1. *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).*

Los diferentes factores que influyen en la calidad de los datos son [1]:

* Exactitud: un ejemplo sería cuando tratamos de representar números decimales, es decir, en España por ejemplo usamos la coma para separar la parte entera de la parte decimal, y en el sistema inglés se usa el punto. Por lo que, si tenemos por ejemplo la variable peso medida en Kg dicho valor viene como en el sistema inglés con un punto y nos encontramos en España, ambos formatos serían válidos pero no exactos ya en este caso tendría que representarse con una coma.
* Completitud: siguiendo con el ejemplo anterior sería el porcentaje de pesos que sí que tienen valor, es decir, que tienen datos sin valores en blanco.
* Consistencia: por ejemplo, a la hora de sacarse el carnet de conducir una persona es apta si y solo si ha pasado el test psicotécnico, ha aprobado el examen teórico y el examen práctico. Por lo tanto, hay que comprobar que estos tres campos sean consistentes, ya que de lo contrario la persona no sería apta.
* Atemporalidad: siguiendo con el ejemplo anterior, podríamos tener un atributo temporal que indicase cuánto ha tardado una persona en sacarse el carnet desde que hizo el examen psicotécnico hasta que aprobó el examen práctico.
* Unicidad: se correspondería con el porcentaje de valores duplicados respecto a todos los datos que tenemos en tráfico sobre las personas que quieren sacarse el carnet de conducir.
* Validez: por ejemplo, en el test psicotécnico, en el examen teórico y en el examen práctico solo es válido una cadena que indique “A” si es apto o “NA” si es no apto, de lo contrario el resto de los valores no estarían permitidos.

1. *¿En qué tipos de bases 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).*

Las bases de datos que a priori no necesitan conocer los datos que se van a almacenar son las bases de datos no relacionales.

Un ejemplo de uso sería *Facebook App*, dicha aplicación puede estar instalada tanto en *Android* como en *IOS* y *Windows Phone*, por lo que a la hora de capturar los datos de los usuarios se puede recibir la información de forma diferente, es decir, en estructuras diferentes.

Otro ejemplo, el uso de sensores en un coche por parte de un fabricante, cada modelo puede llevar sensores diferentes y no sabes cómo vas a recibir la información.

Finalmente, ahora que está de moda el planeta Marte, el *Perseverance* [2] aterrizó hace poco en dicho planeta, no sabemos lo que se va a encontrar por lo que la información que envíe debe tener cierta flexibilidad.

1. *Pon ejemplos visuales con imágenes de las 7 visualizaciones que permiten un nivel más alto de abstracción (adjuntar las 7 imágenes)*

## Ejercicio 2

|  |
| --- |
| Bibliografía |

[1] «PID\_00265705.pdf». Accedido: feb. 25, 2021. [En línea]. Disponible en: http://materials.cv.uoc.edu/daisy/Materials/PID\_00265705/pdf/PID\_00265705.pdf.

[2] M. Interesante, «Perseverance: este es el primer vídeo de un aterrizaje en Marte», *MuyInteresante.es*, feb. 23, 2021. https://www.muyinteresante.es/ciencia/video/primeros-videos-del-perseverance-en-marte-251614022782 (accedido feb. 25, 2021).