



# Tablón - M2.851 - Tipología y ciclo de vida de los datos aula 1

## Recibidos

### Mireia Calvo Gonzalez

Inicio de la Práctica 1

**A:** Tablón Tipología y ciclo de vida de los datos aula 1

**Fecha:** 09.03.2018 00:32

Hola a [tod@s](mailto:tod@s),

(<mailto:tod@s>). Ya tenéis disponible el enunciado de la Práctica 1 en el aula, cuyo principal objetivo es la extracción de datos procedentes de un sitio web. La idea es que trabajéis en grupos de 2-3 personas, pero si lo preferís, también podéis realizar la práctica de manera individual. Para organizar los grupos, podéis utilizar la carpeta Práctica1/GruposPráctica1 del foro.

Durante la semana del 19 de marzo, el grupo podrá mandarme por correo electrónico ([mcalvogonza@uoc.edu](mailto:mcalvogonza@uoc.edu) (<mailto:mcalvogonza@uoc.edu>)) una entrega parcial opcional, con el repositorio Github con el que hayáis avanzado. Esta entrega parcial es muy recomendable para recibir asesoramiento sobre la dirección tomada sin que esto cuente para la nota final de la práctica.

Como recursos para realizar la práctica tenéis el libro "Web scraping with Python", del que os recomiendo leer el capítulo 2 ya que en él se habla de 3 aproximaciones para hacer web scraping. No obstante, tenéis disponible el libro entero por si deseáis profundizar más. Si preferís realizar la práctica en R, os recomiendo que leáis "Part Three A Bag of Case Studies" del libro "Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining". Otros recursos extra que os pueden ser útiles son los siguientes:

- <https://www.datacamp.com/courses/free-introduction-to-r>  
(<https://www.datacamp.com/courses/free-introduction-to-r>).
- <https://www.datacamp.com/courses/intro-to-python-for-data-science>  
(<https://www.datacamp.com/courses/intro-to-python-for-data-science>).

Además, en el mismo enunciado encontraréis ejemplos de 2 alumnos del semestre pasado que os pueden ayudar a enfocar la práctica.

No dudéis en enviar vuestras preguntas al foro (carpeta Práctica1/DudasPráctica1) o en contactarme directamente si pensáis que vuestras preguntas son muy concretas.

El documento final deberá entregarse en el espacio de Entrega y Registro de AC del aula antes de las 23:59 del lunes 16 de abril. Por último, recordaros que, para aprobar la asignatura, es obligatorio entregar

todas las actividades.

¡Mucho ánimo, y seguimos!

Mireia