**UOC\_PRAT1\_WEB\_SCRAPING**

**UOC Máster en ciencia de datos**

**Practica 1 Web Scraping**

**Descripción:**

Establecer los números y estrellas más repetidos durante los sorteos de Euromillones, previa generación de un dataset con los datos de los sorteos desde el 10 de Mayo de 2004 y hasta el 25 de Octubre de 2019.

**Bibliotecas:**

Las siguientes bibliotecas son necesarias para la ejecución de la aplicación Python:

pip install requests

pip install beautifulsoup4

pip install pandas

pip install matplotlib

pip install datetime

**Ficheros:**

* /EuroMill.py: Código en Python para la realización del web scraping
* /EuroMillones.csv: fichero csv con el dataset obtenido
* /Gonzalez\_Martinez\_Diego\_TCVD\_PRACTICA1.pdf: Respuestas a la práctica 1
* /Gráfico Repeticiones números.jpg
* /Gráfico Repeticiones estrellas.jpg

**Dataset**

El dataset está compuesto por 9 atributos:

* Fecha
* dia\_sorteo
* numero1
* numero2
* numero3
* numero4
* numero5
* estrella1
* estrella2

**Recursos**

* Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC. Lawson, R. (2015).
* Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
* Sitio web con la historia de Euromillones: <https://www.euro-millions.com/es/historia>
* Sitio web con los datos de Euromillones: <https://www.combinacionganadora.com/euromillones/>