

# Tipología i cicle de vida de les dades

## Pràctica 1: Web Scraping

Alumne: Antonio Milán Otero

# Introducció

Amb aquest projecte extreurem les ofertes de treball llistades a la web oficial de l'estat espanyol i les transformarem en un dataset que pot ser utilitzat per estudiar les característiques de la oferta laboral espanyola.

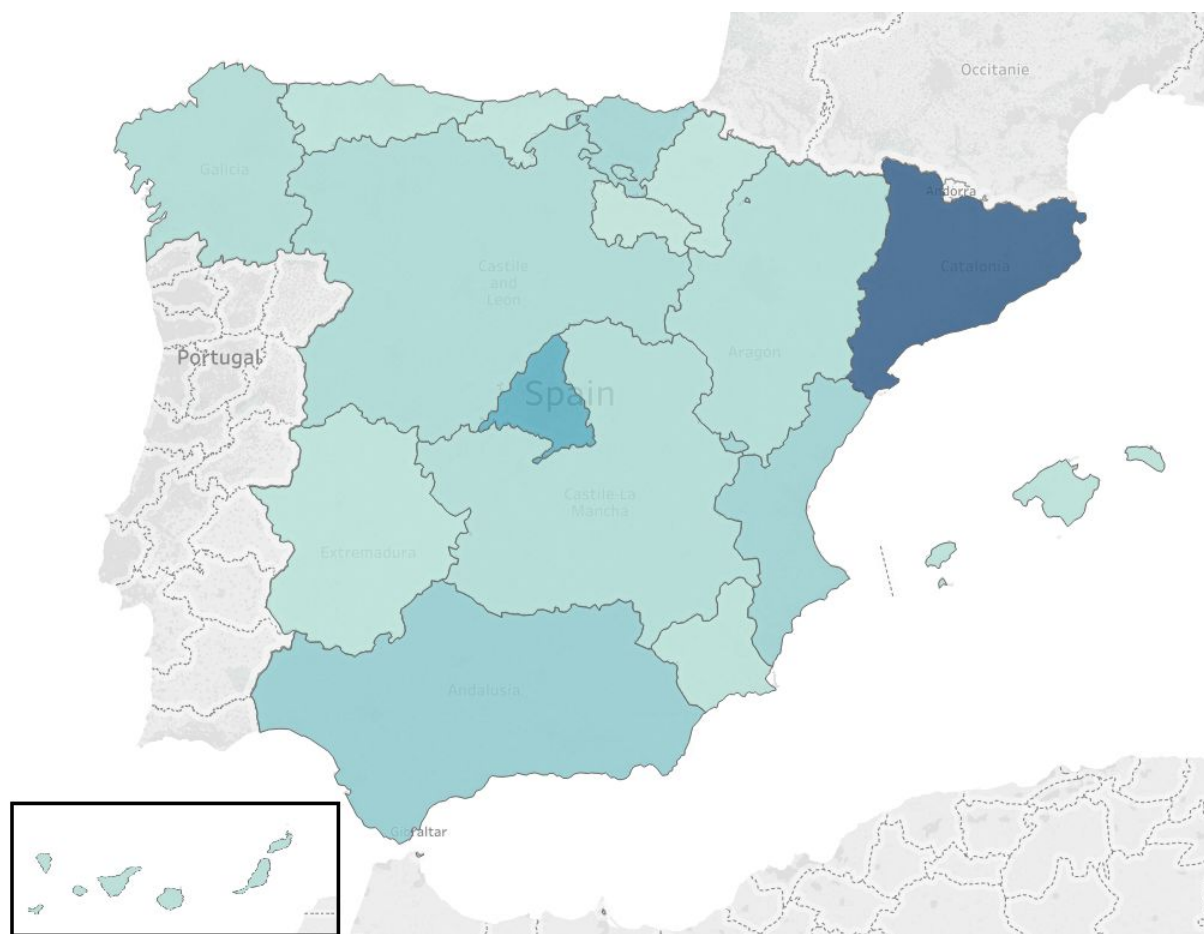


Figura 1: Distribució del nombre d'ofertes per comunitat, durant les setmanes 43-44 de 2018.

## Descripció del dataset

Aquest data set proporciona els següents camps (entre altres):

- Nom de la oferta
- Titulació necessària
- Lloc
- Salari
- Descripció
- Data

Aquestes dades s'han recollit durant les setmanes 43-44 de 2018 fent servir un procés de web scraping sobre el portal de treball que proporciona el govern espanyol [1].

# Agraïments

Aquestes dades són facilitades pel Servicio Público de Empleo Estatal i com podem llegir a la seva política de privacitat:

*Utilitzación de la información contenida en la Web.*

*El Servicio Público de Empleo Estatal facilita la consulta libre y gratuita de la información contenida en la Web, permitiendo la copia y distribución de páginas siempre que se citen la fuente y la fecha en la que se ha realizado la copia, no se manipulen ni alteren los contenidos y no se utilice directamente con fines comerciales.*

En aquest cas farem un ús merament acadèmic, les fonts de cada oferta estan incloses a la mateixa fila que la oferta i a més a més distribuïrem les dades amb llicència CC BY-NC-SA 4.0 License ja que es la que millor s'adapta a les demandes de la política de privacitat de la web originaria. Per més detalls sobre aquesta llicència consultar [2]

# Inspiració

La inspiració per portar a cap aquest projecte es la de facilitar una millor comprensió de l'estat laboral a Espanya.

Aquestes dades poden servir per fer diferents estudis sobre el mercat laboral a l'estat Espanyol. Algunes idees d'ús serien:

- Analitzar els diferents requeriments professionals que tenen les diferents autonomies d'Espanya.
- Identificar el tipus i la qualitat del treball actual al país.
- Analitzar les regions amb més i menys ofertes de treball.
- Analitzar la distribució de les diferents professions en funció de la regió.

També es podria expandir el data set amb informació sobre el cost de la vida per tal de donar una visió sobre a quina part d'Espanya rendeix mes un mateix salari.

# Llicència

Aquests dades han estat extretes de la web [1] i com comentem amunt, les dades son públiques sempre que es compleixin les condicions esmentades. Per tal motiu, aquests dades les oferim sota la llicència CC0: Public Domain License.

## Referències

[1]: <https://empleate.gob.es/empleo/#/>

[2]: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

## Bibliografia

- Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC.
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.