

PRA 1 - Tipología y ciclo de vida de los datos

Maria Dolores Moyano Guerrero y Víctor Cáncer Castillo

Índice

Contexto	2
Título	3
Descripción del dataset	3
Representación gráfica	5
Contenido	5
Agradecimientos	7
Inspiración	7
Licencia	8
Código	9
Dataset	9
Vídeo	9

Contexto

Estos datos se han recogido para practicar el *web scraping* en la asignatura de *Tipología y ciclo de vida de los datos* del Máster de ciencia de datos de la UOC.

Cómo (futuros) científicos de datos hemos tenido la curiosidad de estudiar cómo está el mercado laboral actualmente en varias ciudades europeas y americanas. Además hemos querido averiguar en qué lugares el trabajo de científico de datos está más reconocido por las empresas y por lo tanto mejor remunerados.

Para ello hemos obtenido los sueldos que se ofrecen por diferentes empresas utilizando la web <https://glassdoor.es/>, donde los trabajadores pueden informar de su sueldo de manera anónima. Por otro lado hemos extraído datos de la web <https://datosmacro.expansion.com/> donde hay múltiples datos económicos, entre ellos el salario medio, lo cual nos puede mostrar si el trabajo del científico de datos está mejor/peor remunerado que el resto de trabajos en ese país o ciudad.



Título

Obtendremos dos datasets de los cuales extraeremos nuestras conclusiones:

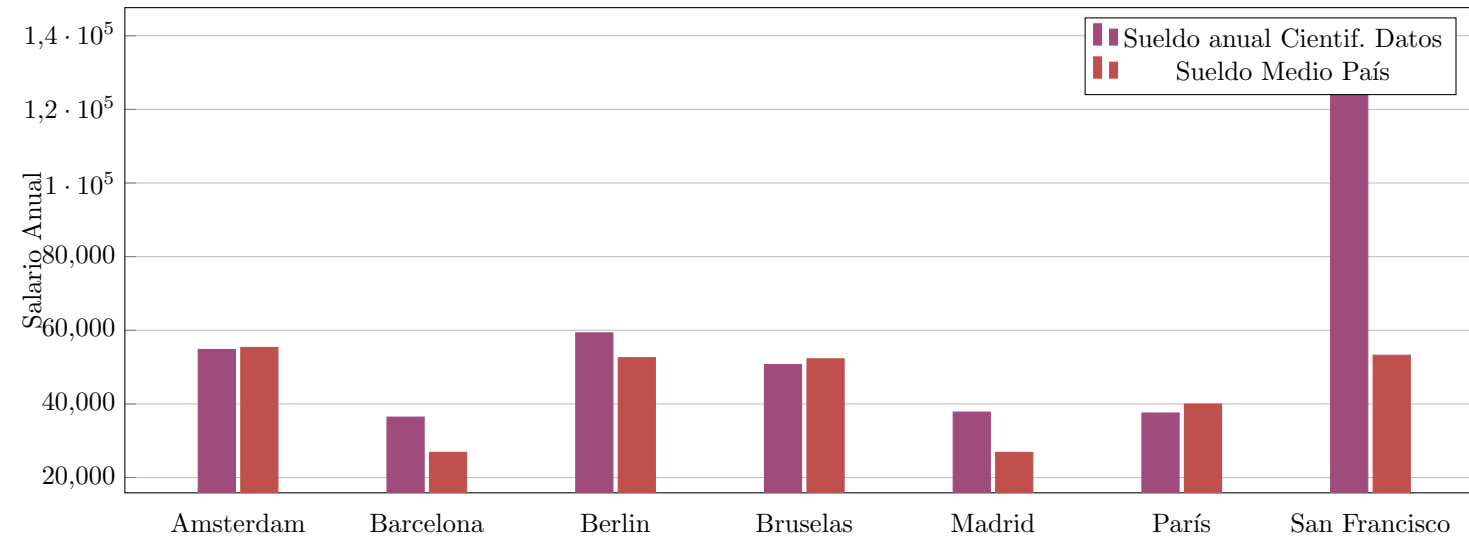
- *Sueldos_DataScientist_ciudades.csv*
- *SMI.csv*

Descripción del dataset

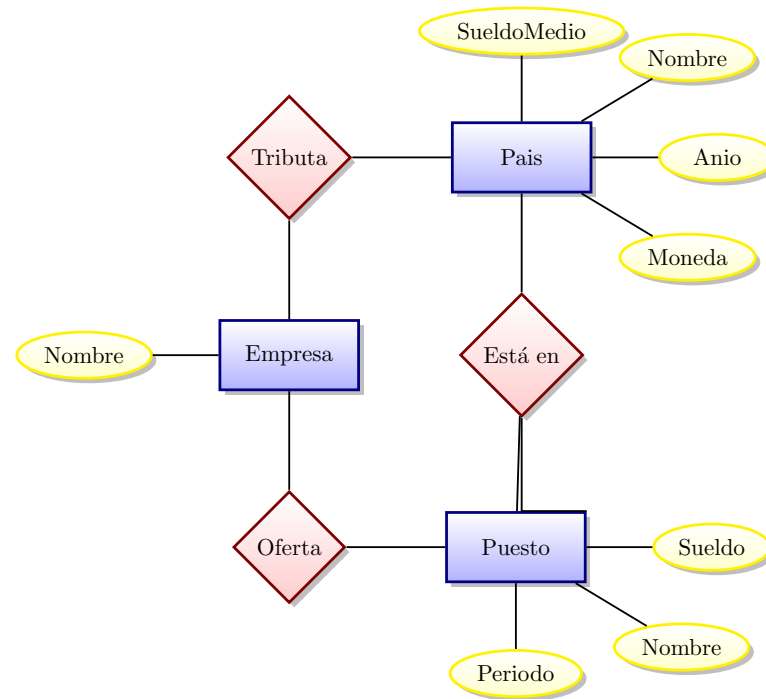
Para las ciudades descritas se recogen los datos:

- Dataset_sueldos.csv
- Dataset_sueldos_clean.csv
- SMI.csv
- Sueldos_DataScientist_ciudades.csv

Los datos obtenidos recogen los datos siguientes, a fecha 3/04/2022, gráficamente:



Representación gráfica



Contenido

Dataset_sueldos.csv

- **Ciudad:** Ciudad donde está radicado el empleo.
- **Empresa:** Empresa que proporciona el empleo.

- **Sueldo:** Sueldo bruto correspondiente al empleo, ciudad y empresa.
- **Periodo:** Empleo en el periodo indicado, en su caso.

Dataset_sueldos_clean.csv

- **Ciudad:** Ciudad donde está el empleo.
- **Empresa:** Empresa que proporciona el empleo.
- **Sueldo:** Sueldo bruto correspondiente al empleo, ciudad y empresa.
- **Sueldo Anual:** Sueldo ajustado a un año (algunas empresas, no lo proporcionan anualizado).

SMI.csv

- **País:** País que se corresponde al sueldo medio.
- **Año:** Año de cálculo del sueldo medio.
- **SalMed Local:** Salario medio en moneda local.
- **Moneda:** Moneda del salario local.
- **Salmed \$:** Salario medio en dólares.
- **Salmed €:** Salario medio en euros.
- **Var:** Variación del salario medio.

Sueldos_DataScientist_ciudades.csv

- **Ciudad:** Ciudad donde está el empleo.
- **Sueldo anual:** Sueldo anual de los empleos descargados.
- **Suelvo vs Sueldo Medio:** Diferencia entre el sueldo del puesto de Científico de Datos y el sueldo medio del país.

Agradecimientos

Agradecemos a los propietarios de las páginas <https://glassdoor.es/> y <https://datosmacro.expansion.com/>, por los datos de los que nos hemos podido alimentar en este proyecto de Web Scraping.

Inspiración

Los datos recogidos son interesantes porque permiten comparar los sueldos de los científicos de datos en caliente, utilizando el factor de corrección del sueldo medio del país, para poder analizar si efectivamente esos puestos de trabajo son valorados en la ciudad donde se demandan.

Licencia

Se ha escogido la licencia CC BY-SA 4.0 ya que es la más correcta en cuanto a las libertades que se han de cumplir:



Reconocimiento-CompartirIgual 4.0 Internacional (CC BY-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#). Advertencia.

Usted es libre de:

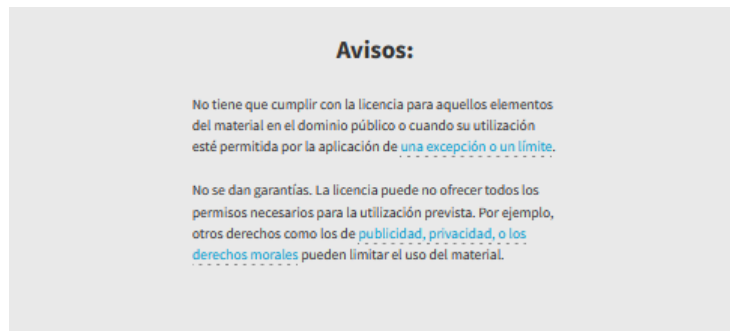
- Compartir** — copiar y redistribuir el material en cualquier medio o formato
- Adaptar** — remezclar, transformar y crear a partir del material para cualquier finalidad, incluso comercial.

El licenciador no puede revocar estas libertades mientras cumpla con los términos de la licencia.

Bajo las condiciones siguientes:

- Reconocimiento** — Debe [reconocer adecuadamente](#) la autoría, proporcionar un enlace a la licencia e [indicar si se han realizado cambios](#). Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
- CompartirIgual** — Si remezcla, transforma o crea a partir del material, deberá difundir sus contribuciones bajo la [misma licencia que el original](#).

No hay restricciones adicionales — No puede aplicar [términos legales o medidas tecnológicas](#) que legalmente restrinjan realizar aquello que la licencia permite.



Código

Para hacer tanto scraping como pre-procesado de datos hemos utilizado Python. El código está disponible en GitHub: https://github.com/mdmoyano/web_scraping/src.

Dataset

El Dataset está publicado en Zenodo, disponible en [10.5281/zenodo.6408002](https://zenodo.org/record/105281)

Vídeo

– Link al video de cada uno –