

**AUTORES: María Sánchez Oliva y Cayetano Bautista Pérez**

**1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.**

Este dataset recoge las ofertas de trabajo relacionadas con los datos publicadas en la página web <https://www.domestika.org>. Debido a la gran cantidad de ofertas que podemos encontrar en esta página, que no es ni mucho menos la número 1 en portales de empleo, nos damos cuenta de la importancia que tienen los datos hoy en día para todas las empresas de cualquier ámbito y de cualquier magnitud.

Podemos encontrar ofertas de cualquier ciudad de España, y también de las ciudades de Los Ángeles y Nueva York de Estados Unidos. Para cada oferta de trabajo se da el detalle del puesto a desempeñar, nombre de la empresa, tipo de jornada que se realizaría, si la modalidad es de teletrabajo o presencial (en cuyo caso se especifica la ciudad a la que habría que acudir a desempeñar el puesto de trabajo) y la fecha de la publicación de la oferta, aparte de una descripción detallada de cada oferta donde podemos encontrar las funciones a desempeñar, las titulaciones requeridas, el salario, etc.

Esta página web, aparte de publicar ofertas de empleo de distintas categorías y empresas, tiene una comunidad en la que participan multitud de profesionales con diferentes perfiles creativos, donde se pueden encontrar una gran variedad de cursos online de todo tipo de ramas.

**2. Definir un título para el dataset. Elegir un título que sea descriptivo.**

Data Jobs Offers.

**3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).**

Ofertas de trabajos relacionadas con los datos publicadas en el portal web <https://www.domestika.org>.

**4. Representación gráfica. Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido**



##### 5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

- **Title:** Título de la oferta, que corresponde con el puesto de trabajo a desempeñar.
- **Link:** Link de la oferta completa publicada en la web de Domestika.
- **Company:** Nombre de la empresa.
- **Contract:** Tipo de jornada (Tiempo completo, Freelance, Prácticas, etc).
- **Location:** Modalidad de trabajo, en remoto o presencial (en este caso se especifica la ciudad donde habría que acudir a desempeñar el puesto de trabajo).
- **Date:** Fecha de publicación de la oferta en formato dd/mm/aa.
- **Description:** Descripción detallada de la oferta donde podemos encontrar las funciones principales a desarrollar, los conocimientos y competencias requeridas, una breve descripción de la empresa, herramientas informáticas necesarias, etc.

**6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.**

A parte de los fundadores de la página web <https://www.domestika.org> donde se publican las ofertas, los propietarios también son las personas de las empresas que publican cada oferta. Les agradecemos que den a conocer estas ofertas de trabajo en una web pública, en la que puede tener acceso a ella cualquier persona, y no lo hagan sólo de manera corporativa.

**7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.**

Este dataset puede ser muy interesante para analizar el tipo de ofertas que se ofrecen en trabajos relacionados con el dato.

Al poder obtener todas las ofertas que se han publicado en la página <https://www.domestika.org/es/jobs/> desde hace más de 10 años, podemos realizar un análisis comparativo por línea temporal.

Podríamos analizar, por ejemplo, en qué ciudades se suelen ofertar este tipo de trabajos y hacer un ranking de las ciudades en las que más ofertas se encuentran, o si el tipo de jornada para este perfil de trabajador suele ser a jornada completa, o sin embargo son trabajos en los que se contrata más de “Freelance”.

Una línea de investigación interesante, dadas las circunstancias en las que nos encontramos ahora con la pandemia sería analizar en el transcurso de los años, la tasa de ofertas en las que la modalidad de trabajo es presencial y en las que es en forma de teletrabajo. Aquí no hay ninguna duda de que a medida que avanzamos en el tiempo, y más concretamente en este último año, la mayoría de las ofertas de trabajo apuestan por la modalidad de teletrabajo al no querer poner en riesgo a sus trabajadores y más aún, si tenemos en cuenta que muchos de los perfiles relacionados con el dato, son perfiles de trabajo que se pueden desempeñar sin ningún problema desde el domicilio, sin necesidad de acudir a ninguna oficina, evitando costes también a la empresa en cuanto a instalaciones y todo lo que el mantenimiento de ellas conlleva.

Por otra parte, también se podría estudiar el salario que se ofrece en este tipo de ofertas, y seguramente nos encontraríamos, en que en los últimos años, estos puestos se revalorizan, en comparación con los años anteriores, debido al boom del big data, y el descubrimiento de los grandes avances y mejoras que esta técnica puede aportar a una compañía.

Como vemos, se podrían investigar análisis verdaderamente interesantes con este dataset, dada la importancia que existe hoy en día en torno a los datos.

**8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:**

- ☐ Released Under CC0: Public Domain License
- ☐ Released Under CC BY-NC-SA 4.0 License
- ☐ Released Under CC BY-SA 4.0 License
- ☐ Database released under Open Database License, individual contents under Database Contents License
- ☐ Other (specified above)

La licencia escogida para la publicación de este conjunto de datos ha sido CC BY-SA 4.0 License. Los motivos que han llevado a la elección de esta licencia tienen que ver con la idoneidad de las cláusulas que esta presenta en relación con el trabajo realizado:

Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado. De esta manera, se reconoce el trabajo ajeno y en qué medida se han realizado aportaciones en relación con el trabajo original.

Se permite un uso comercial. Esto haría que incrementen las probabilidades de que una empresa utilice los datos generados y realicen trabajos de calidad que reporten cierto reconocimiento al autor original.

Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma. Esto hace que el trabajo del autor original continúe distribuyéndose bajo los términos que él mismo planteó.

**9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.**

**10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.**

DOI: 10.5281/zenodo.4682153

| Contribuciones              | Firma          |
|-----------------------------|----------------|
| Investigación previa        | M.S.O., C.B.P. |
| Redacción de las respuestas | M.S.O., C.B.P. |
| Desarrollo código           | M.S.O., C.B.P. |