

Conjuntos de datos y fuentes abiertas

En este mundo impulsado por datos, algunos conjuntos de datos están disponibles de forma gratuita para que cualquiera los acceda, use, modifique y comparta. Estos se llaman **conjuntos de datos abiertos**.

Los conjuntos de datos abiertos incluyen una licencia pública y son muy útiles para tu camino como Científico de Datos. Algunas de las fuentes de conjuntos de datos abiertos más informativas se enumeran a continuación.

Datos del Gobierno:

- <https://www.data.gov/>
- <https://www.census.gov/data.html>
- <https://data.gov.uk/>
- <https://www.opendatanetwork.com/>
- <https://data.un.org/>

Fuentes de Datos Financieros:

- <https://data.worldbank.org/>
- <https://www.globalfinancialdata.com/>
- <https://comtrade.un.org/>
- <https://www.nber.org/>
- <https://fred.stlouisfed.org/>

Datos del Crimen:

- <https://www.fbi.gov/services/cjis/ucr>
- <https://www.icpsr.umich.edu/icpsrweb/content/NACJD/index.html>
- <https://www.drugabuse.gov/related-topics/trends-statistics>
- <https://www.unodc.org/unodc/en/data-and-analysis/>

Datos de Salud:

- <https://www.who.int/gho/database/en/>
- <https://www.fda.gov/Food/default.htm>
- <https://seer.cancer.gov/faststats/selections.php?series=cancer>

- <https://www.opensciencedatacloud.org/>
- <https://pds.nasa.gov/>
- <https://earthdata.nasa.gov/>
- <https://www.sgim.org/communities/research/dataset-compendium/public-datasets-topic-grid>

Datos Académicos y Empresariales:

- <https://scholar.google.com/>
- <https://nces.ed.gov/>
- <https://www.glassdoor.com/research/>
- <https://www.yelp.com/dataset>

Otros Datos Generales:

- <https://www.kaggle.com/datasets>
- <https://www.reddit.com/r/datasets/>

Conjuntos de datos y fuentes de propiedad

Los conjuntos de datos de propiedad contienen datos que son principalmente de propiedad y controlados por individuos u organizaciones específicas. Estos datos tienen una distribución limitada porque se venden con un acuerdo de licencia. Algunos datos de fuentes privadas no pueden ser fácilmente divulgados, como los datos públicos.

Los datos de seguridad nacional, geológicos, geofísicos y biológicos son ejemplos de datos de propiedad. Las leyes de derechos de autor o patentes suelen vincular este tipo de datos. Los conjuntos de datos de propiedad que contienen principalmente información sensible están menos disponibles que los conjuntos de datos abiertos.

A continuación se enumeran algunas fuentes estándar de conjuntos de datos de propiedad.

Cuidado de la salud:

<https://www.sgim.org/communities/research/dataset-compendium/proprietary-datasets>

Datos del mercado financiero:

<https://datarade.ai/data-categories/proprietary-market-data>

Conjuntos de datos basados en Google Cloud:

<https://cloud.google.com/datasets>

Licencias de conjuntos de datos

Cuando seleccionas un conjunto de datos, es necesario revisar la licencia. Una licencia explica si puedes usar ese conjunto de datos o no; o explica si debes aceptar ciertas pautas para usar ese conjunto de datos. Los diferentes tipos de licencias se enumeran a continuación.

1. **MARCAS DE DOMINIO PÚBLICO - DOMINIO PÚBLICO**

Cuando un conjunto de datos tiene una licencia de Dominio Público, todos los derechos para usar, acceder, modificar y compartir el conjunto de datos están abiertos a todos. Aquí técnicamente no hay licencia.

2. **Dedicación y Licencia de Dominio Público de Open Data Commons – PDDL**

La licencia de Open Data Commons tiene las mismas características que la licencia de Dominio Público, pero la diferencia es que la licencia PDDL utiliza un mecanismo de licencia para otorgar los derechos al conjunto de datos.

3. **CREATIVE COMMONS ATRIBUCIÓN 4.0 INTERNACIONAL CC-BY**

Esta licencia permite a los usuarios compartir y modificar un conjunto de datos, pero solo si dan crédito al creador o creadores del conjunto de datos.

4. **ACUERDO DE LICENCIA DE DATOS DE LA COMUNIDAD – CDLA PERMISIVA-2.0**

Al igual que la mayoría de las licencias de código abierto, esta licencia permite a los usuarios usar, modificar, adaptar y compartir el conjunto de datos, pero solo si también se incluye un descargo de responsabilidad sobre garantías y responsabilidades.

5. **LICENCIA DE ATRIBUCIÓN DE OPEN DATA COMMONS - ODC-BY**

Esta licencia permite a los usuarios compartir y adaptar un conjunto de datos, pero solo si dan crédito al creador o creadores del conjunto de datos.

6. **CREATIVE COMMONS ATRIBUCIÓN-SHAREALIKE 4.0 INTERNACIONAL - CC-BY-SA**

Esta licencia permite a los usuarios usar, compartir y adaptar un conjunto de datos, pero solo si dan crédito al conjunto de datos y muestran cualquier cambio o transformación que hayan realizado en el conjunto de datos. Los usuarios pueden no querer usar esta licencia porque tienen que compartir el trabajo que hicieron en el conjunto de datos.

7. ACUERDO DE LICENCIA DE DATOS DE LA COMUNIDAD – CDLA-SHARING-1.0

Esta licencia utiliza el principio de ‘copyleft’ : los usuarios pueden usar, modificar y adaptar un conjunto de datos, pero solo si no añaden restricciones de licencia en el nuevo trabajo o trabajos que crean con el conjunto de datos.

8. LICENCIA DE BASE DE DATOS ABIERTA DE OPEN DATA COMMONS - ODC-ODBL

Esta licencia permite a los usuarios usar, compartir y adaptar un conjunto de datos, pero solo si dan crédito al conjunto de datos y muestran cualquier cambio o transformación que realicen en el conjunto de datos. Los usuarios pueden no querer usar esta licencia porque tienen que compartir el trabajo que hicieron en el conjunto de datos.

9. CREATIVE COMMONS ATRIBUCIÓN-NONCOMERCIAL 4.0 INTERNACIONAL - CC BY-NC

Esta licencia es una licencia restrictiva. Los usuarios pueden compartir y adaptar un conjunto de datos, siempre que den crédito a sus creadores y aseguren que el conjunto de datos no se utilice para ningún propósito comercial.

10. CREATIVE COMMONS ATRIBUCIÓN-NO DERIVADOS 4.0 INTERNACIONAL - CC BY-ND

Esta licencia también es una licencia restrictiva. Los usuarios pueden compartir un conjunto de datos si dan crédito a sus creadores. Esta licencia no permite adiciones, transformaciones o cambios en el conjunto de datos.

11. CREATIVE COMMONS ATRIBUCIÓN-NONCOMERCIAL-SHAREALIKE 4.0 INTERNACIONAL - CC BY-NC-SA

Esta licencia permite a los usuarios compartir un conjunto de datos solo si dan crédito a sus creadores. Los usuarios pueden compartir adiciones, transformaciones o cambios en el conjunto de datos, pero no pueden usar el conjunto de datos para fines comerciales.

12. CREATIVE COMMONS ATRIBUCIÓN-NONCOMERCIAL-NODERIVATIVES 4.0 INTERNACIONAL - CC BY-NC-ND

Esta licencia permite a los usuarios compartir un conjunto de datos solo si dan

crédito a sus creadores. Los usuarios no pueden modificar el conjunto de datos y no pueden usarlo para fines comerciales.