Open datasets and sources
In this data-driven world, some datasets are freely available for anyone to access, use, modify, and share. These are called open datasets.
Open datasets include a public license and are very useful for your journey as a Data Scientist. Some of the most informative open dataset sources are listed below.

Government Data:

https://www.data.gov/
https://www.census.gov/data.html
https://data.gov.uk/
https://www.opendatanetwork.com/
https://data.un.org/
Financial Data Sources:

https://data.worldbank.org/
https://www.globalfinancialdata.com/
https://comtrade.un.org/
https://www.nber.org/
https://fred.stlouisfed.org/
Crime Data:

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/
Health Data:

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
Academic and Business Data:

https://scholar.google.com/
https://nces.ed.gov/
https://www.glassdoor.com/research/
https://www.yelp.com/dataset
Other General Data:

https://www.kaggle.com/datasets
https://www.reddit.com/r/datasets/
Propriety datasets and sources
Proprietary datasets contain data primarily owned and controlled by specific individuals or organizations. This data is limited in distribution because it is sold with a licensing agreement.

Some data from private sources cannot be easily disclosed, like public data.

National security data, geological, geophysical, and biological data are examples of propriety data. Copyright laws or patents usually bind this type of data. Proprietary datasets that mainly contain sensitive information are less widely available than open datasets.

Some standard propriety dataset sources are listed below.

Health Care:

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

Financial Market data:

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

Google Cloud based datasets:

https://cloud.google.com/datasets

Dataset licenses
When you select a dataset, it is necessary to look into the license. A license explains whether you can use that dataset or not; or explains if you have to accept certain guidelines to use that dataset. The different license types are listed below.

PUBLIC DOMAIN MARK - PUBLIC DOMAIN
When a dataset has a Public Domain license, all the rights to use, access, modify and share the dataset are open to everyone. Here there is technically no license.

OPEN DATA COMMONS PUBLIC DOMAIN DEDICATION AND LICENSE – PDDL
Open Data Commons license has the same features as the Public Domain license, but the difference is the PDDL license uses a licensing mechanism to give the rights to the dataset.

CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL CC-BY
This license allows users to share and modify a dataset, but only if they give credit to the creator(s) of the dataset.

COMMUNITY DATA LICENSE AGREEMENT – CDLA PERMISSIVE-2.0
Like most open-source licenses, this license allows users to use, modify, adapt, and share the dataset, but only if a disclaimer of warranties and liability is also included.

OPEN DATA COMMONS ATTRIBUTION LICENSE - ODC-BY
This license allows users to share and adapt a dataset, but only if they give credit to the creator(s) of the dataset.

CREATIVE COMMONS ATTRIBUTION-SHAREALIKE 4.0 INTERNATIONAL - CC-BY-SA

This license allows users to use, share, and adapt a dataset, but only if they give credit to the dataset and show any changes or transformations, they made to the dataset. Users might not want to use this license because they have to share the work they did on the dataset.

COMMUNITY DATA LICENSE AGREEMENT – CDLA-SHARING-1.0
This license uses the principle of ‘copyleft’: users can use, modify, and adapt a dataset, but only if they don’t add license restrictions on the new work(s) they create with the dataset.

OPEN DATA COMMONS OPEN DATABASE LICENSE - ODC-ODBL
This license allows users to use, share, and adapt a dataset but only if they give credit to the dataset and show any changes or transformations they make to the dataset. Users might not want to use this license because they have to share the work they did on the dataset.

CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL 4.0 INTERNATIONAL - CC BY-NC
This license is a restrictive license. Users can share and adapt a dataset, provided they give credit to its creator(s) and ensure that the dataset is not used for any commercial purpose.

CREATIVE COMMONS ATTRIBUTION-NO DERIVATIVES 4.0 INTERNATIONAL - CC BY-ND
This license is also a restrictive license. Users can share a dataset if they give credit to its creator(s). This license does not allow additions, transformations, or changes to the dataset.

CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE 4.0 INTERNATIONAL - CC BY-NC-SA
This license allows users to share a dataset only if they give credit to its creator(s). Users can share additions, transformations, or changes to the dataset, but they cannot use the dataset for commercial purposes.

CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 INTERNATIONAL - CC BY-NC-ND
This license allows users to share a dataset only if they give credit to its creator(s). Users are not allowed to modify the dataset and are not allowed to use it for commercial purposes.

Note: Additional license types exist. Any dataset you use will include details about its license.

Author(s)