Obraz zawierający Czcionka, Grafika, logo, krąg

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Czytanie: Dodatkowe Źródła Zbiorów Danych**

**Szacowany czas: 5 minut**

W tym czytaniu dowiesz się o:

* Otwarte zbiory danych i źródła
* Własnościowe zbiory danych i źródła
* Licencja na zbiory danych

**Otwarte zbiory danych i źródła**

W tym oparciu o dane świecie, niektóre zbiory danych są ogólnodostępne dla każdego, kto chce je przeglądać, używać, modyfikować i udostępniać. Nazywane są **otwartymi zbiorami danych**.  
Otwarte zbiory danych zawierają publiczną licencję i są bardzo przydatne w Twojej drodze jako Data Scientist. Poniżej znajduje się lista niektórych z najbardziej informacyjnych źródeł otwartych zbiorów danych.

**Dane rządowe:**

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

**Źródła danych finansowych:**

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

**Dane o przestępczości:**

* <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/>

**Dane zdrowotne:**

* <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>

**Dane akademickie i biznesowe:**

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

**Inne ogólne dane:**

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

**Zbiory danych i źródła własnościowe**

Zbiory danych własnościowych zawierają dane, które są głównie własnością i kontrolą określonych osób lub organizacji. Te dane są ograniczone w dystrybucji, ponieważ są sprzedawane na podstawie umowy licencyjnej. Niektóre dane z prywatnych źródeł nie mogą być łatwo ujawnione, jak dane publiczne.

Dane dotyczące bezpieczeństwa narodowego, dane geologiczne, geofizyczne i biologiczne to przykłady danych własnościowych. Tego typu dane zazwyczaj są objęte prawami autorskimi lub patentami. Zbiory danych własnościowych, które głównie zawierają wrażliwe informacje, są mniej powszechnie dostępne niż zbiory danych otwartych.

Niektóre standardowe źródła zbiorów danych własnościowych są wymienione poniżej.

**Opieka zdrowotna:**

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

**Dane rynków finansowych:**

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

**Zbiory danych oparte na Google Cloud:**

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

**Licencje zbiorów danych**

Kiedy wybierasz zbiór danych, konieczne jest zapoznanie się z licencją. Licencja wyjaśnia, czy możesz używać danego zbioru danych, czy musisz zaakceptować określone zasady, aby go używać. Różne typy licencji są wymienione poniżej.

1. **PUBLIC DOMAIN MARK - PUBLIC DOMAIN**  
   Kiedy zbiór danych ma licencję Public Domain, wszystkie prawa do używania, dostępu, modyfikacji i udostępniania zbioru danych są otwarte dla wszystkich. Technicznie rzecz biorąc, nie ma tutaj licencji.
2. **OPEN DATA COMMONS PUBLIC DOMAIN DEDICATION AND LICENSE – PDDL**  
   Licencja Open Data Commons ma te same cechy, co licencja Public Domain, ale różnica polega na tym, że licencja PDDL wykorzystuje mechanizm licencjonowania, aby przyznać prawa do zbioru danych.
3. **CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL CC-BY**  
   Ta licencja pozwala użytkownikom na udostępnianie i modyfikowanie zbioru danych, ale tylko pod warunkiem, że podają uznanie dla twórcy(ów) zbioru danych.
4. **COMMUNITY DATA LICENSE AGREEMENT – CDLA PERMISSIVE-2.0**  
   Podobnie jak większość licencji open-source, ta licencja pozwala użytkownikom na używanie, modyfikowanie, dostosowywanie i udostępnianie zbioru danych, ale tylko pod warunkiem, że dołączony jest również zrzeczenie się gwarancji i odpowiedzialności.
5. **OPEN DATA COMMONS ATTRIBUTION LICENSE - ODC-BY**  
   Ta licencja pozwala użytkownikom na udostępnianie i dostosowywanie zbioru danych, ale tylko pod warunkiem, że podają uznanie dla twórcy(ów) zbioru danych.
6. **CREATIVE COMMONS ATTRIBUTION-SHAREALIKE 4.0 INTERNATIONAL - CC-BY-SA**  
   Ta licencja pozwala użytkownikom na używanie, udostępnianie i dostosowywanie zbioru danych, ale tylko pod warunkiem, że podają uznanie dla zbioru danych i pokazują wszelkie zmiany lub przekształcenia, które wprowadzili w zbiorze danych. Użytkownicy mogą nie chcieć korzystać z tej licencji, ponieważ muszą dzielić się pracą, którą wykonali nad zbiorem danych.
7. **COMMUNITY DATA LICENSE AGREEMENT – CDLA-SHARING-1.0**  
   Ta licencja wykorzystuje zasadę „copyleft”: użytkownicy mogą używać, modyfikować i dostosowywać zbiór danych, ale tylko pod warunkiem, że nie dodają ograniczeń licencyjnych do nowej pracy, którą tworzą na podstawie zbioru danych.
8. **OPEN DATA COMMONS OPEN DATABASE LICENSE - ODC-ODBL**  
   Ta licencja pozwala użytkownikom na używanie, udostępnianie i dostosowywanie zbioru danych, ale tylko pod warunkiem, że podają uznanie dla zbioru danych i pokazują wszelkie zmiany lub przekształcenia, które wprowadzają w zbiorze danych. Użytkownicy mogą nie chcieć korzystać z tej licencji, ponieważ muszą dzielić się pracą, którą wykonali nad zbiorem danych.
9. **CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL 4.0 INTERNATIONAL - CC BY-NC**  
   Ta licencja jest licencją restrykcyjną. Użytkownicy mogą udostępniać i dostosowywać zbiór danych, pod warunkiem, że podają uznanie dla jego twórcy(ów) i zapewniają, że zbiór danych nie jest używany do celów komercyjnych.
10. **CREATIVE COMMONS ATTRIBUTION-NO DERIVATIVES 4.0 INTERNATIONAL - CC BY-ND**  
    Ta licencja jest również licencją restrykcyjną. Użytkownicy mogą udostępniać zbiór danych, jeśli podają uznanie dla jego twórcy(ów). Ta licencja nie zezwala na dodatki, przekształcenia ani zmiany w zbiorze danych.
11. **CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE 4.0 INTERNATIONAL - CC BY-NC-SA**  
    Ta licencja pozwala użytkownikom na udostępnianie zbioru danych tylko pod warunkiem, że podają uznanie dla jego twórcy(ów). Użytkownicy mogą udostępniać dodatki, przekształcenia lub zmiany w zbiorze danych, ale nie mogą używać zbioru danych do celów komercyjnych.
12. **CREATIVE COMMONS ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 INTERNATIONAL - CC BY-NC-ND**  
    Ta licencja pozwala użytkownikom na udostępnianie zbioru danych tylko pod warunkiem, że podają uznanie dla jego twórcy(ów). Użytkownicy nie mają prawa modyfikować zbioru danych ani używać go do celów komercyjnych.

***Uwaga: Istnieją dodatkowe typy licencji. Każdy zbiór danych, którego używasz, będzie zawierał szczegóły dotyczące swojej licencji.***

**Autor(zy)**

[Lakshmi Holla](https://www.linkedin.com/in/lakshmi-holla-b39062149/)

**Inni Współtwórcy**

[Malika Singla](https://www.linkedin.com/in/malika-goyal-04798622/)