Experiment No.: 08

Aim: Explore different dataset finders e.g. Google Dataset Search, Kaggle, mendeley etc.

Course Outcome: Analyze different forms of data with respect to different phases of Machine Learning.

Theory:

A dataset finder is a specialized online platform or search engine designed to help users locate datasets from a variety of fields—such as education, healthcare, business, environment, and technology. These platforms index datasets stored across the web, allowing users to search, preview, and download data in multiple formats like CSV, JSON, XML, and Excel.

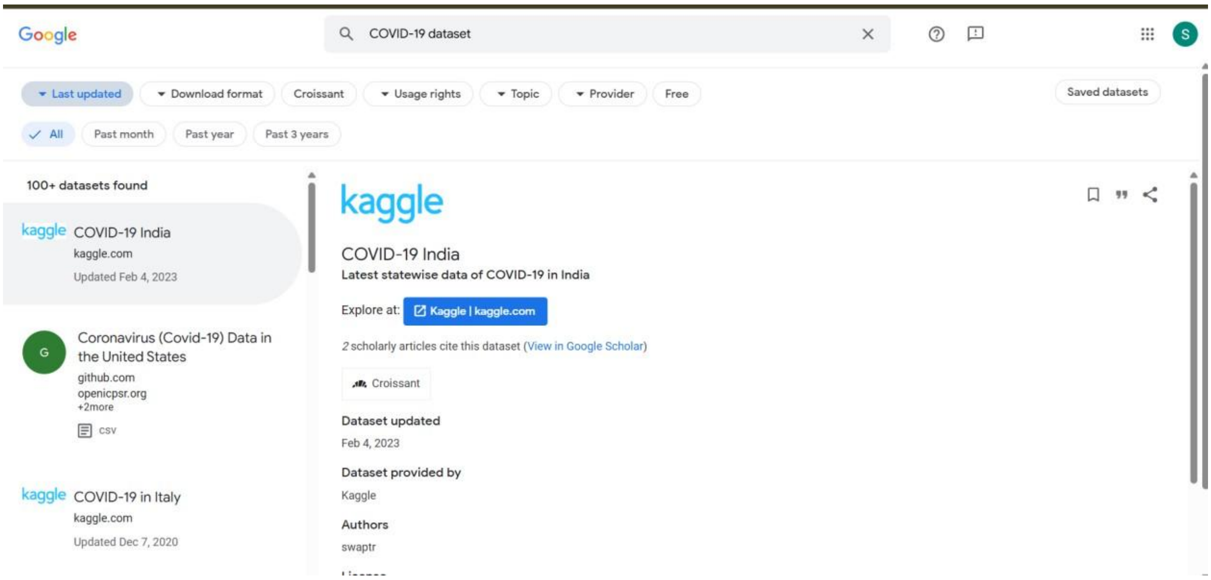
Dataset finders play a vital role in data science and research because they:

- Save time by aggregating datasets from various sources.
- Ensure datasets are well-documented and accessible.
- Promote open data sharing and collaboration.
- Encourage reproducibility in research and data-driven decision-making.

Below are some commonly used dataset finders:

1. Google Dataset Search

- A search engine developed by Google for datasets.
- Works similarly to Google Search but specifically indexes datasets.
- Displays dataset title, publisher, description, and file format.
- Offers filters for file type, date, and license.
- Widely used by researchers for academic and professional data discovery.
- URL: https://datasetsearch.research.google.com



100+ datasets found

COVID-19 India
kaggle.com
Updated Feb 4, 2023

Coronavirus (Covid-19) Data in the United States
github.com
openicpsr.org
+2more
csv

COVID-19 in Italy
kaggle.com
Updated Dec 7, 2020

COVID-19 India
Latest statewise data of COVID-19 in India

2 scholarly articles cite this dataset (View in Google Scholar)

Dataset updated
Feb 4, 2023

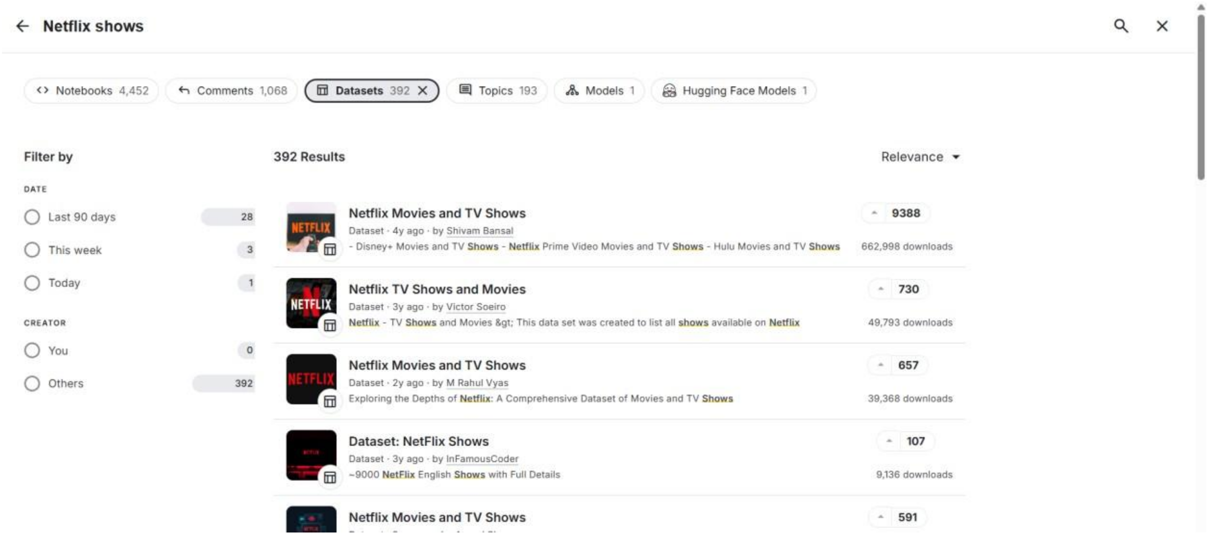
Dataset provided by
Kaggle

Authors
swaptr

2. Kaggle

- One of the largest online communities for data scientists.
- Offers thousands of public datasets across multiple domains.

- Datasets are available in clean, structured formats.
- Provides cloud-based notebooks for analysis and visualization.
- Encourages collaboration and learning through competitions.
- URL: https://www.kaggle.com/datasets



← Netflix shows

<> Notebooks 4,452 | ↶ Comments 1,068 | Datasets 392 ✕ | Topics 193 | Models 1 | Hugging Face Models 1

Filter by

DATE

○ Last 90 days 28

○ This week 3

○ Today 1

CREATOR

○ You 0

○ Others 392

392 Results — Relevance ▾

Netflix Movies and TV Shows — 9388
Dataset · 4y ago · by Shivam Bansal
- Disney+ Movies and TV Shows - Netflix Prime Video Movies and TV Shows - Hulu Movies and TV Shows — 662,998 downloads

Netflix TV Shows and Movies — 730
Dataset · 3y ago · by Victor Soeiro
Netflix - TV Shows and Movies &gt; This data set was created to list all shows available on Netflix — 49,793 downloads

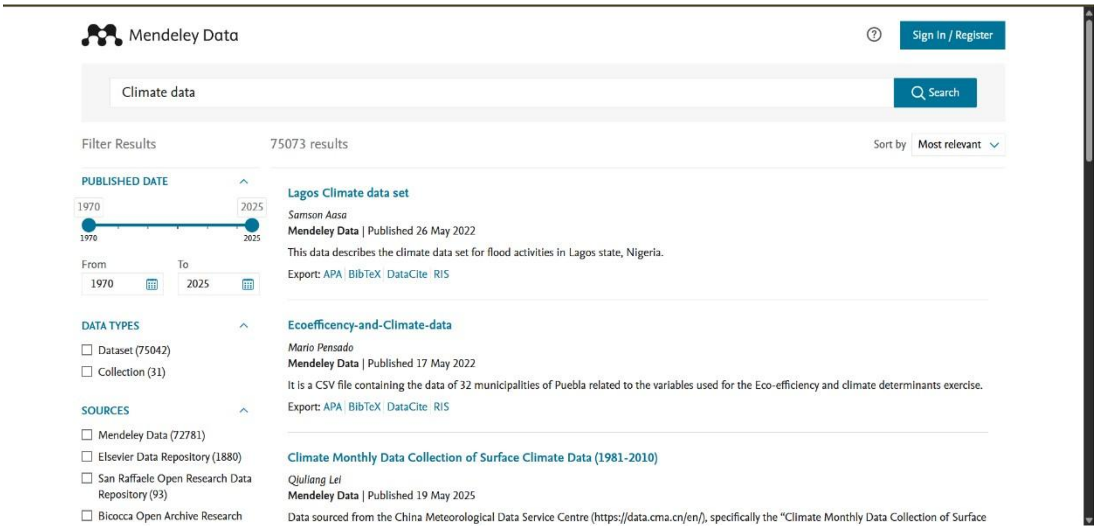
Netflix Movies and TV Shows — 657
Dataset · 2y ago · by M Rahul Vyas
Exploring the Depths of Netflix: A Comprehensive Dataset of Movies and TV Shows — 39,368 downloads

Dataset: NetFlix Shows — 107
Dataset · 3y ago · by InFamousCoder
~9000 NetFlix English Shows with Full Details — 9,136 downloads

Netflix Movies and TV Shows — 591

3. Mendeley Data

- A research data repository developed by Elsevier.
- Allows researchers to publish, store, and share datasets securely.
- Every dataset is assigned a **DOI (Digital Object Identifier)** for citation.
- Used mostly in academic and scientific research.
- Supports multiple file formats and ensures long-term data preservation.
- U RL: https://data.mendeley.com



Mendeley Data — Sign In / Register

Climate data — Search

Filter Results — 75073 results — Sort by Most relevant

PUBLISHED DATE

1970 – 2025

From 1970 To 2025

DATA TYPES

□ Dataset (75042)

□ Collection (31)

SOURCES

□ Mendeley Data (72781)

□ Elsevier Data Repository (1880)

□ San Raffaele Open Research Data Repository (93)

□ Bicocca Open Archive Research

Lagos Climate data set
Samson Aasa
Mendeley Data | Published 26 May 2022
This data describes the climate data set for flood activities in Lagos state, Nigeria.
Export: APA BibTeX DataCite RIS

Ecoefficency-and-Climate-data
Mario Pensado
Mendeley Data | Published 17 May 2022
It is a CSV file containing the data of 32 municipalities of Puebla related to the variables used for the Eco-efficiency and climate determinants exercise.
Export: APA BibTeX DataCite RIS

Climate Monthly Data Collection of Surface Climate Data (1981-2010)
Qiuliang Lei
Mendeley Data | Published 19 May 2025
Data sourced from the China Meteorological Data Service Centre (https://data.cma.cn/en/), specifically the "Climate Monthly Data Collection of Surface

Platform Summary

Platform Name	**Website Link**	**Data Formats**	**Area of Use**	**Key Feature**
Google Dataset Search	datasetsearch.research.google.com	CSV, XLSX, JSON	General research	Aggregates datasets from various websites
Kaggle	kaggle.com/datasets	CSV, JSON, ZIP	Data science, ML	Community-based data sharing
Mendeley Data	data.mendeley.com	ZIP, CSV, PDF	Academic & research	DOI and citation support
Platform Name	**Website Link**	**Data Formats**	**Area of Use**	**Key Feature**
UCI ML Repository	archive.ics.uci.edu	CSV, ARFF	Machine learning	Ready-to-use research datasets

Feature Comparison

Feature	**Google Dataset Search**	**Kaggle**	**Mendeley Data**
Login Required	No	Yes	Yes
Community Support	Moderate	High	Low
Dataset Size Range	Small to Large	Small to Very Large	Moderate
Ideal For	Quick search and discovery	Data science projects	Academic research
Dataset Citation (DOI)	No	Optional	Yes

Conclusion: I have successfully explored different dataset finders such as Google Dataset Search, Kaggle, and Mendeley. These platforms help users easily find, access, and use datasets for research, analysis, and learning purposes.