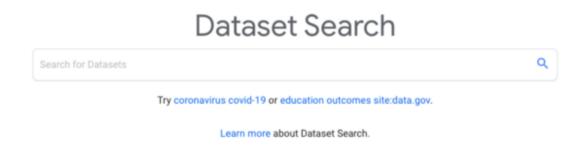
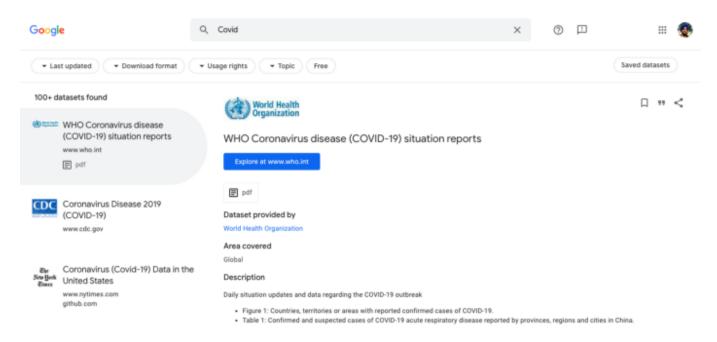
Google 25 million free datasets





<u>Datasetsearch</u>, a free tool for searching for 25 million publicly accessible datasets, was recently published by Google.



The search tool includes filters to limit results based on their license (free or paid), format (CSV, images, etc), and update time.

Note-Google does not curate or provide direct access to the 25 million datasets directly.

Google depends on publishers of datasets to use <u>schema.org</u>'s open standards to define the metadata of their datasets. Then Google indexes the metadata and makes it searchable through publishers.

Since publishers are still expected to host the datasets themselves, Google will also have their datasets indexed by for-profit publishers that adhere to schema.org standards. I found that half of the datasets in the search results in my anecdotal experience came from for-profit aggregators, with an even higher percentage when looking for market-related datasets.

To publish your own datasets, you can simply use the open-standards of schema.org. The number of publicly available datasets is likely to continue growing as more publishers conform to the standard.

At this time, Google does not provide an API for searching or downloading the free datasets.

You can find more information on the same at Google Dataset Blog.

Thanks for reading! I write on Medium about data science-related topics, so you can follow me to get notified when I write more articles like this one.

And don't forget to press the clap button if you liked the article:)

Follow me on Linkedin if you'd like to stay in touch!

Sign up for Analytics Vidhya News Bytes

By Analytics Vidhya

Latest news from Analytics Vidhya on our Hackathons and some of our best articles! Take a look

Get this newsletter

Emails will be sent to byteharman0101@gmail.com. Not you?

Data Science Machine Learning Python Artificial Intelligence Google

About Help Legal

Get the Medium app



