

# Public Data Sets

Created by Igor ARAMBASIC, last modified just a moment ago

Google data set search: <https://toolbox.google.com/datasetsearch>

## Links to various data sets:

- [Kaggle Datasets](#)
- [30 Amazing \(And Free\) Public Data Sources For 2018](#)
- [Awesome Public Datasets](#)
- [Amazon Public Data Sets](#)

## Unsorted links

1. [World Bank Open Data](#) Datasets covering population demographics and a huge number of economic and development indicators from across the world.
2. [IMF Data](#) The International Monetary Fund publishes data on international finances, debt rates, foreign exchange reserves, commodity prices and investments.
3. [The US National Center for Education Statistics](#) Data on educational institutions and education demographics from the US and around the world.
4. [The UK Data Centre](#) The UK's largest collection of social, economic and population data.
5. <https://data.gov.uk/> U.K. Government's open data
6. <https://www.data.gov/> U.S. Government's open data
7. [FiveThirtyEight](#) A large number of polls providing data on public opinion of political and sporting issues.
8. [FBI Uniform Crime Reporting](#) The FBI is responsible for compiling and publishing national crime statistics, with free data available at national, state and county level.
9. [Bureau of Justice](#) Here you can find data on law enforcement agencies, jails, parole and probation agencies and courts.
10. [Click Data Market](#) Offers a free package with access to datasets covering world population, currencies, development indicators and weather data.
11. [NASA Exoplanet Archive](#) Public datasets covering planets and stars gathered by NASA's space exploration missions.
12. [UN Comtrade Database](#) Statistics compiled and published by the United Nations on international trade. Includes Comtrade Lab which is a showcase of how cutting edge analytics and tools are used to extract value from the data.
13. [Financial Times Market Data](#) Up to date information on financial markets from around the world, including stock price indexes, commodities and foreign exchange.
14. [Google Trends](#) Examine and analyze data on internet search activity and trending news stories around the world.
15. [Twitter](#) The advantage Twitter has over the others are that most conversations are public. This means that huge amounts of data is available through their API on who is talking about what, where, when and why.
16. [Google Scholar](#) Entire texts of academic papers, journals, books and legal case law.
17. [Instagram](#) As with Twitter, Instagram posts and conversations are public by default. Their APIs allow likes, mentions and business details to be analyzed.
18. [OpenCorporates](#) The world's largest open database of companies.
19. [Glassdoor API](#) Information about job vacancies, candidates, salaries and employee satisfaction is available through their developer API.
20. [IMDB Datasets](#) Datasets in a number of formats drawn from the web's largest resource on movies, television and people working in those industries.
21. [OpenLibrary Data Dumps](#) Datasets on books including catalogs from libraries around the world
22. [Labelled Faces in the Wild](#) 13,000 collated and labeled images of human faces, for use in developing applications involving facial recognition.
23. [Microsoft Marco](#) Microsoft's open machine learning datasets for training systems in reading comprehension and question answering.
24. [Machine Learning Dataset Repository](#) Collection of open datasets contributed by data scientists involved in machine learning projects.
25. [eBay Market Data Insights](#) Data on millions of online sales and auctions from eBay
26. [Natural History Museum Data Portal](#) Information on nearly 4 million historical specimens in the London museum's collection, as well as scientific sound recordings of the natural world.
27. [CERN Open Data](#) More than one petabyte of data from particle physics experiments carried out by CERN.
28. [One Million Audio Cover Images](#) Dataset hosted at [archive.org](http://archive.org) covering music released around the world, for use in image processing research
29. [Complete Public Reddit Comments Corpus](#) Over one billion public comments posted to Reddit between 2007 and 2015, for training language algorithms
30. [Microsoft Azure Data Markets Free Datasets](#) Freely available datasets covering everything from agriculture to weather
31. [Irish Electric Vehicle Charge Point Status](#) Collates data from the body which oversees the network of EV charge points across the Republic of Ireland and Northern Ireland.
32. [LondonAir](#) Pollution and air quality data from across London
33. <http://datos.gob.es/> Open data initiative of the Government of Spain
34. <https://datos.madrid.es/portal/site/egob/> Open data of Ayuntamiento de Madrid
35. <http://opendata.emtmadrid.es/> information about the different services provided by Empresa Municipal de Transportes de Madrid (EMT) like bus, public bike, movility, parkings
36. <http://www.dtic.upf.edu/~ocelma/MusicRecommendationDataset/lastfm-1K.html> Music recommendation data set
37. <https://labrosa.ee.columbia.edu/millionsong/> Million Song Dataset
38. <https://recsys-challenge.spotify.com/> music dataset from Spotify
39. <https://www.kaggle.com/ahmedelnaggar/fifa-worldcup-2018-dataset> fifa-worldcup-2018-dataset

No labels

