

1- UCI Machine Learning Repository

(<https://archive.ics.uci.edu/ml/datasets.php>)

This is a useful website with about 394 datasets from various application areas. All datasets are detailed and their suitability for various analysis tasks are explained.

2- Stanford Large Network Dataset Collection (<https://snap.stanford.edu/data>)

This repository includes several datasets for interesting application areas: social networks, Bitcoin, Flickr, movies, etc.

3- Movielens (<https://grouplens.org/datasets/>)

This is an excellent repository for datasets related to movies and user ratings for them. It is usually used for implementing new recommender systems.

4- Amazon product data (<http://jmcauley.ucsd.edu/data/amazon/>)

This dataset contains product reviews and metadata from Amazon, including 142.8 million reviews spanning May 1996 - July 2014. The dataset includes reviews (ratings, text, helpfulness votes), product metadata (descriptions, category information, price, brand, and image features), and links (also viewed/also bought graphs).

5- Fuel Economy Data (<http://fueleconomy.gov/feg/download.shtml>)

Fuel economy data are the result of vehicle testing done at the Environmental Protection Agency's National Vehicle and Fuel Emissions Laboratory in Ann Arbor, Michigan, and by vehicle manufacturers with oversight by EPA.

6- Kaggle Datasets (<https://www.kaggle.com/datasets>)

Kaggle is a repository that allows people to share data with each other. There are many datasets already available there. You can also create a new dataset yourself. They award 10,000\$ in prizes to the best 3 datasets published each month.

7- Million Song Dataset

(<https://labrosa.ee.columbia.edu/millionsong/pages/gettingdataset>) The Million Song Dataset is a freely-available collection of audio features and metadata for a million contemporary popular music tracks.

8- GitHub public datasets (<https://github.com/caesar0301/awesome-public-datasets>)

GitHub has provided a collection of datasets from blogs, answers, and user responses. Most of the datasets are free, however, some are not.

- 9- Stat NZ (<http://www.stats.govt.nz/>): Stats NZ's purpose is to empower decisions by adding value to New Zealand's most important data.
- 10- <https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>
- 11- NZTA Open Data Portal (<https://opendata-nzta.opendata.arcgis.com/>)

The NZTA Open Data Portal is a platform where the NZ Transport Agency shares open data. You can explore, analyse, and download data in a range of formats

- 12- Education Counts (<https://www.educationcounts.govt.nz/home>) Datasets from the education sector in New Zealand.
- 13- You can use Google search engine to find your favourite dataset:
<https://toolbox.google.com/datasetsearch>

If you know any other repository or dataset, share it with other students using the following link:

<https://wintecnz.padlet.org/CfIT/e5avzgw941pd>