**Project Data Sets**

* [**UCI Machine Learning Repository**](http://archive.ics.uci.edu/ml/)- A repository of more than 200 data sets for machine learning and data mining
* [**mldata.org**](http://mldata.org/) - A large machine learning data repository
* [**Movie Ratings Data**](http://facweb.cs.depaul.edu/mobasher/classes/ect584/data/movielens.zip) - Real movie ratings data from [**www.movielens.org**](http://www.movielens.org/) Web site. Contains ratings on 1600+ movies by 1000 users
* [**Kaggle.com Competition Data Sets**](http://www.kaggle.com/competitions) - Data sets from a variety of competitions. Also a good source for class project ideas.
* [**Stanford Large Network Dataset Collection**](http://snap.stanford.edu/data/)- A variety of network data sets, including data from social networks, product reviews, online communities, etc.
* [**Yelp Data Set Challenge**](http://www.yelp.com/dataset_challenge/) - Reviews and check-in data on thousands of businesses.
* [**Online Grocery Shopping Data from Instacart.**](https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2)
* [**Million Song Dataset**](http://labrosa.ee.columbia.edu/millionsong/) - Freely-available collection of audio features and metadata for a million contemporary popular music tracks.
* [**All the News**](https://www.kaggle.com/snapcrack/all-the-news) - 143,000 articles from 15 American publications
* [**Public Data sets on Amazon Web Services**](http://aws.amazon.com/publicdatasets/) - Large public data sets (including data sets for US Census, Wikipedia, Freebase, human genome project), ready for big data analytics on the cloud.
* [**Data.gov**](http://catalog.data.gov/dataset)- Publically available data sets from Federal, State, and local government, including economic, geological, demographic and many other types of data sources. This site also includes a list of other [**Open Data Sites**](http://www.data.gov/opendatasites) with similar publicly available data sources from various cities, states, and countries.
* [**KDnugget's list of data sets for data mining**](http://www.kdnuggets.com/datasets/)
* [**Infochimps Data Market**](http://www.infochimps.com/datasets) - Thousands of data sets, including data from various social networks and collaborative tagging sites such as Twitter, Delicious, Last.fm, MusicBrainz, as well as data sets from many other domains.