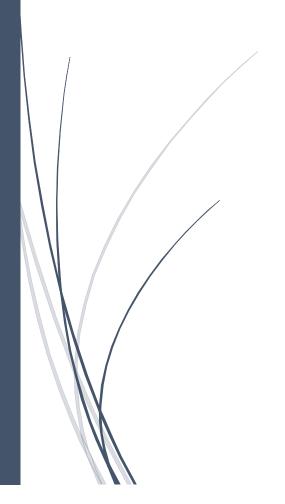1/14/2014

Online Data Sources for Analysis (Verified)

Venkat Reddy

- http://archive.ics.uci.edu/ml/datasets.html
    - More than 250 datasets on Machine Learning and classification. Contains both small and large datasets
    - Gives the details of the papers that cite these data set along with context
- http://www.datawrangling.com/some-datasets-available-on-the-web
    - 400 dataset bookmarks, dataset link along with the tag
    - Good webpage, when you know that you want
    - You may want to Copy these links and save them
- https://www.kaggle.com/
    - Data science challenges
    - Challenges & corresponding data sets
    - Large volume datasets are also available
- http://datamob.org/datasets
    - Links for Lot of datasets
    - Some links are not working though
- http://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html
    - All the sample datasets contained in R datasets package
    - Contains a dataset for almost all types of statistical analysis
- http://vincentarelbundock.github.io/Rdatasets/datasets.html
    - All the sample datasets contained in all R datasets package
    - Contains a dataset for almost all types of statistical analysis
- http://www.aeaweb.org/RFE/toc.php?show=complete
    - U.S. Macro and Regional Data
    - Lot of data sets, real and reliable, be careful with copy rights
- http://www.reddit.com/r/datasets
    - Datasets discussion forum
    - Share, Request, Like, download and discuss about datasets
- http://fimi.ua.ac.be/data/
    - A small list of Frequent Item set Mining Dataset Repository
    - Useful
- http://kdd.ics.uci.edu/
    - Datasets by data type, by application area along with the task files
- http://lib.stat.cmu.edu/datasets/
    - StatLib---Datasets Archive
    - Good collection
- http://robjhyndman.com/TSDL/
    - Very interesting time series datasets
- http://inforumweb.umd.edu/econdata/econdata.html
    - Economic datasets
- http://www.cs.ucr.edu/~eamonn/time_series_data/
    - Time Series and ML datasets
- http://www.rdatamining.com/
    - Many datasets and other resources details also
- http://kevinchai.net/datasets

- Its Kevin's page and he says "I have come across many datasets in my research and thought I'd share my list with everyone."
- Take a backup of the list
- **Other Sources**
  - http://readwrite.com/2008/04/09/where_to_find_open_data_on_the#awesm=~ooRuAlqwbOuVha
  - http://research.stlouisfed.org/fred2/
  - http://www.nyc.gov/html/dcp/html/bytes/applbyte.shtml#lion
  - http://www.sigkdd.org/

Want to update the list with more links? Please write a comment or send it to 21.venkat@gmail.com