Assessment criteria:

* nature of the analysed dataset
* Where/how you acquired it
* primary purpose of the data collection
* where other people can get copies from
* discussion of collection methodology
* controversy and/or quality of the data

Kaggle

This is a data set that follows the daily YouTube videos on the trending page for 2017-2018 for 10 countries (Brazil, Denmark, France, Great Britain, japan, India, Korea, Russia, Canada and Mexico).

Data available on the videos are: video title, channel title, publish time, tags, views, likes and dislikes, description, comment count and a category\_id field varying across regions. Data were originally retrieved from YouTube API and then uploaded on Kaggle, an online “data science and machine learning community” (Kaggle, 2020). Datasets are frequently updated onto this website for users to create different codes, analysis or any data science project and eventually share them with the community.

Available at: <https://www.kaggle.com/datasnaek/youtube-new>

Carnegie Mellon University

The Carnegie Mellon dataset follow viral videos from 2006 to 2011 it is the “largest public viral video dataset (...) where videos were manually selected by experts including editors from Youtube and the Time Magazine”.

The following informations are provided for each video : the video id, video title, view count, category, duration in seconds, upload time, captured time, average rate, maximum rate, minimum rate, number of raters, number of dislikes, number of likes, uploader id/name, uploader url and video description.

A number of analysis were done using this dataset that can be found at <http://www.cs.cmu.edu/~lujiang/camera_ready_papers/ICMR2014-Viral.pdf> showing different things to know while designing a viral video, which content certain countries like better, what appears more to middle-aged males and a representative analysis of viral videos that was last updated in 2014.

Available at: <https://sites.google.com/site/cmuviralvideos/download>