

Raw Datasets

* Note that datasets highlighted in yellow are newly added and were not presented in the Interim Report

Dataset	Original source	Number of features	Number of data points
Movies			
Rotten Tomatoes Movies	https://www.kaggle.com/datasets/andrezaza/clapper-massive-rotten-tomatoes-movies-and-reviews?select=rotten_tomatoes_movie_reviews.csv	16	142052
IMDB Movies 2000-2020	https://www.kaggle.com/datasets/chenyanglim/imdb-v2	24	5487
IMDb Movie	https://ieee-dataport.org/open-access/imdb-movie-reviews-dataset	10	1700
TMDb	https://www.kaggle.com/datasets/ritayandhara/tmdb-data-set	22	10000
The Ultimate 1Million Movies Dataset (TMDb + IMDb)	https://www.kaggle.com/datasets/alanyourch/tmdb-movies-daily-updates	28	1.01m
Reviews			
IMDb Movie Reviews	https://ieee-dataport.org/open-access/imdb-movie-reviews-dataset	7	931565
IMDb Review Dataset - ebD	https://www.kaggle.com/datasets/ebiswas/imdb-review-d	9	5571499

	dataset?select=part-02.json		
Rotten Tomatoes Reviews	https://www.kaggle.com/datasets/andrezaza/clapper-massive-rotten-tomatoes-movies-and-reviews?select=rotten_tomatoes_movie_reviews.csv	11	69263
Movielens 1M (movies.dat)	https://grouplens.org/datasets/movielens/1m/	3	3,883
Movielens 1M (ratings.dat)	https://grouplens.org/datasets/movielens/1m/	4	1,000,209
Movielens 1M (users.dat)	https://grouplens.org/datasets/movielens/1m/	5	6,040
Awards			
The Oscar Award, 1927-2024	https://www.kaggle.com/datasets/unanimad/the-oscar-award	7	10889
Consumer Price Index			
US BLS Monthly CPI, 1915 - 2024	https://data.bls.gov/dataViewer/view/timeseries/CUUR0000SA0	5	1317