**Dataset Details**

**Link:** [**http://grouplens.org/datasets/movielens/**](http://grouplens.org/datasets/movielens/)

**The raw data set was of the form:**

## Character Encoding

The three data files are encoded as [UTF-8](http://en.wikipedia.org/wiki/Utf-8).

## User Ids

* Movielens users were selected at random for inclusion. Their ids have been anonymized.
* Users were selected separately for inclusion in the ratings and tags data sets, which implies that user ids may appear in one set but not the other.
* The anonymized values are consistent between the ratings and tags data files. That is, user id n, if it appears in both files, refers to the same real MovieLens user.

## Ratings Data File Structure

* All ratings are contained in the file ratings.dat. Each line of this file represents one rating of one movie by one user, and has the following format:
* UserID::MovieID::Rating::Timestamp
* The lines within this file are ordered first by UserID, then, within user, by MovieID.
* Ratings are made on a 5-star scale, with half-star increments.
* Timestamps represent seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970.

## Tags Data File Structure

* All tags are contained in the file tags.dat. Each line of this file represents one tag applied to one movie by one user, and has the following format:
* UserID::MovieID::Tag::Timestamp
* The lines within this file are ordered first by UserID, then, within user, by MovieID.
* Tags are user generated metadata about movies. Each tag is typically a single word, or short phrase. The meaning, value and purpose of a particular tag is determined by each user.
* Timestamps represent seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970.

## Movies Data File Structure

* Movie information is contained in the file movies.dat. Each line of this file represents one movie, and has the following format:
* MovieID::Title::Genres
* MovieID is the real MovieLens id.
* Movie titles, by policy, should be entered identically to those found in [IMDB](http://www.imdb.com/), including year of release. However, they are entered manually, so errors and inconsistencies may exist.

Genres are a pipe-separated list, and are selected from the following:

* Action
* Adventure
* Animation
* Children's
* Comedy
* Crime
* Documentary
* Drama
* Fantasy
* Film-Noir
* Horror
* Musical
* Mystery
* Romance
* Sci-Fi
* Thriller
* War
* Western