

# Project Codebook: Dataset of Top 1000 Movies on IMDb

Ephets Head

**This is the codebook for my final project, a data analysis and machine-learning project using the kaggle dataset, “IMDB TOP 1000 MOVIES”. The data can be downloaded for free from the following url:**

[<https://www.kaggle.com/datasets/mysarahmadbhat/imdb-top-1000-movies>]

**The set is fairly easy to understand. When first downloaded, it is made up of 1000 observations and 16 variables. These variables are:**

**Poster\_Link** : A url link of the form *https://...jpg* which takes you to the image of the movie’s poster.

**Series\_Title** : The full movie title.

**Released\_Year** : A four-digit integer, representing the year in which the movie was released.

**Certificate** : The certification of the movie, with 14 possible factored levels:

1. `A` : Restricted to adults (due to abusive language, intimacy, etc.)
2. `UA` : Parental Guidance recommended for children below the age of 12.
3. `PG-13` : Parents strongly cautioned, inappropriate for children under the age of 13.
4. `PG` : Parental Guidance Suggested: some material unsuited for children.
5. `G` : General Audiences: suitable for all ages.
6. `R` : Restricted, no children under the age of 17 without parent or guardian.
7. `U` : Unrestricted public exhibition, anybody can watch.
8. `Passed` : Outdated certificate, meaning the movie passed the Production Code Administration's criteria.
9. `TV-14` : Some material unsuitable for children under the age of 14.
10. `TV-MA` : TV-Mature Audience only, unsuitable for children under 17.
11. `Unrated` : Film is not rated/ has no certificate.
12. `GP` : All ages admitted, parental guidance suggested.
13. `Approved` : Approved for exhibition; a classification that predates the more current certificates.
14. `TV-PG` : Parental guidance suggested.

**Runtime** : The runtime of a movie in minutes, stored initially as a character.

**Genre** : A list of strings representing the genres of the movie, with 21 categories:

1. `Drama`
2. `Crime`
3. `Action`
4. `Adventure`
5. `Biography`
6. `History`
7. `Sci-Fi`
8. `Romance`
9. `Western`
10. `Fantasy`
11. `Thriller`
12. `Family`
13. `War`
14. `Mystery`
15. `Music`
16. `Horror`
17. `Animation`
18. `Film-Noir`
19. `Sport`
20. `Comedy`
21. `Musical`

**IMDB\_Rating:** A 2-digit float, representing the average of the ratings the movie has received on the IMDb website, a score between 1 and 10.

**Overview:** A one-sentence string summarizing the plot of the movie.

**Meta\_Score:** A numeric rating of the film out of 100, as calculated from the average of a large group of respected critics' reviews.

**Director:** The name of the director of the movie.

**Star1:** The main leading actor in the movie.

**Star2:** The second leading actor in the movie.

**Star3:** The third leading actor in the movie.

**Star4:** The fourth leading actor in the movie.

**No\_of\_Votes:** The number of votes (ratings) the movie received on the IMDb website.

**Gross:** The gross revenue of the movie.