Project Data Description

Dataset for Rotten Tomatoes movies 1970 - 2024 (https://www.reddit.com/r/datasets/comments/1ecj6m2/dataset_for_rotten_tomatoes_movies_1970_2024/)

This dataset was compiled by a Reddit user on the r/datasets community (2024). It was collected via web scraping of Rotten Tomatoes movie pages (1970–2024).

- **title**: nominal
- **url**: nominal
- **release_date**: datetime
    - Range: 1970-2024
    - Missing value count: 13
- **critic_score**: quantitative
    - Range: 0-1
    - Missing value count: 3036
    - missing data patterns: blank
- **audience_score**: quantitative
    - Range: 0-1
    - Missing value count: 1587
    - missing data patterns: blank

IMDb-based dataset (https://www.kaggle.com/datasets/rajugc/imdb-movies-dataset-based-on-genre).

The IMDb-based dataset was created by Kaggle user Chidambara (2023). It was compiled using publicly available IMDb data, filtered and organized by genre, and cleaned for use in data analysis and visualization projects.

- **movie_id** - IMDB Movie ID
    - Missing value count: 0
- **movie_name** - Name of the movie
    - Missing value count: 4
- **year** - Release year: ordinal
    - Range: 1880-2025
    - Missing value count: 53248
- **certificate** - Certificate of the movie: categorical
    - Missing value count: 264109
- **run_time** - Total movie run time: quantitative
    - Range: 1-5538
    - Missing value count: 109154
- **genre** - Genre of the movie: categorical
    - Missing value count: 0
- **rating** - Rating of the movie: quantitative
    - Range: 1-10
    - Missing value count: 137362
- **description** - Description of the movie: nominal
    - Missing value count: 0
- **director** - Director of the movie: categorical
    - Missing value count: 27369
- **director_id** - IMDB id of the director: categorical
    - Missing value count: 27369
- **star** - Star of the movie: categorical
    - Missing value count: 58695
- **star_id** - IMDB id of the star: categorical
    - Missing value count: 51858
- **votes** - Number of votes in IMDB website: quantitative
    - Range: 5-2675531
    - Missing value count: 137358
- **gross(in $)** - Gross Box Office of the movie: quantitative
    - Range: 1-936662225
    - Missing value count: 343261

The Oscar Award Dataset (https://www.kaggle.com/datasets/unanimad/the-oscar-award/discussion/data?sort=hotness)

The *Oscar Award Dataset* was created by Kaggle user David Lu. It was collected by scraping publicly available data from the official Oscars website. The dataset provides all nominees and winners from 1927 to 2025, including categories, film titles, and individual nominees.

- **Ceremony** - which ceremony the nomination was for (starting at 1): ordinal
  - Range: 1-97
- **Year** - Year(s) from which the films are honored: ordinal
  - Range: 1927-2024
- **Class** - General award category: categorical
- **CanonicalCategory** - Name of the category, consistent across years : categorical
- **Category** - Name of the category (as per official nomination on Oscars.org): categorical
- **NomId** - nomination ID: nominal
  - Missing value count: 532
- **Film** - The title of the film: nominal
  - Missing value count: 1261
- **FilmId** - Unique string representing the IMDb Title ID: nominal
  - Missing value count: 1261
- **Name** - The precise text used for who is being nominated: categorical
  - Missing value count: 1185
- **Nominees** - The names of who is nominated in a comma separated list (without any extra text like "Written by"): categorical
  - Missing value count: 353
- **NomineeIds** - Unique strings (or question marks) representing the IMDb Name ID, separated by commas: categorical
  - Missing value count: 880
- **Winner** -True if the award was won: categorical
  - Missing value count: 8538
- **Detail** - Detail about the nomination, which could be the character name, song title, etc: nominal
  - Missing value count: 8843
- **Note** - Additional information provided about the award/nomination.
  - Missing value count: 11398
- **Citation** - Official text of the award statement, for Scientific/Technical/Honorary awards.

    - Missing value count: 10831
- **MultifilmNomination** - Generally the data is one nomination per row, but for certain early nominations (Ceremonies 1, 2, 3 & 8), people were nominated for multiple films, and so one nomination could be spread over multiple rows: categorical
    - Missing value count: 11974