**Movies-Information-Database**

Movies-Information-Database is a wide database of information consisting of IMDB movies, their rating, Oscar awards of the movies, personal information about the actors | directors | writers of the movies, reviews given by the reviewers, tags given to the movies by users and the genome scores of those tags along with a table to join IMDB ID’s with TMBD ID’s so that user will be able to use this database with other databases by connecting them.

**Datasets used:**

* <https://www.kaggle.com/datasets/mirajshah07/netflix-dataset>
* <https://www.kaggle.com/datasets/dharmikdonga/academy-awards-dataset-oscars>
* <https://www.kaggle.com/datasets/aigamer/movie-lens-dataset?select=tags.csv>
* <https://www.kaggle.com/datasets/komalkhetlani/imdb-dataset>
* <https://www.kaggle.com/datasets/nikosfragkis/imdb-320000-movie-reviews-sentiment-analysis>

**Description about tables:**

**IMDB\_Movies**: This is a table consists of data which includes almost all the movies listed in IMDB with columns

* ‘imdb\_title\_id’ – auto incrementing unique identifier used to identify the movies.
* ‘title’ – title of the movie
* ‘year’ – year of the movie release
* ‘genre’ – comma separated values of genres of the movie
* ‘duration’ – duration of the movie
* ‘country’ – comma separated values of movie origin country
* ‘language’ – comma separated values of languages the movie made in
* ‘Description’ – description about movie
* ‘production\_company’ – Name of the production company.
* ‘budget’ – Budget allocated for each movie.
* ‘director’ – Director name of the movie, can join with ‘person\_details’ and get details about director.
* ‘writer’ – Writer name of the movie, can join with ‘person\_details’ and get details about writer.
* ‘actor’ – Main actor name of the movie, can join with ‘person\_details’ and get details about actor.
* ‘rating’ – Average of the totalvotes.
* ‘total\_votes’ – total count of votes.

**Netflix**: This table consists of all the movies and TV Shows listed in Netflix

* ‘show\_id’ – Auto incrementing unique identifier used to identify each movie/Tv show listed in Netflix.
* ‘type’ – Type of the content (Movie/TV Show).
* ‘date\_added’ – Date the movie/tv show added to Netflix.
* ‘sensor\_rating’ – sensor rating of the movie/Tv show.

**Listed\_in:** Lists the movies of ‘Netflix’ which are listed in ‘imdb\_movies’ table. As there is no link between imdb\_movies and Netflix, we can only link them using this table which maps each ‘show\_id’ to its respective ‘imdb\_title\_id’ if there exists a corresponding movie of Netflix in Imdb\_movies data.

* ‘listed\_in\_id’ – unique identifier to identify each listing.
* ‘imdb\_title\_id’ – referenced from ‘imdb\_movies’ table.
* ‘show\_id’ – referenced from ‘Netflix’ table

**Person\_details:** Some information about few individuals working in film field. Includes their ID which is unique ‘Imdb’ person identifier which can be further used to link this dataset possibly with other datasets.

* ‘person\_id’ – Unique person identifier given by IMDB.
* ‘primary\_name' – name of the individual.
* ‘birth\_year’ – birth year of the individual.
* ‘death\_year’ – death year of the individual.
* ‘primary\_profession’ – comma separated values of the profession of the individual.

**Oscar\_awards:** This table contains information about Oscar awards given to the films for different categories. We can link title and year from ‘imdb\_movies’ table with this table to get information of the awards of the film individually and we can also join them using the names of ‘director or writer or actor’ to see if they got any award from Oscars.

* ‘oscar\_id’ – unique identifier used to identify each entry of the table.
* ‘ceremony\_number’ – the Oscar ceremony number
* ‘category’ – the nomination category of the movie.
* ‘gender’ – gender of the individual nominated.
* ‘name’ – name of the individual nominated
* ‘winner’ – T (if nomination won) or F (if nomination lost)
* ‘imdb\_title\_id’ – referenced from ‘imdb\_movies’ table.

**Reviews:** This table consists of the information about reviews given to the movie by the reviewers.

* ‘imdb\_title\_id’ – referenced from ‘imdb\_movies’ table.
* ‘review\_id’ – unique identifier to identify each review given by the reviewr.
* ‘reviewer\_name' – name of the reviewer.
* ‘review\_date’ – date of the review.
* ‘review’ – review content.

**Imdb\_tmdb\_connector:** the table which links the imdb\_movies with ratings and movies from other sites like ‘tmbd’.

* ‘movie\_id’ – unique identifier for a movie from other site.
* ‘imdb\_id’ – imdb unique movie identifier.
* ‘tmbd\_id’ – tmbd unique movie identifier.

**Movie\_tag\_scores:** This table contains the information about tag scores of each unique tag given to a movie.

* ‘movie\_tag\_id’ - unique identifier used to identify each entry of the table.
* ‘movie\_id’ – movie identifier referenced from ‘link’ table.
* ‘tag’ – tag given to the movie by the user.
* ‘tag\_score’ – score given to the tag assigned to the movie based on its relevance with the movie.

**Tag:** This table contains information about tag and their id’s.

* ‘tag\_id’ – unique identifier of the tag.
* ‘tag’ – name of the tag.

**User\_tag:** This table contains information about the tags assigned to the movie by the users.

* ‘user\_tag\_id’ - unique identifier used to identify each entry of the table.
* ‘user\_id’ – unique identifier used to identify the user.
* ‘movie\_id’ - movie identifier referenced from ‘link’ table.
* ‘tag’ – name of the tag given to the movie by the user.