**DATASET**

## TMDb\_updated.CSV

<https://www.kaggle.com/datasets/sankha1998/tmdb-top-10000-popular-movies-dataset?resource=download>

TMDB.org is a crowd-sourced movie information database used by many film-related consoles, sites and apps, such as XBMC, MythTV and Plex. Dozens of media managers, mobile apps and social sites make use of its API.  
TMDb lists some 80,000 films at time of writing, which is considerably fewer than IMDb. While not as complete as IMDb, it holds extensive information for most popular/Hollywood films.  
This is dataset of the 10,000 most popular movies across the world has been fetched through the read API.  
TMDB's free API provides for developers and their team to programmatically fetch and use TMDb's data.  
Their API is to use as long as you attribute TMDb as the source of the data and/or images. Also, they update their API from time to time.

This data set is fetched using exception handling process so the data set contains some null values as there are missing fields in the tmdb database. Thought it's good for a young analyst to deal with messing value.  
Hey analyst are you all excited?

Useful for:

1. Text
2. Language

(θα ήταν χρήσιμο να γίνει ένας αρχικός διαχωρισμός σχετικά με την γλώσσα για να ενημερώνονται οι χρήστες)

Title= Name

Description

Votes

**NetflixOriginals.csv**

[**https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores**](https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores)

### **Context**

This dataset consists of all Netflix original films released as of June 1st, 2021. Additionally, it also includes all Netflix documentaries and specials. The data was webscraped off of [this](https://en.wikipedia.org/wiki/Lists_of_Netflix_original_films) Wikipedia page, which was then integrated with a dataset consisting of all of their corresponding IMDB scores. IMDB scores are voted on by community members, and the majority of the films have 1,000+ reviews.

### **Content**

Included in the dataset is:

* Title of the film
* Genre of the film
* Original premiere date
* Runtime in minutes
* IMDB scores (as of 06/01/21)
* Languages currently available (as of 06/01/21)

Useful for:

1. Genre of movie
2. Runtime
3. Language

(θα μπορούσε να γίνει διαχωρισμός με βάση πάλι αρχικά την γλώσσα και στην συνέχεια με την διάρκεια της ταινίας κάνοντας ένα έλεγχο σε κάθε είδος της ταινίας ποιος είναι ο μέσος όρος διάρκειας για να βοηθήσει στην κατηγοριοποίηση)

ΑΛΛΑΓΕΣ

Title = Name

IMDB score= Rating

## IMDB-Movie-Data.csv

<https://www.kaggle.com/datasets/PromptCloudHQ/imdb-data?resource=download>

## About Dataset

Here's a data set of 1,000 most popular movies on IMDB in the last 10 years. The data points included are:

Title, Genre, Description, Director, Actors, Year, Runtime, Rating, Votes, Revenue, Metascrore

Feel free to tinker with it and derive interesting insights.

Useful for:

* 1. Multiple genres of the movies, combination
  2. Runtime
  3. Rating

(θα μπορούσε να βοηθήσει στην κατηγοριοποίηση των ειδών των ταινιών και να γίνει καλύτερος ο συνδυασμός και η ονομασία της νέας ομάδας που θα συμπεριληφθούν.

Επίσης θα μπορούσε στο τελικό αποτέλεσμα η κατηγοριοποίηση να γίνεται με φθίνουσα φορά των βαθμολογιών κάθε ταινίας)

Αλλαγή

Runtime(minutes)=Runtime

Title= Name

## imdb.csv

<https://www.kaggle.com/datasets/mazenramadan/imdb-most-popular-films-and-series>

## Context

IMDB is a popular website for rating films and series I always go there if I want to watch something new and many many users trust it's rankings  
The data is about most popular 7k Films and series on IMDB with rates, The Data is Ideal for Exploratory Data Analysis and You can also use Regression to predict Rate

## Content

These are our columns:  
**Name**: Name of the film/series  
**Data**: Creation date  
**Rate**: IMDB's Rate  
**Votes**: Number of voters  
**Genre**: Genres , Actions , Drama, Romance, etc…  
**Duration**: Duration of the episode , film  
**Type**: whether it's film or series  
**Certificate**:  
TV-Y: Designed to be appropriate for all children  
TV-Y7: Suitable for ages 7 and up  
G: Suitable for General Audiences  
TV-G: Suitable for General Audiences  
PG: Parental Guidance suggested  
TV-PG: Parental Guidance suggested  
PG-13: Parents strongly cautioned. May be Inappropriate for ages 12 and under.  
TV-14: Parents strongly cautioned. May not be suitable for ages 14 and under.  
R: Restricted. May be inappropriate for ages 17 and under.  
TV-MA: For Mature Audiences. May not be suitable for ages 17 and under.  
NC-17: Inappropriate for ages 17 and under  
**Episodes**: Number of Episodes only for series  
**Nudity, violence..** :How much does it have of these

Useful for:

1. Combination of the movies’ genres
2. Duration
3. Certificate
4. Content of the movie

(θα είναι χρήσιμο γιατί βοηθάει πάλι γιατί έχει συνδυασμό των ειδών της ταινίας, έχει την διάρκεια, αλλά έχει και τα περιεχόμενα των σκηνών σε τι ποσοστό είναι αλλά και την καταλληλόλητα των ταινιών ως προς τους ανήλικους οπότε θα βοηθήσει να διαχωριστεί πχ μια οικογενειακή ταινία με βάση όλα τα παραπάνω)

ΑΛΛΑΓΕΣ

Duration=Runtime

Rate = Rating

## imdb\_top\_1000.csv

<https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows>

### **Content**

Data:-

* **Poster\_Link** - Link of the poster that imdb using
* **Series\_Title** = Name of the movie
* **Released\_Year** - Year at which that movie released
* **Certificate** - Certificate earned by that movie
* **Runtime** - Total runtime of the movie
* **Genre** - Genre of the movie
* **IMDB\_Rating** - Rating of the movie at IMDB site
* **Overview** - mini story/ summary
* **Meta\_score** - Score earned by the movie
* **Director** - Name of the Director
* **Star1,Star2,Star3,Star4** - Name of the Stars
* **No**of**votes** - Total number of votes
* **Gross** - Money earned by that movie

Useful for:

1. Runtime
2. Text
3. Genres

## imdb (1000 movies) in june 2022.csv

<https://www.kaggle.com/datasets/ramjasmaurya/top-250s-in-imdb>

Useful for:

1. Text
2. Duration
3. Combination of genres

ΑΛΛΑΓΕΣ

Title = Name

Movie = Description

Κεφαλαία στα ονόματα

## IMDB\_movie\_reviews\_details.csv

<https://www.kaggle.com/datasets/preetviradiya/imdb-movies-ratings-details>

Useful for:

1. Text
2. Duration
3. Runtime

Αλλαγές

Όλα με κεφαλαία

Runtime=Description1