**IMDb Movies Dataset**

imdb\_title\_id - alphanumeric unique identifier of the title in IMDb

title – title name

original\_title – original title name

year – year of release

date\_published - date of release

genre – movie genre

duration – duration in minutes

country - movie country

language – movie language

director - director name

writer – writer name

production\_company – production company

actors – actor names

description – plot description

avg\_vote – average vote

votes – number of votes received

budget – movie budget

usa\_gross\_income – Gross income earned in USA

worlwide\_gross\_income - Gross income earned in worldwide

metascore – a weighted average in that Metacritic assigns more importance, or weight, to some critics and publications than others, based on movie quality and overall stature.

reviews\_from\_users - number of reviews from IMDb users

reviews\_from\_critics - number of reviews from movie critics

**IMDb Names Dataset**

imdb\_name\_id – name ID on IMDb (alphanumeric unique identifier of the name/person)

name – cast member name

birth\_name – birth name

height – height

bio – biographical information

birth\_details – birth details

birth\_year – birth year in YYYY format

date\_of\_birth – date of birth

place\_of\_birth – place of birth

death\_details – death details

date\_of\_death – date of death

place\_of\_death – place of death

reason\_of\_death – cause of death

spouses\_string – spouse name, birth date to death date, divorced and number of children with the cast

spouses – number of spouses

divorces – number of divorces

spouses\_with\_children - number of spouses with which he/she had children

children – number of children

**IMDb ratings dataset**

imdb\_title\_id - alphanumeric unique identifier of the title in IMDb

weighted\_average\_vote - total weighted average rating of all the individual user ratings

total\_votes - total votes received

mean\_vote - total mean vote

median\_vote - total median vote

votes\_10 - number of votes with rating equal to 10

votes\_9 - number of votes with rating equal to 9

votes\_8 - number of votes with rating equal to 8

votes\_7 - number of votes with rating equal to 7

votes\_6 - number of votes with rating equal to 6

votes\_5 - number of votes with rating equal to 5

votes\_4 - number of votes with rating equal to 4

votes\_3 - number of votes with rating equal to 3

votes\_2 - number of votes with rating equal to 2

votes\_1 - number of votes with rating equal to 1

allgenders\_0age\_avg\_vote - average rating from users with age lower than 18

allgenders\_0age\_votes - number of votes from users with age lower than 18

allgenders\_18age\_avg\_vote - average rating from users with age greater or equal to 18 and lower than 30

allgenders\_18age\_votes - number of votes from users with age greater or equal to 18 and lower than 30

allgenders\_30age\_avg\_vote - average rating from users with age greater or equal to 30 and lower than 45

allgenders\_30age\_votes - number of votes from users with age greater or equal to 30 and lower than 45

allgenders\_45age\_avg\_vote - average rating from users with age greater or equal to 45

allgenders\_45age\_votes - number of votes from users with age greater or equal to 45

males\_allages\_avg\_vote - average rating from all male users with demographic data available (all ages)

males\_allages\_votes - number of votes from all male users with demographic data available (all ages)

males\_0age\_avg\_vote - average rating from male users with age lower than 18

males\_0age\_votes - number of votes from male users with age lower than 18

males\_18age\_avg\_vote - average rating from male users with age greater or equal to 18 and lower than 30

males\_18age\_votes - number of votes from male users with age greater or equal to 18 and lower than 30

males\_30age\_avg\_vote - average rating from male users with age greater or equal to 30 and lower than 45

males\_30age\_votes - number of votes from male users with age greater or equal to 30 and lower than 45

males\_45age\_avg\_vote - average rating from male users with age greater or equal to 45

males\_45age\_votes - number of votes from male users with age greater or equal to 45

females\_allages\_avg\_vote - average rating from all female users with demographic data available (all ages)

females\_allages\_votes - number of votes from all female users with demographic data available (all ages)

females\_0age\_avg\_vote - average rating from female users with age lower than 18

females\_0age\_votes - number of votes from female users with age lower than 18

females\_18age\_avg\_vote - average rating from female users with age greater or equal to 18 and lower than 30

females\_18age\_votes - number of votes from female users with age greater or equal to 18 and lower than 30

females\_30age\_avg\_vote - average rating from female users with age greater or equal to 30 and lower than 45

females\_30age\_votes - number of votes from female users with age greater or equal to 30 and lower than 45

females\_45age\_avg\_vote - average rating from male users with age greater or equal to 45

females\_45age\_votes - number of votes from male users with age greater or equal to 45

top1000\_voters\_rating - average rating from top 1000 voters

top1000\_voters\_votes - number of votes from top 1000 voters

us\_voters\_rating - average rating from US voters

us\_voters\_votes - number of votes from US voters

non\_us\_voters\_rating - average rating from non-US voters

non\_us\_voters\_votes - number of votes from non-US voters

**IMDb title\_principals**

imdb\_title\_id – - alphanumeric unique identifier of the title in IMDb

ordering - order of importance in the movie (a number to uniquely identify rows for a given title Id)

imdb\_name\_id - name ID on IMDb (alphanumeric unique identifier of the name/person)

category - category of job done by the cast member

job - specific job done by the cast member, else ‘\N’

characters - name of the character played

**Rotten\_tomatoes\_critic\_reviews Dataset**

rotten\_tomatoes\_link - link from which the movies data have been scraped - e.g. the record "m/0814255" has been scraped from "https://www.rottentomatoes.com/m/0814255"

critic\_name - name of critic who rated the movie

top\_critic - boolean value that clarifies whether the critic is a top critic or not

publisher\_name - name of the publisher for which the critic works

review\_type - was the review fresh or rotten?

review\_score - review score provided by the critic

review\_date - date of the review

review\_content - content of the review

**Rotten\_tomatoes\_movies dataset**

rotten\_tomatoes\_link - link from which the movies data have been scraped - e.g. the record "m/0814255" has been scraped from "https://www.rottentomatoes.com/m/0814255"

movie\_title - title of the movie as displayed on the Rotten Tomatoes website

movie\_info - brief description of the movie

critics\_consensus - comment from Rotten Tomatoes

content\_rating - category based on the movie suitability for audience

genres - movie genres separated by commas, if multiple

directors - name of director(s)

authors - name of author(s)

actors - name of actors

original\_release\_date - date in which the movie has been released

streaming\_release\_date - date in which the movie has been released for streaming

runtime - movie runtume (in minutes)

production\_company - name of the production company

tomatometer\_status - tomatometer value of "Rotten" (less than 60% positive reviews), "Fresh" (at least 60% of positive reviews), and "Certified Fresh" (at least 75% of positive reviews, at least 80 reviews of which at least 5 from top critics)

tomatometer\_rating - percentage of positive critic ratings

tomatometer\_count - critic ratings counted for the calculation of the tomatomer status

audience\_status - audience value of "Spilled" (less than 60% of users gave a rating of at least 3.5) or "Upright" (at least 60% of users gave a rating of at least 3.5)

audience\_rating - percentage of positive user ratings

audience\_count - user ratings counted for the calculation of the audience status

tomatometer\_top\_critics\_count - count of top critic ratings

tomatometer\_fresh\_critics\_count - count of fresh critic ratings

tomatometer\_rotten\_critics\_count - count of rotten critic ratings