

IMDB Movies Data Analysis

Gerald Yuen

2022

PROJECT SUMMARY

The data is from IMDB with around 5000 movies from 1916 to 2016. The dataset is from Kaggle back in 2017.

My goal is to find interesting facts with the dataset while showing analytic skills.

DATA SOURCE

<https://www.kaggle.com/datasets/carolzhangdc/imdb-5000>

LIMITATIONS AND ETHICS

- Data was created 5 years ago, so data might not be up to date.
- Only 5000 movies are included.
- Data is not official from IMDB so result might vary.

DATA CLEANING AND DATA CONSISTENCY CHECKS

- Removing unnecessary columns
- Updated "movie_title" rows
- Check for blank space and replaced with "N/A"
- Removed duplicate values

DATA PROFILE

- Number of rows and columns in the cleaned dataset: 4920 x 19

COLUMN DETAILS:

<u>Variable Name</u>	<u>Description</u>	<u>Datatype</u>	<u>Time Variant</u>
movie_title	Title of the Movie	Qualitative/Nominal	No
duration	Duration in minutes	Quantitative/Continuous	No
title_year	Year of the movie	Quantitative/Discrete	Yes
director_name	Name of the Director of the Movie	Qualitative/Nominal	No
actor_1_name	Primary actor starring in the movie	Qualitative/Nominal	No
actor_2_name	Other actor starring in the movie	Qualitative/Nominal	No
actor_3_name	Other actor starring in the movie	Qualitative/Nominal	No
language	English, Arabic, Chinese, French, German, Danish, Italian, Japanese etc	Qualitative/Nominal	No
country	Country where the movie is produced	Qualitative/Nominal	No
content_rating	Content rating of the movie	Quantitative/Discrete	No
num_critic_for_reviews	Number of reviews	Quantitative/Continuous	No
director_facebook_likes	Number of likes of the Director on his Facebook Page	Quantitative/Continuous	No
budget	Budget of the movie in Dollars	Quantitative/Continuous	No

<u>Variable Name</u>	<u>Description</u>	<u>Datatype</u>	<u>Time Variant</u>
gross	Gross earnings of the movie in Dollars	Quantitative/Continuous	No
genres	Film categorization like 'Animation', 'Comedy', 'Romance', 'Horror', 'Sci-Fi', 'Action', 'Family'	Qualitative/Nominal	No
cast_total_facebook_likes	Total number of facebook likes of the entire cast of the movie	Quantitative/Continuous	No
imdb_score	IMDB Score of the movie on IMDB	Quantitative/Continuous	No
color	Film colorization. 'Black and White' or 'Color'	Qualitative/Nominal	No
Plot_keywords	Keywords describing the movie plot	Qualitative/Nominal	No

QUESTIONS TO EXPLORE:

- What is the average time of a movie?
- Who are the top directors and how much money does their movies made?
- What genre are the most popular?
- For the highest gross earning of the movies, how are the IMDB scores?