

IMDB Movie Analysis

Project Description

The IMDB Movie Analysis project aims to analyze a dataset of movies from the Internet Movie Database (IMDB) to gain valuable insights into various aspects of the film industry. The analysis will involve exploring key attributes such as movie ratings, genres, budgets, and more to understand trends and patterns within the dataset. By leveraging data visualization and statistical techniques, this project will provide valuable information for movie genres, duration, directors, Budget.

Approach

For this project, first we'll get understanding of the given data. Then We'll clean the data as per our requirement by removing null values, delete unnecessary columns, etc. After the cleaning we'll using pivot table, various functions and charts for desired answers for the questions. We'll continues ask Whys to data to get in depth of the root of the problem. At the end we'll present our answers with proper formatting in table and graphs.

Tech-Stack Used

- I have used Microsoft Excel (2021) for data analysis.

Insights

- Most common genres of movies is Drama, Action, Thriller, comedy.
- The average duration of movies is 109 minutes and there is upward trendline between IMDB score and movie duration.
- Most common movies language is english, french and spanish and languages who have high average IMDB score is Telugu and Persian.
- I have identify the directors with the highest scores and there are 186 directors who are in top 10% of total directors and have higher average IMDB score compared to mean of all director's IMDB score.
- I have identified movies with highest profit margin and there is positive correlation between movie budgets and gross earnings and movies with highest profit margin is Avatar.

Result

In this Project of IMDB Movies Analysis, I have gained a vast knowledge of statistics and excel functions. Through this I learn how to apply statistics in real world problems and how to visualise the insights. By enhancing my portfolio, this project will enable me to take advantage of intriguing opportunities in the future.

A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

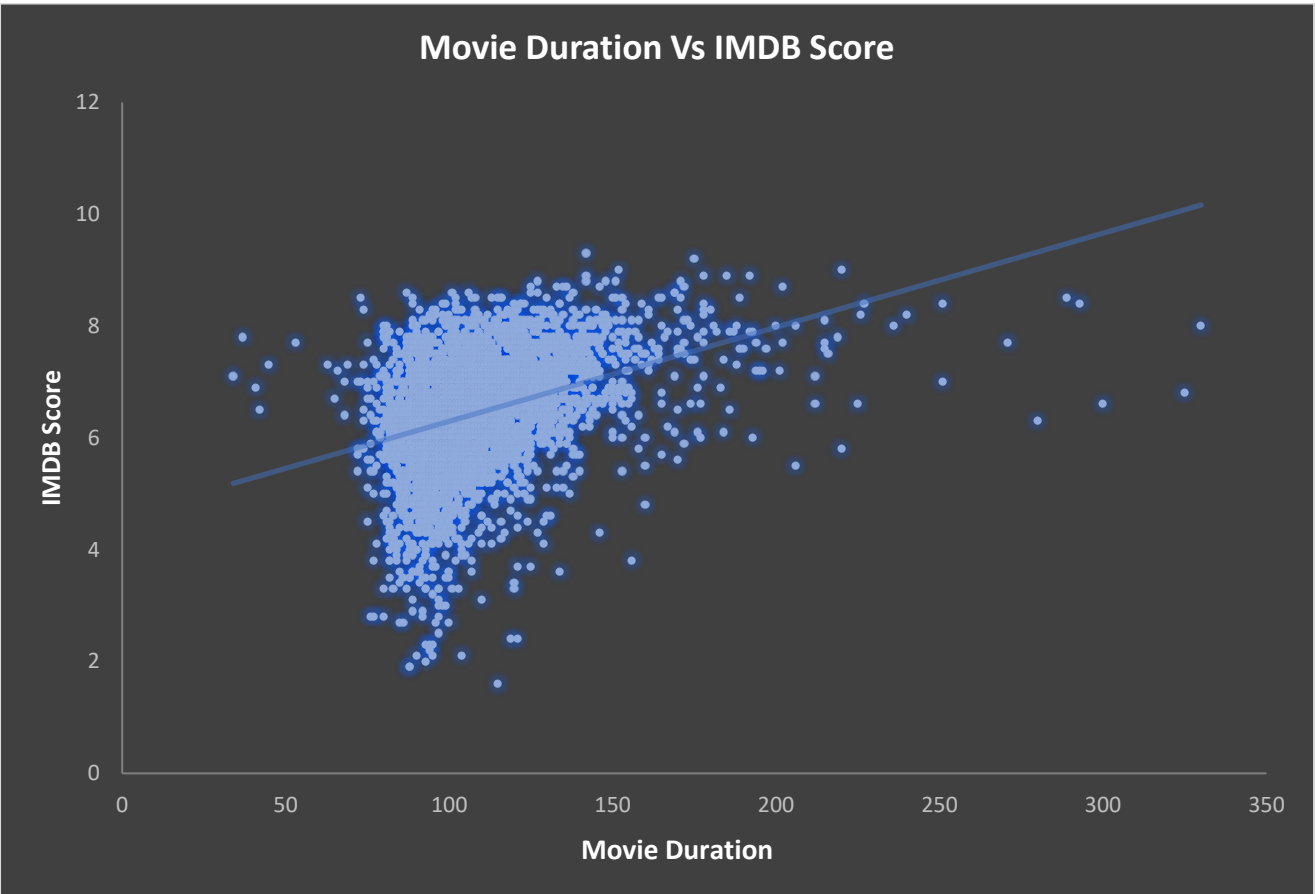
Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Genres	Count of Movies	Mean of IMDB Score	Median of IMDB Score	Mode of IMDB Score	Range of IMDB Score	Variance of IMDB Score	Standard Deviation of IMDB Score
Drama	1910	6.789005236	6.9	6.7	7.2	0.793973356	0.891051826
Comedy	1491	6.182763246	6.3	6.7	6.9	1.080983914	1.039703763
Thriller	1084	6.374815498	6.4	6.5	6.3	0.938378656	0.968699466
Action	935	6.285989305	6.3	6.6	6.9	1.077033647	1.037802316
Romance	865	6.425433526	6.5	6.5	6.4	0.93842828	0.96872508
Adventure	766	6.454960836	6.6	6.7	6.6	1.245895756	1.116197006
Crime	702	6.548148148	6.6	6.6	6.9	0.967083465	0.983404019
Fantasy	496	6.285080645	6.4	6.7	6.7	1.297922574	1.139264049
Sci-Fi	483	6.325879917	6.4	7	6.9	1.361379919	1.166781865
Family	441	6.2	6.3	5.4	6.7	1.364807256	1.168249655
Horror	378	5.901058201	5.9	6.2	6.3	0.978940679	0.989414311
Mystery	376	6.471808511	6.5	6.6	5.5	1.012822261	1.00639071
Biography	242	7.140082645	7.2	7	4.4	0.502153712	0.708628049
Animation	197	6.700507614	6.8	7.3	5.8	0.982284006	0.99110242
War	159	7.048427673	7.1	7.1	4.3	0.648283691	0.805160662
Music	159	6.371698113	6.5	6.5	6.9	1.464670701	1.210235804
History	152	7.131578947	7.2	7.7	3.4	0.448608033	0.669782079
Sport	147	6.601360544	6.8	7.2	6.4	1.091290666	1.044648585
Musical	102	6.550980392	6.7	7.1	6.4	1.29485198	1.13791563
Documentary	64	6.99375	7.2	6.6	6.9	1.476210937	1.214994213
Western	58	6.765517241	6.8	6.8	4.8	0.979845422	0.989871417
Short	2	6.8	6.8	#N/A	0.6	0.09	0.3
Film-Noir	1	7.7	7.7	#N/A	0	0	0

B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Movie Duration	
Mean	109.8236
Median	105
Variance	518.4045
Standard Deviation	22.7685



C. Language Analysis: Situation: Examine the distribution of movies based on their language.

Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Language	Count of Movies	Mean of IMDB Score	Median of IMDB Score	STDEV of IMDB Score
English	3602	6.420849528	6.5	1.052458611
French	37	7.286486486	7.2	0.553691378
Spanish	26	7.05	7.15	0.810151933
Mandarin	14	7.021428571	7.25	0.737930089
German	13	7.692307692	7.7	0.615769111
Japanese	12	7.625	7.8	0.861321659
Hindi	10	6.76	7.05	1.05470375
Cantonese	8	7.2375	7.3	0.412121038
Italian	7	7.185714286	7	1.069617517
Portuguese	5	7.76	8	0.875442745
Korean	4	7.875	7.9	0.414578099
Norwegian	4	7.15	7.3	0.497493719
Danish	3	7.9	8.1	0.43204938
Dutch	3	7.566666667	7.8	0.329983165
Hebrew	3	7.5	7.3	0.355902608
Persian	3	8.133333333	8.4	0.449691252
Thai	3	6.633333333	6.6	0.368178701
Aboriginal	2	6.95	6.95	0.55
Dari	2	7.5	7.5	0.1
Indonesian	2	7.9	7.9	0.3
Arabic	1	7.2	7.2	0
Aramaic	1	7.1	7.1	0
Bosnian	1	4.3	4.3	0
Czech	1	7.4	7.4	0
Dzongkha	1	7.5	7.5	0
Filipino	1	6.7	6.7	0
Hungarian	1	7.1	7.1	0
Icelandic	1	6.9	6.9	0
Kazakh	1	6	6	0
Maya	1	7.8	7.8	0
Mongolian	1	7.3	7.3	0
None	1	8.5	8.5	0
Romanian	1	7.9	7.9	0
Russian	1	6.5	6.5	0
Swedish	1	7.6	7.6	0
Telugu	1	8.4	8.4	0
Vietnamese	1	7.4	7.4	0
Zulu	1	7.3	7.3	0

D. Director Analysis: Influence of directors on movie ratings.

Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Directors	Average IMDB Score	Directors with Highest Scores
Charles Chaplin	8.6	Charles Chaplin
Tony Kaye	8.6	Tony Kaye
Alfred Hitchcock	8.5	Alfred Hitchcock
Damien Chazelle	8.5	Damien Chazelle
Majid Majidi	8.5	Majid Majidi
Ron Fricke	8.5	Ron Fricke
Sergio Leone	8.433333333	Sergio Leone
Christopher Nolan	8.425	Christopher Nolan
Asghar Farhadi	8.4	Asghar Farhadi
Marius A. Markevicius	8.4	Marius A. Markevicius
Richard Marquand	8.4	Richard Marquand
S.S. Rajamouli	8.4	S.S. Rajamouli
Billy Wilder	8.3	Billy Wilder
Fritz Lang	8.3	Fritz Lang
Lee Unkrich	8.3	Lee Unkrich
Lenny Abrahamson	8.3	Lenny Abrahamson
Pete Docter	8.233333333	Pete Docter
Hayao Miyazaki	8.225	Hayao Miyazaki
Elia Kazan	8.2	Elia Kazan
George Roy Hill	8.2	George Roy Hill
Joshua Oppenheimer	8.2	Joshua Oppenheimer
Juan José Campanella	8.2	Juan José Campanella
Quentin Tarantino	8.2	Quentin Tarantino
Victor Fleming	8.15	Victor Fleming
Milos Forman	8.133333333	Milos Forman
Akira Kurosawa	8.1	Akira Kurosawa

E. Budget Analysis: Explore the relationship between movie budgets and their financial success.

Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Movie Title	Budget	Profit
Avatar	\$237M	\$524M
Jurassic World	\$150M	\$502M
Titanic	\$200M	\$459M
Star Wars: Episode IV - A New Hope	\$11M	\$450M
E.T. the Extra-Terrestrial	\$11M	\$424M
The Avengers	\$220M	\$403M
The Lion King	\$45M	\$378M
Star Wars: Episode I - The Phantom Menace	\$115M	\$360M
The Dark Knight	\$185M	\$348M
The Hunger Games	\$78M	\$330M

Correlation Coefficient	0.222901783
Movie with Highest Profit (Avatar)	\$52,35,05,847