

IMDB Movie Analysis Report

Project Description:

Overview

The primary objective of this project is to analyze the factors influencing the success of movies on IMDB, with success being defined by high IMDB ratings. Understanding these factors is crucial for movie producers, directors, and investors who aim to make data-driven decisions to maximize audience reception and financial returns. Through this analysis, we explore the relationship between different movie attributes such as genre, duration, language, director, and budget with IMDB scores.

Problem Statement

Given a dataset of IMDB movies, we aim to investigate the key elements that contribute to a movie's success. This analysis will provide insights into how different factors, such as genre, budget, language, and director, influence movie ratings. The findings from this project can help stakeholders in the movie industry make informed choices when selecting scripts, allocating budgets, and targeting audiences.

Approach:

Data Cleaning:

To ensure accurate and reliable analysis, the dataset underwent preprocessing, which involved the following steps:

- Handling Missing Values: Filling missing values where applicable and removing incomplete records.
- Removing Duplicates: Eliminating redundant entries.
- Standardizing Formats: Converting data types for consistency.
- Feature Engineering: Creating new relevant columns, if necessary, for enhanced insights.

Data Analysis:

Each factor influencing movie success was examined through descriptive statistics, correlation analysis, and data visualization. The key analytical tasks included:

- Movie Genre Analysis: Examining the most common genres and their impact on IMDB ratings.
- Movie Duration Analysis: Studying the relationship between movie duration and IMDB scores.
- Language Analysis: Identifying the most common movie languages and their effect on ratings.

- Director Analysis: Determining top directors based on their average IMDB scores and percentile rankings.
- Budget Analysis: Investigating the correlation between movie budgets and gross earnings, along with identifying movies with the highest profit margins.

Five 'Whys' Approach:

To dig deeper into relationships and uncover root causes, the Five 'Whys' technique was employed. For example, if higher budgets were linked to better ratings, we questioned why this occurs at multiple levels until reaching the fundamental reason.

Tech-Stack Used:

WPS Office: Used for data cleaning, descriptive statistics, correlation analysis, visualization (charts, graphs), and calculations.

Insights:

A. Movie Genre Analysis:

- The dataset contains a diverse range of movie genres, with Comedy, Action, and Drama being the most common.
- Genres like Western and Biography tend to have higher average IMDB scores.

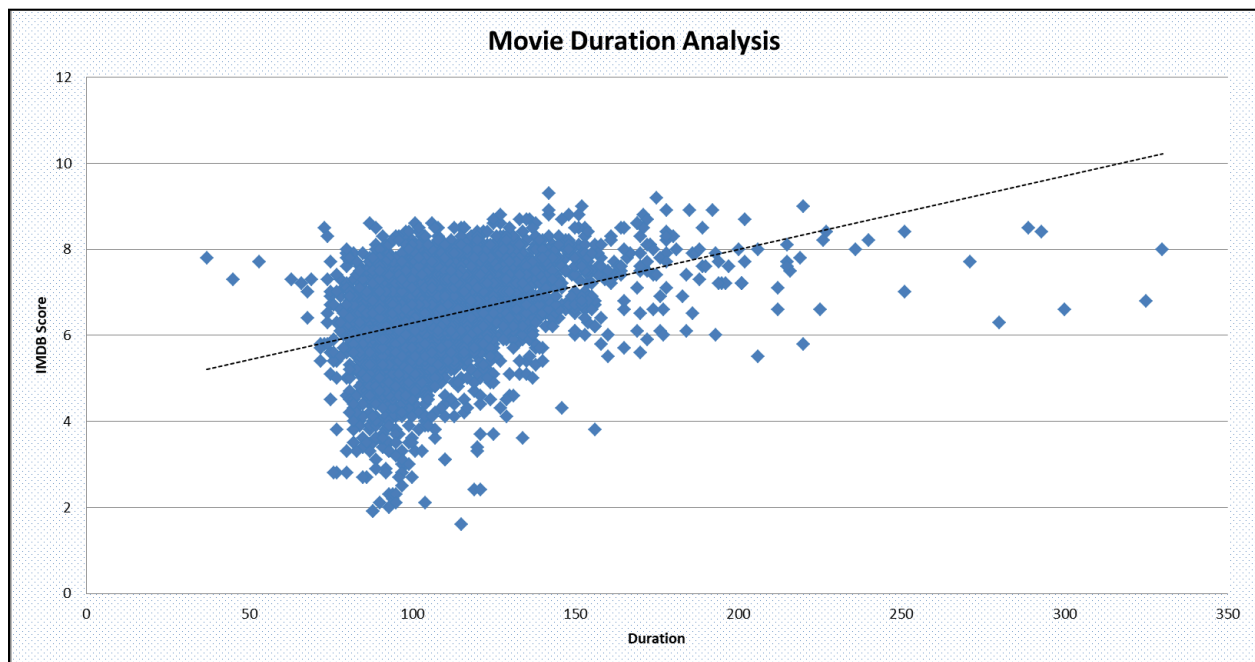
With PIVOT Table	
Genres	Count of genres
Action	959
Adventure	369
Animation	45
Biography	205
Comedy	989
Crime	255
Documentary	26
Drama	668
Family	3
Fantasy	37
Horror	164
Musical	2
Mystery	23
Romance	1
Sci-Fi	7
Thriller	1
Western	2
Grand Total	3756

With COUNTIF Function			
Genres	Values	Average (IMDB Score)	
Action	959	6.29	
Adventure	369	6.55	
Animation	45	6.74	
Biography	205	7.16	
Comedy	989	6.17	
Crime	255	6.94	
Documentary	26	6.8	
Drama	668	6.83	
Family	3	6.5	
Fantasy	37	6.28	
Horror	164	5.85	
Musical	2	6.75	
Mystery	23	6.65	
Romance	1	7.1	
Sci-Fi	7	6.63	
Thriller	1	4.8	
Western	2	8.1	
Total	3756		

Descriptive	Values
Mean	220.94
Median	37
Mode	2
Max	989
Min	1
Var	111809.06
StdDev	324.39

B. Movie Duration Analysis:

- The average movie duration in the dataset is approximately 100–120 minutes.
- Movies with moderate durations (90–150 minutes) tend to receive higher IMDB scores, whereas extremely short or long movies generally have lower scores comparatively.
- The scatter plot analysis and trendline suggest a weak positive correlation between duration and IMDB ratings.



Descriptive Statistics	Values
Mean	110.26
Median	106
StdDev	22.64
Correlation	0.37

C. Language Analysis:

- English is the dominant language in most movies, followed by French, Spanish, and Japanese, but with a huge margin.
- English-language films show a broad range of IMDB scores, indicating varying content quality, whereas foreign-language films tend to have higher average ratings due to niche audiences.

With Pivot Table

Row Labels	Count of language
Aboriginal	2
Arabic	1
Aramaic	1
Bosnian	1
Cantonese	7
Czech	1
Danish	3
Dari	2
Dutch	3
English	3598
Filipino	1
French	34
German	10
Hebrew	1
Hindi	5
Hungarian	1
Indonesian	2
Italian	7
Japanese	10
Kazakh	1
Korean	5
Mandarin	15
Maya	1
Mongolian	1
None	1
Norwegian	4
Persian	3
Portuguese	5
Romanian	1
Russian	1
Spanish	23
Thai	3
Vietnamese	1
Zulu	1
Grand Total	3756

With COUNTIF Function					
Language: ▼	Count ▼	Mean ▼	Median ▼	Std Dev. ▼	
Aboriginal	2	6.95	6.6	1.06	
Arabic	1	7.2	6.6	1.06	
Aramaic	1	7.1	6.6	1.06	
Bosnian	1	4.3	6.6	1.06	
Cantonese	7	7.34	6.6	1.06	
Czech	1	7.4	6.6	1.06	
Danish	3	7.9	6.6	1.06	
Dari	2	7.5	6.6	1.06	
Dutch	3	7.57	6.6	1.06	
English	3598	6.43	6.6	1.06	
Filipino	1	6.7	6.6	1.06	
French	34	7.36	6.6	1.06	
German	10	7.77	6.6	1.06	
Hebrew	1	8	6.6	1.06	
Hindi	5	7.22	6.6	1.06	
Hungarian	1	7.1	6.6	1.06	
Indonesian	2	7.9	6.6	1.06	
Italian	7	7.19	6.6	1.06	
Japanese	10	7.66	6.6	1.06	
Kazakh	1	6	6.6	1.06	
Korean	5	7.7	6.6	1.06	
Mandarin	15	7.08	6.6	1.06	
Maya	1	7.8	6.6	1.06	
Mongolian	1	7.3	6.6	1.06	
None	1	8.5	6.6	1.06	
Norwegian	4	7.15	6.6	1.06	
Persian	3	8.13	6.6	1.06	
Portuguese	5	7.76	6.6	1.06	
Romanian	1	7.9	6.6	1.06	
Russian	1	6.5	6.6	1.06	
Spanish	23	7.08	6.6	1.06	
Thai	3	6.63	6.6	1.06	
Vietnamese	1	7.4	6.6	1.06	
Zulu	1	7.3	6.6	1.06	
Total	3756				

D. Director Analysis:

- Top directors based on average IMDB ratings were identified using percentile calculations.
- Directors in the 90th percentile and above have consistently high-rated movies.
- The impact of a director on a movie's success was found to be significant, with renowned directors contributing to higher audience expectations and reception.

Director_name	Average of imdb_score
Akira Kurosawa	8.7
Seven Samurai	8.7
Alfred Hitchcock	8.5
Psycho	8.5
Charles Chaplin	8.6
Modern Times	8.6
Damien Chazelle	8.5
Whiplash	8.5
Majid Majidi	8.5
Children of Heaven	8.5
Ron Fricke	8.5
Samsara	8.5
Tony Kaye	8.6
American History X	8.6

PERCENTILE	8.66
Mean	8.59
Median	8.6
SD	0.08

E. Budget Analysis:

- A moderately positive correlation was found between movie budgets and gross earnings.
- High-budget movies generally earn more, but some low-budget films achieve remarkable success, proving that quality content can overcome budget constraints.
- The highest profit margins were observed in movies with low-to-mid budgets but high audience engagement, suggesting that an effective marketing strategy and strong storytelling can lead to financial success.

Top 10 Profitable Movies:

movie_title	gross	budget	Profit
Avatar	760505847	237000000	523505847
The Avengers	623279547	220000000	403279547
Titanic	658672302	200000000	458672302
Jurassic World	652177271	150000000	502177271
The Dark Knight	533316061	185000000	348316061
Star Wars: Episode I - The Phantom Menace	474544677	115000000	359544677
The Lion King	422783777	45000000	377783777
The Avengers	623279547	220000000	403279547
Star Wars: Episode IV - A New Hope	460935665	11000000	449935665
E.T. the Extra-Terrestrial	434949459	10500000	424449459

CORRELATION	0.099496423
Highest Profit Margin	523505847

Result:

Through this analysis, we identified critical factors influencing IMDB movie ratings. The findings suggest that while budget, duration, and director reputation play a role, the genre and storytelling quality have a substantial impact on audience reception. The correlation between financial investment and earnings highlights the importance of balancing production costs with content quality. These insights are valuable for stakeholders looking to optimize their movie production strategies.

Drive Link:

The final report has been uploaded to Google Drive. Please find the

Dataset with solutions at the following link:

 IMDB_Movies_Modified.xlsx