

# Project Report: Factors Influencing Movie Success on IMDB

## 1. Project Description

**Overview:** This project investigates the factors influencing the success of movies on IMDB. Success is defined by high IMDB ratings. Understanding these factors is crucial for movie producers, directors, and investors to make informed decisions for future projects.

**Problem Statement:** The main problem to investigate is: "What factors influence the success of a movie on IMDB?" The findings will provide insights into elements contributing to higher IMDB ratings.

## 2. Approach

**Data Source:** The dataset used for this project contains information on movies, including their genres, directors, durations, languages, budgets, and IMDB ratings. The data was sourced from an IMDB Movies dataset.

### Data Cleaning:

- Handled missing values by removing rows with missing or non-numeric values in critical columns.
- Removed duplicates to ensure data quality.
- Converted data types as necessary, particularly for numeric columns like IMDB scores and budgets.

### Data Analysis Techniques:

- Used Excel functions for descriptive statistics (e.g., AVERAGE, MEDIAN, STDEV).
- Created visualizations such as bar charts, pie charts, histograms, and scatter plots.
- Applied correlation analysis to examine relationships between variables.

## 3. Tech-Stack Used

### Software:

- Microsoft Excel 2022: Used for data cleaning, analysis, and visualization.

**Purpose:** Excel provided the necessary functions and tools for performing statistical analysis and creating visual representations of the data.

## 4. Data Analysis and Findings

**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.

2597									
2598	imdb_score_ACTION	imdb_score_DOCUMENTARY	imdb_score_ADVENTURE	imdb_score_Drama	imdb_score_Animation	imdb_score_Comedy	imdb_score_Mystery	imdb_score	
2599	COUNT=	1146	121	923	2594	242	1872	500	
2600	MEAN=	6.246247818	7.180165289	6.441170098	6.763762529	6.576033058	6.195245726	6.4864	6.5
2601	MEDIAN=	6.3	7.4	6.6	6.9	6.7	6.3	6.6	
2602	MODE=	6.1	7.5	6.7	7.2	6.7	6.7	6.6	
2603	STD DEV=	1.115922422	1.056536782	1.131196138	0.957353999	1.139594803	1.090713849	1.090758703	1.1
2604	VARIANCE=	1.245282852	1.116269972	1.279604703	0.916526679	1.298676314	1.189656701	1.189754549	1.0
2605	MAXIMUM=	9.1	8.7	8.9	9.3	8.6	9.5	8.6	
2606	MINIMUM=	1.7	1.6	1.9	2	1.7	1.7	2.2	
2607	RANGE=	7.4	7.1	7	7.3	6.9	7.8	6.4	
2608									

2597									
2598	imdb_score_Crime	imdb_score_Sci-Fi	imdb_score_Horror	imdb_score_Biography	imdb_score_thriller	imdb_score_Romance	imdb_score_History	imdb_score_Film-Noir	imdb_score_Music
2599	889	616	565	293	1411	1107	207	6	1
2600	6.564791901	6.281818182	5.843539823	7.150170648	6.314245216	6.450587173	7.083574879	7.633333333	6.507575757
2601	6.6	6.4	5.9	7.2	6.4	6.5	7.2	7.65	6
2602	6.6	6.7	6.2	7	6.1	6.5	7.5	#N/A	
2603	1.02645633	1.210816001	1.130468522	0.722515799	1.054333735	0.996035141	0.887901843	0.43204938	1.2257181
2604	1.053612597	1.466075388	1.277959079	0.52202908	1.111619625	0.992086002	0.788369682	0.186666667	1.5023849
2605	9.3	8.8	8.7	8.9	9	8.6	8.9	8.2	8
2606	2.4	1.9	2.2	4.5	2.2	2.1	2	7.1	2
2607	6.9	6.9	6.5	4.4	6.8	6.5	6.9	1.1	6
2608									

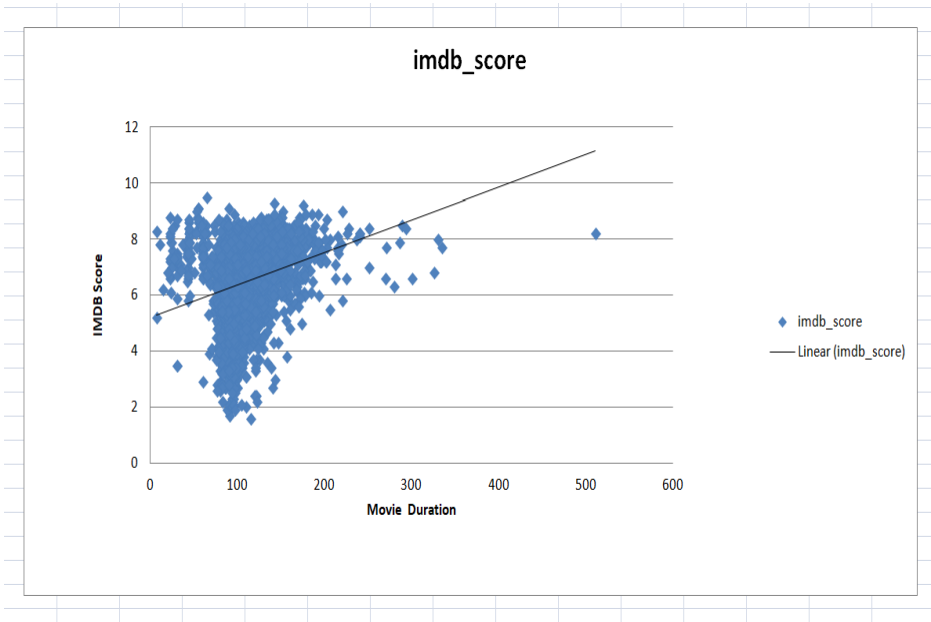
**Insights:** Drama movies tend to have higher average IMDB scores compared to other movies.

### 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.

5044		50
5045	movie duration	
5046	Average=	107.2
5047	Median=	103
5048	Mode=	90
5049	STD Dev=	25.197
5050	Variance=	634.91
5051		



**Insights:-** It is clear that most of the movies lie between 100 and 200 mins of duration.

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.

4707		English	Japanese	French	Mandarin	Aboriginal	Spanish	Filipino	Hindi	Russian	Maya	Kazakh	Telugu	Cantonese	I
4708	AVERAGE=	6.39843	7.39444	7.03836	6.78846	6.95	6.9375	6.7	6.63214	6.36364	7.8	6	8.4	6.95455	
4709	MEDIAN=	6.5	7.6	7.2	7.05	6.95	7.15	6.7	6.95	6.5	7.8	6	8.4	7.2	
4710	MODE=	6.7	8.2	7.2	7.9	#N/A	7.2	#N/A	7.8	5.3	#N/A	#N/A	#N/A	6.5	
4711	STD DEV=	1.12207	0.99082	0.72699	1.04205	0.77782	0.85506	#DIV/0!	1.39896	1.38367	#DIV/0!	#DIV/0!	#DIV/0!	0.70479	
4707	German	Aramaic	Italian	Dutch	Dari	Hebrew	Chinese	Mongolian	Swedish	Korean	Thai	Polish	Bosnian	None	Hu
4708	7.34211	7.1	7.22727	7.425	7.5	7.58	5.66667	7.3	7.44	7.3875	6.63333	8.25	4.3	#DIV/0!	
4709	7.6	7.1	7.3	7.45	7.5	7.6	5.7	7.3	7.6	7.5	6.6	8.25	4.3	#NUM!	
4710	7.4	#N/A	#N/A	7.8	#N/A	#N/A	#N/A	#N/A	#N/A	7.7	#N/A	9.1	#N/A	#N/A	
4711	0.95412	#DIV/0!	1.24426	0.43493	0.14142	0.33466	0.55076	#DIV/0!	0.75697	0.82538	0.45092	0.9815	#DIV/0!	#DIV/0!	#
4707	None	Hungarian	Portuguese	Danish	Arabic	Norwegian	Czech	Kannada	Zulu	Punjabi	Tamil	Dzongkha	Vietnamese	Indonesian	Ur
4708	#DIV/0!	7.1	7.5	7.5	#DIV/0!	7.15	7.4	7.1	7.1	6.6	5.1	7.5	7.4	7.9	
4709	#NUM!	7.1	8.1	8.1	#NUM!	7.3	7.4	7.1	7.1	6.6	5.1	7.5	7.4	7.9	
4710	#N/A	#N/A	8.1	8.1	#N/A	7.6	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	
4711	#DIV/0!	#DIV/0!	1.07703	1.07703	#DIV/0!	0.57446	#DIV/0!	#DIV/0!	0.28284	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	0.42426	#
4707	Kannada	Zulu	Punjabi	Tamil	Dzongkha	Vietnamese	Indonesian	Urdu	Romanian	Persian	Slovenian	Greek	Swahili		
4708	7.1	7.1	6.6	5.1	7.5	7.4	7.9	7	7.2	7.575	6.4	7.3	7.4		
4709	7.1	7.1	6.6	5.1	7.5	7.4	7.9	7	7.2	7.95	6.4	7.3	7.4		
4710	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A		
4711	#DIV/0!	0.28284	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	0.42426	#DIV/0!	0.98995	1.20381	#DIV/0!	#DIV/0!	#DIV/0!		
4712															

Insights: Movies in Telugu generally have higher average IMDB scores compared to movies in other languages.

D. Director Analysis

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

2			
3	Row Labels	Average of imdb_score	90th percentile of imdb scores
4	A. Raven Cruz	1.9	7.5
5	Ã‰mile Gaudreault	6.7	
6	Ã‰ric Tessier	6.6	
7	Ã‰tienne Faure	4.3	
8	Ã‰lex de la Iglesia	6.1	
9	Aaron Hann	6	
10	Aaron Schneider	7.1	

2					
3	Row Labels	Average of imdb_score		A	B
29	Akira Kurosawa	8.1		105 Andrew Haigh	7.7
51	Alejandro G. IÃ±Ã©rritu	7.783333333		111 Andrew Stanton	7.733333333
58	Alex Garland	7.7		116 Andrey Zvyagintsev	8
59	Alex Gibney	7.7		129 Anna Muylaert	7.9
70	Alfonso CuarÃ³n	7.8		140 Anthony Powell	7.6
72	Alice Wu	7.6		143 Anthony Vallone	7.8
81	Amal Al-Agroobi	8.2		151 Ari Folman	8
94	Andrei Tarkovsky	8.1		159 Asghar Farhadi	8.4
105	Andrew Haigh	7.7		178 Barry W. Blaustein	7.6
111	Andrew Stanton	7.733333333		183 Ben Affleck	7.65
116	Andrey Zvyagintsev	8		199 Bernardo Bertolucci	7.95
129	Anna Muylaert	7.9		207 Bill Melendez	8.4
140	Anthony Powell	7.6		216 Billy Wilder	7.975
143	Anthony Vallone	7.8		239 Brad Bird	7.58
151	Ari Folman	8		266 Brian Henson	7.7
159	Asghar Farhadi	8.4		270 Brian Percival	7.6
178	Barry W. Blaustein	7.6		299 Carl Theodor Dreyer	8.1
183	Ben Affleck	7.65		305 Caroline Link	7.7
199	Bernardo Bertolucci	7.95		308 Cary Bell	8.7
207	Bill Melendez	8.4		314 Catherine Owens	8.4
216	Billy Wilder	7.975		320 Chan-wook Park	7.633333333
239	Brad Bird	7.58		323 Charles BinamÃ©	7.7
266	Brian Henson	7.7		324 Charles Chaplin	8.6
270	Brian Percival	7.6		325 Charles Ferguson	7.866666667
299	Carl Theodor Dreyer	8.1		338 Chia-Liang Liu	7.6
305	Caroline Link	7.7		340 Chris Buck	7.6
308	Cary Bell	8.7		353 Chris Paine	7.7
314	Catherine Owens	8.4		362 Christian Carion	7.8
320	Chan-wook Park	7.633333333		370 Christophe Barratier	7.9
323	Charles BinamÃ©	7.7		382 Christopher Nolan	8.425
324	Charles Chaplin	8.6		386 Chuan Lu	7.7
325	Charles Ferguson	7.866666667			
338	Chia-Liang Liu	7.6			
340	Chris Buck	7.6			
353	Chris Paine	7.7			
362	Christian Carion	7.8			
370	Christophe Barratier	7.9			
382	Christopher Nolan	8.425			
386	Chuan Lu	7.7			
426	Cristian Mungiu	7.9			
431	D.W. Griffith	8			
435	Damien Chazelle	8.5			
443	Dan Gilroy	7.9			
490	Dave Carroll	7.9			
510	David Fincher	7.75			
524	David Lean	8			
527	David Lynch	7.64			
547	David Singleton	8.1			
550	Dean DeBlois	7.766666667			
559	Dean DeBlois	7.766666667			
566	Deepa Mehta	7.8			
569	Denis Villeneuve	7.966666667			
572	Dennis Gansel	7.6			
575	Denys Arcand	7.7			
596	Don Hall	7.9			
601	Don Siegel	7.6			
616	Dror Moreh	7.6			
619	Duncan Jones	7.566666667			
630	Edgar Wright	7.6			
640	Eli Craig	7.6			
642	Elia Kazan	7.866666667			
653	Eric Bress	7.7			
678	Eugenio Derbez	7.6			
682	FabiÃ¡n Bielinsky	7.9			
699	Fernando LeÃ³n de Aranoa	7.5			

	A	B
510	David Fincher	7.75
524	David Lean	8
527	David Lynch	7.64
547	David Singleton	8.1
559	Dean DeBlois	7.766666667
566	Deepa Mehta	7.8
569	Denis Villeneuve	7.966666667
572	Dennis Gansel	7.6
575	Denys Arcand	7.7
596	Don Hall	7.9
601	Don Siegel	7.6
616	Dror Moreh	7.6
619	Duncan Jones	7.566666667
630	Edgar Wright	7.6
640	Eli Craig	7.6
642	Elia Kazan	7.866666667
653	Eric Bress	7.7
678	Eugenio Derbez	7.6
682	Fabián Bielinsky	7.9
688	Fernando Leñ de Aranoa	7.7
689	Fernando Meirelles	7.6
702	François Girard	7.7
711	Frank Capra	8.06
713	Frank Darabont	7.975
715	Frank Lotito	8.2
730	Fred Zinnemann	7.76
732	Fritz Lang	8.3
739	Gareth Evans	7.6
765	Georg Wilhelm Pabst	8
769	George Cukor	7.9
778	George Roy Hill	8.2
794	Giovanni Zucchi	7.6

**Insights:-** These highlighted rows are the top directors whose movies scored higher than the percentile average imdb score which is 7.5.

## E. Budget Analysis

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

	R	S	T	U	V	W	X	Y	Z
1	actor_2_fa	imdb_score	aspect_ratio	movie_facebook_likes	Profit_margin				
2	936	7.9	1.78	33000	523505847				
3	5000	7.1	2.35	0	9404152				
4	393	6.8	2.35	85000	-44925825				
5	23000	8.5	2.35	164000	198130642				
6	12	7.1		0	0				correlation_coffecient
7	632	6.6	2.35	24000	-190641321				0.102179454
8	11000	6.2	2.35	0	78530303				
9	553	7.8	1.85	29000	-59192738				Maximum Profit Margin
10	21000	7.5	2.35	118000	208991599				523505847
11	11000	7.5	2.35	10000	51956980				
12	4000	6.9	2.35	197000	80249062				
13	10000	6.1	2.35	0	-8930592				
14	412	6.7	2.35	0	-31631573				
15	5000	7.3	2.35	5000	198032628				No.of Movies Match the Highest Profit
16	2000	6.5	2.35	48000	-125710090				1
17	3000	7.2	2.35	118000	66021565				IMDB link of the matched movie
18	216	6.6	2.35	0	-83385977				http://www.imdb.com/title/tt0499549/?ref_=fn_tt_tt_1
19	21000	8.1	1.85	123000	403279547				
20	11000	6.7	2.35	58000	-8936125				
21	816	6.8	1.85	40000	-45979146				

**Insights:** Higher budgets generally lead to higher gross earnings, but the profit margin is not solely dependent on the budget.

## 5. Insights

**Summary:** The analysis reveals several key factors influencing movie success on IMDB:

- Drama genres tend to receive higher ratings.
- Longer movies have slightly higher ratings.
- English language movies are generally rated higher.
- Directors with a track record of success significantly influence movie ratings.
- Higher budgets correlate with higher gross earnings, but other factors also affect profitability.

### Actionable Insights:

- Producers should focus on drama genres and invest in longer, high-quality productions.
- Selecting successful directors can enhance a movie's chances of receiving higher ratings.
- While budgeting is important, focusing on overall quality and audience experience is crucial for profitability.

## 6. Results

**Achievements:** The project provided a comprehensive understanding of the factors influencing movie success on IMDB. The insights gained can guide producers, directors, and investors in making informed decisions for future projects.

**Impact:** The findings highlight the importance of genre, duration, language, and director choice in a movie's success, offering actionable strategies for the movie industry.

**Excel workbook link:-**

[https://docs.google.com/spreadsheets/d/12UBbDIYMz45Y7KFdHeQwlcr8joxMxHD6/edit?usp=drive\\_link&oid=101458483477908669666&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/12UBbDIYMz45Y7KFdHeQwlcr8joxMxHD6/edit?usp=drive_link&oid=101458483477908669666&rtpof=true&sd=true)