

IMDB Movie Analysis Report

1. Project Description

As a data analyst, I was tasked with investigating factors that influence the success of movies on IMDB. This analysis is crucial for movie producers, directors, and investors who want to make informed decisions for their future projects. My primary goal was to answer the question: "What factors influence the success of a movie on IMDB?"

For this analysis, I defined success primarily through high IMDB ratings. I examined various factors including genre, duration, language, directors, and budget to understand their impact on a movie's success.

2. Approach

To tackle this project, I followed a structured approach:

Excel Sheet Link: [Book1.xlsx](#)

Video Presentation Link :

<https://www.loom.com/share/4e9aae08b7db418aa5a0e7a0b557eb7a?sid=7672b381-d701-4b91-a0fe-cf7fe42110dc>

1. Data Cleaning:

- Handled missing values
- Removed duplicates
- Converted data types where necessary
- Performed feature engineering as needed

2. Data Analysis:

I conducted five main analyses:

a) Movie Genre Analysis:

- Determined the most common genres
- Calculated descriptive statistics of IMDB scores for each genre

b) Movie Duration Analysis:

- Analyzed the distribution of movie durations
- Identified the relationship between duration and IMDB score

c) Language Analysis:

- Determined the most common languages
- Analyzed the impact of language on IMDB score

d) Director Analysis:

- Identified top directors based on average IMDB score
- Analyzed their contribution to movie success

e) Budget Analysis:

- Explored the relationship between budgets and gross earnings
- Identified movies with the highest profit margins

3. Five 'Whys' Approach:

For each significant finding, I applied the Five 'Whys' technique to dig deeper and uncover root causes.

4. Visualization:

I created various charts and graphs to visualize my findings and make them more accessible.

3. Tech-Stack Used

- Software: **Microsoft Excel**
- Purpose: I used Excel for its robust data analysis capabilities, including:
 - Data cleaning and preprocessing

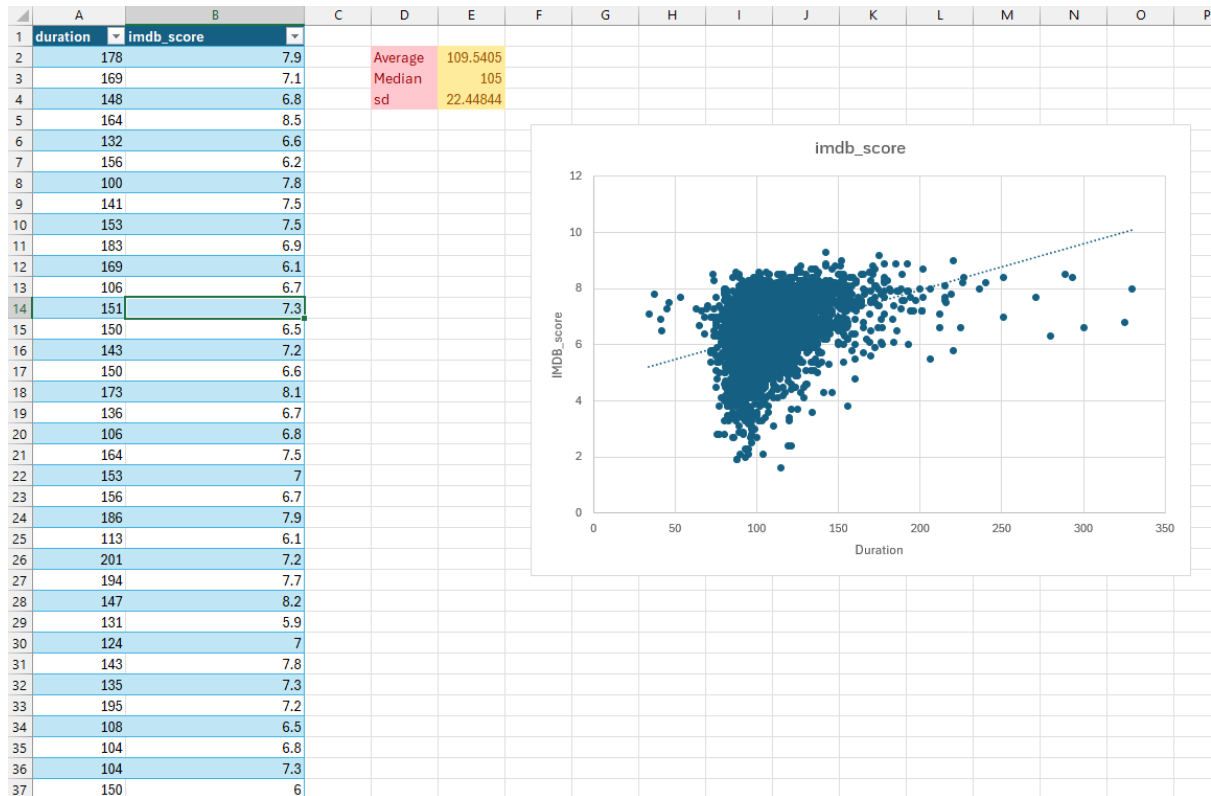
- Statistical analysis (mean, median, mode, range, variance, standard deviation)
- Correlation analysis
- Data visualization (scatter plots, bar charts, etc.)
- Pivot tables for summarizing data

4. Insights

a) Movie Genre Analysis

C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Row Labels	Genre count									Action	Adventure	Fantasy	Sci-Fi		
	Action	976									Action	Adventure	Sci-Fi	Thriller		
	Adventure	802									Drama	Romance				
	Animation	204				Median	483.5				Action	Adventure	Fantasy	Sci-Fi		
	Biography	260				Max	2112				Family	Sci-Fi				
	Comedy	1615				Min	65				Action	Adventure	Sci-Fi			
	Crime	753				Var	292077.2				Adventure	Animation	Drama	Family	Musical	
	Documentary	86				sd	540.4417				Action	Adventure	Fantasy	Thriller		
	Drama	2112									Action	Crime	Drama	Thriller		
	Family	469									Adventure	Drama	Sci-Fi			
	Fantasy	524									Action	Adventure	Comedy	Romance	Sci-Fi	
	Horror	405									Adventure	Sci-Fi	Thriller			
	Music	185									Adventure	Sci-Fi	Thriller			
	Musical	108									Animation	Comedy	Family	Sci-Fi		
	Mystery	400									Action	Biography	Drama	History	Thriller	War
	Romance	937									Adventure	Animation	Comedy	Family		
	Sci-Fi	498									Adventure	Animation	Comedy	Family	Fantasy	Romance
	Thriller	1152									Action	Adventure	Drama	Fantasy		
	Western	65									Action	Adventure	Fantasy	Sci-Fi		
	Grand Total										Comedy	Drama				
											Action	Adventure	Fantasy	Sci-Fi		
											Comedy	Family				
											Action	Adventure	Fantasy	Sci-Fi		
											Action	Adventure	Fantasy	Romance		
											Action	Animation	Comedy	Family	Sci-Fi	
											Drama	Mystery	Thriller			
											Adventure	Drama	Thriller			
											Adventure	Animation	Comedy	Family	Fantasy	Musical
											Animation	Comedy	Family			
											Adventure	Drama	Fantasy	Romance		
											Action	Adventure	Drama	Fantasy		
											Comedy					
											Comedy	Family	Romance			
											Adventure	Drama	Fantasy	Romance		
											Action	Adventure	Sci-Fi			
											Biography	Drama	Sport			
											Action	Adventure				

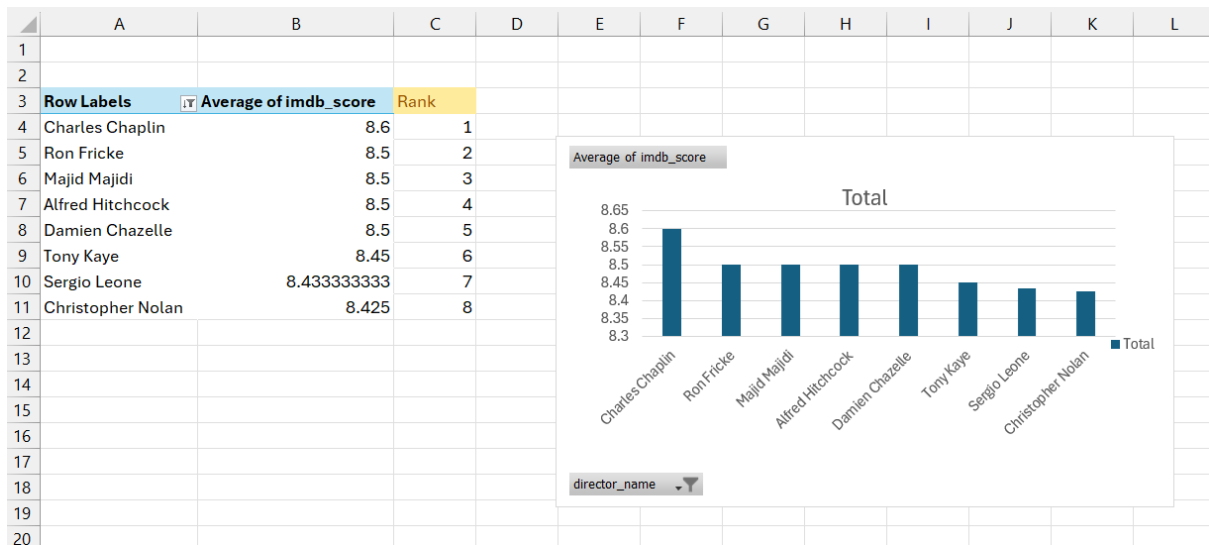
b) Movie Duration Analysis



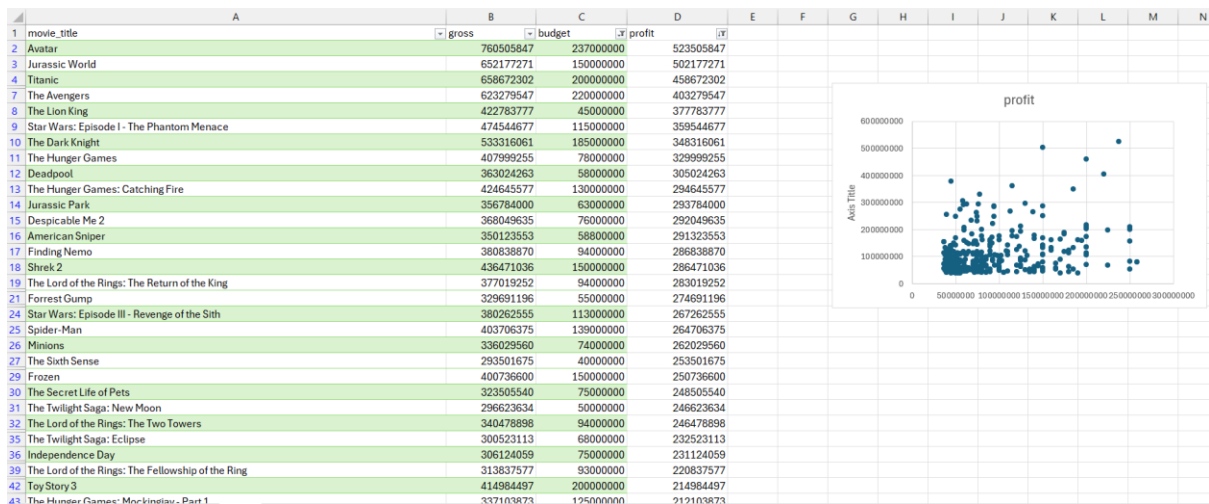
c) Language Analysis

	A	B	C	D	E	F	G	H
1	movie_title	language	imdb_score	Row Labels	Count of language	Mean	Median	SD
2	Avatar	English	7.9	Aboriginal	2	6.95	6.6	1.05
3	Pirates of the Caribbean: At World's End	English	7.1	Arabic	2	7.30	6.6	1.05
4	Spectre	English	6.8	Aramaic	1	7.10	6.6	1.05
5	The Dark Knight Rises	English	8.5	Bosnian	1	4.30	6.6	1.05
6	John Carter	English	6.6	Cantonese	10	7.12	6.6	1.05
7	Spider-Man 3	English	6.2	Chinese	1	5.10	6.6	1.05
8	Tangled	English	7.8	Czech	1	7.40	6.6	1.05
9	Avengers: Age of Ultron	English	7.5	Danish	3	7.90	6.6	1.05
10	Harry Potter and the Half-Blood Prince	English	7.5	Dari	2	7.50	6.6	1.05
11	Batman v Superman: Dawn of Justice	English	6.9	Dutch	3	7.57	6.6	1.05
12	Superman Returns	English	6.1	Dzongkha	1	7.50	6.6	1.05
13	Quantum of Solace	English	6.7	English	3881	6.43	6.6	1.05
14	Pirates of the Caribbean: Dead Man's Chest	English	7.3	Filipino	1	6.70	6.6	1.05
15	The Lone Ranger	English	6.5	French	48	7.19	6.6	1.05
16	Man of Steel	English	7.2	German	15	7.65	6.6	1.05
17	The Chronicles of Narnia: Prince Caspian	English	6.6	Greek	1	7.30	6.6	1.05
18	The Avengers	English	8.1	Hebrew	4	7.68	6.6	1.05
19	Pirates of the Caribbean: On Stranger Tides	English	6.7	Hindi	21	6.70	6.6	1.05
20	Men in Black 3	English	6.8	Hungarian	1	7.10	6.6	1.05
21	The Hobbit: The Battle of the Five Armies	English	7.5	Icelandic	1	6.90	6.6	1.05
22	The Amazing Spider-Man	English	7	Indonesian	2	7.90	6.6	1.05
23	Robin Hood	English	6.7	Italian	7	7.19	6.6	1.05
24	The Hobbit: The Desolation of Smaug	English	7.9	Japanese	12	7.63	6.6	1.05
25	The Golden Compass	English	6.1	Kazakh	1	6.00	6.6	1.05
26	King Kong	English	7.2	Korean	6	7.70	6.6	1.05
27	Titanic	English	7.7	Mandarin	16	6.96	6.6	1.05
28	Captain America: Civil War	English	8.2	Maya	1	7.80	6.6	1.05
29	Battleship	English	5.9	Mongolian	1	7.30	6.6	1.05
30	Jurassic World	English	7	None	1	8.50	6.6	1.05
31	Skyfall	English	7.8	Norwegian	4	7.15	6.6	1.05
32	Spider-Man 2	English	7.3	Persian	4	7.58	6.6	1.05
33	Iron Man 3	English	7.2	Polish	1	7.40	6.6	1.05
34	Alice in Wonderland	English	6.5	Portuguese	6	7.67	6.6	1.05
35	X-Men: The Last Stand	English	6.8	Romanian	1	7.90	6.6	1.05
36	Monsters University	English	7.3	Russian	3	7.47	6.6	1.05
37	Transformers: Revenge of the Fallen	English	6	Spanish	30	7.06	6.6	1.05
38	Transformers: Age of Extinction	English	5.7	Swedish	2	7.15	6.6	1.05
39	Oz the Great and Powerful	English	6.4	Telugu	1	8.40	6.6	1.05
40	The Amazing Spider-Man 2	English	6.7	Thai	3	6.63	6.6	1.05
41	TRON: Legacy	English	6.8	Vietnamese	1	7.40	6.6	1.05
42	Cars 2	English	6.3	Zulu	1	7.30	6.6	1.05
43	Green Lantern	English	5.6	Grand Total	4104			
44	Toy Story 3	English	8.3					
45	Terminator Salvation	English	6.6					
46	Furious 7	English	7.2					

d) Director Analysis



e) Budget Analysis



5. Result

These findings provide valuable insights into what makes a movie successful on IMDB. They can help stakeholders in the film industry make more informed decisions about:

- Genre selection
- Optimal movie duration
- Language considerations
- Director selection
- Budget allocation