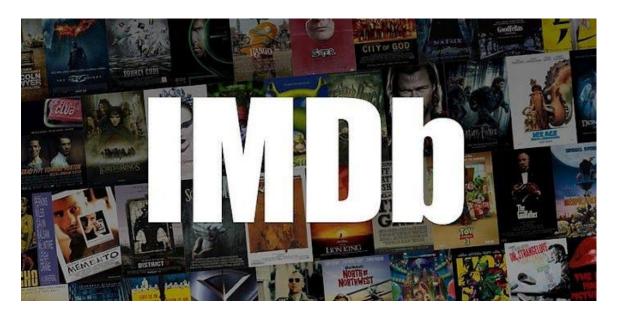
IMDB Movie Analysis



Project Description:

The objective of this project is to analyze the factors influencing the success of movies on IMDB, with success defined by high IMDB ratings. I aim to provide insights into the impact of various factors such as movie genre, duration, language, director, and budget on IMDB ratings. By conducting a thorough analysis, I aim to help movie producers, directors, and investors make informed decisions for their future projects.

Approach:

Link to the excel file: **Excel File**

Data Cleaning:

Α	В	C	D	E	F	G	Н	1	J	K	L	M	N	0	Р	Q	R
color	director_nam∈nu	um_critic_f	duration	director_fa	actor_3_fa	actor_2_n	actor_1_fa	gross	genres	facenumb	eplot_keyw	movie_im	num_user	language	country	content_r	budget
Color	James Camero	723	178	0	855	Joel David	1000	760505847	Action Adventure Fant		avatar fut	http://wv	3054	English	USA	PG-13	237000000
Color	Gore Verbinski	302	169	563	1000	Orlando Bl	40000	309404152	Action Adventure Fant	0	goddess n	http://wv	1238	English	USA	PG-13	300000000
Color	Sam Mendes	602	148	0	161	Rory Kinne	11000	200074175	Action Adventure Thri	1	bomb esp	http://wv	994	English	UK	PG-13	245000000
Color	Christopher No	813	164	22000	23000	Christian B	27000	448130642	Action Thriller	0	deception	http://wv	2701	English	USA	PG-13	250000000
Color	Andrew Stanto	462	132	475	530	Samantha	640	73058679	Action Adventure Sci-F	1	alien ame	http://wv	738	English	USA	PG-13	263700000
Color	Nathan Greno	324	100	15	284	Donna Mu	799	200807262	Adventure Animation	1	17th centu	http://wv	387	English	USA	PG	260000000
Color	Joss Whedon	635	141	0	19000	Robert Do	26000	458991599	Action Adventure Sci-F	4	artificial in	http://wv	1117	English	USA	PG-13	250000000
Color	David Yates	375	153	282	10000	Daniel Rad	25000	301956980	Adventure Family Fant	3	blood boc	http://wv	973	English	UK	PG	250000000
Color	Zack Snyder	673	183	0	2000	Lauren Col	15000	330249062	Action Adventure Sci-F	C	based on c	http://wv	3018	English	USA	PG-13	250000000
Color	Bryan Singer	434	169	0	903	Marlon Bra	18000	200069408	Action Adventure Sci-F	0	crystal epi	http://wv	2367	English	USA	PG-13	209000000
Color	Marc Forster	403	106	395	393	Mathieu A	451	168368427	Action Adventure	1	action here	http://wv	1243	English	UK	PG-13	200000000
Color	Gore Verbinski	313	151	563	1000	Orlando Bl	40000	423032628	Action Adventure Fant	2	box office	http://wv	1832	English	USA	PG-13	225000000
Color	Gore Verbinski	450	150	563	1000	Ruth Wilsc	40000	89289910	Action Adventure Wes	1	horse out	http://wv	711	English	USA	PG-13	215000000
Color	Zack Snyder	733	143	0	748	Christophe	15000	291021565	Action Adventure Fant		based on c	http://wv	2536	English	USA	PG-13	225000000
Color	Andrew Adams	258	150	80	201		22000	141614023	Action Adventure Fam	4	brother br	http://wv	438	English	USA	PG	225000000
Color	Rob Marshall	448	136	252	1000	Sam Claflir	40000	241063875	Action Adventure Fant	4	blackbeard	http://wv	484	English	USA	PG-13	250000000
Color	Barry Sonnenf	451	106	188	718	Michael St	10000	179020854	Action Adventure Com	1	alien crim	http://wv	341	English	USA	PG-13	225000000
Color	Peter Jackson	422	164	0	773	Adam Brov	5000	255108370	Adventure Fantasy	0	army elf l	http://wv	802	English	New Zeal	PG-13	250000000
Color	Marc Webb	599	153	464	963	Andrew Ga	15000	262030663	Action Adventure Fant	0	lizard outc	http://wv	1225	English	USA	PG-13	230000000
Color	Ridley Scott	343	156	0	738	William Hu	891	105219735	Action Adventure Dram	0	1190s arc	http://wv	546	English	USA	PG-13	200000000
Color	Peter Jackson	509	186	0	773	Adam Brov	5000	258355354	Adventure Fantasy	6	dwarf elf	http://wv	951	English	USA	PG-13	225000000
Color	Chris Weitz	251	113	129	1000	Eva Green	16000	70083519	Adventure Family Fant	2	children e	http://wv	666	English	USA	PG-13	180000000
Color	James Camero	315	194	0	794	Kate Winsl	29000	658672302	Drama Romance	0	artist love	http://wv	2528	English	USA	PG-13	200000000
Color	Anthony Russo	516	147	94	11000	Scarlett Jo	21000	407197282	Action Adventure Sci-F	0	based on c	http://wv	1022	English	USA	PG-13	250000000
Color	Peter Berg	377	131	532	627	Alexander	14000	65173160	Action Adventure Sci-F	0	box office	http://wv	751	English	USA	PG-13	209000000
Color	Colin Trevorro	644	124	365	1000	Judy Greer	3000	652177271	Action Adventure Sci-F	0	dinosaur c	http://wv	1290	English	USA	PG-13	150000000
Color	Sam Raimi	300	135	0	4000	James Fran	24000	373377893	Action Adventure Fant	1	death doc	http://wv	1303	English	USA	PG-13	200000000
Color	IMDB Movies- r	aw IME	DB Movies	clean N	lovie Genr	e Analysis		Juration An	Action Advantural Sci. I		Di (1)	httn://usi	1107	English	LICA	DG 13	200000000

I started by preprocessing the dataset to make it suitable for analysis. This involved handling missing values and removing duplicates. All the missing values in numerical fields are replaced with the mean of their respective column and the missing values in textual columns are replaced with a value that has the highest frequency. Both Director_name and movie_name columns are not filled with values of the highest frequency as that doesn't work in this case as the movie name and director name are distinct.

Data Analysis: I analyzed the distribution of movie genres, durations, languages, and directors, and their impact on IMDB scores. I calculated descriptive statistics such as mean, median, mode, range, variance, and standard deviation to understand the relationship between these factors and IMDB ratings. I also used visualization techniques such as scatter plots and trendlines to identify trends and patterns in the data.

Report and Data Story: Finally, I created a detailed report summarizing our findings and insights. I used visualizations to help tell the story and make our findings more understandable.

Tech-Stack Used:

I utilized Microsoft Excel 2021 for data analysis and visualization. Excel's built-in functions and features such as COUNTIF, AVERAGE, MEDIAN, MODE, VAR, STDEV, CORREL, and PERCENTILE facilitated our analysis. Excel was chosen for its versatility, familiarity, and ease of use, making it an ideal choice for conducting comprehensive data analysis and generating actionable insights for our project.

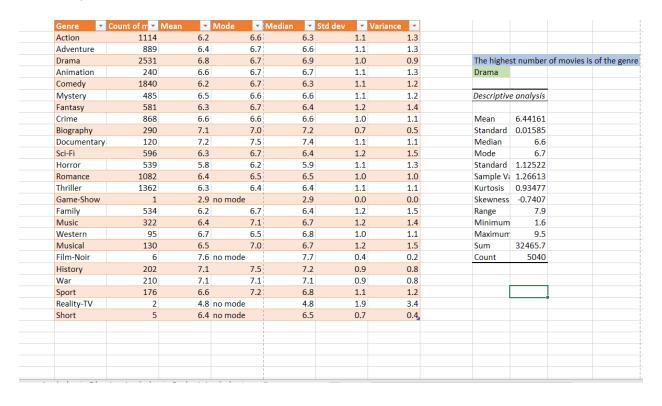


Insights:

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.

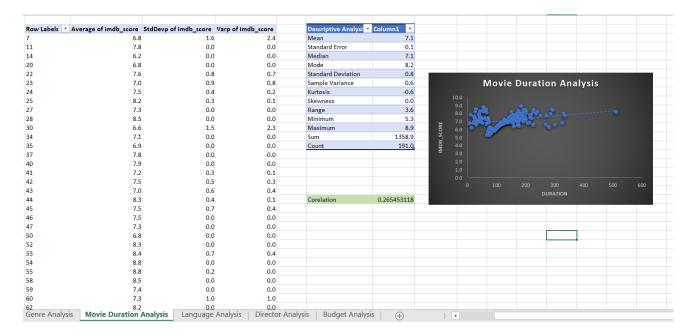


It is revealed that Drama is the genre with the most films-2531 movies in total

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.

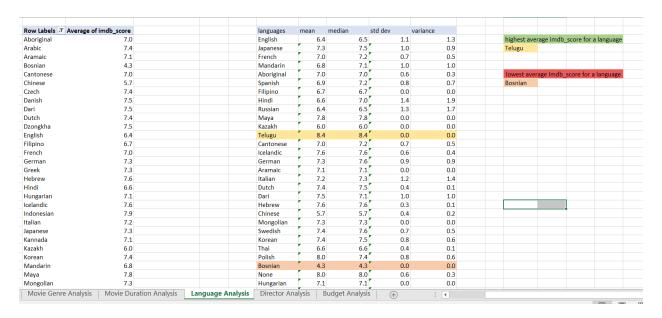


Movie Duration Analysis showed a weak positive correlation between movie duration and IMDB ratings. While longer movies may have slightly higher ratings on average, the relationship is not significant.

Language Analysis:

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.

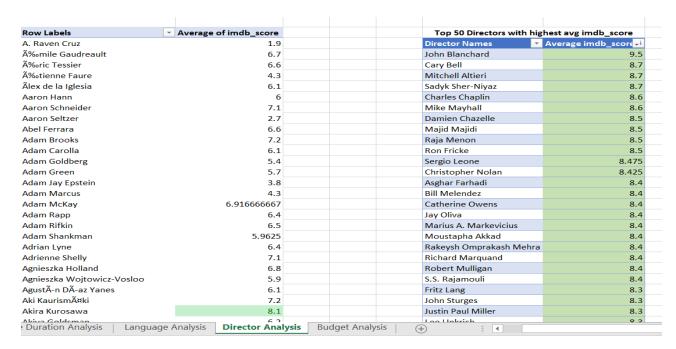


It is indicated that movies in 'Telugu' tend to have higher average IMDB ratings compared to other languages. It also shows that the lowest average IMDB_score for a language is 4.3 for 'Bosnian'.

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.

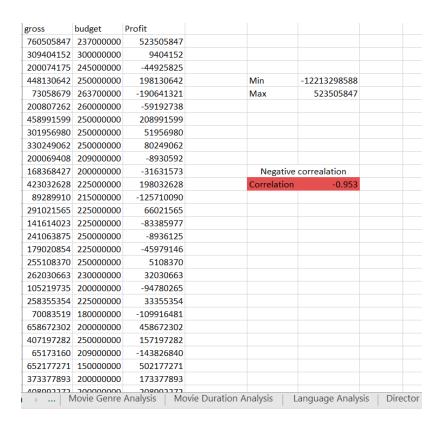


Identified top directors based on their average IMDB scores. Certain directors have a tendency to make films with higher reviews, which suggests that they have an impact on the box office success of films.

Budget Analysis:

Explore the relationship between movie budgets and their financial success.

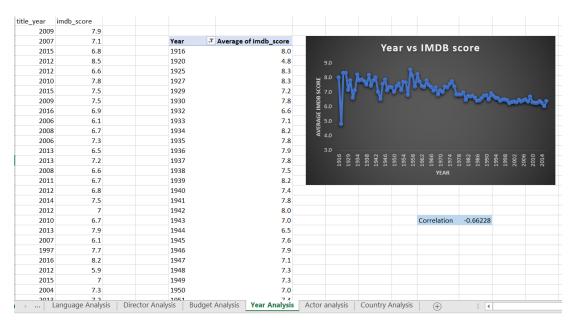
• Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.



It is revealed that there is a negative correlation between movie budgets and gross earnings. This implies that as movie budgets increase, gross earnings tend to decrease.

Year Analysis:

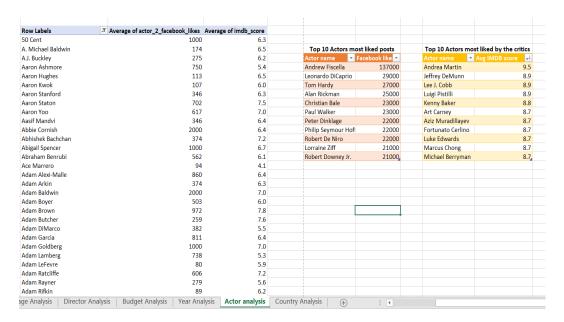
Exploring trends in movie ratings over the years based on the provided release year data.



A negative correlation of -0.66 between the year of release and IMDB scores suggests that as the release year increases, the average IMDB scores tend to decrease. This could imply several factors influencing movie ratings over time, such as changing audience preferences, evolving film industry standards, or increased competition

Actor Analysis:

Analyzing if there is any impact of actors or actresses on movie ratings based on Facebook likes.



Here, we got the Top 10 actors most liked posts by actors which also tells who are the general public favorite, and we also the Top 10 actors that are most liked by the critics.

Country Analysis:

Examining whether the country of origin affects movie ratings.

mdb_score					
7.9					
7.1	Country		ge of imdb_score		
6.8	Afghanis		7.4	Counti	ry with highest ratings
8.5	Argentin	3	7.5	Kyrgyzs	stan
6.6	Aruba		4.8		
7.8	Australia		6.5		
7.5	Bahamas		4.4	Count	ry with lowest ratings
7.5	Belgium		6.0	Bahama	as
6.9	Brazil		7.3	New Lir	ne
6.1	Bulgaria		6.1		
6.7	Cambodi	a	5.6		
7.3	Cameroo	n	7.5		
6.5	Canada		6.1		
7.2	Chile		6.9		
6.6	China		6.6		
6.7	Colombia	1	7.5		
6.8	Czech Re	public	7.0		
7.5	Denmark		7.2		
7	Dominica	ın Republic	6.9		
6.7	Egypt		8.1		
7.9	Finland		7.2		
6.1	France		6.7		
7.7	Georgia		5.6		
8.2	Germany	1	6.3		
5.9	Greece		7.0		
7	Hong Ko	ng	6.7		
7.3	Hungary	-	6.5		
7.2	Iceland		7.3		
6.5	India		6.5		
Language Analysis	Director Analysis	Budget Analysis	Year Analysis	Actor analysis	Country Analysis

We have found out that the country with highest ratings is Kyrgyzstan and the country with lowest ratings is Bahamas, New Line.

Results:

To sum up, our analysis reveals the complex relationship between movie success on IMDB. A variety of characteristics influences ratings; some are more obvious than others, such as director, genre, language, and budget; some associations are more subtle. Stakeholders in the film business can use this data to help them make well-informed decisions about upcoming projects.

Conclusion:

Through this project, I've gained valuable insights into the factors influencing movie ratings on IMDB. By analyzing various aspects such as genre, duration, language, director, and budget, I provided actionable insights that can inform decision-making for future projects. Our analysis contributes to a better understanding of the factors driving the success of movies on IMDB