

IMDb

Movies Analysis

By

Ravi Bhushan Singh

Description

The IMDB Movie Analysis project aims to uncover the key factors that influence a movie's success, with success defined by high IMDB ratings. Using a comprehensive dataset of IMDB movies, the project explores various aspects such as genre, duration, language, director influence, and budget to identify patterns and correlations that contribute to higher ratings. The analysis begins with data cleaning, where the dataset is prepared for detailed exploration by handling missing values, removing duplicates, and converting data types as needed. The core of the project involves performing a series of analyses using Excel, including descriptive statistics, correlation calculations, and visualizations to understand how different factors like genre, budget, and director reputation impact a movie's success. Key techniques such as the "Five Whys" approach are employed to dig deeper into the reasons behind observed patterns, providing a robust understanding of the underlying causes. The project culminates in a report that not only answers the specific questions posed but also offers actionable insights that can guide decision-making for movie producers, directors, and investors. The final deliverable includes a detailed report and a video presentation where the candidate explains their approach, showcases their findings using Excel, and highlights key insights gained from the analysis. The project is judged on the quality of data analytics, Excel proficiency, statistical understanding, and the ability to draw meaningful insights from the data.

Tech Stack Used

- MS Excel
- MS Power Point

Excel File Link

https://docs.google.com/spreadsheets/d/1ySx0BR2-U6L7Np662oNUS0oAw0b_7Ca3/edit?usp=drive_link&ouid=116403904215561153724&rtpof=true&sd=true

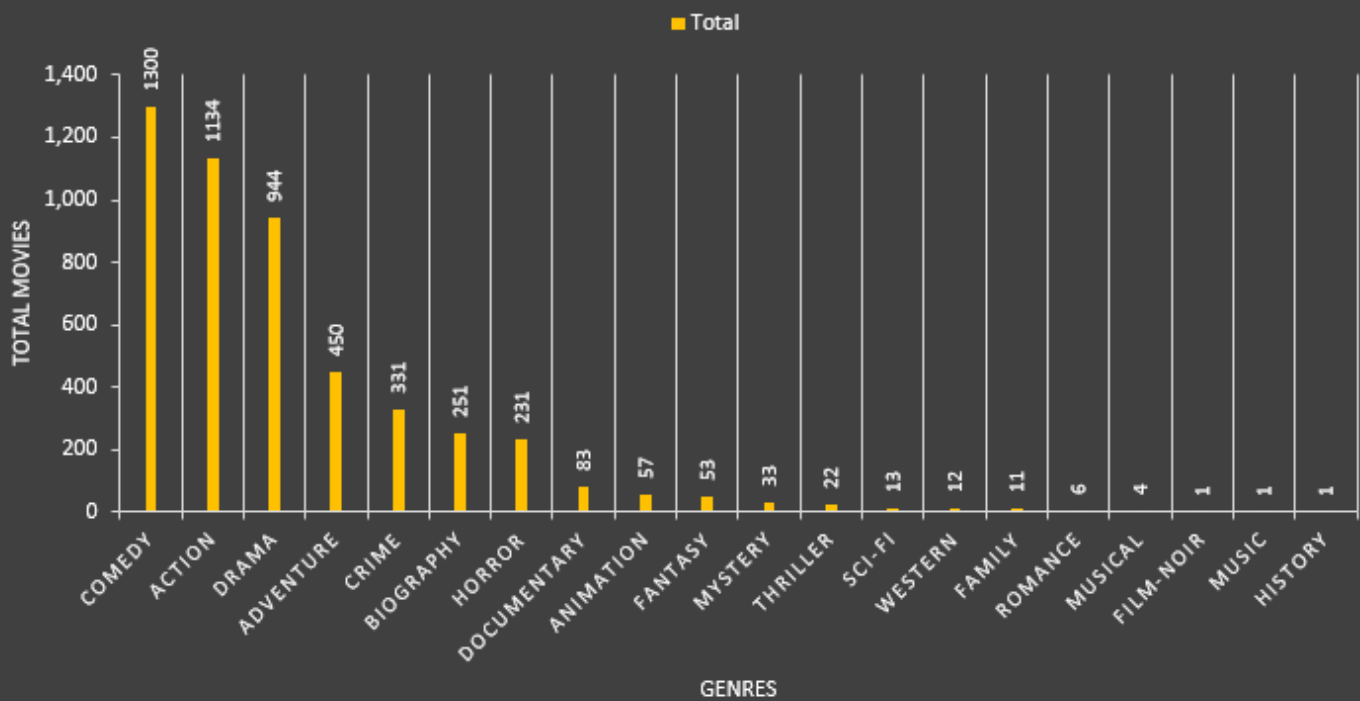
Movie Genre Analysis

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. Use Excel's COUNTIF function to count the number of movies for each genre. You might need to manipulate the 'genres' column to separate multiple genres for a single movie. Use Excel's functions like AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, and STDEV to calculate descriptive statistics. Compare the statistics to understand the impact of genre on movie ratings.

Ans.- The most common genres of movies are comedy, action, drama, adventure and crime etc.

Row Labels	Mean IMDB Score	Median(IMDB Score)	Mode	Max Score	Min Score	Var of imdb_scor	Range
Action	6.2	6.3	6.1	9.1	1.7	1.2	7.4
Adventure	6.5	6.7	6.7	8.6	2.3	1.2	6.3
Animation	6.5	6.9	7.1	8.4	3.7	1.4	4.7
Biography	7.2	7.2	7	8.9	4.5	0.5	4.4
Comedy	6.2	6.3	6.4	9.5	1.9	1.1	7.6
Crime	6.8	6.9	7.4	9.3	3.1	0.9	6.2
Documentary	7.2	7.5	7.5	8.7	1.6	1.4	7.1
Drama	6.7	6.9	6.7	8.8	2	1.0	6.8
Family	5.7	5.7	5.7	8.6	2.8	3.9	5.8
Fantasy	6.4	6.6	7.4	7.9	4.3	0.8	3.6
Film-Noir	7.6	7.6	7.6	7.6	7.6	NA	0
History	7.5	7.5	7.5	7.5	7.5	NA	0
Horror	5.7	5.7	5.9	8.5	2.2	1.3	6.3
Music	7.2	7.2	7.2	7.2	7.2	NA	0
Musical	6.0	6.7	N/A	7.2	3.4	3.2	3.8
Mystery	6.6	6.7	6.6	8.5	3.2	1.3	5.3
Romance	5.9	5.7	5.1	7.1	5.1	0.8	2
Sci-Fi	6.0	6.1	6.1	8.2	2.8	2.2	5.4
Thriller	5.6	5.5	5.8	8.1	3.4	1.7	4.7
Western	6.6	7.1	7.1	8.9	3.8	2.6	5.1

GENRE WISE DISTRIBUTION



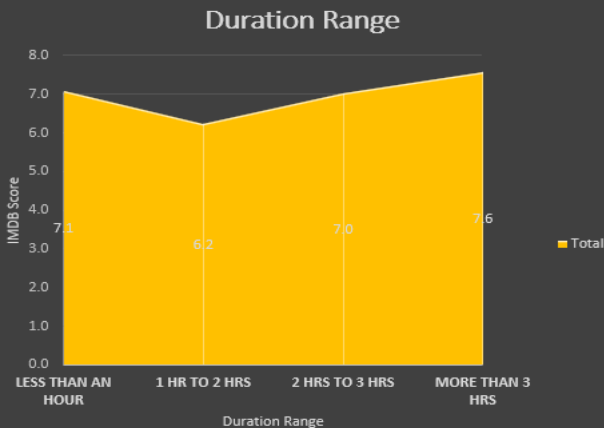
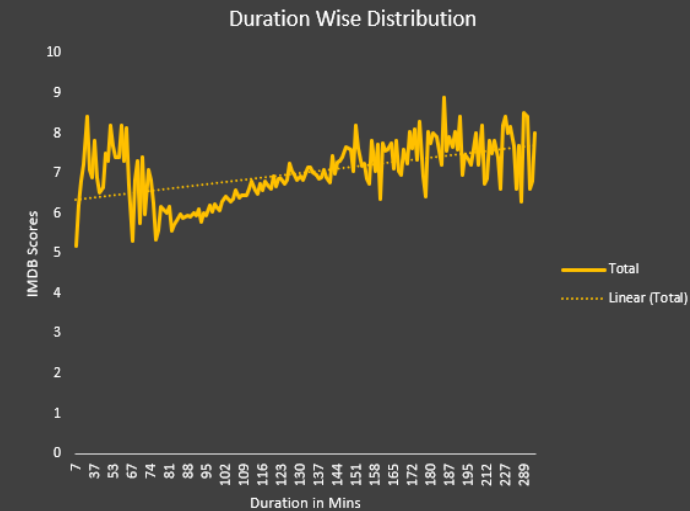
Movie Duration Analysis

Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score. Calculate descriptive statistics such as mean, median, and standard deviation for movie durations. Use Excel's functions like AVERAGE, MEDIAN, and STDEV. Create a scatter plot to visualize the relationship between movie duration and IMDB score. Add a trendline to assess the direction and strength of the relationship.

Ans.- From the below chart we can see that the movies of 2 to 3 hours and more than three hours tends to have more IMDB score than movies of duration less than 2 hours, we can also observe that as duration of the movie increases the IMDB score also increases.

Column1	Column2
Mean	108
Meadian	104
Mode	90
Standard Deviation	23

Duration Rang	Average of imdb_scor	Median	Mode	Var of imdb_scor	StdDev of imdb_scor	Max of imdb_scor	Min of imdb_scor	Range
Less than an hour	7.1	7.2	7.2	0.6	0.8	8.4	5.2	3.2
1 Hr to 2 Hrs	6.2	6.4	6.7	1.2	1.1	9.5	1.6	7.9
2 Hrs to 3 Hrs	7.0	7.1	6.7	0.8	0.9	9.3	2.2	7.1
More Than 3 Hrs	7.6	7.7	7.2	0.6	0.8	9	5.5	3.5

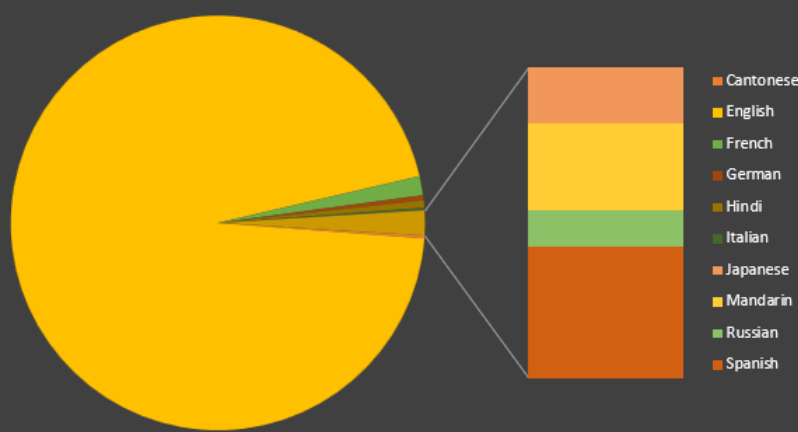


Language Analysis

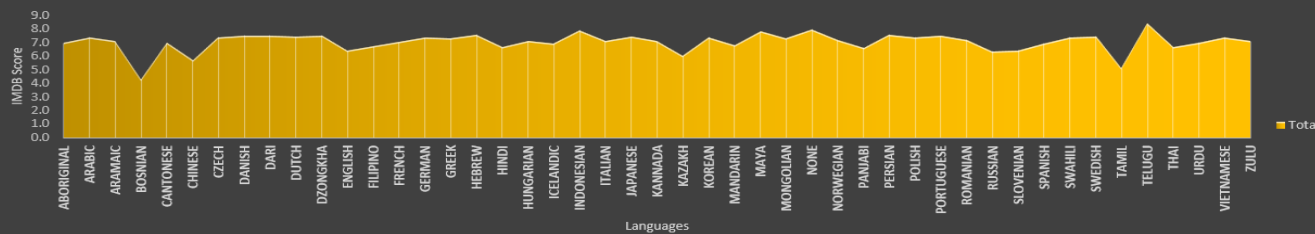
Examine the distribution of movies based on their language. Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics. Use Excel's COUNTIF function to count the number of movies for each language. Calculate the mean, median, and standard deviation of the IMDB scores for each language. Compare the statistics to understand the impact of language on movie ratings.

Ans.- The most common languages used in movies are English, French, Hindi and Manderin etc.

Top 10 Languages



Language's impact on IMDB Score



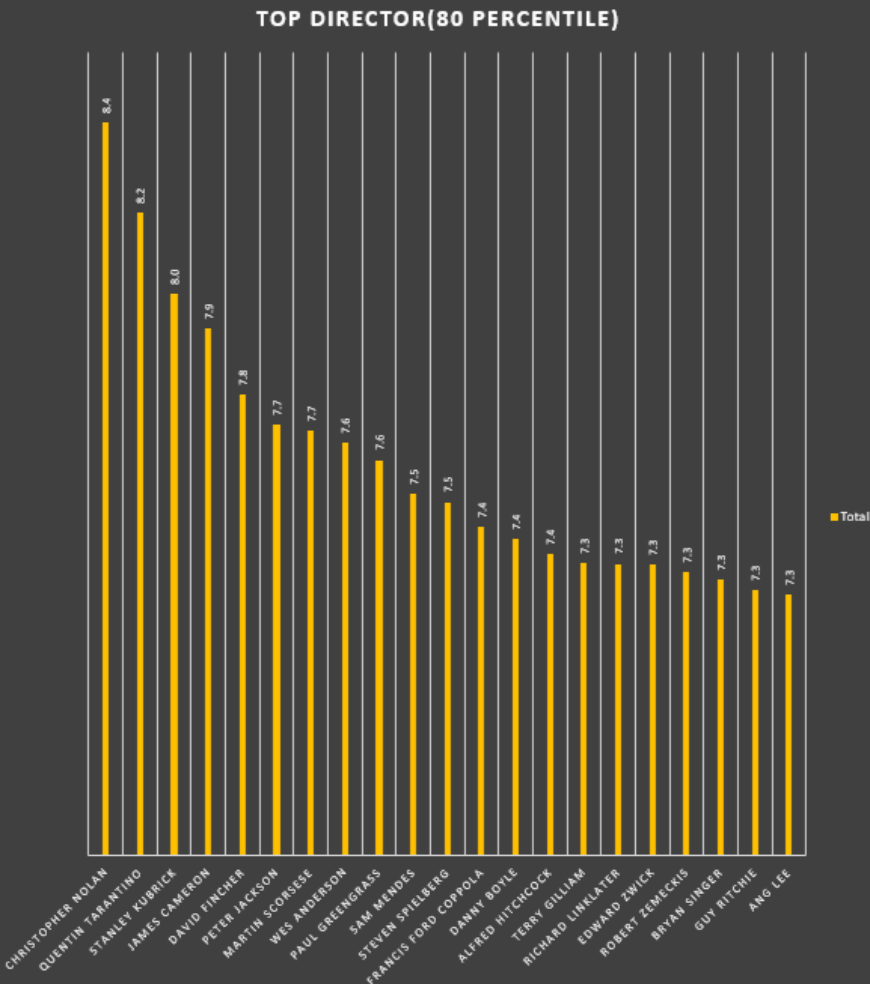
Row Label	Total Movies	Mean Scor	Median	Mode	Max of imdb_score	Min of imdb_score	Range	Variance	Standard Deviation
Cantonese	11	7.0	7.2	7.3	7.8	5.3	2.5	0.5	0.7
English	4619	6.4	6.5	6.7	9.5	1.6	7.9	1.2	1.1
French	72	7.0	7.2	7.2	8.4	4.9	3.5	0.5	0.7
German	19	7.3	7.6	7.8	8.5	4.9	3.6	0.9	1.0
Hindi	28	6.6	7.0	7.8	8.5	2.8	5.7	2.0	1.4
Italian	10	7.1	7.2	NA	8.9	5.1	3.8	1.5	1.2
Japanese	17	7.4	7.7	8.2	8.7	5.6	3.1	1.0	1.0
Mandarin	26	6.8	7.1	7.9	7.9	3.2	4.7	1.1	1.0
Russian	11	6.4	6.5	5.3	8.1	4.1	4	1.9	1.4
Spanish	40	6.9	7.2	7.2	8.2	4.4	3.8	0.7	0.9

Director Analysis

Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculation. Calculate the average IMDB score for each director. Use Excel's PERCENTILE function to identify the directors with the highest scores. Compare the scores of these directors to the overall distribution of scores.

Ans.- Top directors based on their average IMDB score in 90 percentile with Average IMDB Score greater than or equal to 7.7 are Christopher Nolan(8.4), Quentin Tarantino(8.2), Stanley Kubrick(8.0), James Cameron(7.9), David Fincher(7.8), Peter Jackson(7.7), Martin Scorsese(7.7). Top 80 Percentile(with Average IMDB Score greater than or equal to 7.3) are Mention in the below chart.

Approach – To get the Top directors first I grouped the rows based on the movie count in a separate sheet and above of that I grouped them on the basis of average IMDB score so we can Primarily get those director who had done average(more than seven movies) amount of movies with good IMDB score, so I calculated top 150 based on movie count and then sorted this on the basis of average IMDB score, and on The raw data I calculated 90 and 80 Percentile which was 7.7 and 7.3 respectively.



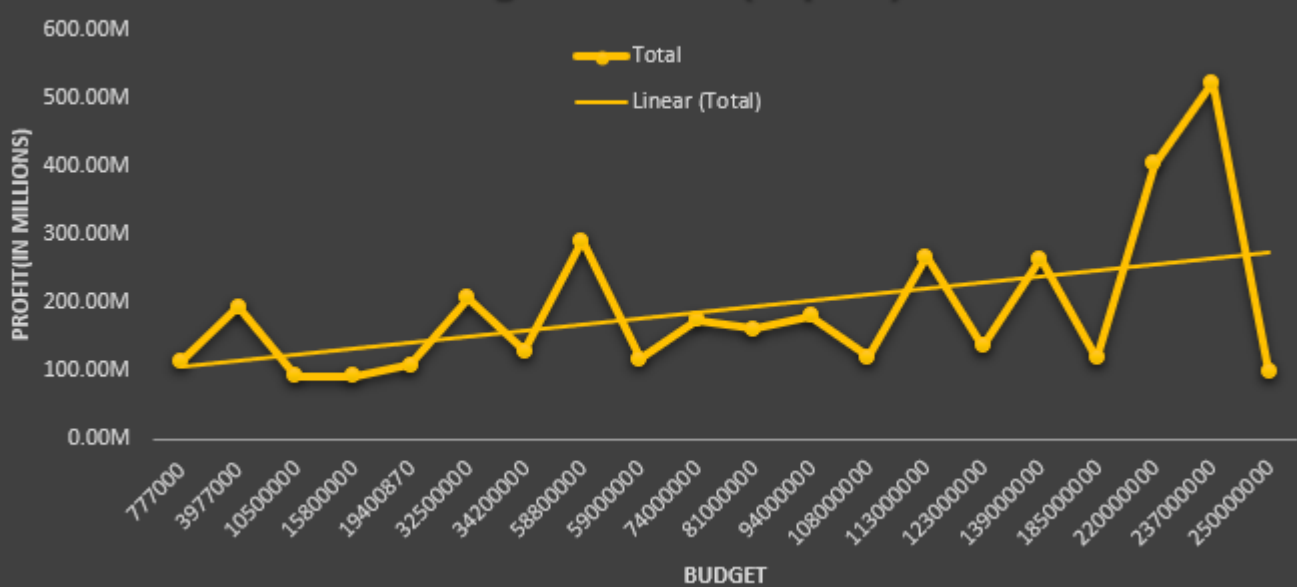
E. Budget Analysis

Explore the relationship between movie budgets and their financial success. Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin. Calculate the correlation coefficient between movie budgets and gross earnings using Excel's CORREL function. Calculate the profit margin (gross earnings - budget) for each movie and identify the movies with the highest profit margin using Excel's MAX function.

Ans.- The correlation between correlation between movie budgets and gross earnings is 0.111235. movies with highest profit are Avatar, Jurassic World, Titanic, Star Wars: Episode IV – A New Hope etc. Top 10 are mentioned in the below table.

Movie Name	Profit(in Millions)
Avatar	523.5
Jurassic World	502.2
Titanic	458.7
Star Wars: Episode IV - A New Hope	449.9
E.T. the Extra-Terrestrial	424.4
The Avengers	403.3
The Lion King	377.8
Star Wars: Episode I - The Phantom Menace	359.5
The Dark Knight	348.3
The Hunger Games	330.0

Budget vs Profit(Top 20)



Insights

One of the key insights from this IMDB movie analysis is that genre plays a significant role in influencing a movie's success, as measured by its IMDB rating. The analysis revealed that certain genres, such as drama and adventure, tend to have higher average ratings compared to genres like horror or comedy. This trend suggests that audiences might value the emotional depth and storytelling complexity often found in drama and adventure films, leading to higher ratings. Another important finding is the impact of budget on financial success. Movies with higher budgets generally showed a positive correlation with gross earnings, indicating that bigger investments in production quality, special effects, and star power can lead to higher financial returns. However, the analysis also uncovered that high budget does not always guarantee the highest profit margin, as several lower-budget films achieved remarkable profitability by keeping costs low while still appealing to a broad audience. Lastly, the influence of directors was evident, with certain directors consistently producing highly rated films. These directors may have developed a unique style or expertise that resonates well with audiences, contributing to the overall success of their movies. Understanding these factors can help producers, directors, and investors make more informed decisions in the highly competitive film industry.

1. Why do drama and adventure genres tend to have higher IMDB ratings?

These genres often focus on complex storytelling and emotional depth, which resonate well with audiences.

2. Why does complex storytelling and emotional depth resonate well with audiences?

Such narratives engage viewers on a deeper level, allowing them to connect with characters and the story more personally.

3. Why does this personal connection lead to higher IMDB ratings?

When viewers feel emotionally invested in a movie, they are more likely to have a positive experience and, consequently, rate the movie higher.

4. Why does a positive viewing experience result in higher ratings?

A positive experience encourages viewers to share their satisfaction with others, leading to more favorable reviews and ratings.

5. Why do favorable reviews and ratings matter for a movie's success?

High ratings and positive reviews attract more viewers, increase the movie's visibility, and ultimately contribute to its overall success, both critically and commercially.