

IMDB MOVIE ANALYSIS

Description:

This project explores a big dataset of movies and their details. We will be looking at different things like what kind of movies are most popular (genres), how long they are (duration), what languages they are in, and who directs them. We will also see if there is a connection between how much a movie costs to make (budget) and how much money it earns, and how these factors might affect a movie's IMDB rating.

By digging into this data, we hope to understand what makes a movie successful on IMDB and learn some interesting things about the movie world!

Approach:

The project involved a comprehensive analysis of the IMDB movie dataset. Data cleaning and preprocessing were conducted to prepare the data for exploration. Key variables such as genre, duration, language, director, and budget were examined to understand their impact on movie success, as measured by IMDB ratings. Statistical analysis and data visualization techniques were employed to derive meaningful insights and identify patterns within the data.

Tech-Stack Used: Microsoft Excel 2022

Insights:

A. Movie Genre Analysis

- **Genre Popularity:** Identified the most prevalent movie genres, providing insights into audience preferences.
- **Genre Impact on Ratings:** Analyzed the correlation between genre and IMDB scores, revealing genres with higher average ratings and those with more consistent performance.

B. Movie Duration Analysis

- **Duration Distribution:** Determined the typical movie length and its variation.
- **Duration-Rating Relationship:** Explored the correlation between movie duration and IMDB scores, identifying if longer or shorter movies tend to be rated higher.

C. Language Analysis

- **Language Prevalence:** Identified the most commonly used languages in movies.

- **Language and Ratings:** Analyzed the impact of language on movie ratings, uncovering potential language-specific preferences or biases.

D. Director Analysis

- **Director Performance:** Ranked directors based on average IMDB scores, identifying top-performing individuals.
- **Director Impact:** Assessed the contribution of directors to movie success using percentile analysis, highlighting those with consistently high-rated films.

E. Budget Analysis

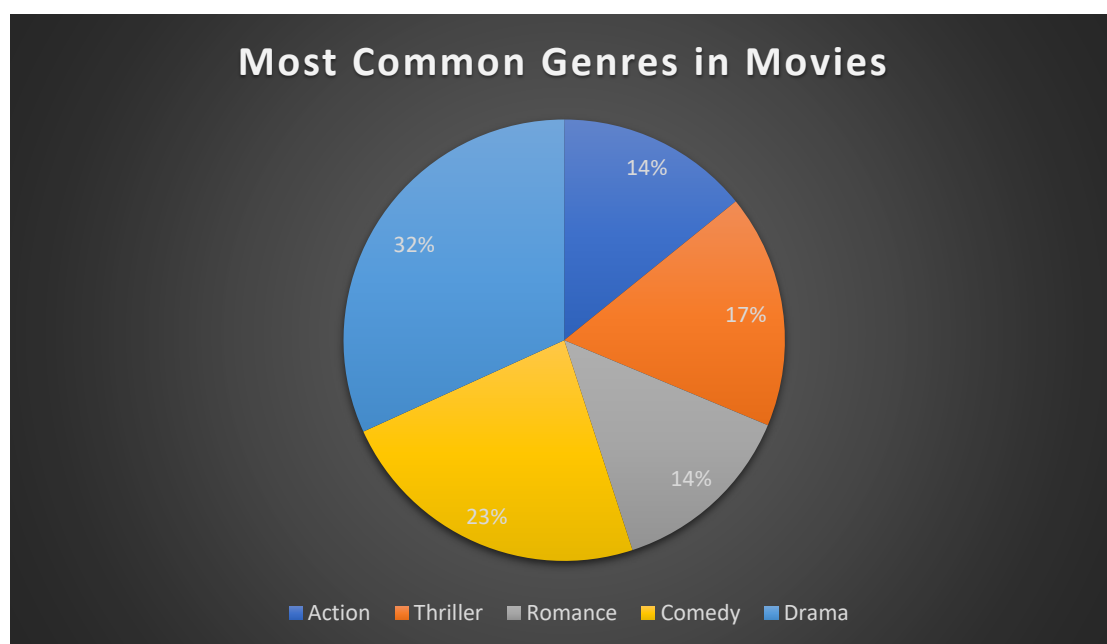
- **Budget-Revenue Correlation:** Determined the relationship between movie budgets and gross earnings.
- **Profitability:** Identified movies with the highest profit margins, revealing successful investment strategies.

These insights provide a foundation for understanding the factors influencing movie success on IMDB and can be used to inform future decision-making in the film industry.

Results:

A. Movie Genre Analysis:

Conclusion: After identifying most common genres in the dataset and analyzed their impact on IMDB scores. Action, Thriller, Romance, Comedy and Drama emerged as the most prevalent genres.

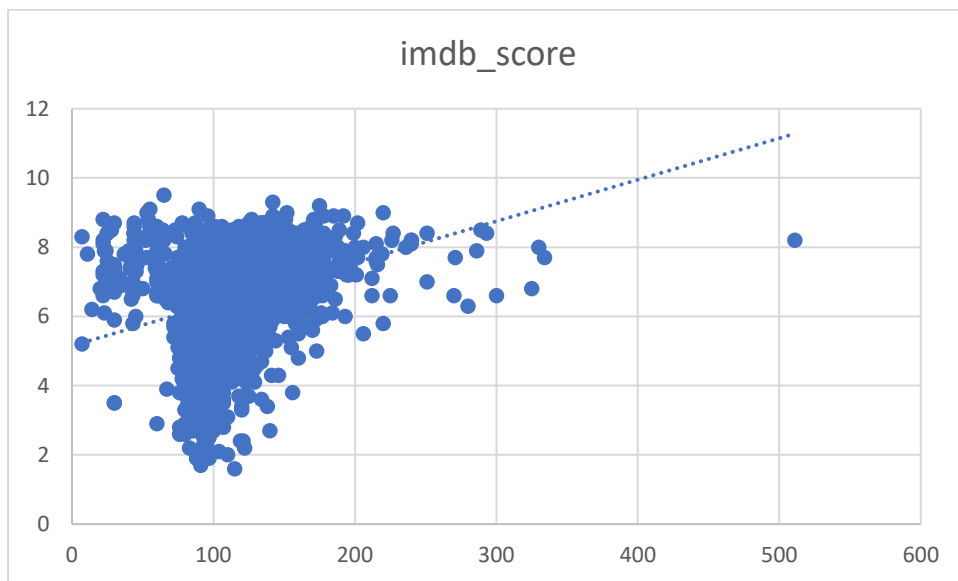


Genres Manipulation	
Mean	6.4408591
Median	6.6
Mode	6.7
Range	7.9
Variance	1.26365812
Standard Deviation	1.12412549

B. Movie Duration Analysis:

Conclusion: Movie Duration Analysis is varied widely, ranging from short films to lengthy epics. However, we observed a weak correlation between movie duration and IMDB score, suggesting that movie length alone does not significantly influence audience ratings.

From the below scatter plot there is a slight possibility to score 6 – 8.5 IMDB score if movie duration is more than 250 minutes. With the help of trend line, we can also predict IMDB score can increase with respect to duration of the movie.



Average of Movie Duration:	107.503
Median	104
Standard Deviation	25.25221

C) Language Analysis:

Conclusion: English emerged as the dominant language in the dataset, followed by French and German. While these languages represented a significant portion of the films analyzed, it's noteworthy that movies produced in languages other than English generally garnered lower average IMDB scores. This suggests a potential bias towards English-language films in terms of viewer ratings or the platform itself. However, further analysis considering factors like genre, release year, and regional audience preferences could provide more nuanced insights into the relationship between language and movie success.

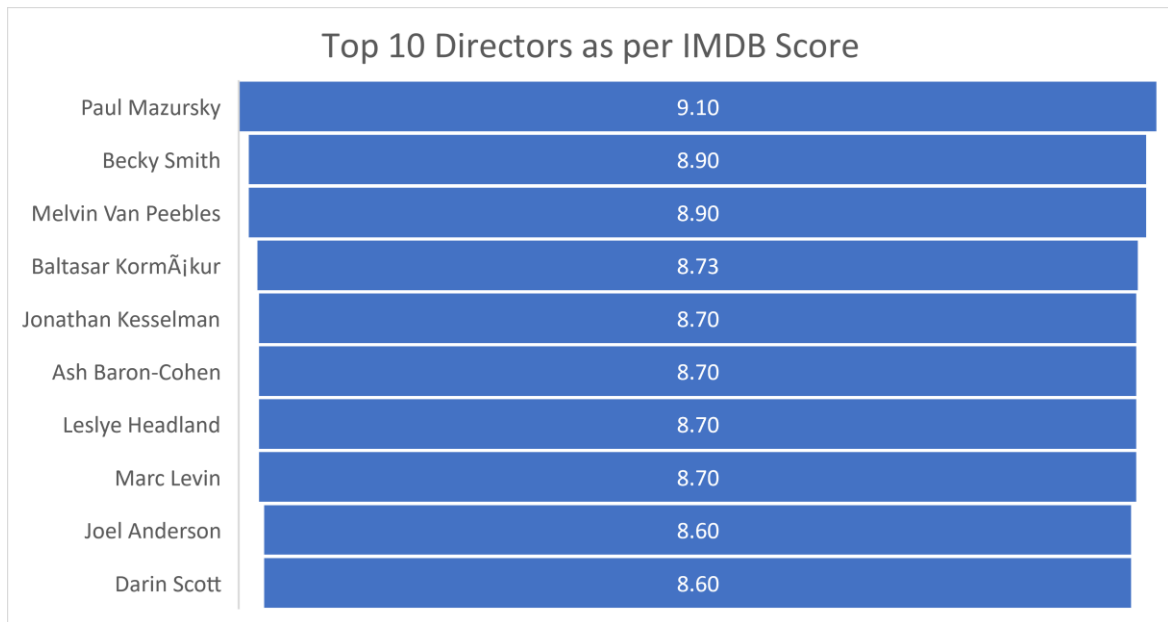
Top 10 common languages that are mentioned below:

language2	Count
English	4504
French	71
Spanish	39
Hindi	28
Mandarin	24
German	19
Japanese	16
Russian	11
Cantonese	11
Italian	11

D. Director Analysis:

Conclusion: The analysis of director performance revealed a clear correlation between directorial talent and movie success. Top-tier directors consistently achieved higher average IMDB scores, suggesting a significant influence on audience perception and critical acclaim. While this analysis highlights the impact of individual directors, further exploration is necessary to understand the interplay between directorial style, genre, and other factors contributing to overall movie success.

Below chart shows the top 10 directors by their IMDB score:



E. Budget Analysis:

Conclusion: The analysis of movie budgets and gross earnings revealed a strong positive correlation, indicating that higher budgets often lead to higher box office returns. However, the relationship between budget and profit margin was less straightforward. While some high-budget films generated substantial profits, others yielded lower returns, emphasizing the importance of effective budget allocation and risk management in the film industry.

CORRELATION

0.095144685

Max Profit Margin

523505847

The correlation of 0.095144685 is close to 0, suggesting a weak positive correlation. This means that there is a slight tendency for movies with higher budgets to have slightly higher gross earning but, the relationship is not strong.

Top 10 highest profit generated movies:

movie_title	gross	budget	Profit
Avatar	760505847	237000000	523505847
Pirates of the Caribbean: At World's End	309404152	300000000	9404152
Spectre	200074175	245000000	-44925825
The Dark Knight Rises	448130642	250000000	198130642
Star Wars: Episode VII - The Force Awakens	47621564	39315590	8305974
John Carter	73058679	263700000	-1.91E+08
Spider-Man 3	336530303	258000000	78530303
Tangled	200807262	260000000	-59192738
Avengers: Age of Ultron	458991599	250000000	208991599
Harry Potter and the Half-Blood Prince	301956980	250000000	51956980

For Excel Sheet Drive Link click [here](#)