

IMDB MOVIE ANALYSIS



PROJECT DESCRIPTION

Introduction

- The movie industry is highly competitive, with thousands of films released each year.
Understanding what drives a movie's success can help industry professionals make informed decisions.
- IMDB ratings serve as a key measure of audience reception and can influence a movie's popularity and financial success.

Purpose of the Project

This project aims to analyze an IMDB movie dataset to uncover patterns and trends that affect movie ratings.

By identifying key factors that contribute to higher ratings, stakeholders can optimize movie production and marketing strategies.

Key Focus Areas

Movie Genre: Examining which genres tend to receive higher ratings.

Movie Duration: Assessing whether longer or shorter movies perform better.

Language: Investigating how the language of a movie influences its ratings.

Director Influence: Identifying top-rated directors and their impact on success.

Budget & Financial Success: Exploring the relationship between movie budgets and profitability.

Expected Impact

Insights from this analysis can help:

- ✓ ☐ **Producers & Investors** – Make data-driven investment decisions.
- ✓ ☐ **Directors & Filmmakers** – Understand audience preferences to create better content.
- ✓ ☐ **Marketing Teams** – Tailor promotional strategies based on successful movie attributes.

APPROACH



Step 1: Data Collection

Used an IMDB dataset containing movie-related attributes such as genre, duration, language, director, budget, and ratings.

Identified **1,300 movies** with missing values that required additional data.



Step 2: Data Enrichment

Utilized **TMDB API** to scrape missing details for **600 movies** using Python scripts.

Filled in gaps for critical attributes like budget, revenue, and release details.



Step 3: Data Cleaning & Preparation

Removed duplicate records to ensure data integrity.

Converted data types (e.g., numerical fields for budget and ratings).

Standardized categorical values (e.g., genres and languages).



Step 4: Exploratory Data Analysis (EDA)

Used **Excel functions** (COUNTIF, AVERAGE, MEDIAN, STDEV, etc.) to compute descriptive statistics.

Created **visualizations** (scatter plots, bar charts, and correlation analysis) to uncover patterns



Step 5: Insights & Reporting

- Analyzed key relationships, such as **genre vs. IMDB rating** and **budget vs. financial success**.
- Compiled findings into a structured report and **presentation with visual storytelling**.

TECH STACK USED

Data Source



IMDB Dataset – Contains movie-related attributes such as genre, duration, language, director, budget, and ratings.



TMDB API – Used to scrape missing data for **600 movies** (out of 1,300 with missing values).



Python Libraries:

requests – To interact with the TMDB API.
csv– For reading and writing csv files.

Tools Used



Microsoft PowerPoint– For presenting insights



Python – For web scraping and data retrieval (Used TMDB API)



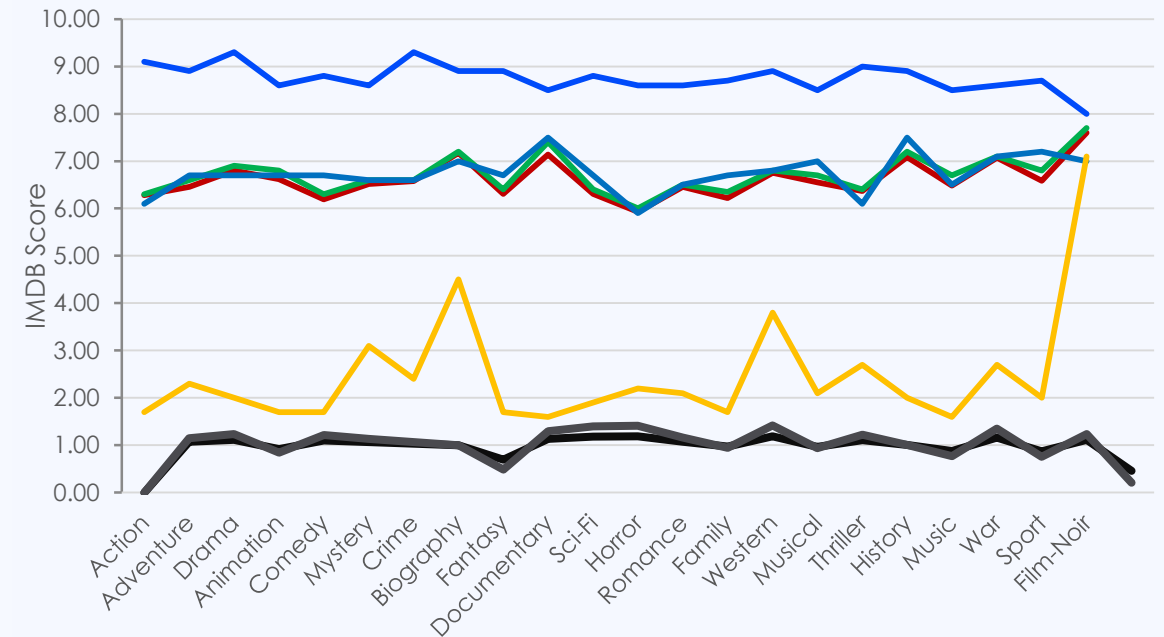
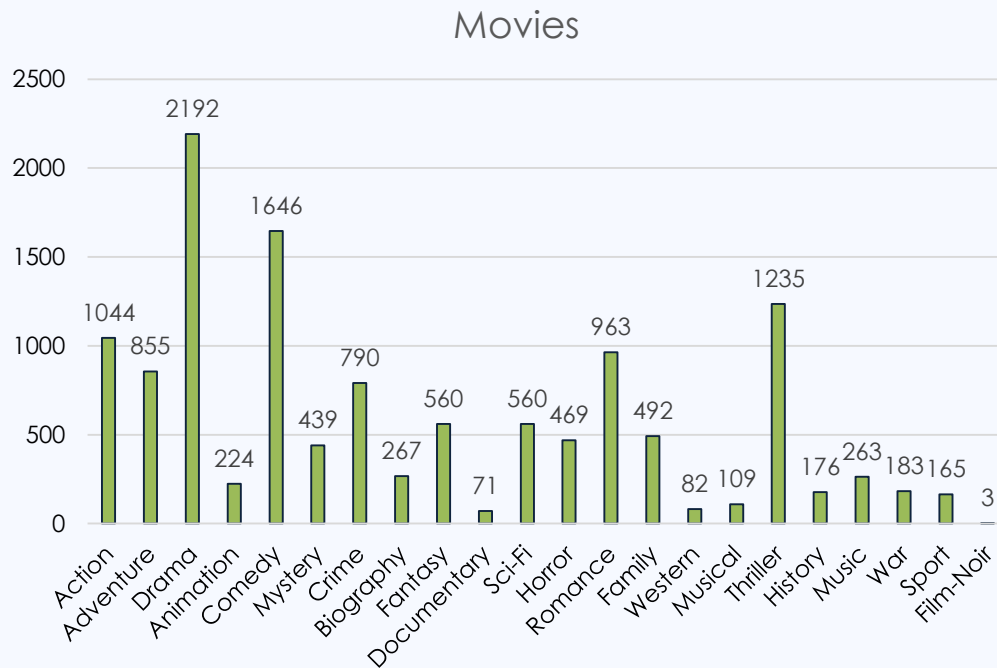
Excel – For data cleaning, exploratory data analysis (EDA), and visualization.

- Functions Used: COUNTIF, AVERAGE, MEDIAN, STDEV, CORREL, PERCENTILE, etc.
- Visualizations: Bar charts, scatter plots, trendlines, correlation analysis.

Excel Report:
[Link](#)

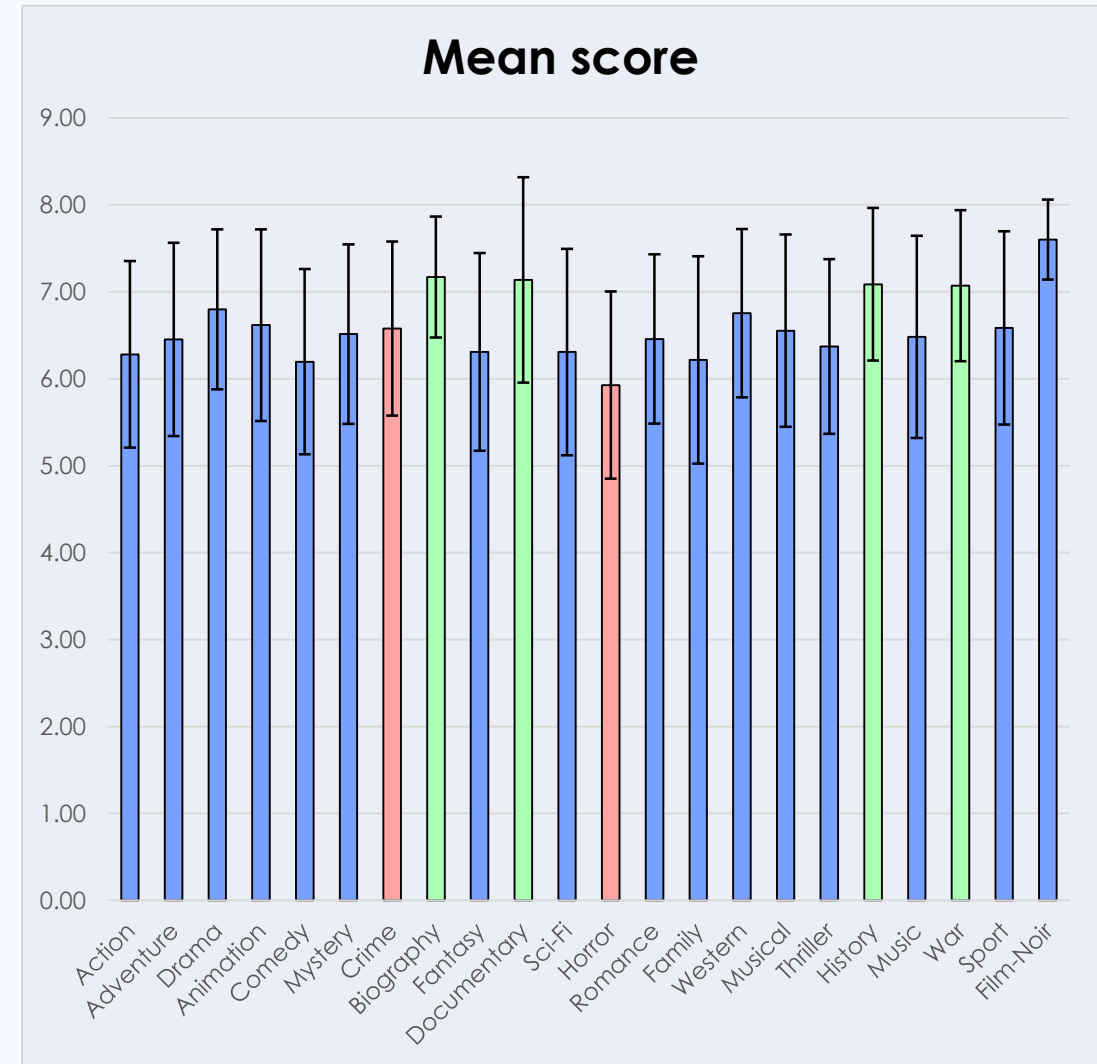
INSIGHTS

Movie Genre Analysis : The most common genres are Drama, Comedy, and Action. Biography, Documentary, and Film-Noir have the highest average IMDB scores, while Horror has the lowest. Variance and standard deviation indicate that Drama and Thriller have diverse ratings, whereas Biography and Documentary have more consistent scores. This suggests certain genres tend to receive higher ratings, possibly due to audience preference and storytelling depth..

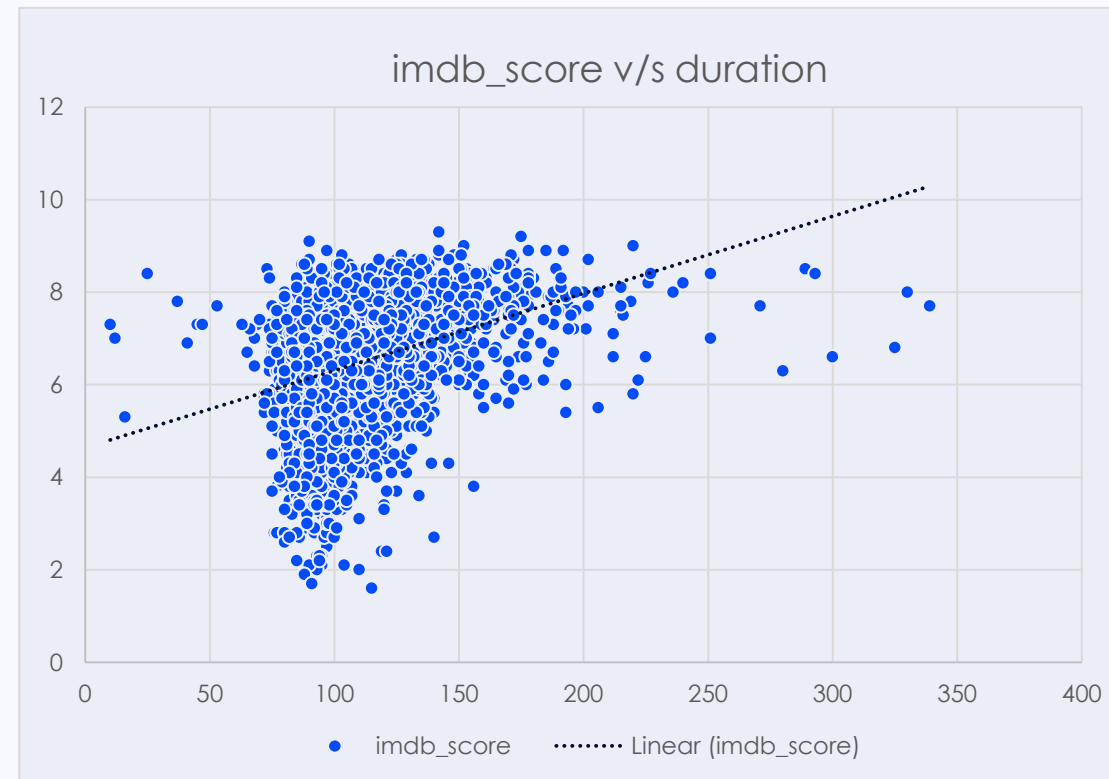


Key Findings:

- **Top-rated genres:** *Biography* (7.17), *Documentary* (7.14), *War* (7.07), and *History* (7.09) have the highest average ratings, suggesting audiences appreciate real-life stories and historical narratives.
- **Lowest-rated genres:** *Horror* (5.93) and *Comedy* (6.20) tend to have lower ratings, possibly due to subjective audience preferences.
- **Most consistent ratings:** *Biography* and *Film-Noir* have the **lowest standard deviations**, indicating less variation in ratings.
- **Most variable ratings:** *Thriller*, *Horror*, and *Sci-Fi* show higher variability, suggesting they are more hit-or-miss for audiences.



Movie Duration Analysis : The average movie duration is approximately 110 minutes, with a median of 105 minutes, indicating that half of the movies are shorter than this length. The mode, 101 minutes, suggests that this duration is the most common. A standard deviation of 22.96 minutes shows moderate variation in movie lengths. This distribution can help assess whether longer or shorter movies tend to receive higher IMDb ratings.



Key Findings:

Longer Movies Perform Better

Movies with **higher durations** generally receive **higher ratings**, suggesting that audiences may perceive longer films as more developed or high-quality.

Story Depth & Complexity

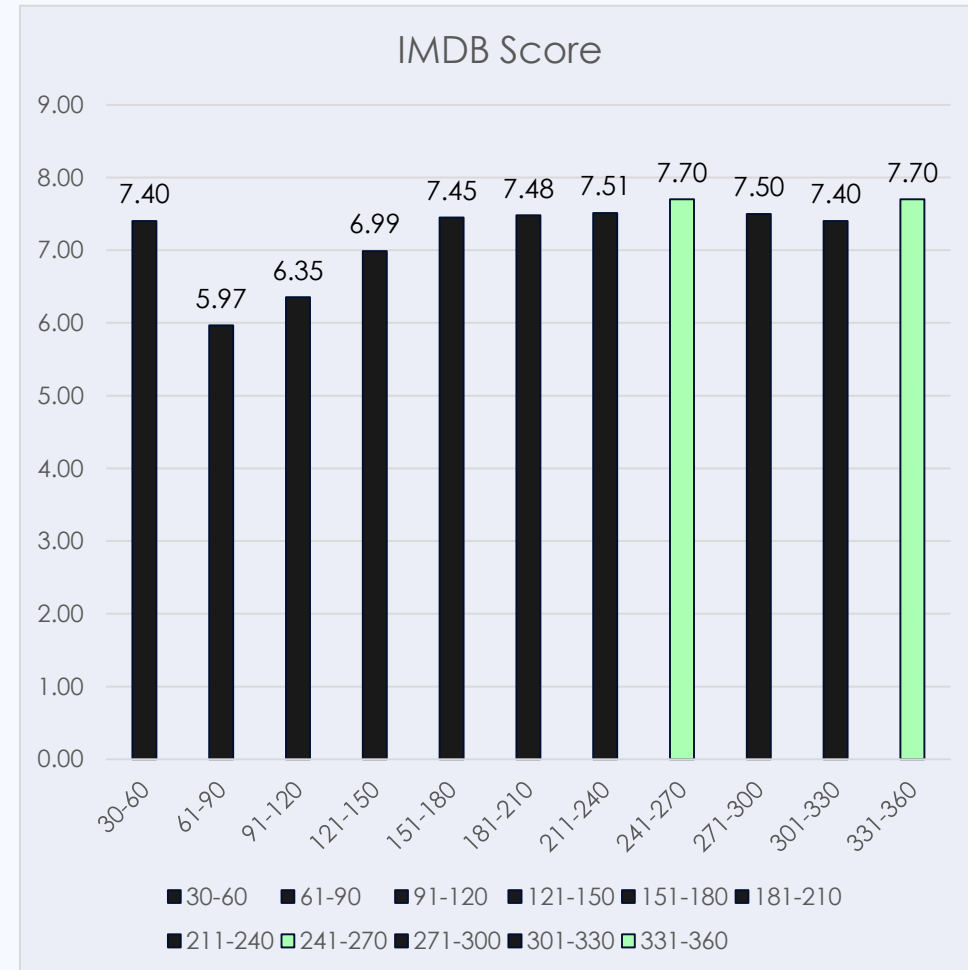
Longer movies often have more time to **develop characters, plots, and themes**, which can lead to **higher audience appreciation** and better ratings.

Director & Production Impact

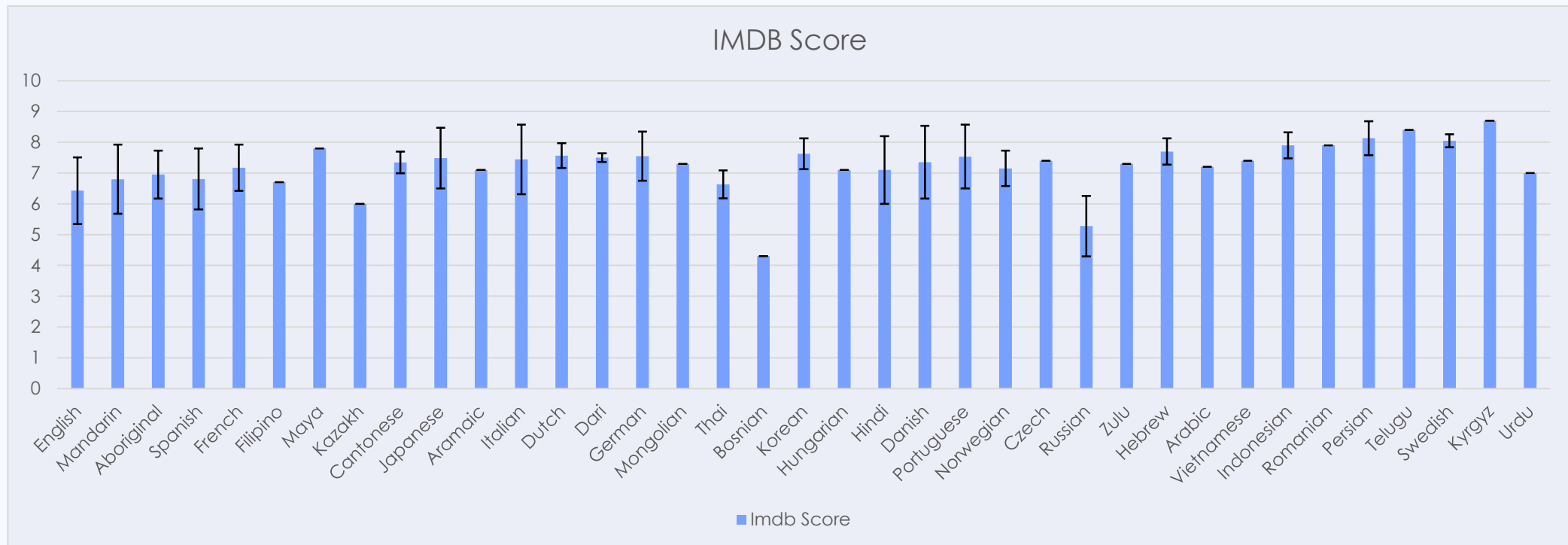
Established directors and studios tend to produce **longer, high-budget films**, which often receive **better critical and audience reception**.

Viewer Engagement vs. Fatigue

While longer movies perform better on average, **overly long durations** might lead to diminishing returns if the content isn't engaging.



Language Analysis : The dataset shows a strong dominance of English-language movies (4,088 films), but their average IMDB score (6.43) is lower compared to some less common languages. Movies in Persian (8.13), Telugu (8.4), and Kyrgyz (8.7) have the highest average scores, though these categories have very few entries, which may lead to sampling bias. Languages like French (7.17), German (7.55), and Japanese (7.49) also exhibit relatively high scores. This suggests that while English movies are the most prevalent, films in other languages often receive higher ratings, potentially due to niche audience appreciation or cultural factors.



Key Findings:

Dominance of Certain Languages

A few languages (likely **English, Hindi, Spanish, etc.**) dominate the dataset, reflecting global film industry trends.

Impact on IMDB Ratings

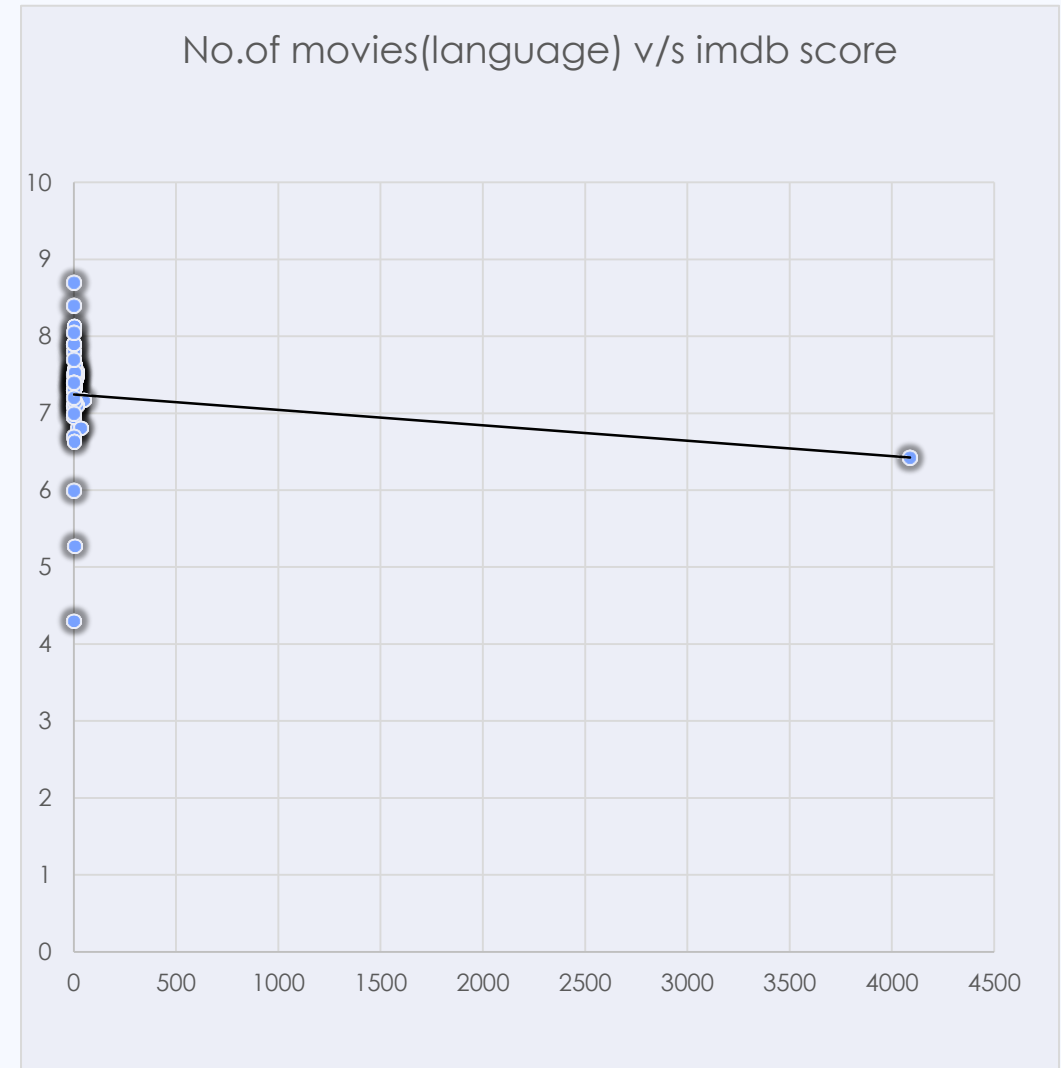
Languages with **higher average ratings** may indicate a preference for certain storytelling styles or production qualities associated with those languages.

Potential Bias

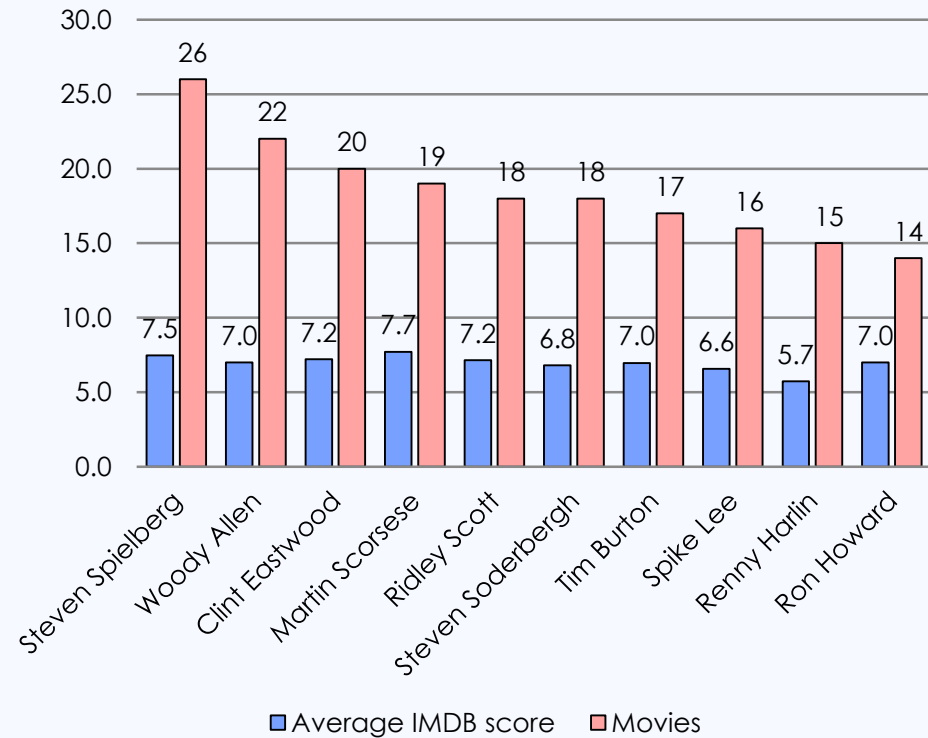
Since IMDB ratings are user-driven, biases may exist where **English-language films receive more ratings and higher visibility** compared to non-English films.

Cultural Influence

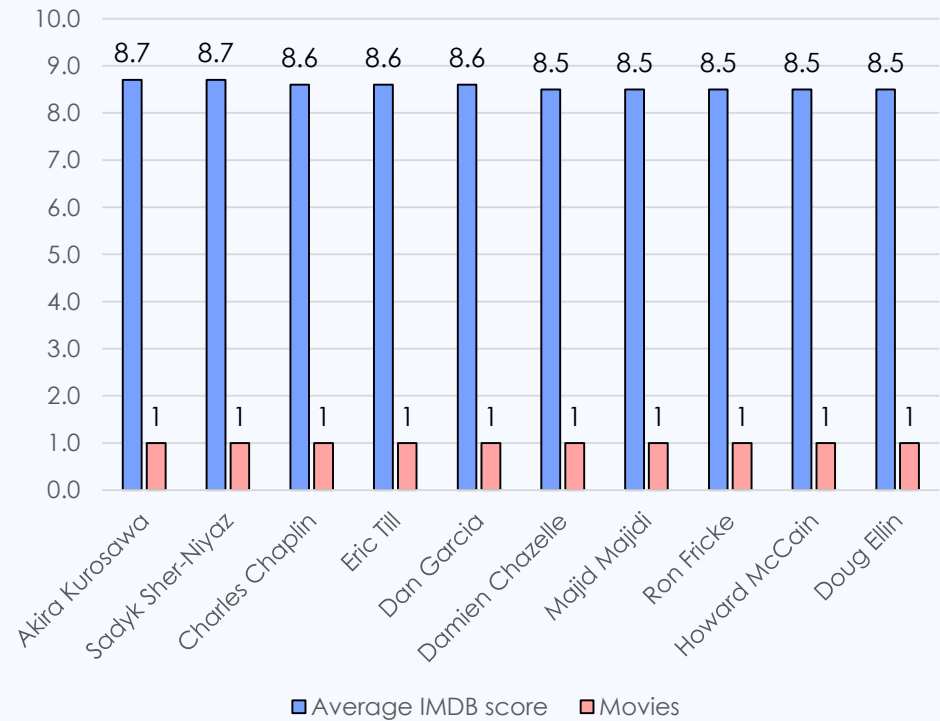
Highly rated films in different languages may highlight **regional storytelling strengths**, such as **emotional depth in dramas** or **action-heavy blockbusters**.



Director Analysis : Top directors are identified by their average IMDB scores, with percentile analysis highlighting the highest performers. Consistently high-rated directors indicate strong influence, though biases like limited filmography and genre preferences may impact results.



Based on IMDB score and no.of movies made



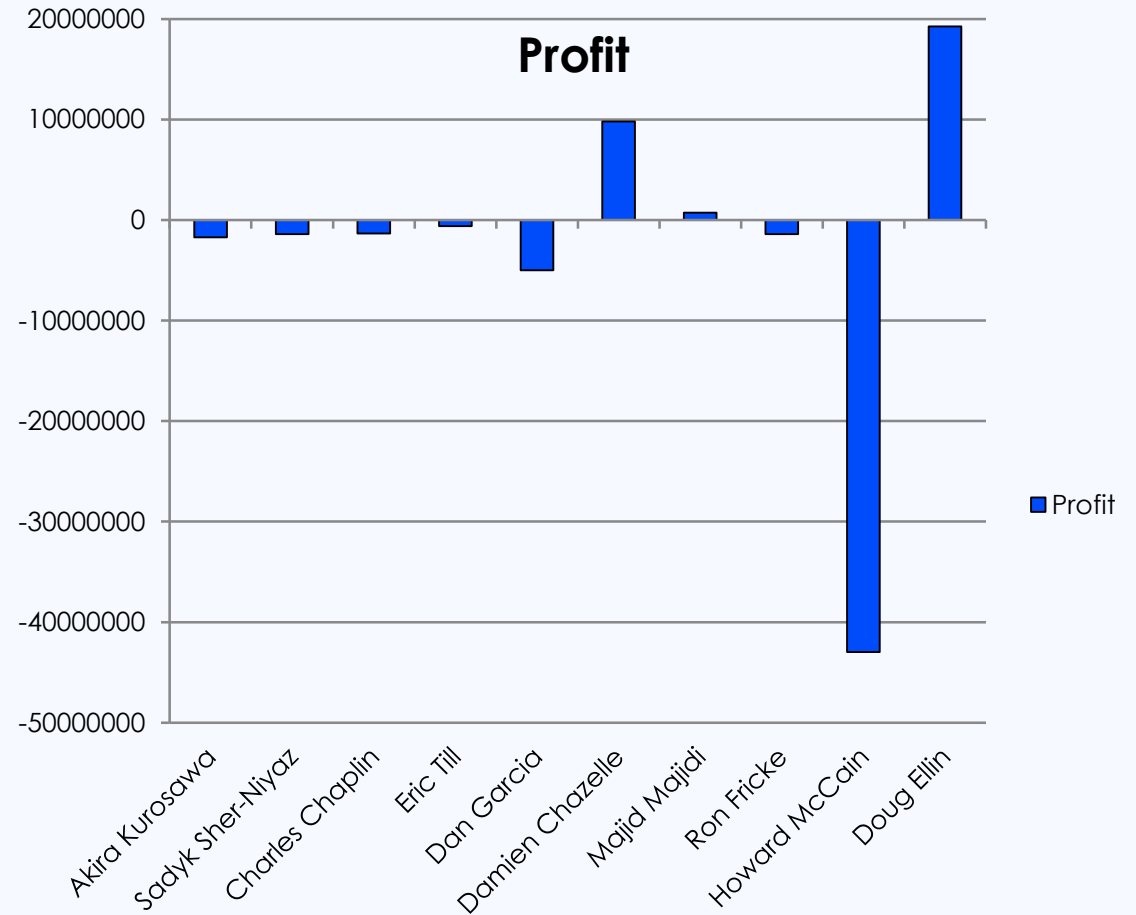
Based on IMDB score

Key Findings:

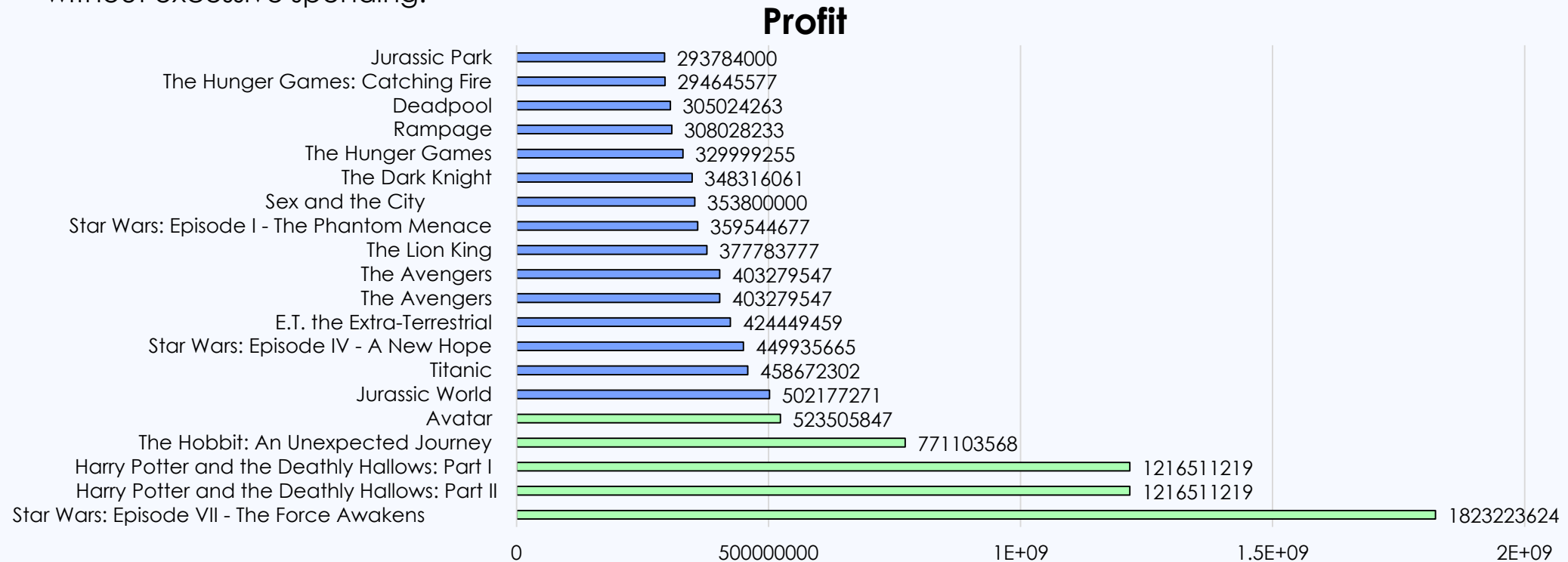
Consistent Ratings: Directors with consistently high ratings indicate a strong influence on movie success.

Quality Filmmaking: Directors in the highest percentiles exhibit a reputation for quality filmmaking.

Potential Biases: Biases such as limited filmography and genre preferences seem to affect results.



Profit Analysis : The budget analysis reveals a weak positive correlation (0.1027) between movie budgets and profits, suggesting that while higher budgets may contribute to financial success, other factors also play a crucial role. The top-grossing films, such as *Star Wars: The Force Awakens* and *Harry Potter and the Deathly Hallows*, achieved massive profits, reinforcing the trend that franchises and established brands tend to perform well. However, the presence of mid-budget films like *Deadpool* and *The Hunger Games* with substantial earnings indicates that strategic marketing, audience engagement, and unique storytelling can lead to profitability, even without excessive spending.



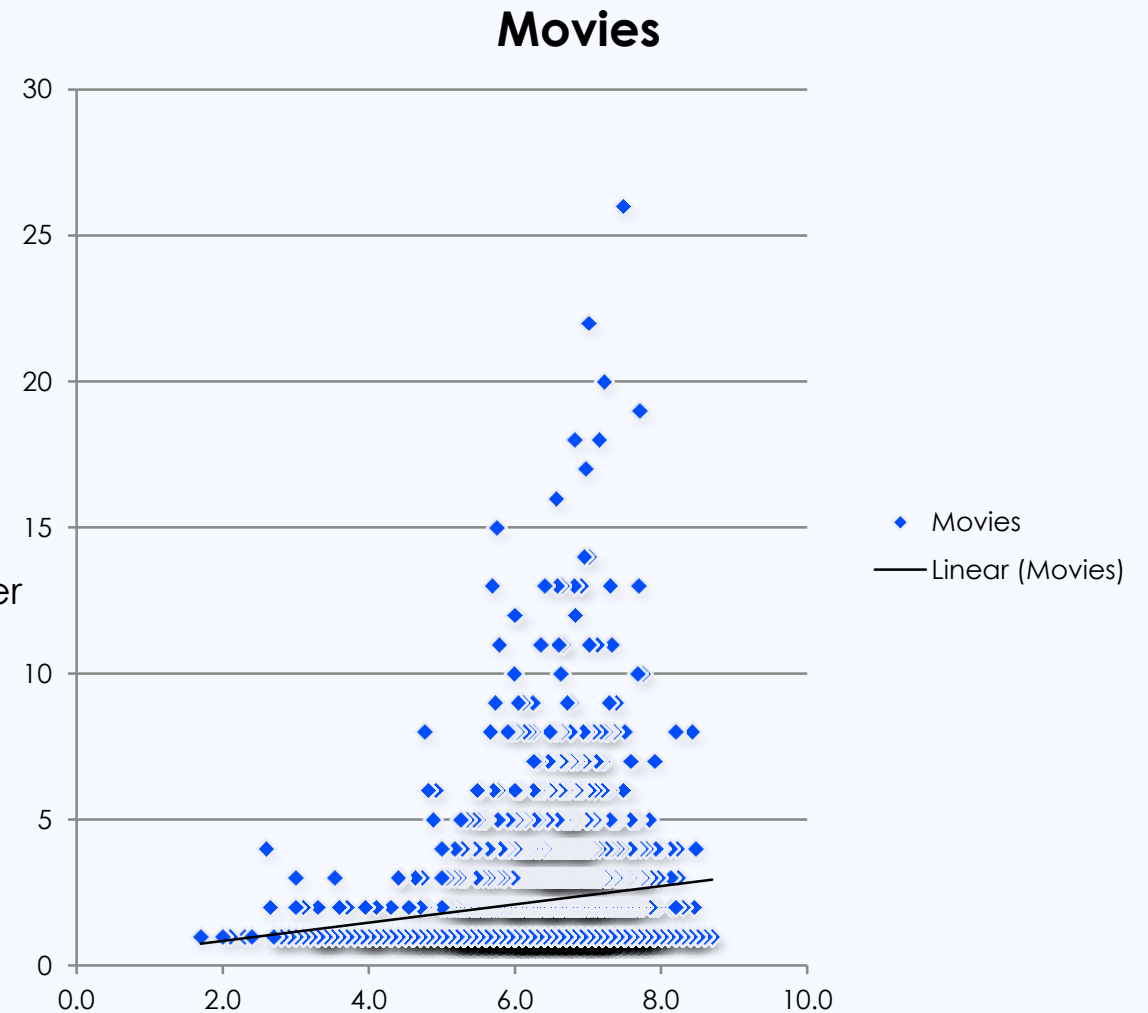
Key Findings:

Weak Positive Correlation: The correlation coefficient (0.1027) suggests that while higher budgets may contribute to financial success, they are not the sole determinant of profitability.

Franchise Dominance: Top-grossing movies are largely from well-established franchises (*Star Wars*, *Harry Potter*, *The Avengers*), indicating brand recognition significantly impacts financial success.

Mid-Budget Success: Movies like *Deadpool* and *The Hunger Games* achieved high profits despite having lower budgets, highlighting the role of audience engagement and strategic marketing.

Other Influencing Factors: Storytelling, marketing strategies, and audience reception likely play a more significant role in financial success than just budget size.



RESULT

Genre Influence

Some genres tend to have higher average IMDB scores.
Variability in ratings suggests storytelling and execution are crucial.

Movie Duration & Ratings

Weak positive correlation between duration and IMDB score.
Longer movies slightly tend to have better ratings, but extreme lengths may not always be favorable.

Language & Ratings

English movies dominate in quantity but not necessarily in quality.
Several non-English movies have higher average IMDB scores, indicating strong storytelling appeal.

Director Influence

Top directors consistently produce highly rated movies.
Their success is affected by film genre, production quality, and audience expectations.

Budget & Profitability

Weak positive correlation (0.1027) between budget and profit.
High budgets don't guarantee success; brand value and audience engagement matter more.

Key Takeaways

No single factor determines success—multiple elements contribute.
Quality storytelling, audience connection, and branding are key drivers.
Understanding these insights can help predict and enhance movie performance.

THANK YOU

