IMDB Movie Analysis

Project Description:

This project aims to explore and analyse the factors that contribute to the success of movies on IMDB, where success is defined by high IMDB ratings. Using a comprehensive dataset of movies, we investigate how various attributions – such as genre, budget, runtime, director, cast, release year, and more – influence audience ratings and overall movie performance.

Through this analysis, we intend to uncover trends and patterns that can help stakeholders in the film industry – such as producers, directors, and investors – make data-driven decisions to enhance the quality and success of their future projects. By identifying the key drivers of high ratings, this project also offers insights into audience preferences and changing dynamics in the entertainment industry.

Key Objective:

- Understand the distribution and trends of IMDB movie ratings.
- Identify relationships between movie features (e.g., genre, budget, cast) and IMDB scores.
- Determine which factors are most strongly correlated with high movie ratings.
- Provide actionable insights and recommendations for the movies industry.

Movie Genre Analysis:

Most Common Genres

From the **pivot table** showing genre counts:

- **Top 5 most common genres:**
    1. **Drama** – 1941 movies
    2. **Comedy** – 1503 movies
    3. **Thriller** – 1117 movies
    4. **Action** – 962 movies
    5. **Romance** – 877 movies

These genres dominate the dataset and are likely the most popular or frequently produced.

Descriptive Statistics Analysis of IMDB Scores

Let's break it down by **popularity (count) and quality (IMDB stats).**

Highest Average IMDB Scores (Top Quality Genres)

Genre	Mean IMDB Score
Film-Noir	7.70 (only 1 movie)
Biography	7.14
History	7.14
War	7.05
Documentary	6.99

- These genres, although **less frequent,** have **consistently higher quality content** based on ratings.
- **Film-Noir** ranks highest but has only 1 entry – so it lacks statistical strength.

Lowest Average IMDB Scores (Low Quality Genres)

Genre	Mean IMDB Score
Horror	5.93
Comedy	6.18
Family	6.21
Fantasy	6.29
Action	6.29

- **Horror** movies have the **lowest average rating,** indicating lower audience reception despite moderate representation.

Genre with Highest Variability (Unpredictable Quality)

Genre	Variance	Std. Dev
Music	1.47	1.21
Documentary	1.50	1.22
Sci-Fi	1.34	1.16
Family	1.35	1.16
Musical	1.30	1.14

These genres show **higher standard deviations**, suggesting more **variation in quality** (some great, some poor).

Genre with Most Consistency (Predictable Quality)

Genre	Variance	Std. Dev
History	0.45	0.67
Biography	0.50	0.71
War	0.65	0.81
Drama	0.79	0.89

These genres have **lower variance**, meaning **audiences can expect consistent quality** from them.

Key Takeaways:

- **Most common** genre: **Drama** (and it's also relatively well-rated: avg 6.79).
- **Highest-rated genres** (excluding single entries): **Biography, History, War** – fewer in number but high in quality.
- **Horror** is **popular but poorly rated,** often a riskier genre.
- **Documentary and Music** genres show **high variation**, possibly due to niche audiences.
- **History and Biography** genres are **both high in quality and consistent**, ideal for viewers seeking critically acclaimed films.

Movie Duration Analysis:

Descriptive Statistics for Duration:

- **Average Duration:** 109.92 minutes
- **Median Duration:** 106 minutes
- **Standard Deviation:** 22.75 minutes

Interpretation:

- The **average movie** is around **110 minutes,** and the **median** is slightly lower at **106 minutes,** indicating a relatively symmetric distribution with a mild right skew.
- A **standard deviation of 22.75** suggests most movies fall within the range of approximately **87 to 132 minutes** (i.e., ±1 SD from the mean).

IMDb Score vs Duration: Scatter Plot Analysis

- The **regression equation** is:
  IMDb Score = 0.0167 x Duration + 4.6295
- **R² = 0.1299**, which indicates that only about **13% of the variation** in IMDb scores is explained by movie duration.

Interpretation:

- There is a **positive correlation** between movie duration and IMDb score – **longer movies tend to have slightly higher IMDb scores.**
- However, the **correlation is weak** (R^2 is quite low), meaning duration alone does **not significantly predict** a movie's rating.
- The data shows **a dense cluster of movies around the 90-120-minute mark** with scores mostly ranging between **5.5 and 8.0.**

Key Insights:

Most movies are within the 90-130 min range, with average duration being around 110 minutes.

Longer movies tend to **receive slightly higher IMDb scores,** but the relationship is **not strong.**

Other factors like **genre, case, direction, storyline, or production quality** may play a **larger role** in determining IMDb ratings.

The scatterplot shows a **wide spread of IMDb scores** across different durations, supporting the idea that **duration alone doesn't define movie quality.**

Language Analysis:

Dominance of English-language Movies

Count: 3668 out of 3848 movies (approx. 95.3%) are in English.

IMDb Mean Score: 6.42 (lowest among high-frequency languages)

Observation: Despite being the most common, English-language movies have relatively **lower average IMDb scores**, suggesting a wider range in quality and more commercial, mainstream content.

Top 5 Most Common Non-English Languages

Language	Mean	Median	Standard Deviation
French	7.29	7.2	0.56
Spanish	7.05	7.15	0.83
Mandarin	7.02	7.25	0.77
German	7.69	7.7	0.64
Japanese	7.63	7.8	0.90

- These languages have **higher IMDb scores** compared to English.
- This could indicate **stronger critical reception,** higher quality, or niche appeal.

Highest Rated Languages (Mean IMDb Score)

Language	Mean IMDb Score	Count
None	8.50	1
Telugu	8.40	1
Persian	8.13	3
Danish	7.90	3
Romanian	7.90	1
Indonesian	7.90	2
Portuguese	7.76	5
Korean	7.70	5
German	7.69	13
Japanese	7.63	12

- These **less-represented languages often have higher scores,** possibly due to being **critically acclaimed films** or **festival selections.**

Languages with Lower IMDb Scores:

Language	Mean IMDb Score	Count
Bosnian	4.30	1
Kazakh	6.00	1
English	6.42	3668
Thai	6.63	3
Filipino	6.70	1

- Bosnian has the **lowest score,** though with only 1 film, it's not statistically significant.
- English's low score reflects **high variance** and broad spectrum content.

Variability in Scores (Standard Deviation)

- **English: 1.05** – High variability, as expected from a large dataset.
- **Italian: 1.16** – Among the highest, suggesting a wide range in film quality.
- **Hindi: 1.11** – Also shows variability, possibly due to mix of commercial and indie content.
- **Japanese (0.90), Spanish (0.83)** – Moderate variation, indicating diversity in content types.

General Insights

- **Language diversity** is low (English-dominated), but **non-English films often have better critical reception.**
- Many non-English films with **limited representation** still perform **exceptionally well** in terms of ratings.
- Films in **European and East Asian languages (French, German, Japanese, Korean)** consistently show **high average ratings** with **lower standard deviation,** indicating consistent quality.
- Languages like **Persian, Danish, and Telugu** have **high scores** but need cautious interpretation due to low sample sizes.

Actionable Insights:

For Content Curation: Include more **non-English films,** especially in **German, Korean, and Persian,** to increase average IMDb scores and offer diverse, high-quality options.

For Platform Strategy: Highlight **high-rated films in less common languages** as hidden gems or critic's picks.

For Filmmakers/Producers: Quality is recognized regardless of language – aim for strong storytelling and production values, even in regional languages.

Director Analysis:

Director	IMDb Scores (Top Movies)	Average Score	Appearances in Top 10%
Christopher Nolan	9.0, 8.8, 8.6, 8.5	8.725	4
Peter Jackson	8.9, 8.8, 8.7	8.8	3
Steven Spielberg	8.9, 8.6	8.75	2
Francis Ford Coppola	9.2, 9.0	9.1	2
David Fincher	8.8, 8.6	8.7	2
Martin Scorsese	8.7, 8.5	8.6	2
Quentin Tarantino	8.9, 8.5	8.7	2

Other notable directors with **1 entry** in the Top 10%.

- Frank Darabont (9.3) – Highest single IMDb score

- Akira Kurosawa, Milos Forman, Charles Chaplin, Hayao Miyazaki, James Cameron, Ridley Scott – all with 8.5-8.7

Insights from Percentile Rankings

- **All directors listed are in the Top 10%** percentile, signifying their movies rank among the best-rated in the dataset.

- **Consistency Matters:** Directors like Nolan, Jackson, and Spielberg have **multiple entries,** showing they repeatedly deliver high-quality films.

- **One-Hit Wonders vs. Legacy:**
    - Frank Darabont has the highest-rated movie (9.3), but fewer entries.
    - Christopher Nolan balances **quantity and quality,** cementing his reputation.

Contribution Analysis

Directors with High Influence:

- **Christopher Nolan:**
    - Most frequent in Top 10% (4 movies).
    - High consistency across all scores (avg. 8.73).
    - Reflects strong storytelling and critical acclaim.
- **Peter Jackson:**
    - Known for *The Lord of the Rings trilogy* – all rated very highly.
    - Solid average of 8.8, suggesting sustained success.
- **Steven Spielberg & Francis Ford Coppola:**
    - Legendary directors with high individual scores and industry- defining films.

Summary Table: Director Impact Metrics

Metric	Leader(s)
Highest IMDb Score	Frank Darabont (9.3)
Most Top 10% Appearances	Christopher Nolan (4)
Highest Avg. IMDb (≥2 films)	Francis Ford Coppola (9.1)
Consistently High Ratings	Peter Jackson, Spielberg, Nolan

Conclusion:

- **Christopher Nolan** stands out for both **volume and quality.**
- **Francis Ford Coppola and Peter Jackson** have near-perfect averages, demonstrating strong **directional influence** on a film's success.
- **Percentile rank analysis** validates that top directors not only make one great film – they shape cinematic excellence consistently.

Correlation Between Budget and Gross

- **Correlation Coefficient = 0.10085**

Interpretation:

A correlation of **0.10** indicates a **very weak positive relationship** between **budget and gross earnings.**

This suggests that **higher budgets do not strongly predict financial success.**

Movies with large budgets don't necessarily generate large profits; **other factors (e.g., storyline, marketing, timing, cast)** play a significant role.

Top Profit Margin Movie:

Movie	Profit Margin ($)
Avatar	$523,505,847

Insight:

Avatar stands out as the **most profitable movie** based on absolute profit margin.

This emphasizes that **a well-executed high-budget movie** (with cutting-edge VFX and global appeal) can achieve **extraordinary returns.**

Key Insights:

Budget ≠ Profit:

Weak correlation confirms that **a high budget doesn't guarantee high returns.**

Some low-to-mid-budget films might deliver **better ROI (return on investment)** that big-budget productions.

Success Drivers Beyond Budget

Factors contributing to financial success may include:

- Strong **franchise/IP** (e.g., Marvel, Star Wars)
- **Director's reputation** (e.g., James Cameron, Nolan)

- Effective **global marketing**
- **Word-of-mouth** and critical reception

Outliers Matter

- Films like Avatar act as **outliers** – massive hits that significantly skew average metrics.
- These outliers **inflate overall gross trends** but don't represent typical budget-performance relationships.

Recommendations:

- **Invest in originality and audience resonance** over pure budget expansion.
- **Analyse ROI alongside absolute gross** to get a more balanced view of financial success
- **Avoid over-reliance on budget** as a predictor of success; instead focus on **storytelling, timing, and execution.**