IMDB MOVIE ANALYSIS

BY ABHISHEK BHAKAT

MAY 2025

Agenda

01. PROJECT DESCRIPTION

02. APPROACH

03. TECH-STACK USED

04. INSIGHTS

05. RESULTS

PROJECT DESCRIPTION

This project, titled **“IMDB Movie Analysis”**, focuses on exploring the key factors that contribute to a movie's success on the IMDB platform. Success, in this context, is defined by high IMDB ratings. By analyzing a real-world dataset of movies, the project aims to uncover patterns and insights that can help filmmakers, producers, directors, and investors make more informed decisions about future film projects.

APPROACH

DATA CLEANING

- Handled missing values and removed duplicates
- Standardized data formats (e.g., genres, duration, budget)
- Converted necessary columns to appropriate data types
- Created new calculated fields like *profit margin*

APPROACH

EXPLORATORY DATA ANALYSIS (EDA)

- Analyzed genre distribution and their influence on IMDB ratings
- Examined the relationship between movie duration and ratings
- Studied language-wise rating trends
- Identified top-performing directors using percentile scores
- Calculated budget-to-earnings correlation and profit margins

APPROACH

STATISTICAL TECHENIQUES

- Applied descriptive statistics: mean, median, mode, range, standard deviation, variance
- Used correlation and trendline analysis to study relationships
- Employed percentile and ranking techniques for deeper insights

APPROACH

VISUALIZATION

- Used Excel charts like bar charts, scatter plots, and pie charts
- Added trendlines to visualize correlation strength
- Highlighted top insights using conditional formatting and visual cues

TECH-STACK USED

MICROSOFT EXCEL 365

- **Purpose:** Data Cleaning, Analysis, Visualization, and Statistical Computation

- **Functions & Features Used:**

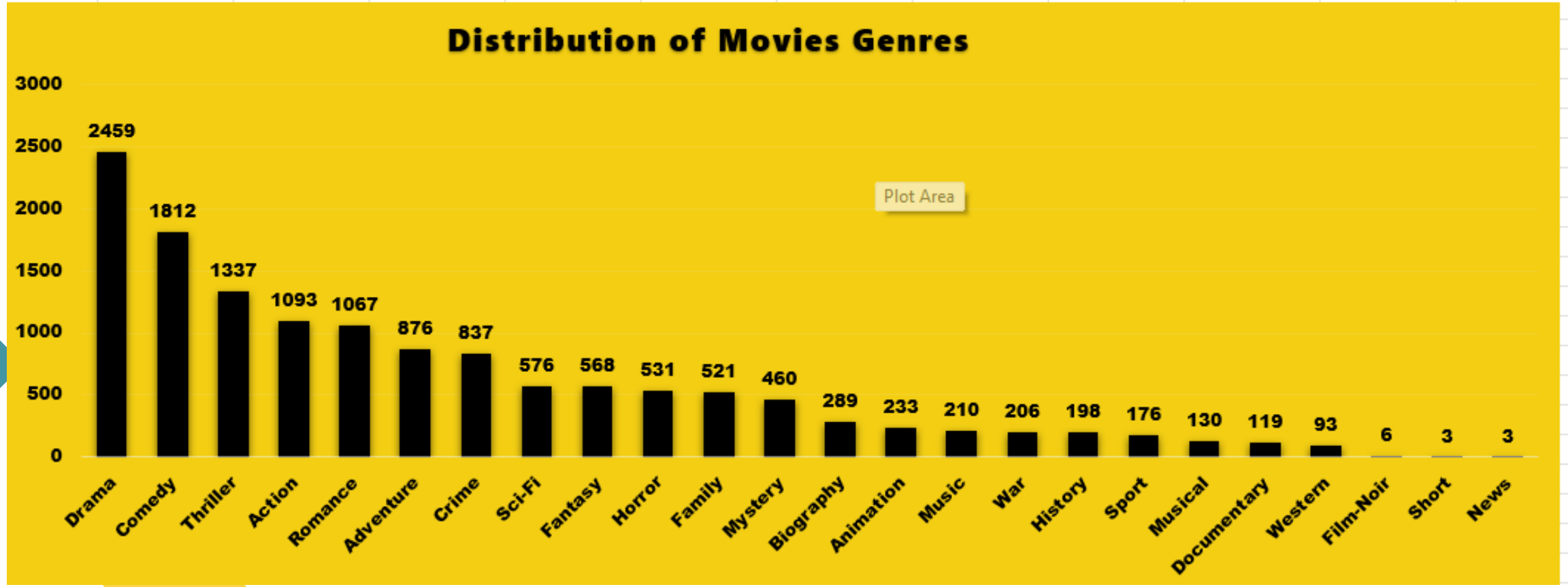
- `COUNTIF`, `AVERAGE`, `MEDIAN`, `MODE`, `MAX`, `MIN`, `STDEV`, `VAR`, `CORREL`, `PERCENTILE`
- Pivot Tables for summarization and cross-tab analysis
- Scatter plots, bar charts, pie charts, and trendlines for visualization
- Conditional formatting for highlighting patterns
- Data Filters and Sort functions for focused exploration


INSIGHTS

MOVIE GENRE ANALYSIS

TASK A : 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.



MOVIE GENRE ANALYSIS

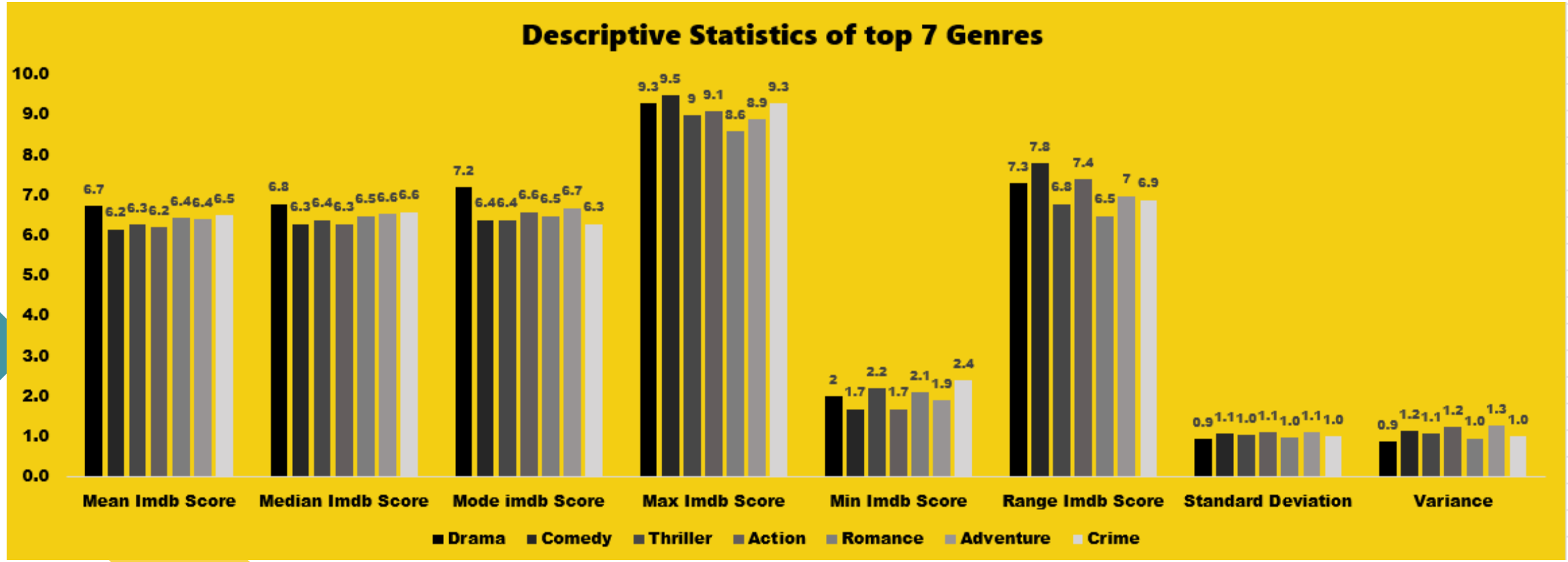
INSIGHTS

Chart 1: Distribution of Movie Genres

- **Drama** is the most produced genre with **2,459 movies**, followed by **Comedy (1,812)** and **Thriller (1,337)**.
- These three genres make up a significant portion of the total movie dataset, indicating their popularity and widespread appeal.
- Less common genres include **Western (93)**, **Film-Noir (6)**, **Short (3)**, and **News (3)**, suggesting niche interest or limited production.
- **Action, Romance, and Adventure** also have substantial counts, highlighting strong viewer demand for excitement and emotional narratives.

MOVIE GENRE ANALYSIS

Here is the descriptive statistics of the top 7 Genres



MOVIE GENRE ANALYSIS

INSIGHTS

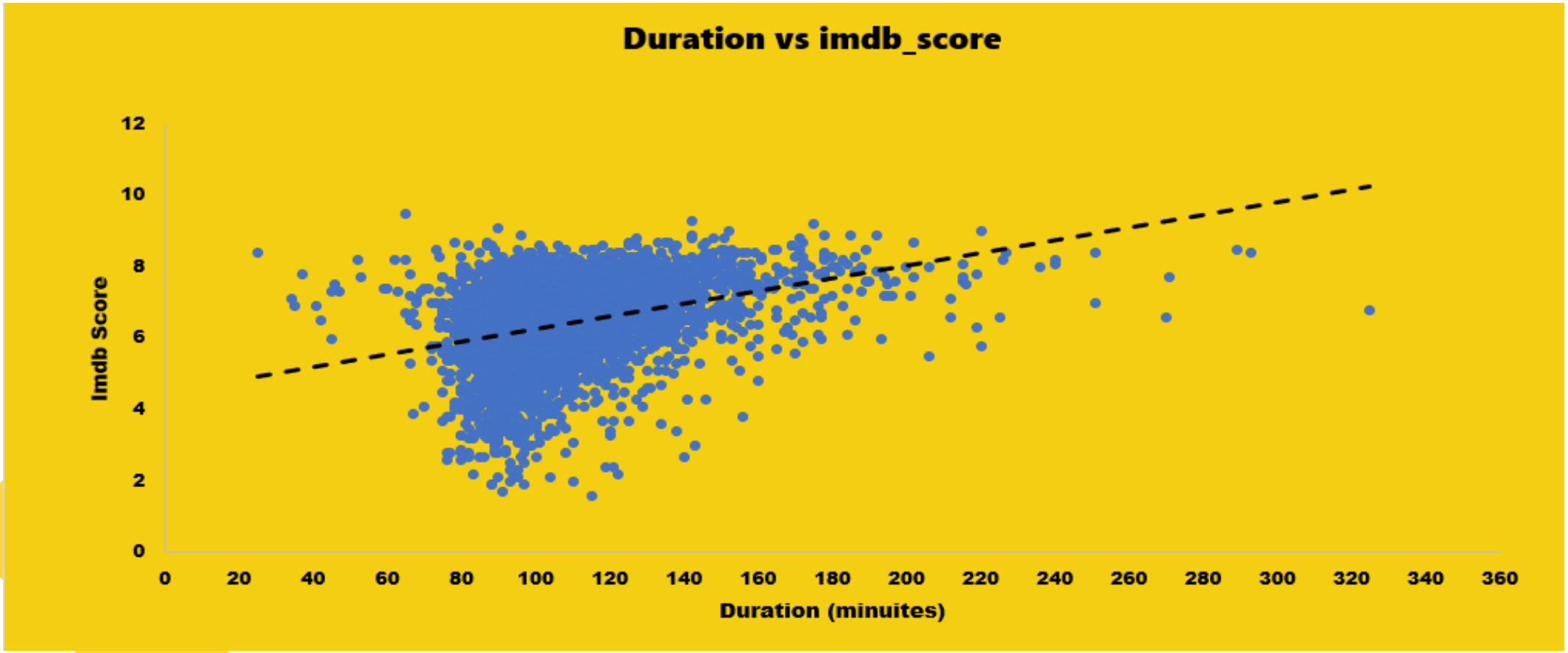
Descriptive Statistics of Top 7 Genres :

- **Drama** has the highest **mean IMDb score (6.7)** and **median score (6.8)**, indicating consistent quality and viewer approval.
- **Comedy** has the highest **mode IMDb score (7.2)** and a relatively high **maximum score (9.5)**, showing that while average ratings are moderate, some comedies perform exceptionally well.
- **Thriller** has the **widest range** in IMDb scores (**7.8**), suggesting varying quality across thriller movies.
- **Adventure** has the **lowest minimum score (1.7)** and **highest variance (1.3)**, indicating high inconsistency in viewer ratings.
- **Crime** and **Action** also show relatively high variation, with **standard deviations around 1.1**.
- Overall, **Drama and Comedy** offer the most reliable viewing experiences in terms of average ratings and consistency.

MOVIE DURATION ANALYSIS

TASK B : Analyze the distribution of movie durations and its impact on the IMDB score.

Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.



MOVIE DURATION ANALYSIS

INSIGHTS

Duration vs Imdb Score :

- The **scatter plot** shows a **positive correlation** between movie duration and IMDb score, as indicated by the upward trend of the regression line.

- This suggests that **longer movies tend to receive higher IMDb scores** on average, although the relationship is not very strong.

- The **majority of movies fall between 80 and 150 minutes**, with scores mostly ranging from 5 to 8.

- A few outliers exist — very short or very long movies with unusually high or low scores, but they are rare.

MOVIE DURATION ANALYSIS

Descriptive Statistics for Duration

MEAN	MEDIAN	STANDARD DEVIATION
108.02	104	21.98

MOVIE DURATION ANALYSIS

INSIGHTS

Descriptive Statistics for Movie Duration :

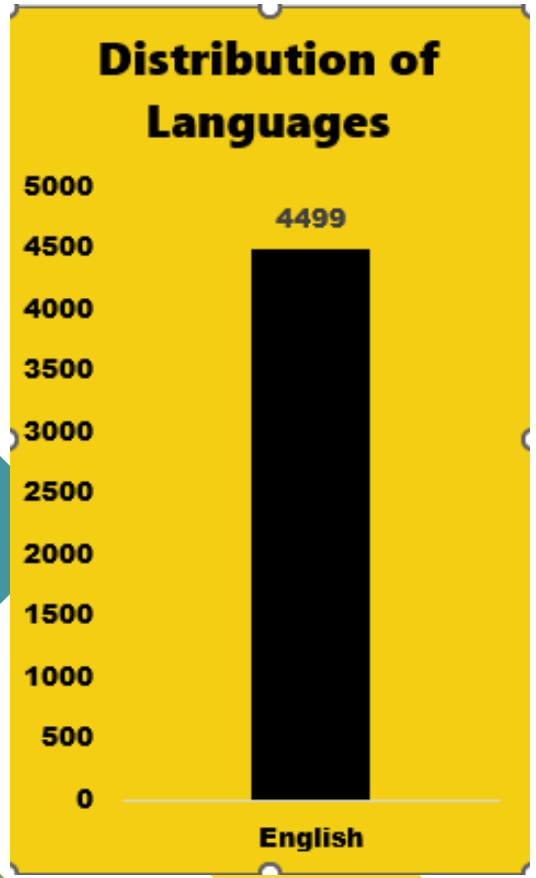
- The **average movie duration** is **108.02 minutes**, with a **median of 104 minutes**, indicating a slightly right-skewed distribution (some longer movies).

- The **standard deviation is 21.98 minutes**, meaning most movies fall within the range of approximately **86 to 130 minutes**.

LANGUAGE ANALYSIS

TASK C : Examine the distribution of movies based on their Language.

Determine the most common language used in the movies and analyze their impact on IMDB score using descriptive statistics.



LANGUAGE ANALYSIS

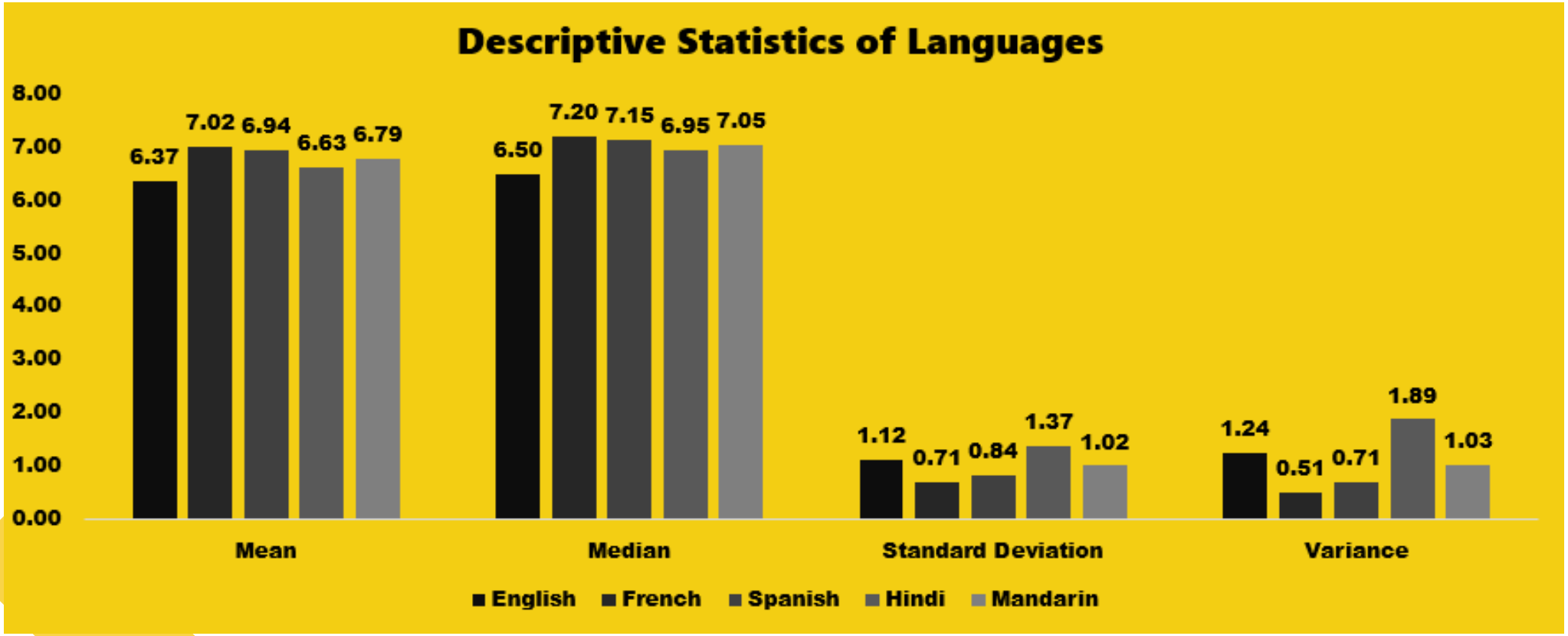
INSIGHTS

Distribution of Languages :

- **English dominates** the dataset, with **4499 movies**, far more than any other language.
- **French (72), Spanish (40), Hindi (28), Mandarin (24)**, and **Japanese (16)** are the next most common languages, but they appear **in significantly lower numbers**.
- Most of the other languages are underrepresented (less than 10 movies), and many (like Tamil, Urdu, Vietnamese, etc.) have **only 1 movie**.
- This skewed distribution suggests that **English-language films are overrepresented**, possibly due to greater global production and/or availability in datasets.

LANGUAGE ANALYSIS

Descriptive statistics of Languages



LANGUAGE ANALYSIS

INSIGHTS

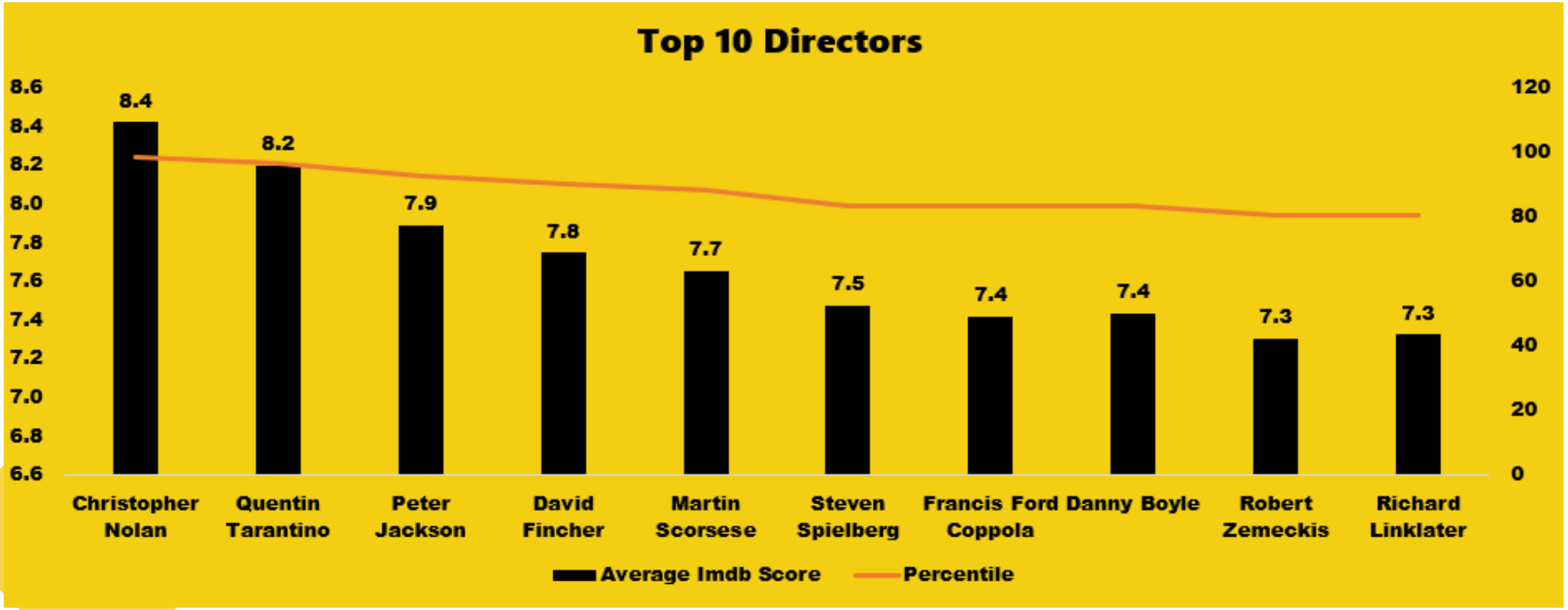
Descriptive Statistics of Languages:

- **French and Spanish movies have the highest average IMDb scores:**
  - **Mean IMDb Score**: French (7.02), Spanish (6.94), Mandarin (6.79)
  - **English movies have a lower mean** of **6.37**.
- **Median scores follow a similar trend**, indicating **consistent audience approval** for French and Spanish movies.
- **Hindi movies have the highest variance (1.89)**, meaning the quality of Hindi movies varies significantly — some perform very well, others poorly.
- **French movies have the lowest standard deviation (0.71)**, suggesting that **they are more consistent in quality**.
- While **English movies are the most produced**, their **average ratings are lower and more variable** than some of the less represented languages.

DIRECTOR ANALYSIS

TASK D : Influence of directors on movie ratings

Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.



DIRECTOR ANALYSIS

INSIGHTS

Top Directors:

•**Christopher Nolan** stands out with the **highest average IMDb score of 8.4** and is placed in the **topmost percentile**, indicating widespread critical and audience acclaim.

•**Quentin Tarantino** follows with an **8.2 average score**, also within a very high percentile, reflecting a consistently strong filmography.

•**Peter Jackson (7.9)**, **David Fincher (7.8)**, and **Martin Scorsese (7.7)** also rank in high percentiles, showing that their work is highly appreciated in the industry.

•As we move down the list to **Steven Spielberg**, **Francis Ford Coppola**, and **Danny Boyle** (7.4–7.5 scores), there's a slight dip in average ratings, but the percentile remains relatively stable.

•**Robert Zemeckis** and **Richard Linklater**, with **7.3 average scores**, still maintain a good percentile, affirming their solid reputations.

•The **orange line (Percentile)** shows a **slight downward trend**, indicating a small decline in relative standing, but all directors remain in the **top quartile** of performance.

DIRECTOR ANALYSIS

IMDB score and percentile insights

Director Name	Average Imbd score	Count of Movies	Percentile
Christopher Nolan	8.4	8	98.7
Quentin Tarantino	8.2	8	96.8
Peter Jackson	7.9	9	92.7
David Fincher	7.8	10	90.7
Martin Scorsese	7.7	20	88.8
Steven Spielberg	7.5	26	83.5
Francis Ford Coppola	7.4	11	83.5
Danny Boyle	7.4	8	83.5
Robert Zemeckis	7.3	13	80.6
Richard Linklater	7.3	11	80.6

DIRECTOR ANALYSIS

INSIGHTS

Top 10 Directors & Percentile:

- **Christopher Nolan and Quentin Tarantino** lead with the highest **IMDb scores (8.4 & 8.2)** and **percentiles (98.7% & 96.8%)**, despite directing only 8 movies each—showing consistent excellence.

- **Peter Jackson, David Fincher, and Martin Scorsese** maintain **high ratings (7.7–7.9)** with solid percentiles, proving both **quality and volume** in their filmography.

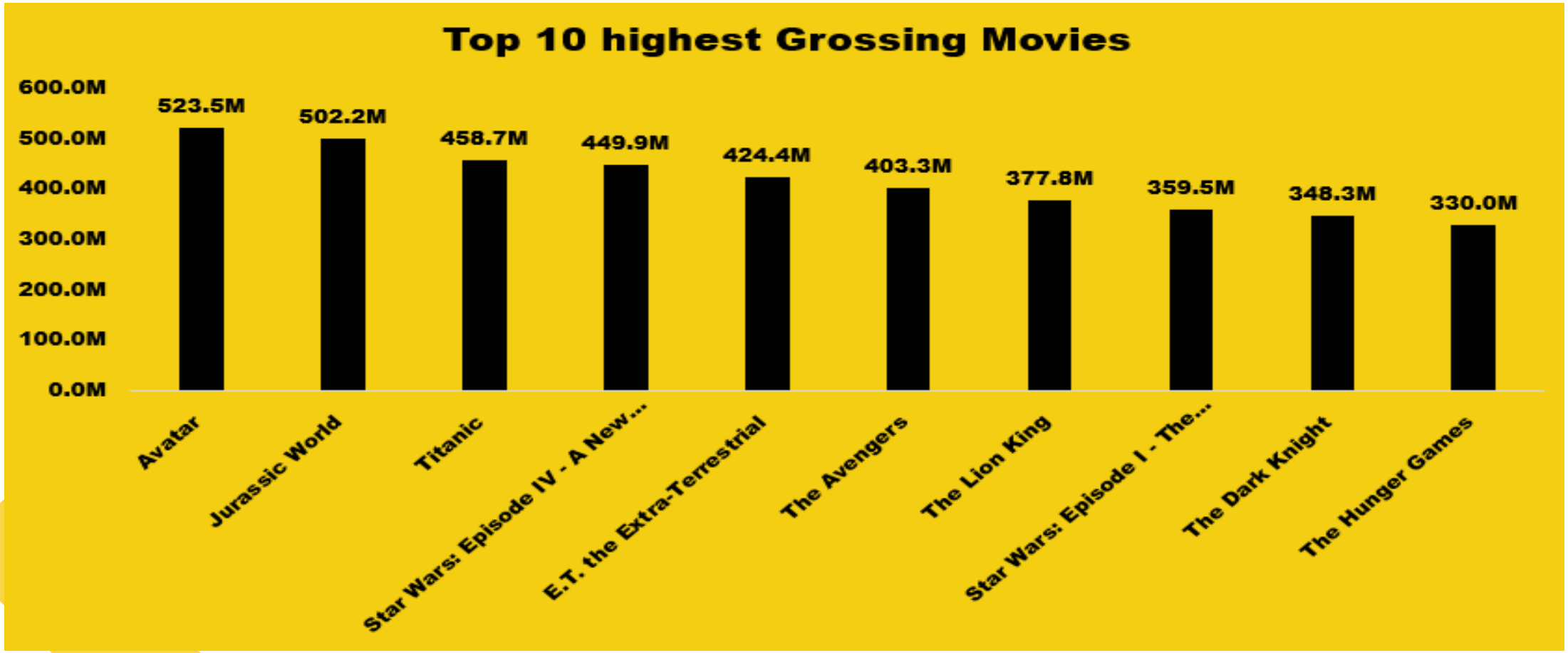
- **Steven Spielberg**, with the **highest movie count (26)**, has a slightly lower average (7.5) but remains in the top **83.5 percentile**, reflecting his **wide-ranging contributions** to cinema.

- All listed directors are in the **top 20% (≥80th percentile)**, reinforcing their positions as **critically respected and influential filmmakers**.

BUDGET ANALYSIS

TASK E : Explore he relationship between movie budget and their financial success

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



BUDGET ANALYSIS

INSIGHTS

Top 10 Highest Grossing Movies:

•**Avatar** tops the chart with a massive **$523.5M**, followed closely by **Jurassic World ($502.2M)** and **Titanic ($458.7M)** — all crossing the **$450M mark**.

•**Franchises dominate** the list, including *Star Wars* (2 entries), *The Avengers*, and *The Dark Knight*, highlighting the **power of cinematic universes**.

•**The Lion King** and **E.T.** show that even **animated and older classics** can compete strongly with modern blockbusters.

•The revenue gap between the top and bottom film is nearly **$194M**, underlining the **varying scale of box office success** even within the top 10.

BUDGET ANALYSIS

Gross vs Budget Correlation

Correlation between Budget and Gross : 0.238730915

Movies with Highest Profit Margin :

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

BUDGET ANALYSIS

INSIGHTS

Gross vs budget Correlation :

•**Weak Positive Correlation**: The correlation of **0.2384** suggests a **very weak positive relationship** — as one variable (e.g., number of movies directed) increases, the other (e.g., average IMDb score) only slightly tends to increase, but **not strongly or consistently**.

•**Quality ≠ Quantity**: Directors who make more movies don't necessarily have higher IMDb scores. This implies that **movie quality (ratings) is not directly dependent on the quantity** of movies produced.

•**Other Factors at Play**: Since the correlation is weak, it indicates that **other variables** — like genre, storytelling, star cast, or audience trends — may have a **greater impact** on a movie's IMDb score or success.

RESULTS

The IMDB Movie Analysis project successfully identified key factors that influence a movie's rating and financial performance. By applying data cleaning, statistical analysis, and Excel-based visualizations, the project delivered:

- A deeper understanding of how **genres, duration, language, director reputation, and budget** impact movie success.
- Clear evidence that **quality direction and mid-to-high budgets** often correlate with higher IMDB ratings and better returns.
- **Actionable insights** that filmmakers, producers, and investors can use to make data-driven decisions for future projects.

This project enhanced my analytical thinking, Excel proficiency, and ability to extract valuable business insights from raw data.

Thank you

Abhishek Bhakat

abhishekbhakat6@gmail.com

Link to dataset : Dataset