

# IMDB MOVIE ANALYSIS

PRESENTED BY SOUMYA SINGHAL | DATE: 5 DECEMBER 2024

The IMDb logo is displayed in a bold, black, sans-serif font. It is positioned within a yellow square, which is part of a larger yellow rectangular area on the right side of the slide. The logo consists of the letters 'IMDb' in a stylized, slightly irregular font.

# PROJECT OVERVIEW

- Objective: Analyze IMDB data to identify factors influencing a movie's success.
- Dataset: Includes variables like genres, duration, language, director, budget, and gross earnings.
- Description: This report investigates the factors influencing a movie's success on IMDB, where success is measured by high IMDB ratings. Through data cleaning, analysis, and visualization, key aspects such as genre, duration, language, director, and budget are explored. The report provides actionable insights into how these factors impact movie ratings, offering valuable guidance for producers, directors, and investors in the film industry. The findings aim to support data-driven decision-making to enhance future movie success.

## TECH-STACK USED

- Microsoft Excel for Analysis
- PowerPoint for Presentation
- Apowersoft Screen Recorder

# APPROACH

- Data Cleaning: Handled missing values and formatted columns.
- Data Analysis: Used Excel functions for descriptive statistics and visualizations.
- Five 'Whys': To identify root causes behind observed trends.

# DATA PREPROCESSING

## ■ Data Cleaning

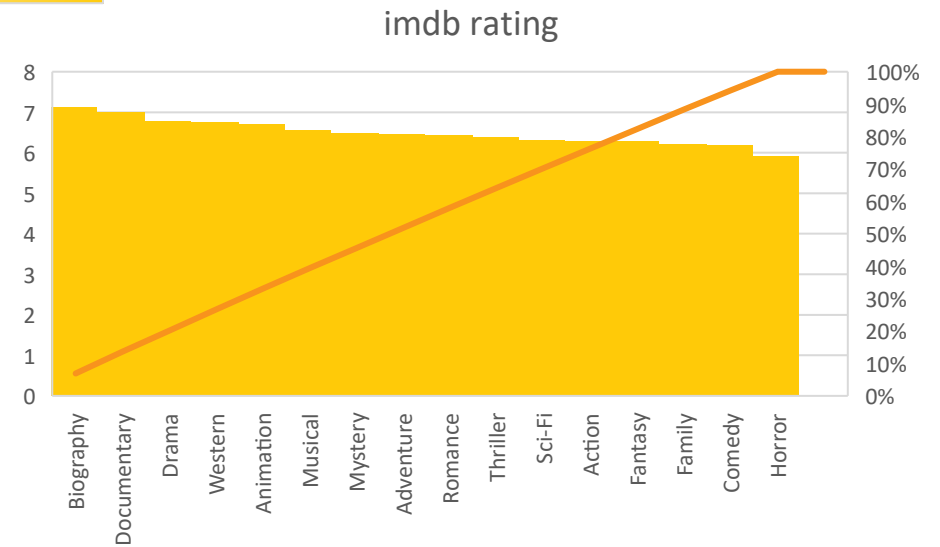
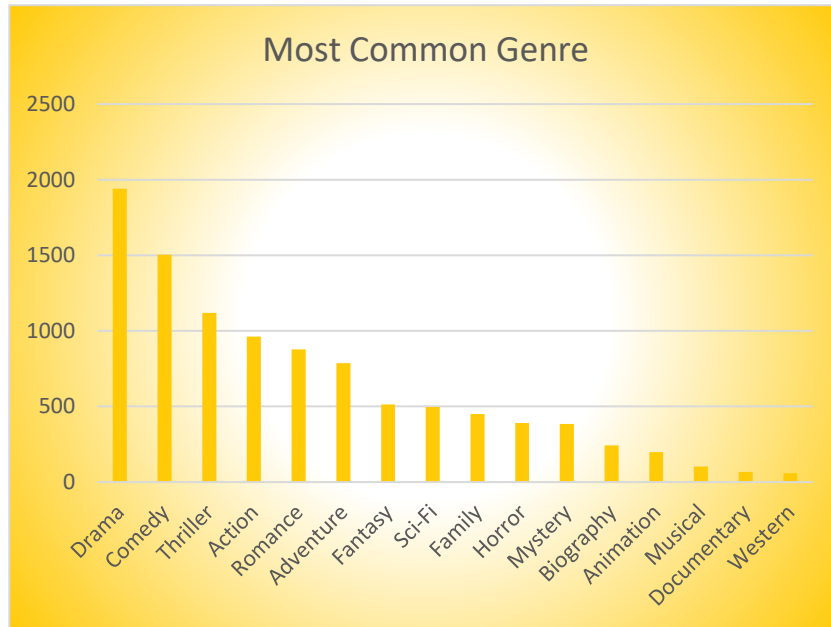
In this step, the data was cleaned by addressing missing values, removing duplicate rows, and ensuring appropriate data types for each column. Columns such as 'budget', 'gross', and 'imdb\_score' were examined for missing values, and numerical columns were filled with median values. Categorical columns were filled with the mode where appropriate.

## ■ Feature Engineering

Splitting the 'genres' column into separate binary columns and then to single column.

## KEY FINDINGS: GENRE ANALYSIS

- Drama leads with 1941 movies, followed by Comedy and Thriller.
- Biography has the highest IMDb average score (7.14), Horror the lowest (5.93).
- Documentary shows the highest variability in scores, while Biography is consistent.

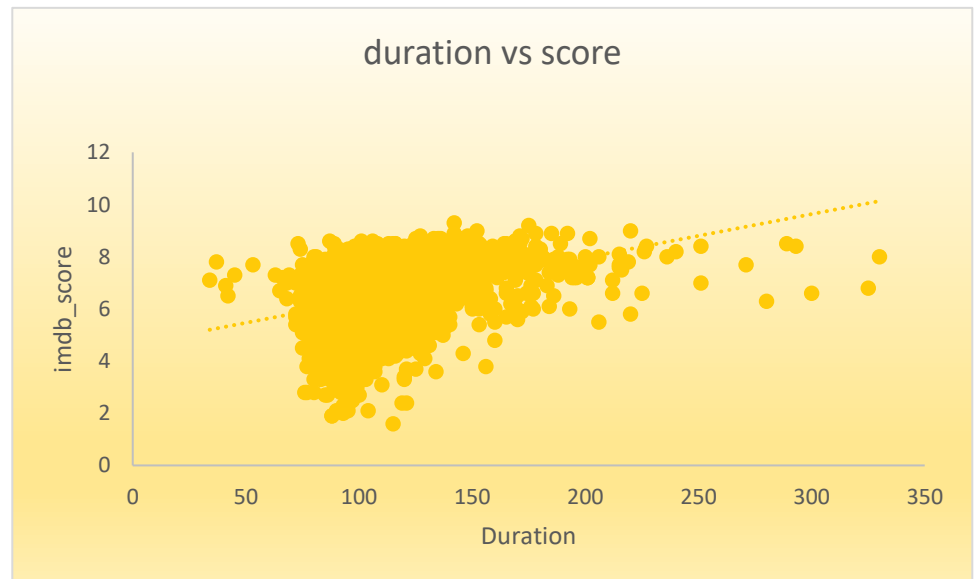


## KEY FINDINGS: DURATION ANALYSIS

- Short movies like 'Evil Dead II' (37 min) scored well (7.8).
- Long movies like 'The Godfather' (175 min) also score high (9.2).
- Positive correlation between longer duration and higher IMDb scores.



DURATION	STATS
Mean	109.9068
Median	106
Mode	101
Variance	517.2472
Standard Deviation	22.74307
Maximum	330
Minimum	34



## KEY FINDINGS: LANGUAGE ANALYSIS

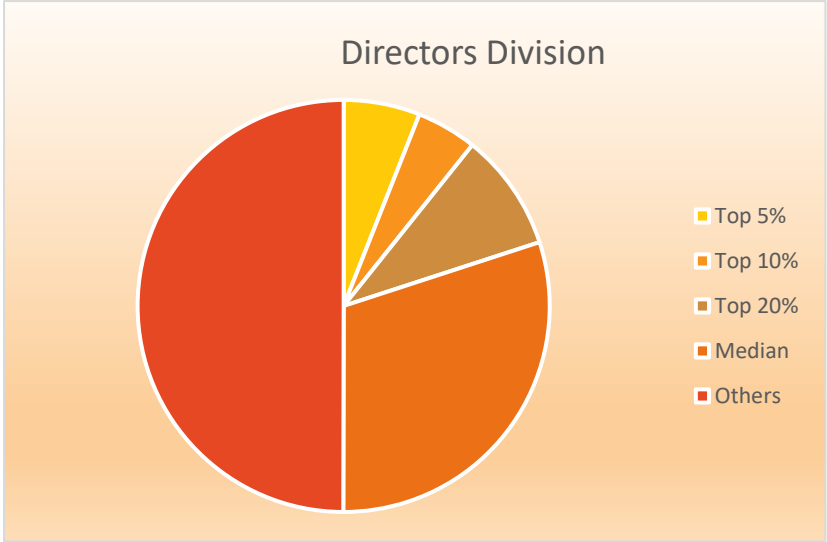
- English dominates with 3673 movies (avg. score 6.42).
- Persian leads in average IMDb score (8.13) but with only 3 movies.
- Cantonese shows consistent scoring with low standard deviation (0.41).

## KEY FINDINGS: DIRECTOR ANALYSIS

- Top 5% of directors (105) achieved the highest IMDb scores.
- A significant number of directors scored below the 50th percentile.
- Successful directors often have consistent high ratings.

95th percentile	7.7
90th percentile	7.5
80th percentile	7.114285714
50th percentile	6.457142857

	count
Top 5%	105
Top 10%	83
Top 20%	162
Median	526
Others	875
total	1751



## KEY FINDINGS: BUDGET ANALYSIS

- Weak correlation (0.1) between budget and gross earnings.
- 'Avatar' achieved highest profit (\$523M), 'The Host' incurred massive losses.
- High-budget movies often struggle with profit margins due to high costs.

	MOVIE	PROFIT
MAX PROFIT	Avatar	523505847
MIN PROFIT	The Host	-12213298588

Correlation	
gross - budget	0.100982454
budget - profit	-0.952791815
gross - profit	0.205857047

## CONCLUSION & KEY INSIGHTS

- Drama and Biography genres consistently perform well.
- Longer movies tend to achieve higher IMDb scores.
- Budget management is crucial for profitability.
- English-language movies dominate but diversity in language yields high scores.

# RECOMMENDATIONS

- Focus on Biography and Documentary genres for critical acclaim.
- Optimize movie duration around 120-180 minutes for better reception.
- Prioritize effective budget management to maximise profitability.
- Encourage diverse language productions for niche audience engagement.



VIDEO PRESENTATION LINK

■ [Click here](#)

EXCEL FILE LINK

■ [Click here](#)