

Movie Metadata Analysis

1. Project Overview

This project analyzes a dataset of movies to identify patterns in genres, budgets, revenue, ratings, and other attributes. The goal is to understand which factors influence a movie's success and provide insights into trends across different years, directors, and countries.

2. Tools Used

- Microsoft Excel (for initial cleaning & preprocessing)
- Microsoft Power BI (for data modeling, visualization, and dashboards)

3. Dataset

- **Source:** Movie Metadata (sample dataset)
- **Key Columns:**
 - movie_title – Name of the movie
 - director_name – Director of the movie
 - genres – Genre(s) of the movie
 - budget – Production budget (USD)
 - gross – Worldwide gross revenue (USD)
 - imdb_score – IMDb rating (1–10 scale)
 - year – Release year
 - language – Movie language
 - country – Country of production
 - duration – Duration of movie (minutes)
 - rating – MPAA rating (PG, R, etc.)

4. Steps Followed

I. Data Collection & Import:

- Imported raw dataset (movies metadata) into Power BI.

- Dataset contained key fields like: movie_title, genres, budget, gross, imdb_score, director_name, year, country, language, duration, rating.

II. Data Cleaning & Transformation (Power Query):

- Removed nulls, duplicates, and irrelevant columns.
- Standardized text fields (e.g., trimmed spaces in movie_title, split multiple genres).
- Converted data types (budget/gross → numeric, year → date hierarchy).
- Created MovieGenres bridge table to handle multiple genres per film.

III. Data Modeling (Star Schema Design):

- Designed a fact table: Movies (budget, gross, imdb_score, duration, etc.).
- Built dimension tables: Genres, Directors, Countries, Languages, Dates.
- Established relationships:
 - Movies → MovieGenres → Genres
 - Movies → Directors
 - Movies → Countries
 - Movies → Languages
 - Movies[Year] → Date[Year]

IV. Calculated Measures (DAX):

- Created KPIs such as:
 - Profit = Gross – Budget
 - ROI = (Profit / Budget) × 100
 - Average Duration, Average IMDB Rating
 - Profitability grouped by Budget Category (Low, Medium, High).

V. Visualization & Dashboard Building:

- Built charts to explore profitability, ROI, IMDB trends, movie duration by decade, genre dynamics.
- Used slicers for Genre, Country, Director, Year for interactive filtering.
- Added KPI cards for Total Profit, Avg ROI, Top Director, Top Genre.

5. Key Insights

I. Movie Trends & Ratings:

- Number of movies released has grown rapidly since the 1970s, peaking around 2010.
- Most movies are rated R (39.9%), followed by PG-13 (28.9%) and PG (16.5%).

- English dominates as the primary language (~13K movies), with other languages showing very low representation.

II. Directors & Production:

- Steven Spielberg, Clint Eastwood, and Martin Scorsese are the top directors by number of movies.
- Directors like John Blanchard (IMDB 9.5) and Sadyk Sher-Niyaz (IMDB 8.7) achieved highest IMDB averages with fewer films.

III. Budget, Duration & Profitability:

- Low and Medium budget movies are far more profitable than High-budget ones.
- Standard-duration movies are the most common (~80%), but Short-duration Low-budget movies have higher ROI.
- High-budget Standard-duration films actually show significant losses (~ -83B total profit).

IV. Box Office Performance:

- Top-grossing films include The Lion King (1994), Avatar (2009), and Jurassic World (2015).
- Despite high budgets, returns vary widely, with ROI ranging from 103% (Inside Out) to 839% (The Lion King).

V. IMDB & Audience Ratings:

- Average IMDB score across all movies is 6.45.
- Distribution shows most films clustered between 5–7 rating.
- Quality trend has declined by decade – IMDB average dropped from ~8.0 (1930s) to ~6.3 (2010s).

VI. Genre Analysis:

- Comedy, Adventure, Thriller, and Family are the most profitable genres.
- Game-Show (1516% ROI) and Film-Noir (410% ROI) stand out with extremely high ROI despite fewer movies.

VII. Country & ROI

- USA dominates production volume, but countries like Kenya (323% ROI) and UAE (39% ROI) outperform in ROI.
- Some small industries generate much higher returns compared to Hollywood.

6. Screenshots

General Overview:



Top Performers & Profitability:

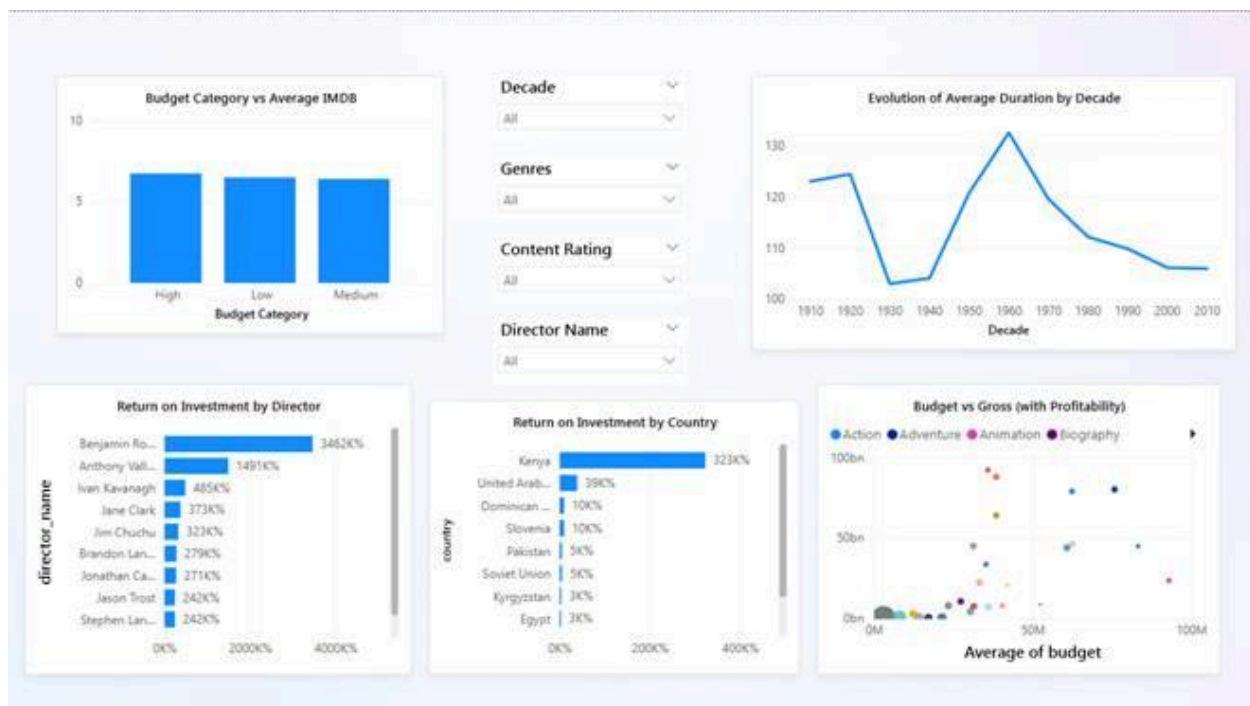
Top 10 highest-grossing movies						
movie_title	year	Sum of budget	Sum of gross	Total Profit	Return on Investment	Avg IMDB Score
The Lion King	1994	22,50,00,000.00	2,11,39,18,885.00	1,88,89,18,885.00	839.52%	8.50
American Sniper	2014	35,26,00,000.00	2,10,07,41,318.00	1,74,79,41,318.00	495.45%	7.30
Jurassic World	2015	60,00,00,000.00	2,60,87,09,084.00	2,00,87,09,084.00	334.78%	7.00
Star Wars: Episode I - The Phantom Menace	1999	46,00,00,000.00	1,89,81,78,708.00	1,43,81,78,708.00	312.65%	6.50
Avatar	2009	94,80,00,000.00	3,04,20,23,388.00	2,09,40,23,388.00	220.89%	7.90
Shrek 2	2004	90,00,00,000.00	2,61,88,26,216.00	1,71,88,26,216.00	190.98%	7.20
The Dark Knight	2008	74,00,00,000.00	2,13,32,64,244.00	1,39,32,64,244.00	188.28%	9.00
Frozen	2013	90,00,00,000.00	2,40,44,19,600.00	1,50,44,19,600.00	167.16%	7.60
Toy Story 3	2010	1,00,00,00,000.00	2,07,49,22,485.00	1,07,49,22,485.00	107.49%	8.30
Inside Out	2015	1,05,00,00,000.00	2,13,87,26,202.00	1,08,87,26,202.00	103.69%	8.30
Total		7,17,58,00,000.00	23,13,37,30,130.00	15,95,79,30,130.00	222.39%	7.76

Total Profit & Avg Profit Per Movie for each Duration Category & Budget Category							
budget_category	High		Low		Medium		Total
	Total Profit	AvgProfitPerMovie	Total Profit	AvgProfitPerMovie	Total Profit	AvgProfitPerMovie	
Standard	-83,13,55,77,232.00	-8,24,75,771.06	90,93,64,77,600.00	2,05,97,163.67	41,93,63,97,871.00	71,32,040.45	49,73,72,98,239.00
Short	-2,43,61,37,402.00	-3,08,37,182.30	21,36,87,80,585.00	2,13,26,128.33	11,49,41,98,888.00	1,08,12,981.08	30,42,68,42,071.00
Long	2,00,00,34,699.00	1,06,95,372.72	5,72,76,24,469.00	3,18,20,135.94	7,60,36,20,755.00	2,38,35,801.74	15,33,12,79,923.00
Total	-83,57,16,79,935.00	-6,55,97,864.94	1,18,03,28,82,654.00	2,10,88,597.94	61,03,42,17,514.00	84,04,601.70	95,49,54,20,233.00

Quality & Distributions:



ROI & Duration Trends:



Profitability by Genre & Decade:

