

# Movie Analytics for Smarter Decisions

- How movie data is transformed into insights
- Audience: Business & Non-Technical Stakeholders

By Aman Verma  
1753 FY AI&ML

# Why Data-Driven Decisions Matter

- Movie investments involve high risk
- Data helps reduce uncertainty
- Predictive models support informed decisions

# Our Dataset Overview

Information includes:

- Movie budget
- Genre
- Audience ratings
- Reviews & votes
- Box office revenue

# Our Dataset Column Descriptions :

Column Name	Description
<b>MovieID</b>	Unique identifier for each movie (integer from 1 to 999,999)
<b>Title</b>	Synthetic movie title written in natural language style
<b>Genre</b>	Primary genre of the movie (Drama, Action, Comedy, etc.)
<b>ReleaseYear</b>	Year the movie was released (1950–2025)
<b>ReleaseDate</b>	Full release date within the release year (YYYY-MM-DD)
<b>Country</b>	Country where the movie was produced
<b>BudgetUSD</b>	Estimated production budget in US dollars (from 100k to 300M)
<b>US_BoxOfficeUSD</b>	Gross box office revenue from the US market
<b>Global_BoxOfficeUSD</b>	Total worldwide box office revenue
<b>Opening_Day_SalesUSD</b>	Estimated US ticket sales on opening day
<b>One_Week_SalesUSD</b>	Estimated US ticket sales during the first week
<b>IMDbRating</b>	IMDb rating on a scale of 1.0 to 10.0
<b>RottenTomatoesScore</b>	Rotten Tomatoes score (0–100 percentage)
<b>NumVotesIMDb</b>	Number of user votes on IMDb
<b>NumVotesRT</b>	Number of user votes on Rotten Tomatoes
<b>Director</b>	Synthetic name of the movie director
<b>LeadActor</b>	Synthetic name of the lead actor

# Business Challenges

- Uncertain movie success
- High production costs
- Difficulty predicting audience response

# Step 1: Data Preparation

- Cleaned missing values
- Standardized numeric fields
- Encoded movie genres
- Ensured data consistency

# Step 2: Predictive Modeling

- Regression models predict revenue & ratings
- Classification models label movies as Hit or Flop

# Models Used

- Linear Regression
- Random Forest
- Logistic Regression
- Decision Tree & KNN

# What Influences Success?

Key factors identified:

- Budget
- Audience ratings
- Reviews & votes
- Movie genre

# Business Impact

- Better budget planning
- Reduced investment risk
- Improved marketing decisions
- Smarter project selection

# Implementation Reference

- The full implementation code demonstrating this data preparation process is available at the following link:

[https://colab.research.google.com/drive/1I\\_ssBEzzdccYNq4AvBgeH8oR66TBlrLz?usp=sharing](https://colab.research.google.com/drive/1I_ssBEzzdccYNq4AvBgeH8oR66TBlrLz?usp=sharing)

- This link is provided for transparency and auditability.

# Key Takeaway

- Machine Learning enables smarter movie investments
- Data-backed insights lead to better outcomes