MOVIE BUSINESS ANALYSIS

By : Class Group 3 DS-PT11

Members:

1. **Irene Kibengo**
2. **Erastus K Njuguna**
3. **Benson Mwihia**
4. **Daniel Akwabi**
5. **Luciana Ndanu**
6. **Sydney Were**

Introduction:

This presentation provides data-driven insights into factors contributing to movie box office success. We aim to guide Flix company's new movie studio in making strategic decisions on genre, budget, release timing, and talent to maximize returns.

Understanding the Business Problem

Industry Trend

The entertainment industry is rapidly transforming, with original content driving audience engagement and revenue. Major players are heavily investing in original films, reaping substantial financial returns and strengthening their brand presence.

Flix Company's Goal

Flix company is launching a new movie studio but lacks data-driven insights on factors contributing to box office success. Our role is to explore public movie performance data to uncover patterns for financial success.

Our Objective

Provide clear, actionable recommendations to guide decisions on genre, budget, release timing, and other production choices, ensuring data-driven success for the new studio.

Stakeholders, Use Case and Project Objectives

Primary Stakeholder

- The Head of the New Movie Studio will leverage data-driven insights to inform strategic decisions on film production.

Use Case

- Insights will identify high-performing genres, optimal budget ranges, ideal release windows, and casting strategies to maximize box office success and ROI.

Project Objectives

- Identify which genres perform best at the box office, considering revenue and profitability.
- Analyze the impact of budget, runtime, cast, and release month on a film’s success.
- Provide actionable recommendations for the types of films the company should produce

Data Sources and Understanding for Analysis

Box Office Mojo

Provides domestic box office revenue data, including title, studio, domestic gross, release date, and year, crucial for financial performance analysis.

TheMovieDB (TMDb)

Offers user-generated popularity and vote data, genres, budget, and revenue, complementing Box Office Mojo with audience traction and cross-comparison metrics.

IMDb

Contains detailed metadata like primary title, genres, runtime, start year, and user ratings (average rating, num votes), used for movie characteristics and audience perception.

The Numbers

Provides film production budgets and worldwide gross, offering a complete financial picture and enabling ROI calculation, a key metric for production decisions.

Data Cleaning and Preparation

To ensure reliable and actionable insights on box office performance, we performed essential data cleaning and integration using datasets from Box Office Mojo, IMDb, and TMDb.

Key Actions Taken:

Focused the dataset on relevant information

We removed data that was not aligned with our analysis goals, such as foreign box office earnings, since our focus is on domestic performance.

Addressed missing values

We carefully filled in missing values in a way that reduces the impact of outliers and ensures the data remains reliable. In cases where data was critically incomplete, we excluded those entries.

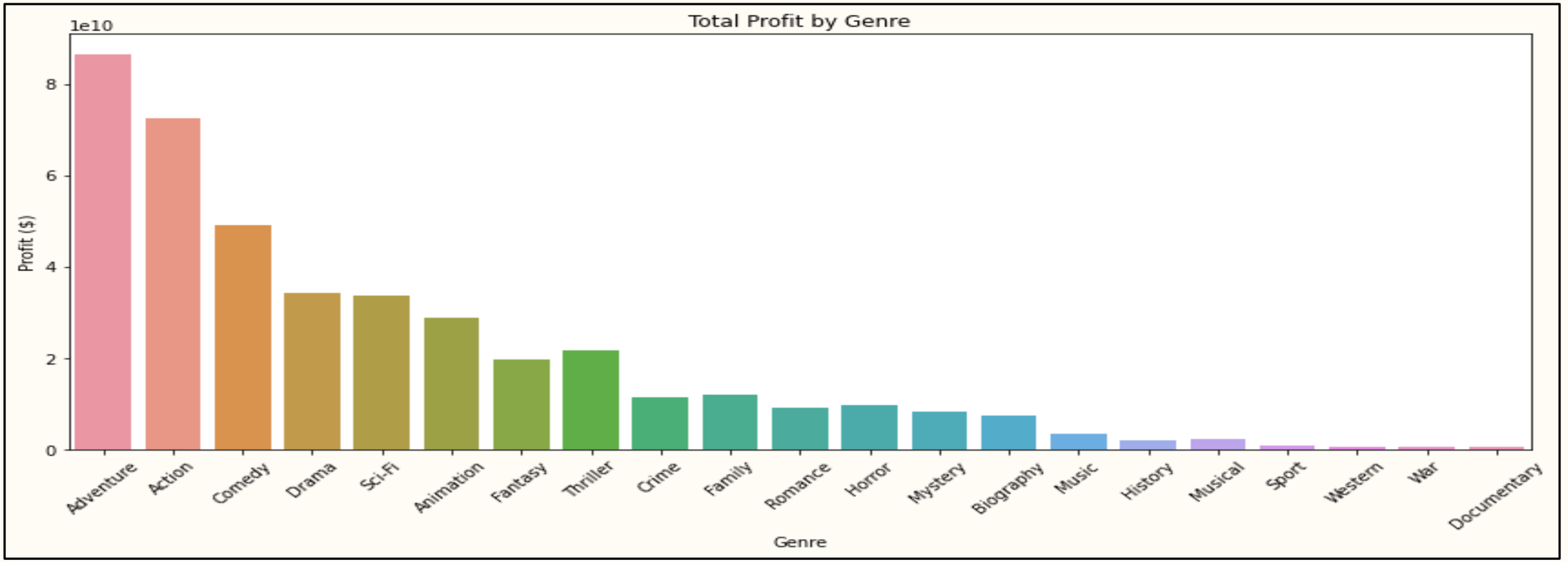
Reviewed the data for quality and consistency

We checked for any inconsistencies, duplicates, or incomplete records to ensure the dataset was clean and accurate.

Final dataset overview

After cleaning, our dataset includes 3,387 movie records with 5 key columns relevant to financial performance.

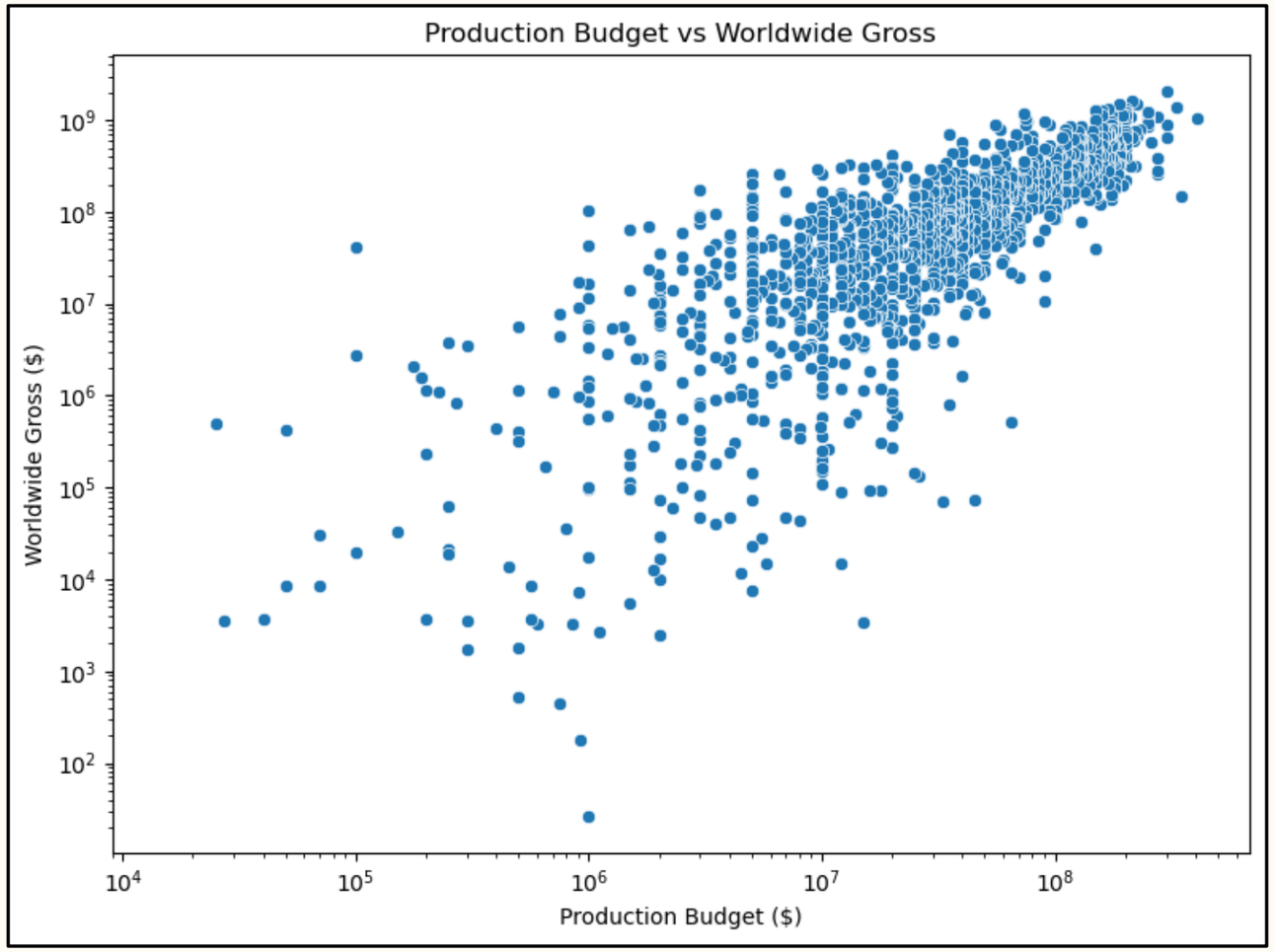
Exploratory Data Analysis: Profit by Genre



Profit Trends by Genre

•**Action, Adventure, comedy and drama** have the highest total profits due to global appeal and larger production/marketing budgets, leading to high box office returns.
•**Documentary, war, and western** genres show the lowest profits, reflecting limited releases, smaller audiences, and lower budgets.

Impact of Budget, Runtime, and Number of Votes



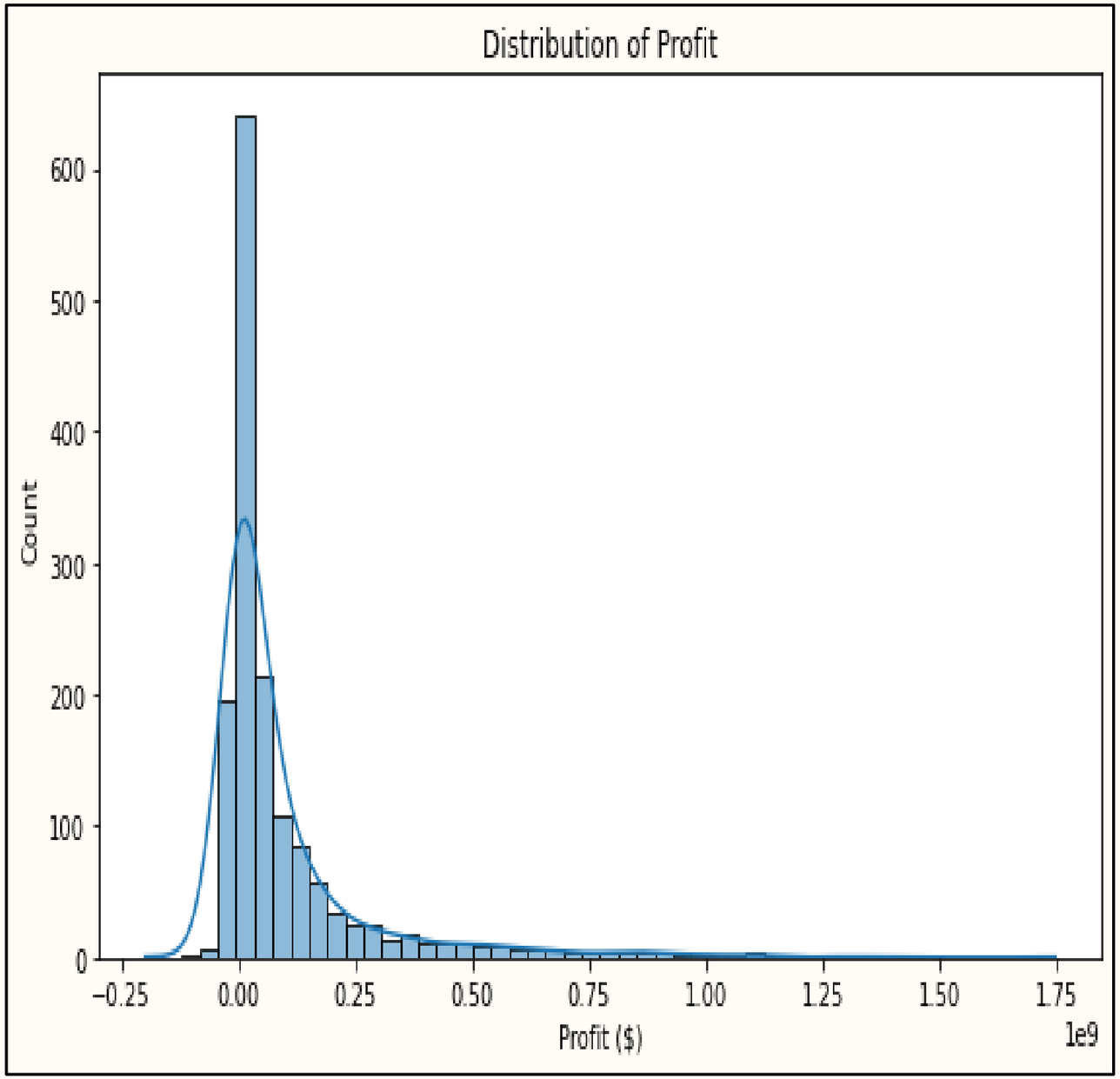
Typical Observations

•**Positive Correlation**: Higher production budgets generally lead to higher worldwide gross.

•**Upward Trend**: Scatter plots often slope upward to the right, showing that spending more usually results in greater revenue but not always proportional.

•**Right-Skewed Distribution**: A long right tail shows a few blockbusters drive most profits, typical of high-risk, high-reward industries like film.

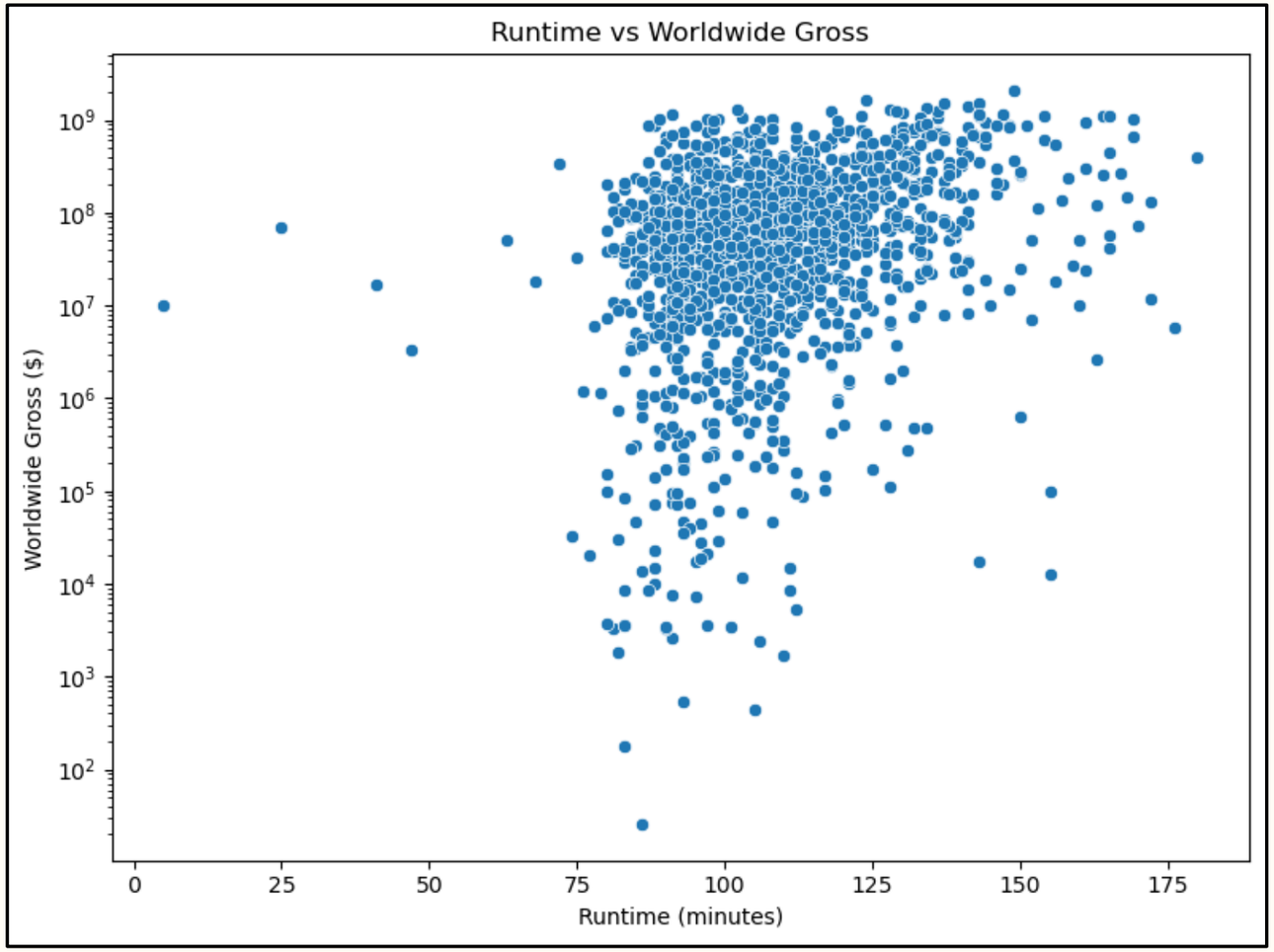
Profit Distribution



Typical Observations include:

- A sharp peak near $0 or negative values, suggests that many films either break even or lose money.
- A long tail toward the right, shows that while only a few films make huge profits, they skew the average upward.
- This is known as a right-skewed distribution, common in industries with high risk and high reward (like the film industry).

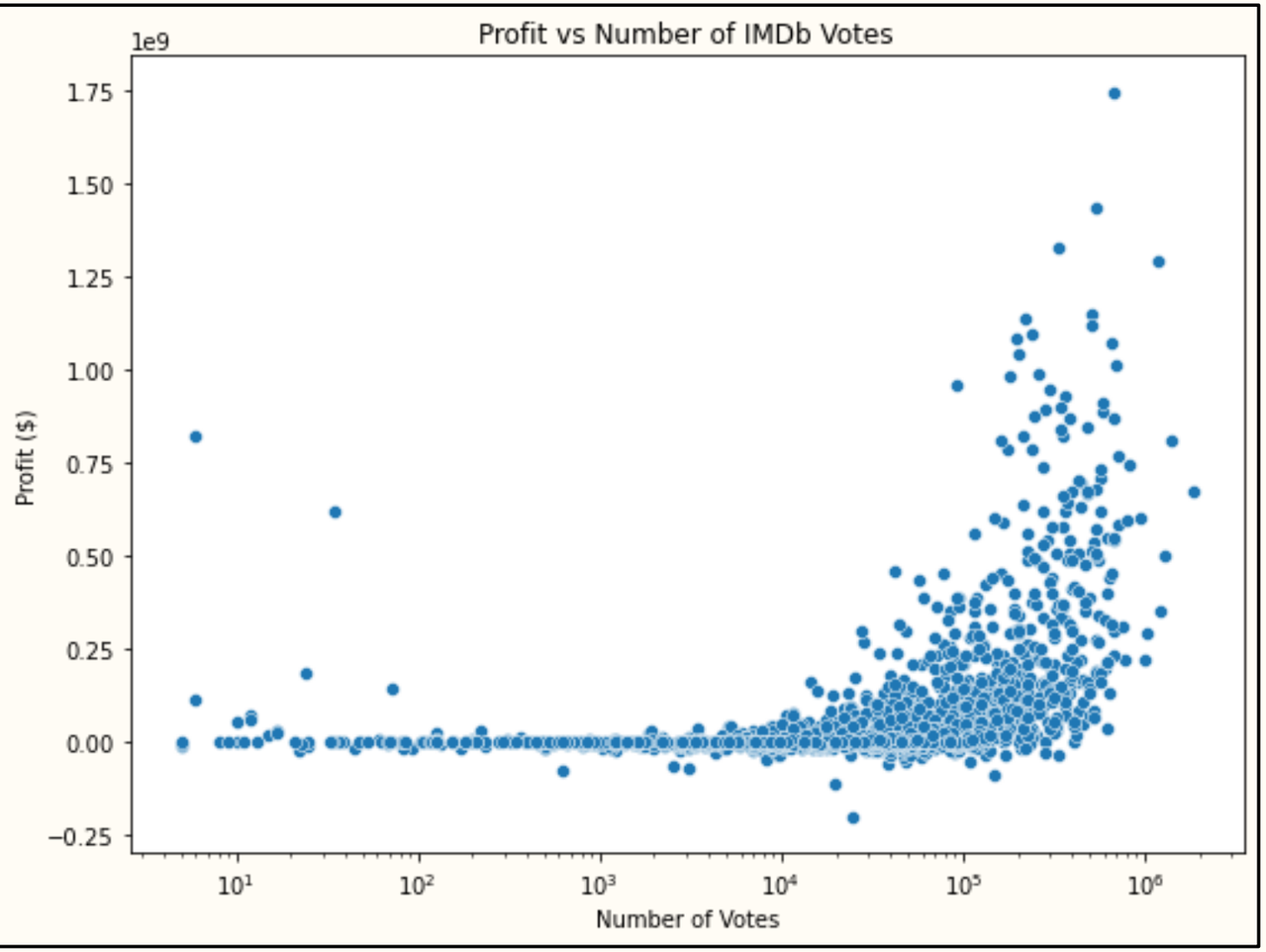
Runtime vs Worldwide Gross



From our scatter plot, we observe that:

- Scatter points are usually spread widely across all runtime values.
- A film's length alone rarely predicts financial success.
- Many high-grossing films tend to cluster around 100–140 minutes. This is the sweet spot where blockbusters, action, and adventure films typically fall.

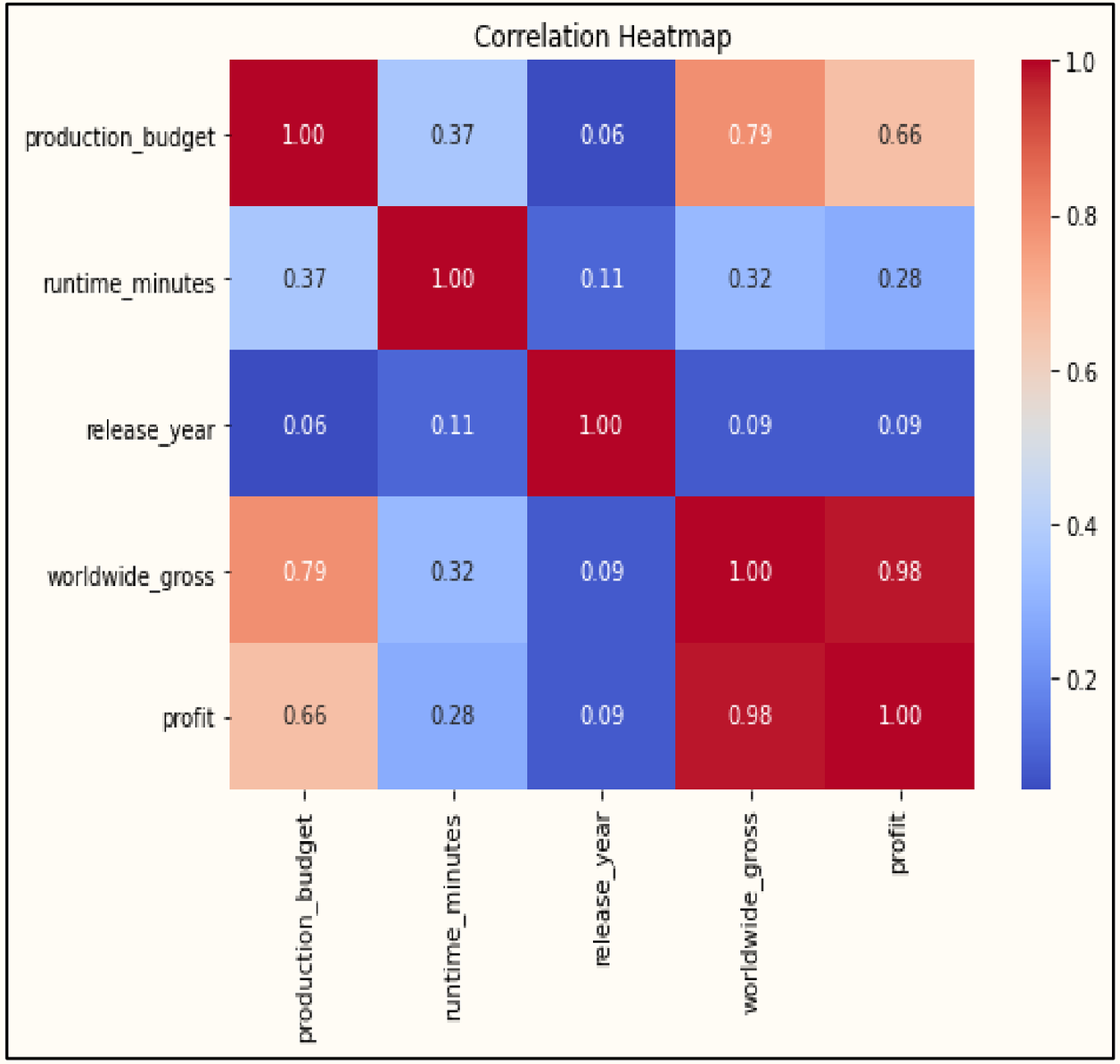
Profit vs Number of IMDb votes



From our observations, we find that:

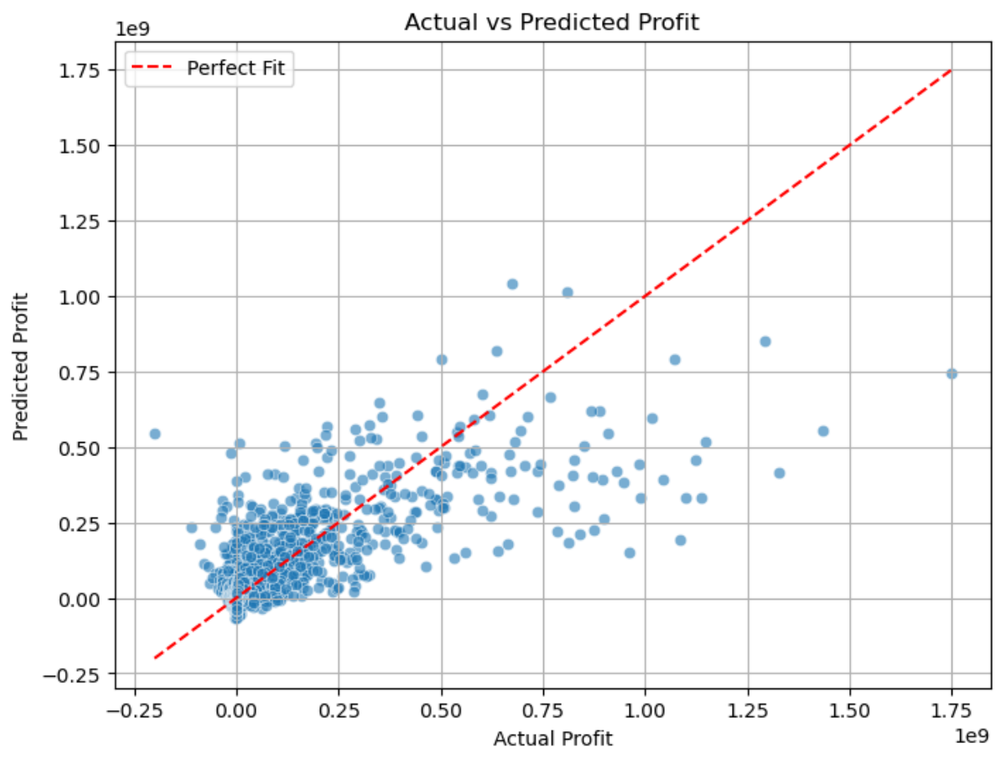
- Films with more IMDb votes tend to have higher profits.
- More audience engagement (votes) often reflects wider viewership and box office success.
- The points usually trend upwards as the number of votes increases.
- Films with fewer votes (low popularity) show mixed profit outcomes — some lose money, while some gain.
- Films with very high IMDb votes (i.e., >100,000) almost always have positive profits.

Correlation Heatmap



Variable Pair	Correlation Strength	Insight
Production Budget vs Worldwide Gross	Strong Positive(0.79)	Bigger budgets often lead to higher gross due to global releases, marketing, and big casts.
Production Budget vs Profit	Moderate Positive(0.66)	Higher budgets *may* result in higher profit, but outcomes vary based on cost and reception.
Worldwide Gross vs Profit	Very Strong Positive (0.98)	Profit is mostly driven by gross — since **Profit = Gross - Budget**.
Runtime vs Profit/Gross	Weak/No Correlation (0.28)	Runtime does not significantly impact revenue or profit.
Release Year vs Production Budget	Weak (0.06)	Slight trend suggests budgets may have gradually increased over time.

Linear Regression



Intercept: $4,925,123
R-squared: 0.546
RMSE: 129332520.437

Feature	Coefficient	Coefficient ($)
production_budget	1.62E+00	$1.62
runtime_minutes	-7.25E+05	($724,806.83)
num_votes	4.45E+02	$445.46
average_rating	7.26E+06	$7,259,763.08

The goal is to predict Movie profit using linear regression by testing multiple numeric features i.e production budget , run time, average rating and number of votes

Linear Regression Summary

Intercept: $4.93M — baseline profit if all features are zero.
R^2 = 0.55 — model explains ~55% of profit variation (moderate fit).
RMSE ≈ $129M — high error, indicating profit is influenced by many other factors.

Coefficient Insights

Budget: +$1.62 profit per $1 spent — strong positive link.
Votes: +$445 per vote — more audience engagement boosts returns.
Rating: +$7.26M per point — critical acclaim strongly impacts profit.
Runtime: -$724K per extra minute — longer films may slightly reduce profit.

Final Summary

Higher budgets, ratings, and votes boost profits, while longer runtimes may hurt. Model has moderate predictive power, but high error is expected in such a complex domain.

Hypotheses Testing Results

This is a one-tailed independent t-test comparing the average ratings of Action vs. non-Action movies.

Null Hypothesis (H_0):
Action movies do not have significantly higher average ratings than non-Action movies.

Alternative Hypothesis (H_1):
Action movies have significantly higher average ratings than non-Action movies.

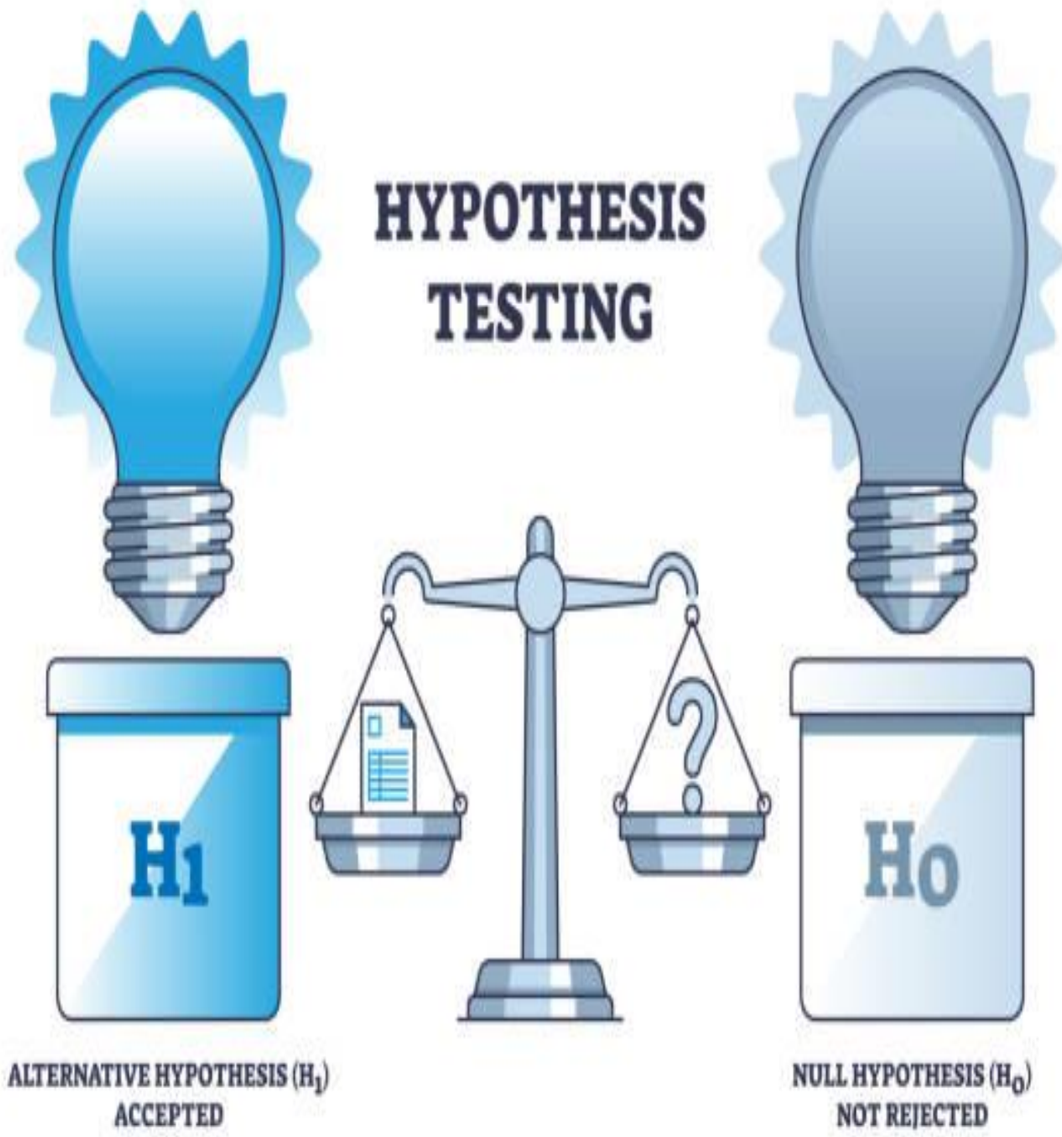
Hypotheses Testing Results

T-statistic: -0.7174 One-tailed P-value: 0.2367

Fail to reject the null hypothesis: No significant evidence that Action movies have higher ratings.

The p-value is less than alpha and the t-statistic is positive (i.e., Action movies have a higher mean), it rejects the null hypothesis, indicating statistical evidence that Action movies are rated higher.

Otherwise, it fails to reject the null, meaning there's no strong evidence for a difference in the expected direction.



Recommendations for Flix Company

Genre Selection

Focus on Action, Adventure, and Sci-Fi genres, which consistently show the highest gross and profit.

Budget Planning

Allocate production budgets based on historically successful investment ranges, considering that higher budgets generally earn more, but profit must be carefully managed.

Optimal Runtime

Optimal runtimes appear around 120-150 minutes for blockbuster success.

Hype

Build hype early to raise IMDb vote count, improving engagement and potentially profit.

Thank You!

We value your input! Please take a moment to provide feedback on the analysis using the following questions:

1. **Did the analysis of the data resonate with the company's goal?**

2. **Any other insight on the data that you find useful?**

"Thank you for your valuable feedback! Your input will help us refine the analysis.

Contacts

1. **Irene.Kibengo@student.moringaschool.com**
2. **Erastus.Njuguna@student.moringaschool.com**
3. **Benson.Mwihia@student.moringaschool.com**
4. **Daniel.Akwabi@student.moringaschool.com**
5. **Luciana.Ndanu@student.moringaschool.com**
6. **Sydney.Were@student.moringaschool.com**