Do Big Budgets Mean Big Profits? Analyzing Movie ROI and Ratings

S.M.Amin Mohammadi + Guillermo Fiallo-Montero

DataScience: Mini Project - SQL: From Data to Insight

Project Overview

1 Datasets

Kaggle Movies Metadata (23k movies) + IMDb Ratings.

2 Business Problem

How can studios allocate budgets wisely to maximize profits?

3 Hypotheses

Higher budgets → Higher profits?

Genres like Action/Adventure yield the best ROI.

Higher IMDb ratings → Higher ROI

Modern Films (2000s-2010s) → Higher Budgets → Lower ROI

Data Acquisition & Challenges

Data Pipeline

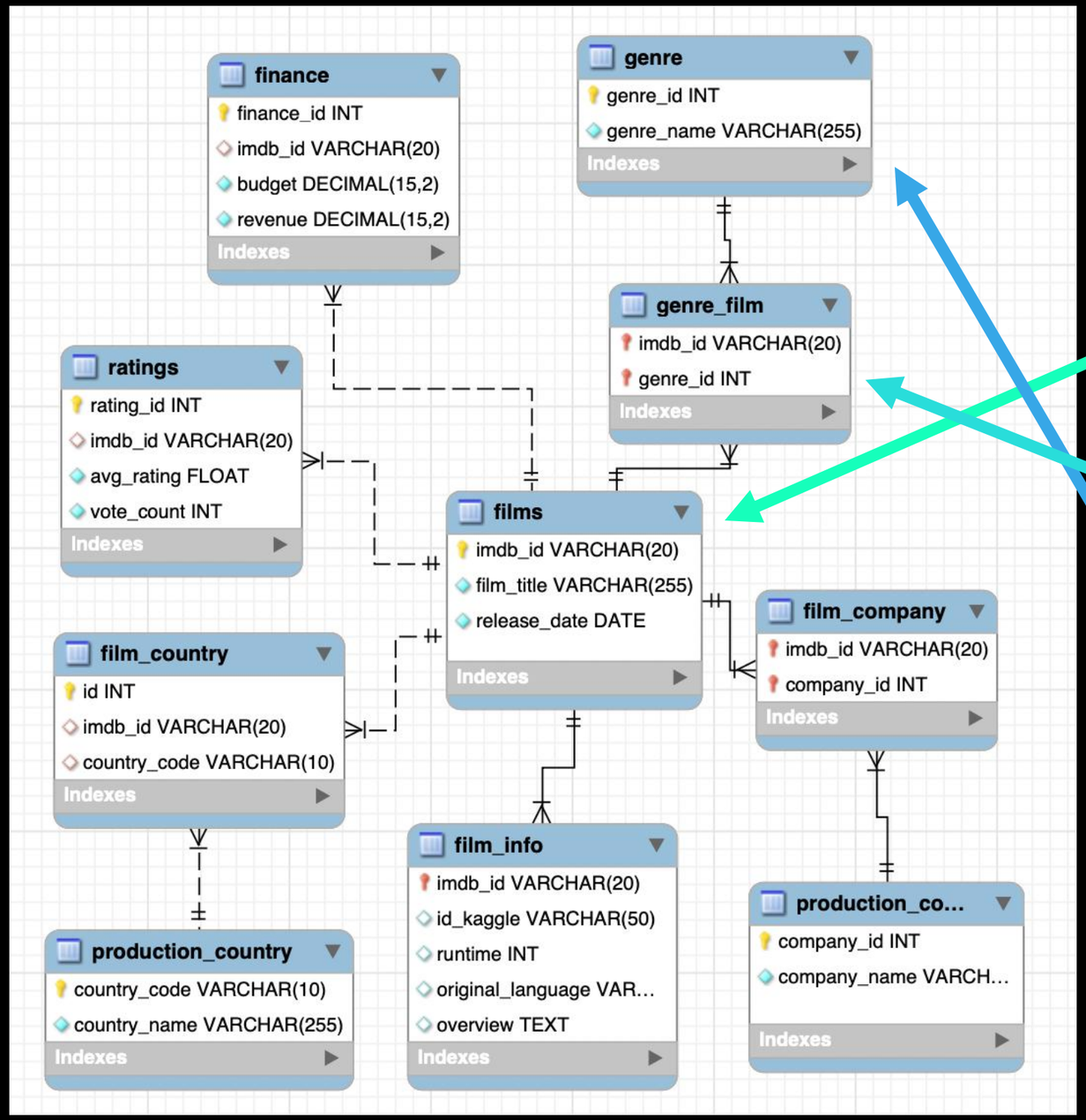
Raw CSV/TSV → Cleaned (Python) → MySQL Database.

Key Challenges

JSON-formatted genres, encoding issues.

Complementary Datasets

IMDb ratings enriched profitability analysis.



Database Design (ERD)

films Table

Central node.

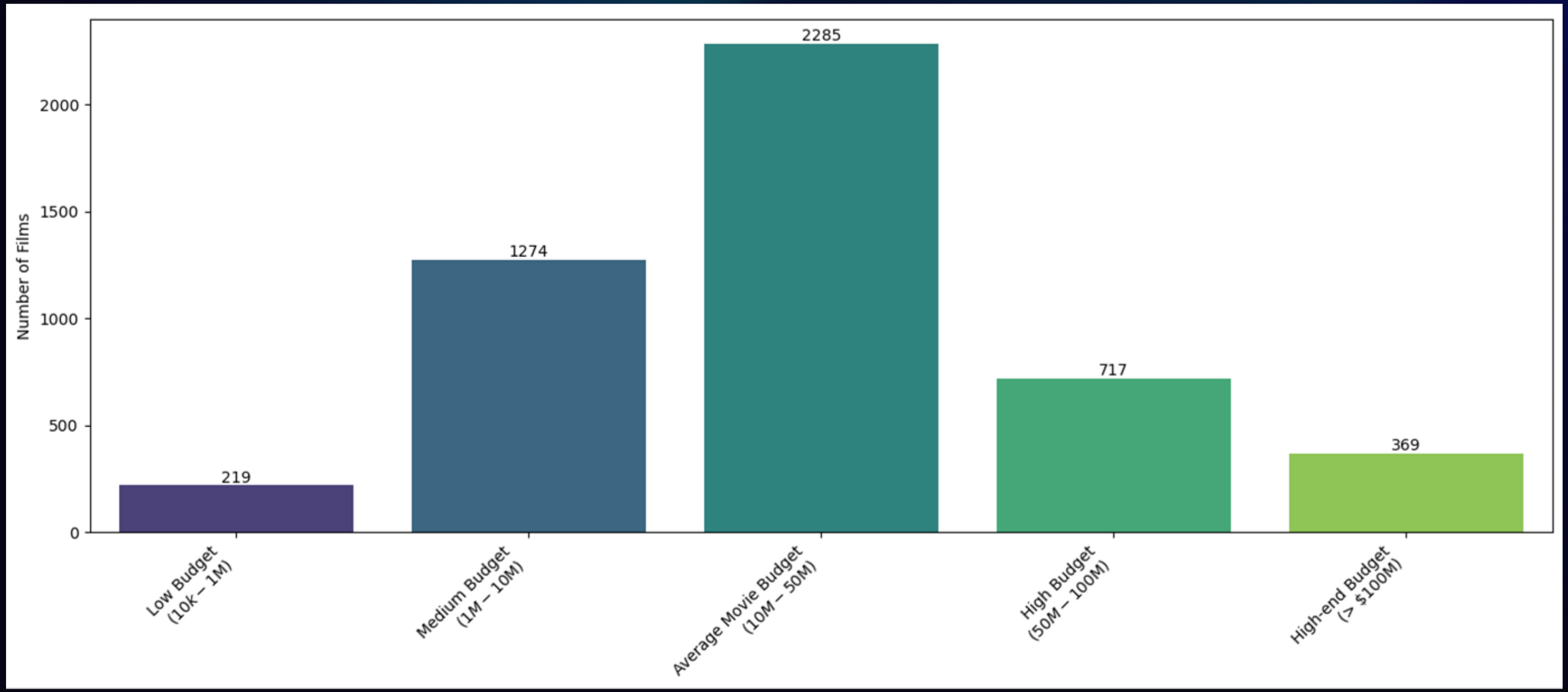
genres_film Bridge Table

Many-to-many relationships.

Key Decision

Separating genres allowed flexible analysis.

Budget Tiers – Distribution



SQL Insights (Budget Efficiency)

Key Query Example

```
SELECT budget_tier,
ROUND(AVG((revenue - budget) / budget), 2)
AS avg_roi
FROM film_fin_temp
GROUP BY budget_tier;
```

Result

High-end Budget	1.95 avg ROI
Low Budget:	1.93 avg ROI
Medium Budget	1.57 avg ROI
Average Budget	1.29 avg ROI
High Budget	1.26 avg ROI

Recap of Questions

"Do higher budgets guarantee higher profits?" 💰

🎬 *"Which genres give the best ROI?*

Visualization 1 – Budget vs. ROI

1

Plot

Budget Distribution by tiers

Budget Tiers – Distribution & Risk

Low Budget:

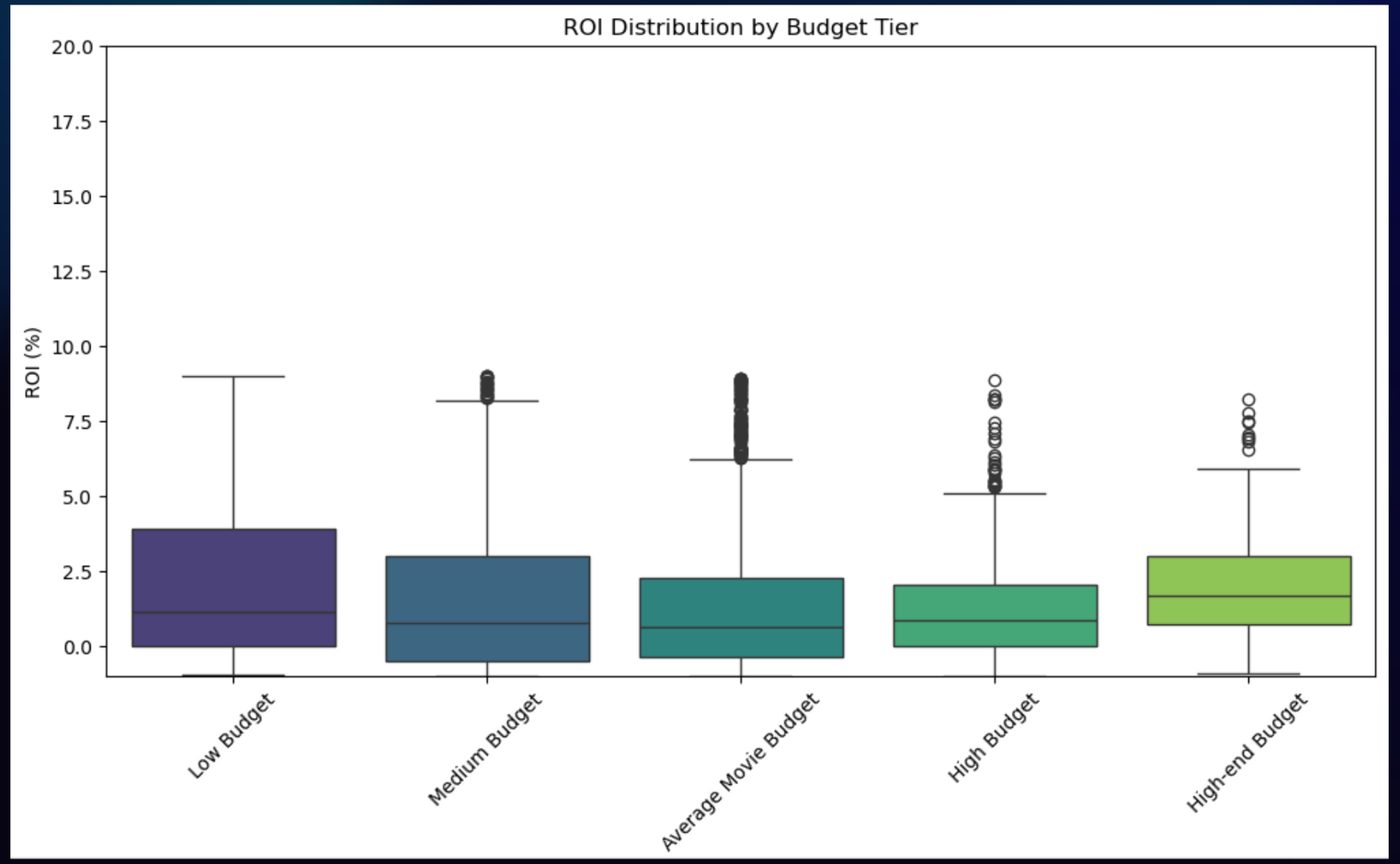
Wide spread = High risk, high reward.

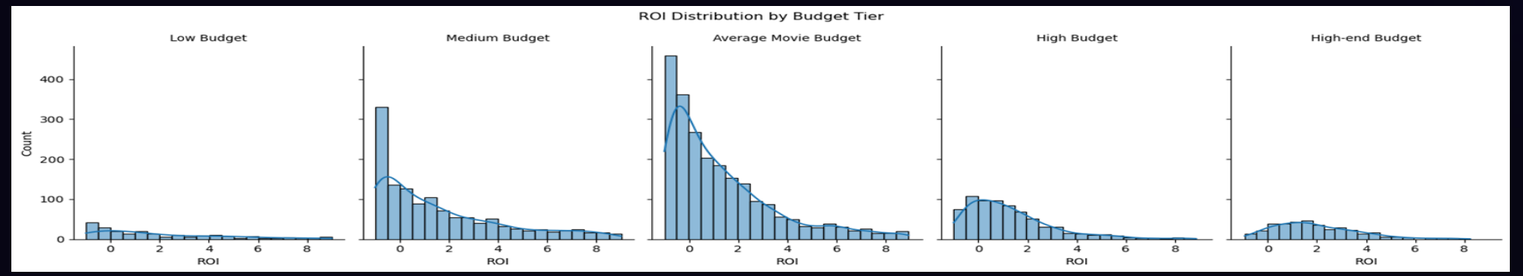
High Budget:

Tight cluster = Safer & high ROI.

Mid Budget:

Many flops (negative ROI).





Visualization 1 – Budget vs. ROI

1

Plot

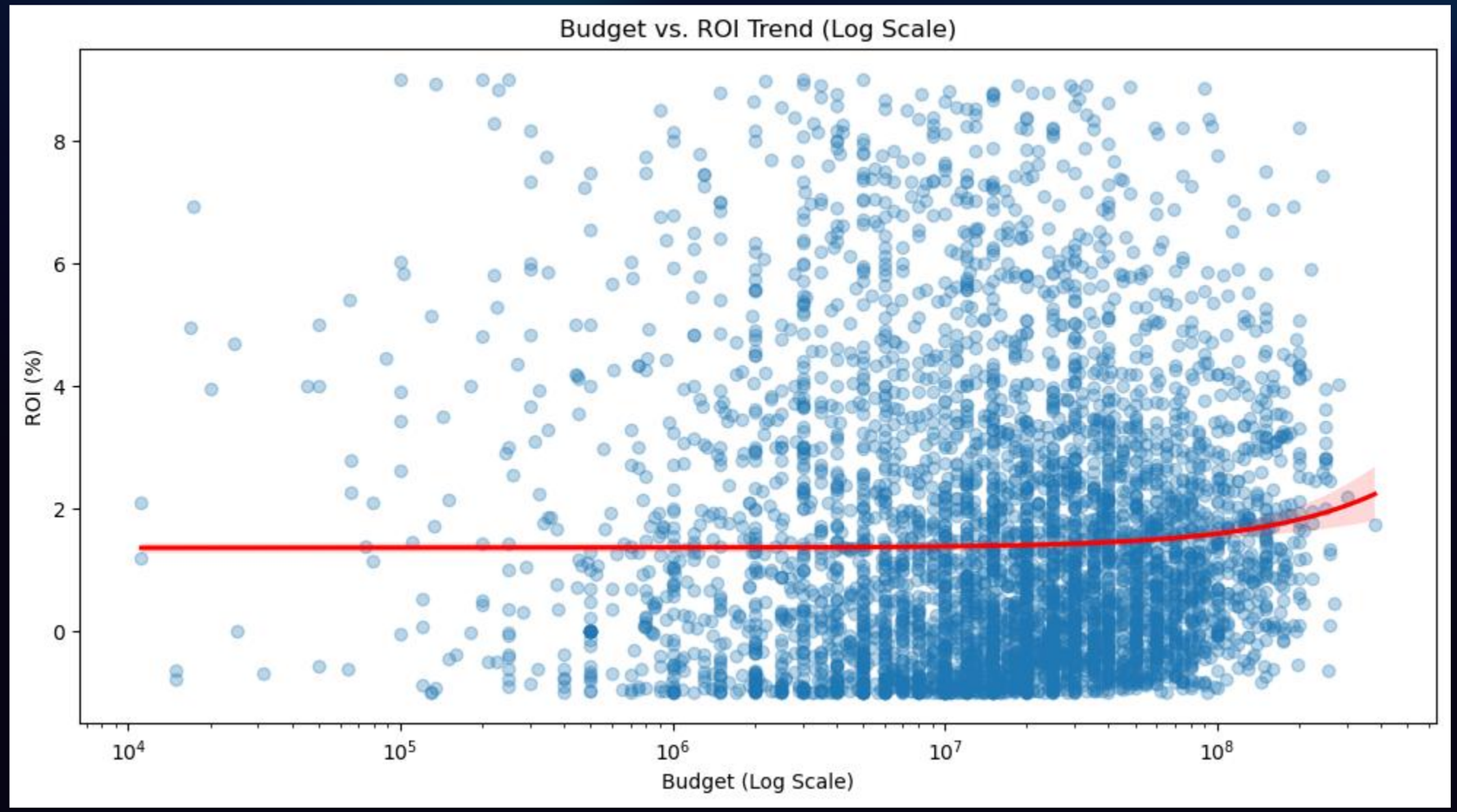
Budget Distribution by tiers

2

Budget vs. ROI

Horizontal trend line with few outliers

Budget vs. ROI – The Reality



Visualization 1 – Budget vs. ROI

1

Plot

Budget Distribution by tiers

2

Budget vs. ROI

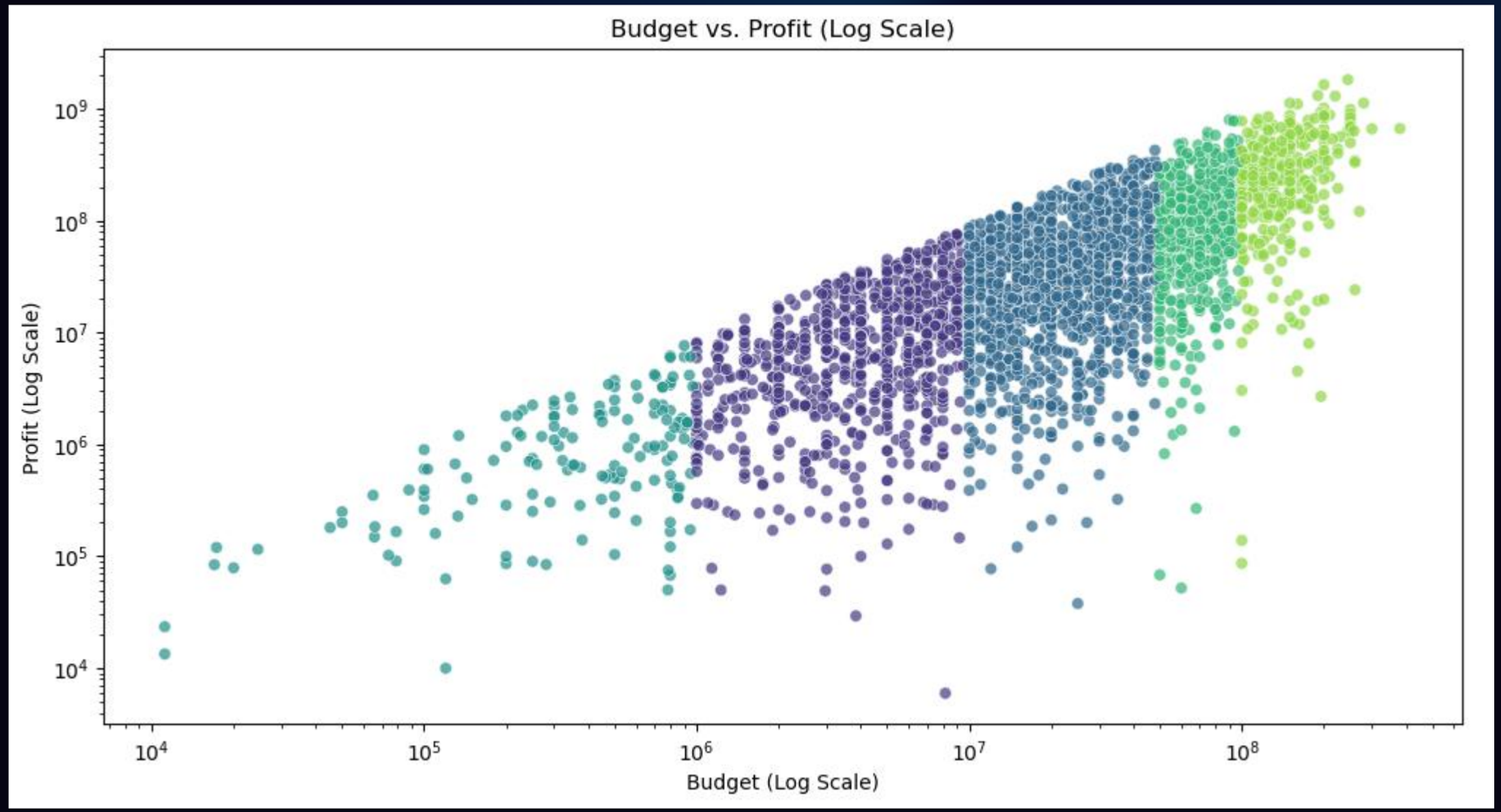
Horizontal trend line with few outliers

3

Budget vs. Profit

Showing positive correlation.

Budget vs. Profit



High Budget	$ 299,9M	195 % avg ROI
Low Budget:	$ 788k	193 % avg ROI

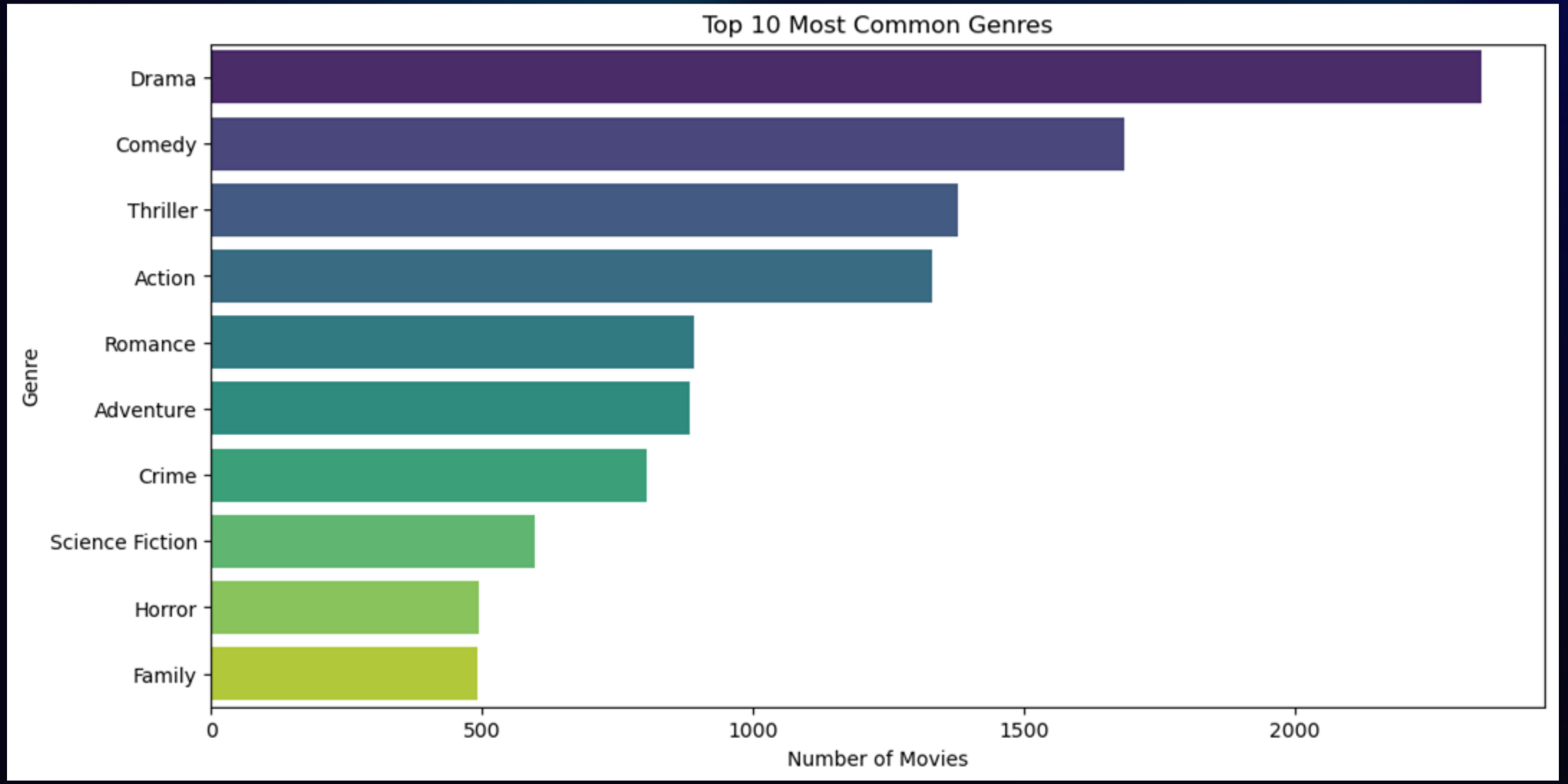
Visualization 2 – Top Genres by Number of Movies

1

Plot

Top 10 Genres by Number of Movies (bar chart).

Top 10 Genres by Number of Movies



Visualization 2 – Top Genres by ROI

1

Plot

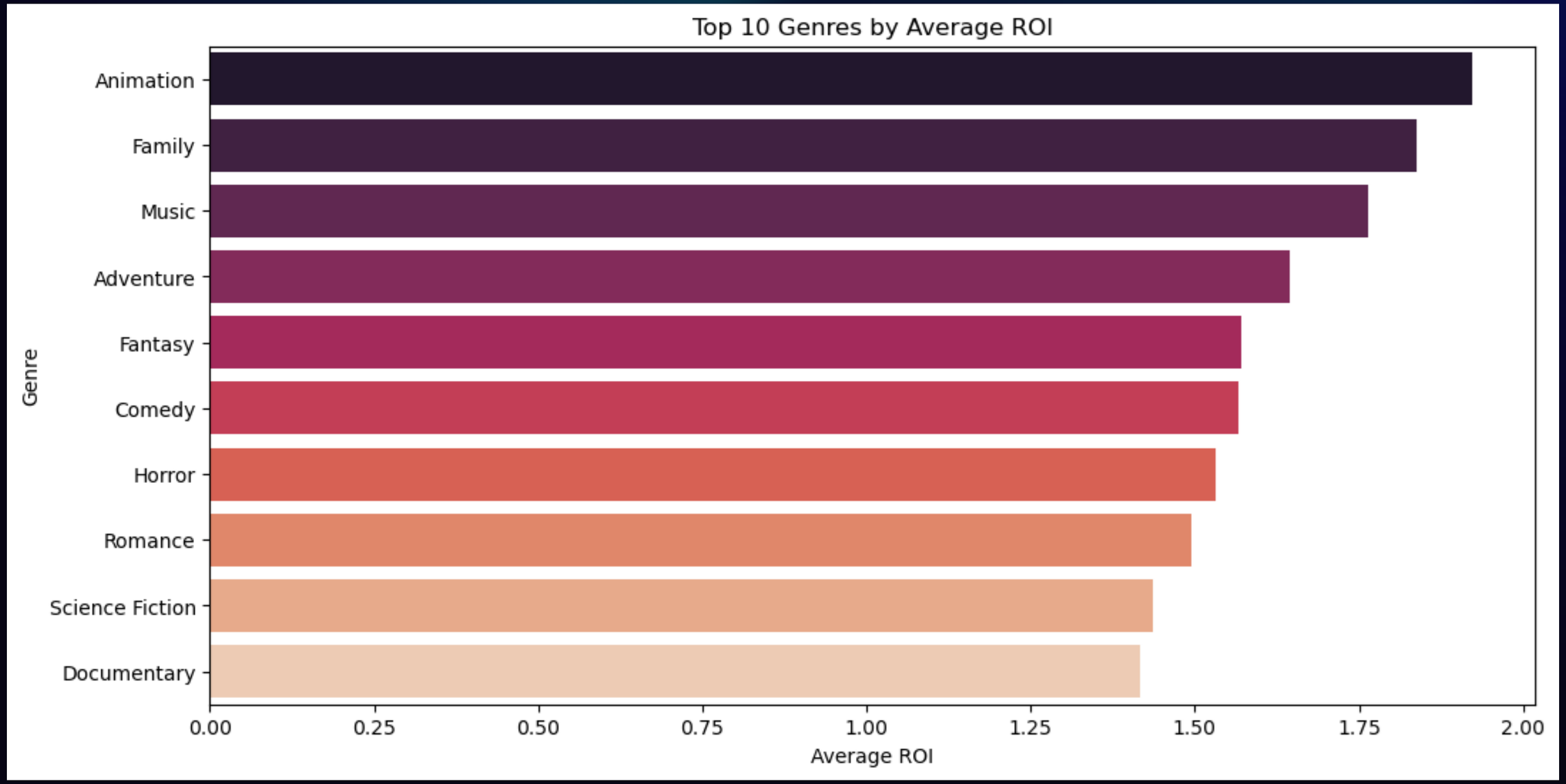
Top 10 Genres by Average ROI (bar chart).

2

Highlight

Animation (1.92), Family (1.84), Adventure (1.64)

Top Genres by ROI



Visualization 2 – Top Genres by ROI

1

Plot

Top 10 Genres by Average ROI (bar chart).

2

Highlight

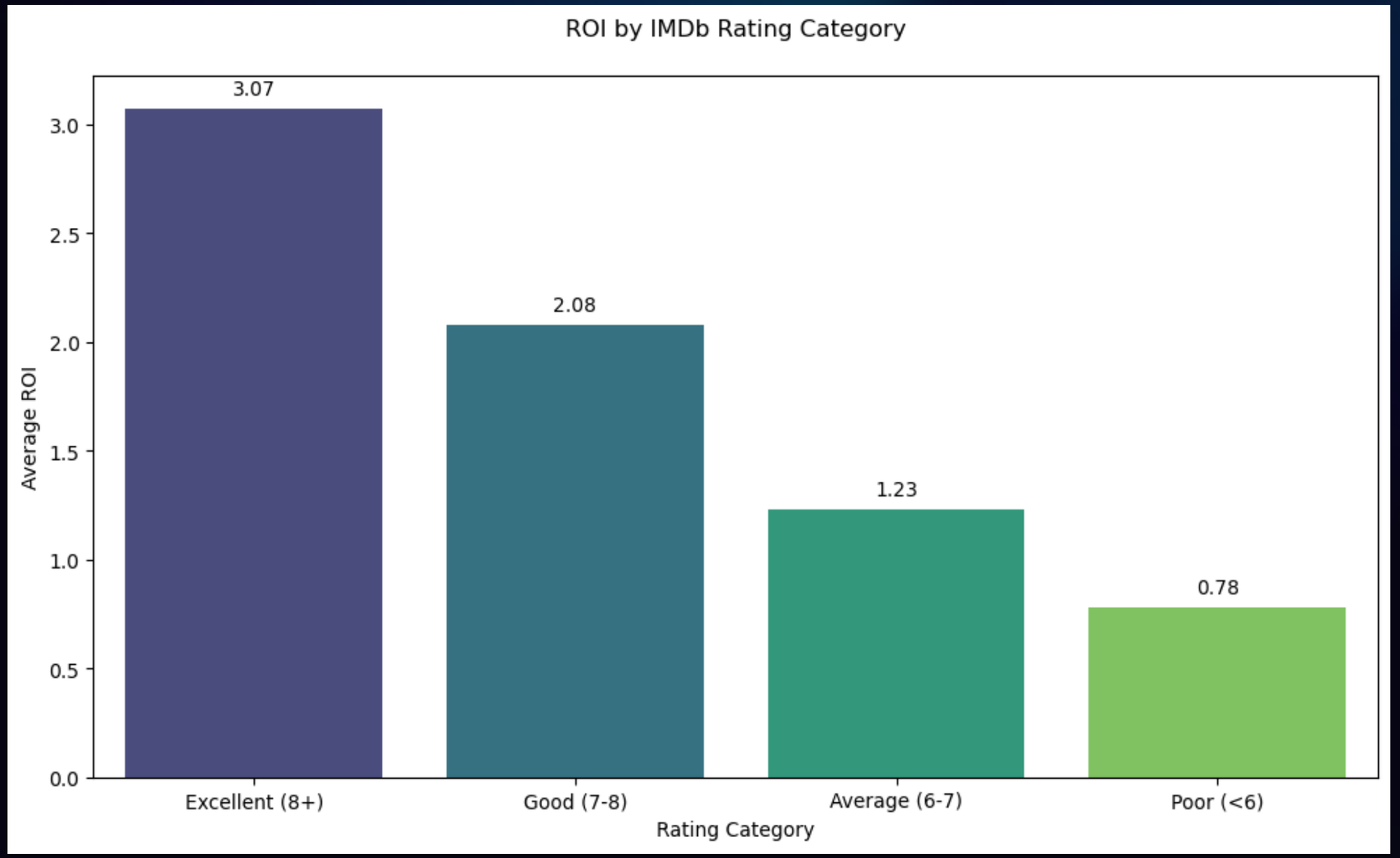
Animation (1.92), Family (1.84), Adventure (1.64)

3

Callout

Family films combine profitability.

IMDb Ratings & ROI



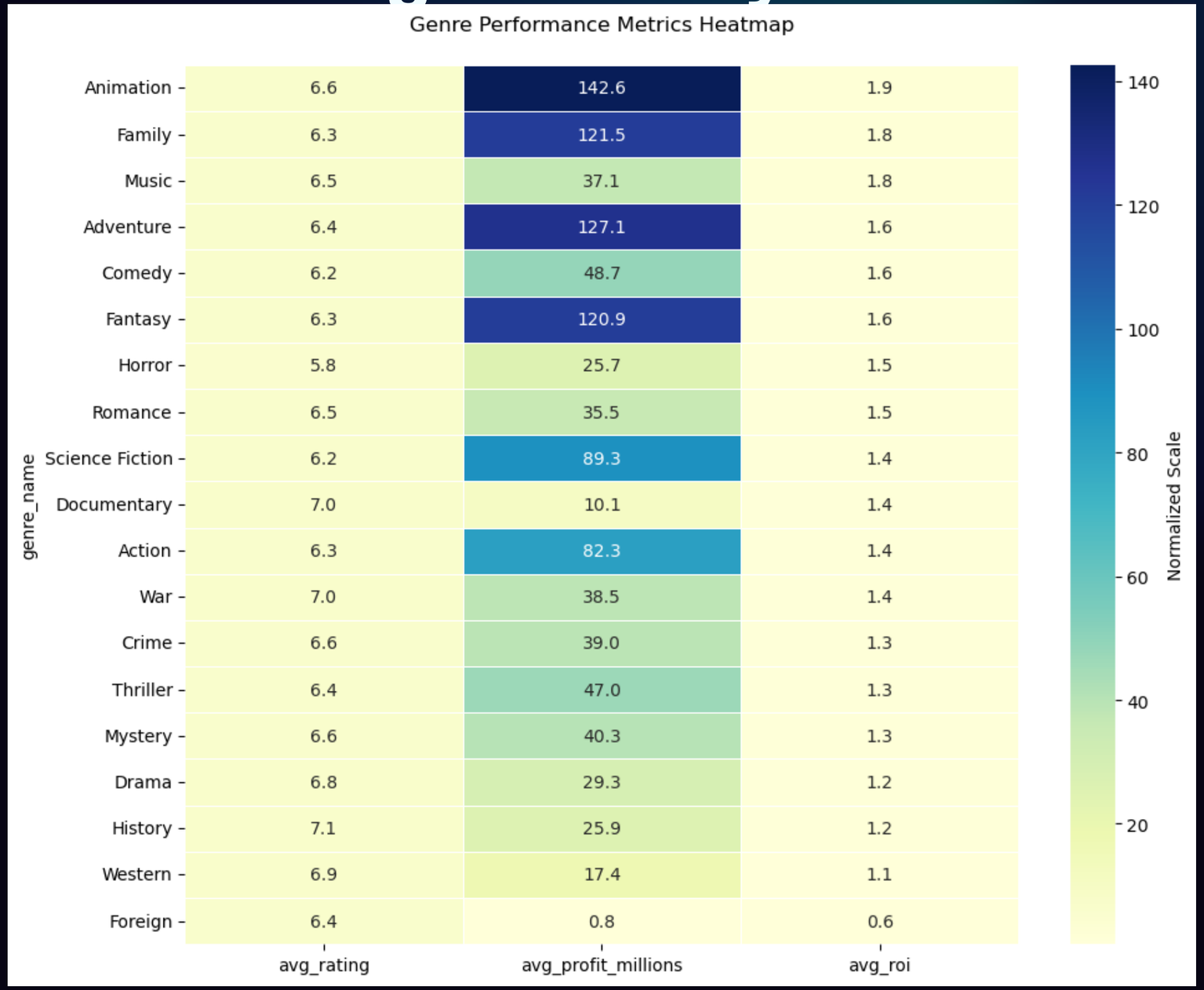
Key Takeaway:

"Films rated 8+ generate 3.1x ROI (vs. 0.8x for <6), confirming that quality drives profitability."

Recommendation:

Studios should focus on producing 7+ rated film- they deliver 2-3x higher ROI than sub-6 rated ones.

IMDb Ratings & ROI by Genre



Key Takeaway:

1. **Quality ≠ Profitability**

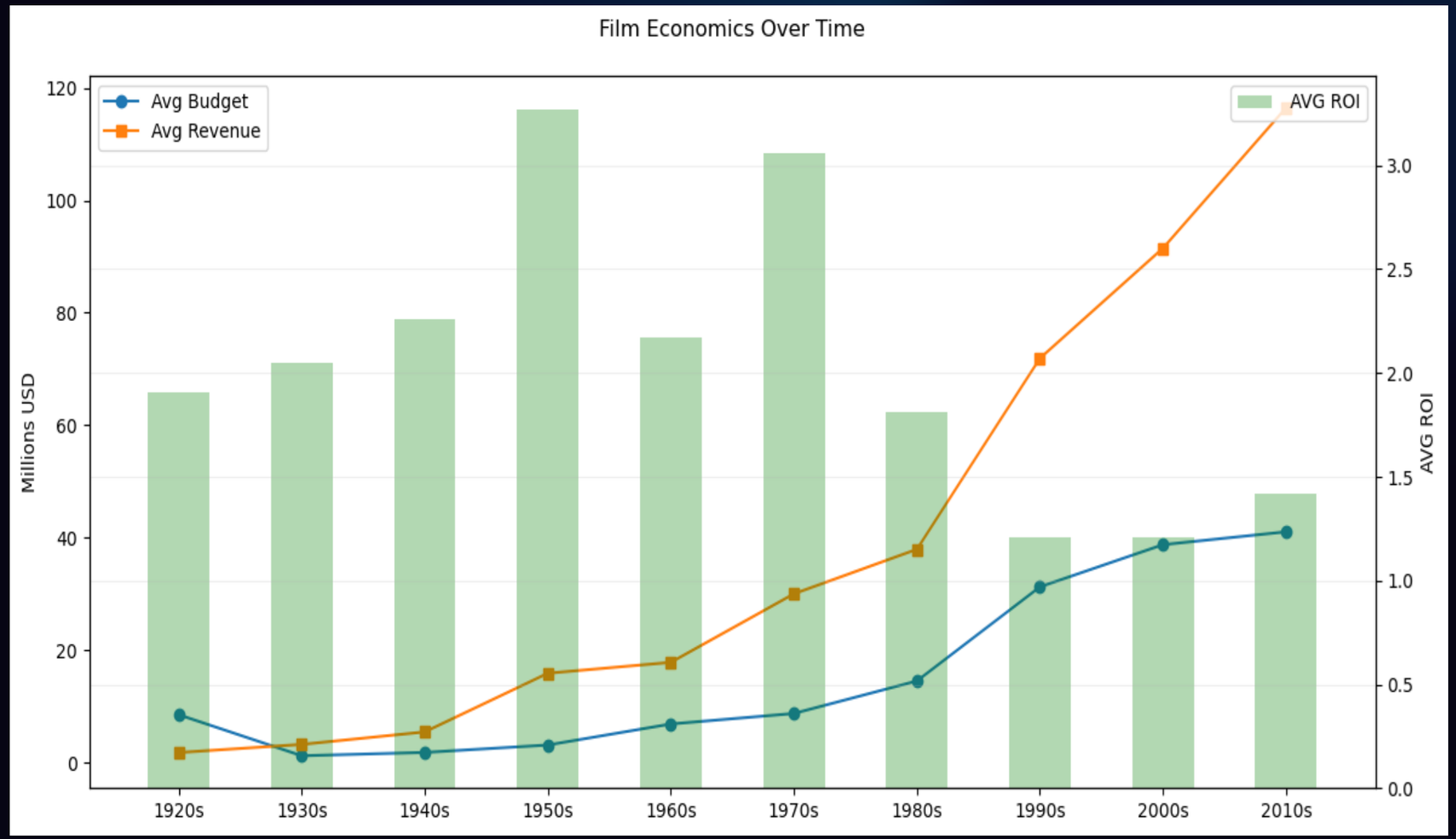
   Top-rated genres (Documentary/War at 7.0) underperform high ROI genres (Animation at 6.6)

1. **Best ROI Bets**
   - Animation: 1.9x ROI ($142M profit)
   - Family: 1.8x ROI ($121M profit)
   - Music: 1.8x ROI (Efficient niche)

Recommendation:

Studios should prioritize Animation/Family films—they deliver blockbuster ROI (1.8-1.9x) without requiring 7+ ratings.

Decoding 100 Years of Film Economics: Budgets, Revenue & ROI Trends



Key Trends:

- 1950s Peak: 3M→16M (4.1x ROI)
- Modern Era: 41M→116M (1.8x RO)

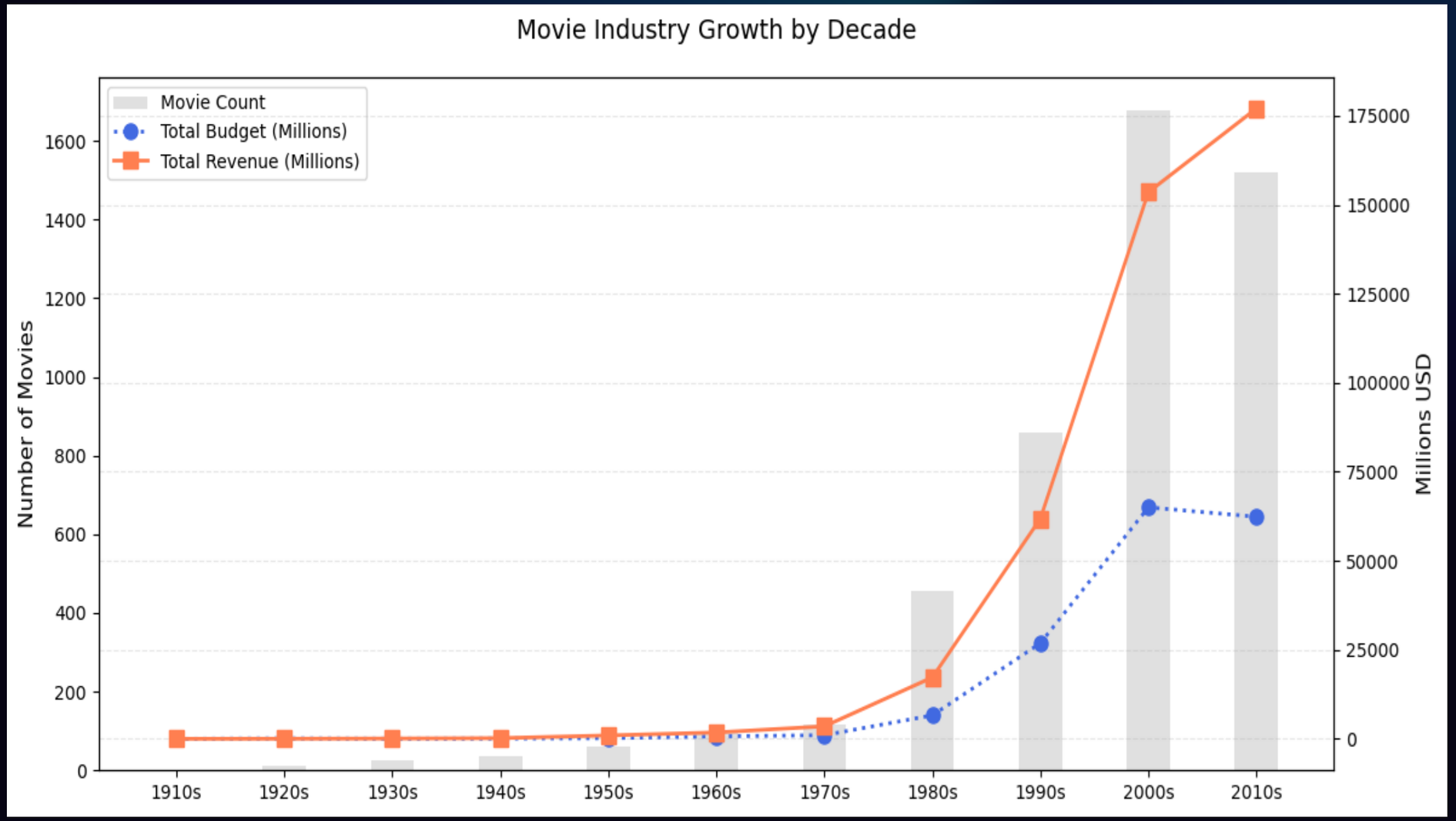
Conclusion:

1. Budgets ↑ 32% (2000s: 38.7M→2010s:41M)
2. ROI ↓ 56% vs. 1950s (4.1x → 1.8x)
3. Profit Margins Shrank:
   - 1950s: 1 budget→5 revenue
   - 2010s: 1 budget→2.8 revenue

Critical Nuance:

- Post-2000 ROI is stable (1.4x–1.8x) due to franchises/globalization.

Movie Industry Growth: Trends & Insights



Key Insights:

📈 Movie production peaked in the 2000s.

💰 Budgets surged after the 1980s.

🎬 Blockbusters drove massive revenue growth. (1990s and 2000s)

📉 2010s: Fewer movies, higher revenues (streaming impact).

⚖️ Revenues outpaced budgets—better monetization.

Challenges & Lessons Learned

Biggest Obstacle

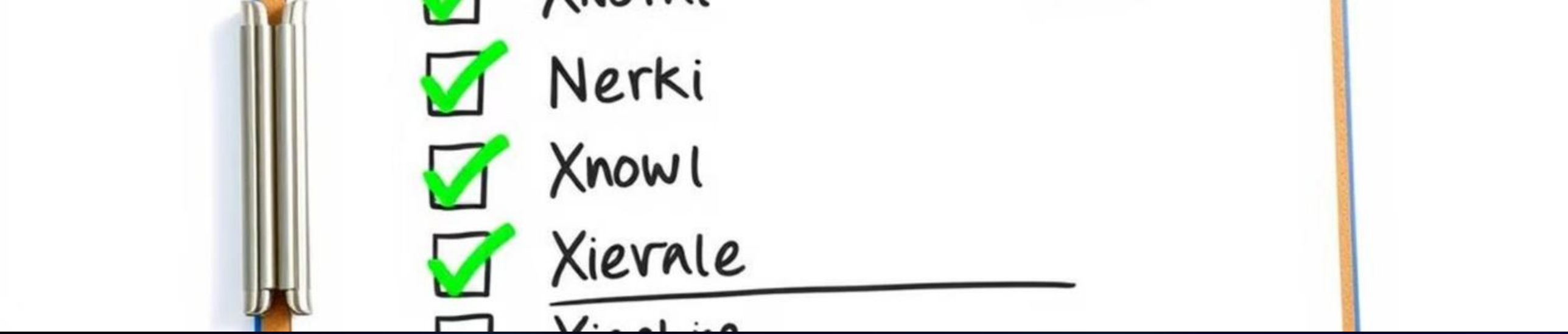
Untangling JSON genres.

Solution

Python scripts to flatten JSON.

Lesson

Clean data early!



Conclusions & Business Implications

✅ Higher budgets → Higher profits.

✅ Animation/Family = Best ROI.

✅ Creativity beats budget size.

Thank You!

S.M.Amin Mohammadi

Guillermo Fiallo-Montero

GitHub repo QR code