

Movie Data Visualization with Tableau

by Grit, 2018.02

1. Overview

In this project, we'll explore data from The Movie Database to draw out and display interesting patterns. Let's start by defining some leading questions:

Question 1: How have movie genres changed over time? (Budgeted over \$100m)

Question 2: How do the attributes differ between Universal Pictures and Paramount Pictures?
(Budgeted over \$100m)

Question 3: How have movies based on novels performed relative to movies not based on novels?
(Budgeted over \$100m)

Question 4: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

Then we'll explore the data to find answers, which will be all displayed by dashboards.

For Q4, in addition to creating a dashboard, we'll create a narrative told through visualizations and text using Tableau Stories feature.

2. Data Wrangling with Pandas, NumPy and IPython

The raw data should be handled with Python modules and packages:

- Dealing with duplicates
- Removing the 'Unnamed' columns
- Normalizing the data types and formats
- Comparing the two "release" columns to make them match
- Handling the Null/NA/Zero values
- Saving the tidy dataset as a CSV file

[Here is the codes for data cleanup.](#)

3. Data Visualization with Tableau

The following pictures are our dashboards and story. You can check the [on-line Tableau dashboard](#) or download the [original Tableau file](#) for more details. However, the layout of Question 1 dashboard may be disorganized, and the layout from the [original Tableau file](#) is okay.

4. Summary

The answers to each question, which have been embedded in the charts, will be listed as follows:

Question 1: How have movie genres changed over time? (Budgeted over \$100m)

- The 5 most popular genres are Drama, Comedy, Thriller, Action and Adventure, and the percentage of total is 60%. Among the top genres, Drama is always the most popular one, with the average percentage of 17%.
- Clearly for all genres, Q1 is the weakest of the year, followed by a full rise in Q2, with some genres, such as Action and Adventure, reaching their peak. Then the differences come out in Q3, which is the outbreak period for Thriller, but a growth period before the next for Drama. However, on the whole, Q3 is still the prime of the year with a slight advantage over Q4.

Question 2: How do the attributes differ between Universal Pictures and Paramount Pictures? (Budgeted over \$100m)

- Both studios have shown growth in production, despite the lack of stability in the early years. However, Universal Pictures has got a bit advantage over its rival since around 2005.
- The movies, located in whether "In - Top Genres" or "Out- Top Genres", bring Paramount Pictures higher median profit.

Question 3: How have movies based on novels performed relative to movies not based on novels? (Budgeted over \$100m)

- Compared with movies not based on novel, although the number of movies adapted is only 5% of the latter, the average rating is 0.3 higher, and the median profit is actually 31% higher than the latter.
- In almost all genres, except for comedies, the movies adapted from novels have a higher median profit.

Question 4 - Story: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

- The median or average rating is 6.2. The median profit of profitable movies is \$66.24m.
- The Pearson correlation between the ratings and profits is 0.26, implying a weak positive linear relationship between the two variables.
- Among the non-profitable movies, the movies above the average rating accounts for 68%, twice the size of ones not higher than the average rating. And the top grossing movies with more than \$800m, basically have a rating above the average.
- 27% of releases are in the "High Profits, High Ratings" category, but they make up 67% of profit.

Q1: How have movie genres changed over time? (Budgeted over \$100m)

1. Genre Ranking By Years

| Release_year | Simple_genre | | | | | | | | | | | | | | | | | | | TV Movie |
|--------------|--------------|--------|----------|--------|-----------|-------|---------|-----------------|--------|--------|---------|---------|-----------|-------|---------|-----|---------|-------------|---------|----------|
| | Drama | Comedy | Thriller | Action | Adventure | Crime | Romance | Science Fiction | Horror | Family | Fantasy | Mystery | Animation | Music | History | War | Western | Documentary | Foreign | |
| 2015 | 19% | 13% | 12% | 10% | 9% | 6% | 5% | 5% | 4% | 4% | 3% | 3% | 2% | 2% | 1% | 1% | 1% | | | |
| 2010 | 18% | 13% | 12% | 12% | 7% | 6% | 6% | 5% | 4% | 4% | 4% | 3% | 3% | 1% | 1% | 1% | 0% | 0% | 0% | |
| 2005 | 17% | 13% | 12% | 9% | 7% | 6% | 7% | 4% | 4% | 5% | 4% | 4% | 2% | 1% | 1% | 1% | 0% | 0% | 0% | |
| 2000 | 15% | 14% | 12% | 11% | 7% | 7% | 7% | 4% | 3% | 4% | 3% | 4% | 2% | 1% | 1% | 1% | 1% | 0% | 0% | |
| 1995 | 17% | 13% | 13% | 10% | 7% | 7% | 7% | 6% | 4% | 3% | 4% | 5% | 1% | 1% | 1% | 1% | 0% | 0% | | 0% |
| 1990 | 15% | 12% | 13% | 12% | 7% | 8% | 6% | 5% | 4% | 5% | 4% | 3% | 1% | 1% | 1% | 1% | 1% | 0% | 0% | |
| 1985 | 14% | 16% | 10% | 10% | 8% | 6% | 6% | 6% | 8% | 4% | 5% | 2% | 1% | 1% | 1% | 2% | | | 0% | |
| 1980 | 13% | 10% | 10% | 11% | 8% | 4% | 7% | 9% | 8% | 3% | 6% | 3% | 1% | 3% | 3% | 1% | 1% | 0% | 0% | |
| 1975 | 18% | 7% | 12% | 10% | 11% | 4% | 3% | 9% | 7% | 3% | 4% | 3% | 2% | 4% | 2% | 3% | 1% | | | |
| 1970 | 20% | 7% | 13% | 15% | 7% | 13% | 4% | 6% | 1% | 2% | 1% | 4% | 1% | 1% | 2% | 2% | 2% | | | |
| 1965 | 19% | 6% | 7% | 10% | 14% | 6% | 8% | 4% | 1% | 4% | 1% | 5% | 1% | 4% | 3% | 5% | 4% | | | |
| 1960 | 23% | 10% | 8% | 10% | 10% | 3% | 6% | | 3% | 4% | 1% | 2% | 1% | 4% | 6% | 4% | 4% | | | |
| Grand.. | 17% | 13% | 12% | 11% | 7% | 6% | 6% | 5% | 4% | 4% | 4% | 3% | 2% | 1% | 1% | 1% | 1% | 0% | 0% | 0% |

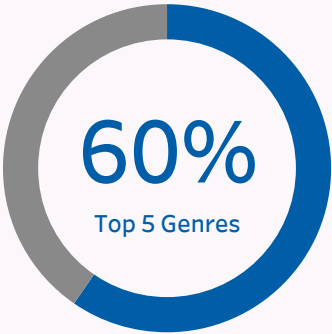
Ranking of genres over y..
1 20

Simple_genre
All

Ranking of genres over y..
To 20

Top N Genres
5

Highlight Release_year (..
No items highlighted

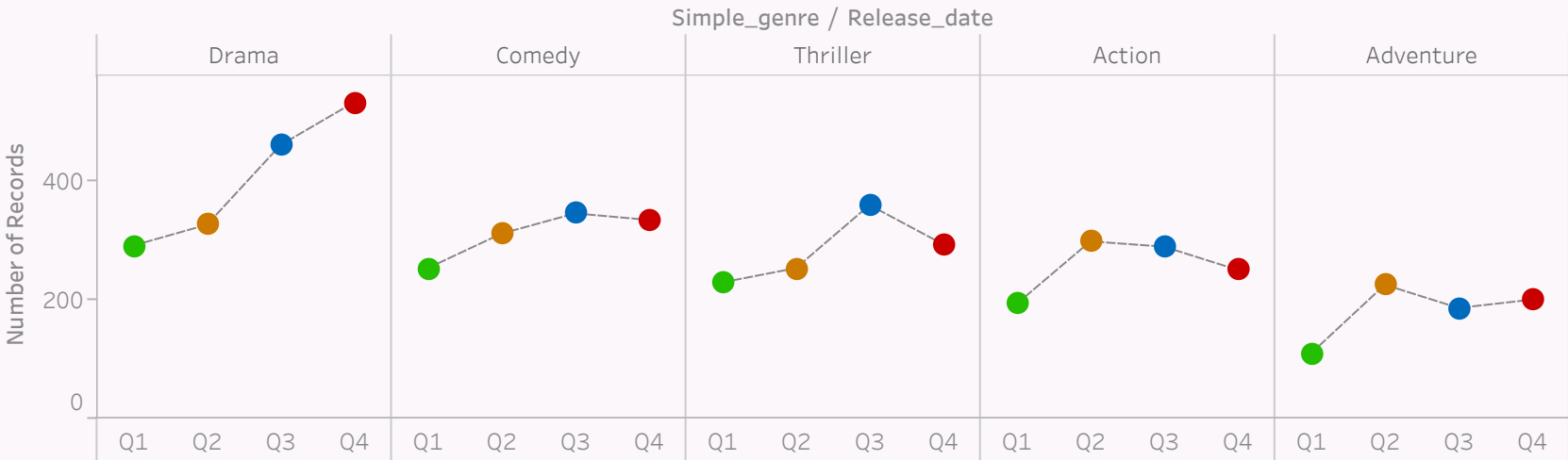


The 5 most popular genres are **Drama, Comedy, Thriller, Action and Adventure**, and the percentage of total is **60%**. Among the top genres, **Drama** is always the most popular one, with the average percentage of **17%**.

Q3

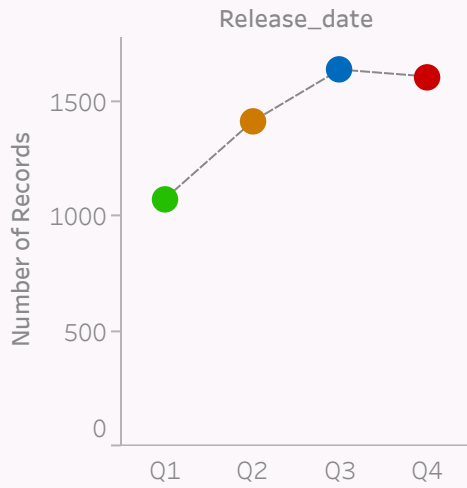
Clearly for all genres, **Q1** is the weakest of the year, followed by a full rise in Q2, with some genres, such as **Action and Adventure**, reaching their peak. Then the differences come out in Q3, which is the outbreak period for Thriller, but a growth period before the next for Drama. However, on the whole, **Q3 is still the prime of the year** with a slight advantage over Q4.

2. Numbers By Genres & Quarters



Release Quarter
Q1 Q2 Q3 Q4

3. Numbers by Quarters





VS



Q2: How do the attributes differ between Universal Pictures and Paramount Pictures? (Budgeted over \$100m)

Paramount, Out - Top Genres Paramount, In - Top Genres Universal, In - Top Genres Universal, Out - Top Genres

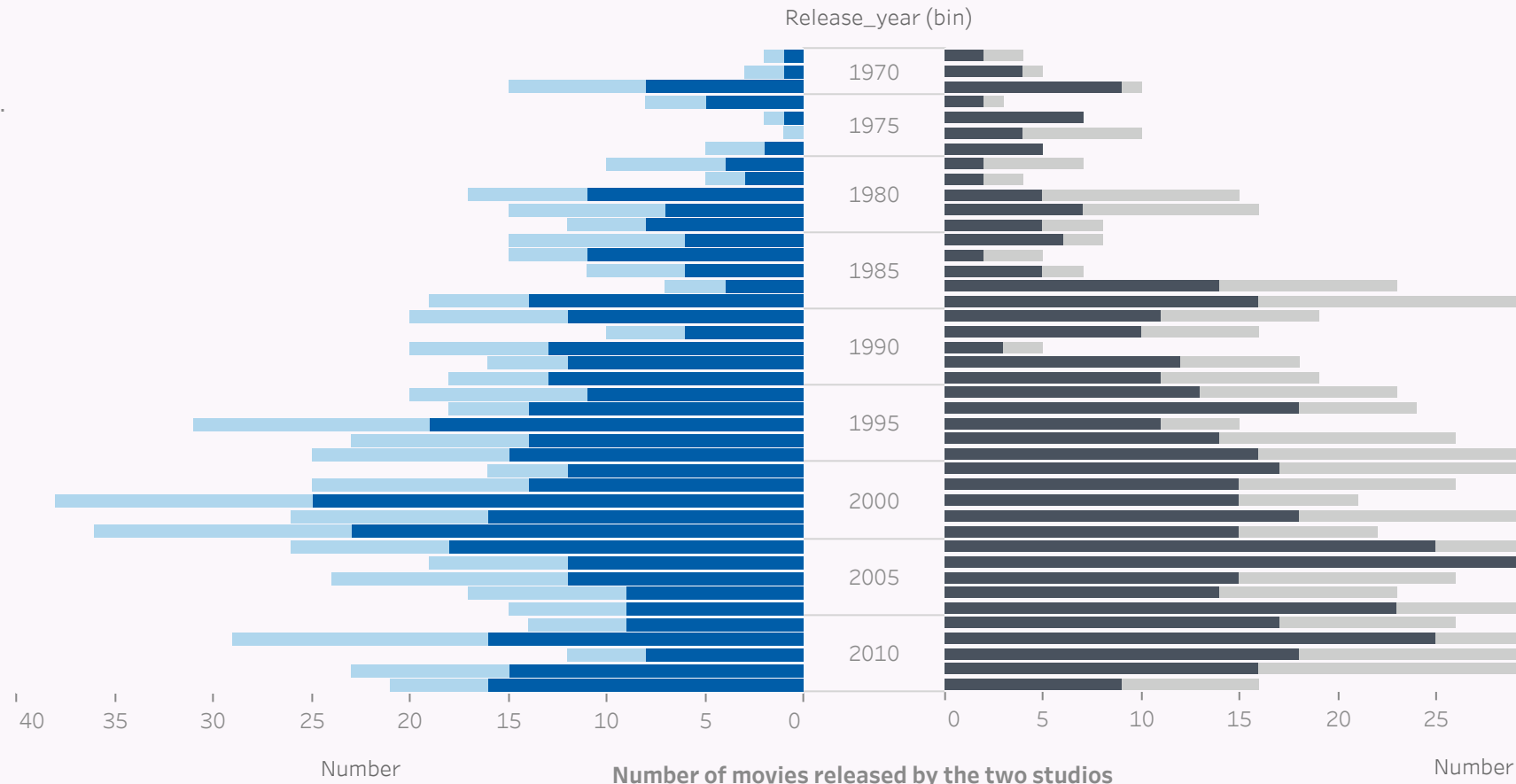
Highlight Top N Genres
No items highlighted

Highlight Release_year (..
No items highlighted

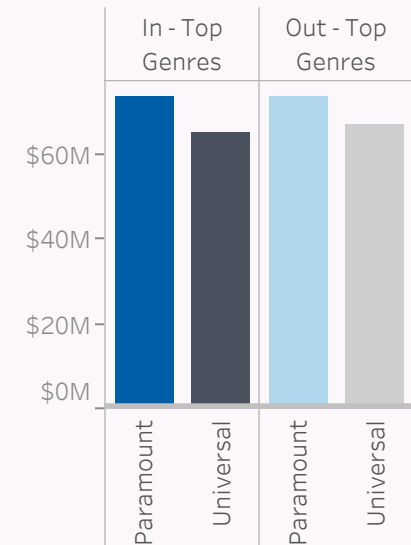
Top N Genres
5

Year of Release_date
1971 to 2015

Vote_average
3.0 to 8.3



Median Profit by Studios



Summary1:
Both studios have shown growth in production, despite the lack of stability in the early years. However, **Universal Pictures** has got a bit advantage over its rival since around 2005.

| | | |
|-----|------------------|-----|
| 435 | In - Top Genres | 498 |
| 269 | Out - Top Genres | 318 |
| 704 | Grand Total | 816 |

Summary2:
The movies, located in whether "In - Top Genres" or "Out- Top Genres", bring **Paramount Pictures** higher median profit.

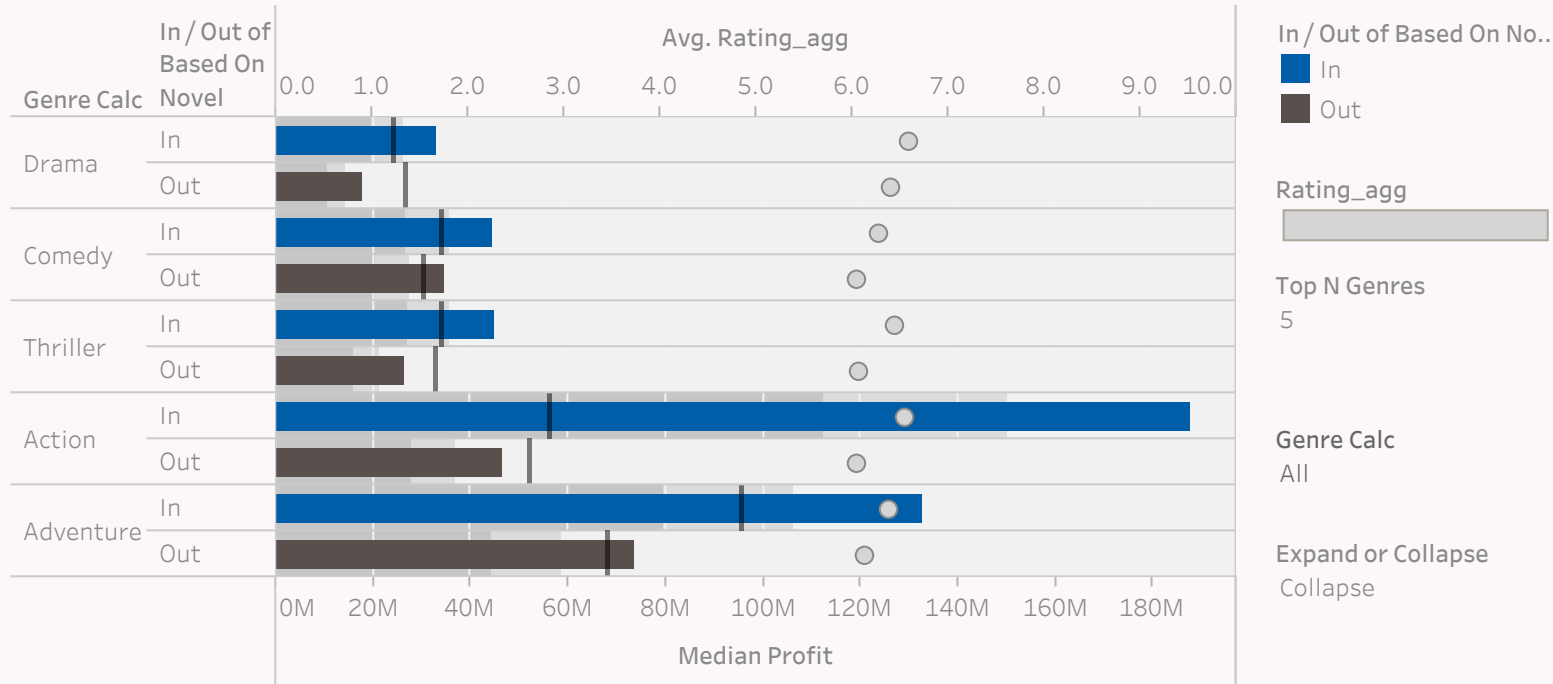


VS



Q3: How have movies based on novels performed relative to movies not based on novels? (Budgeted over \$100m)

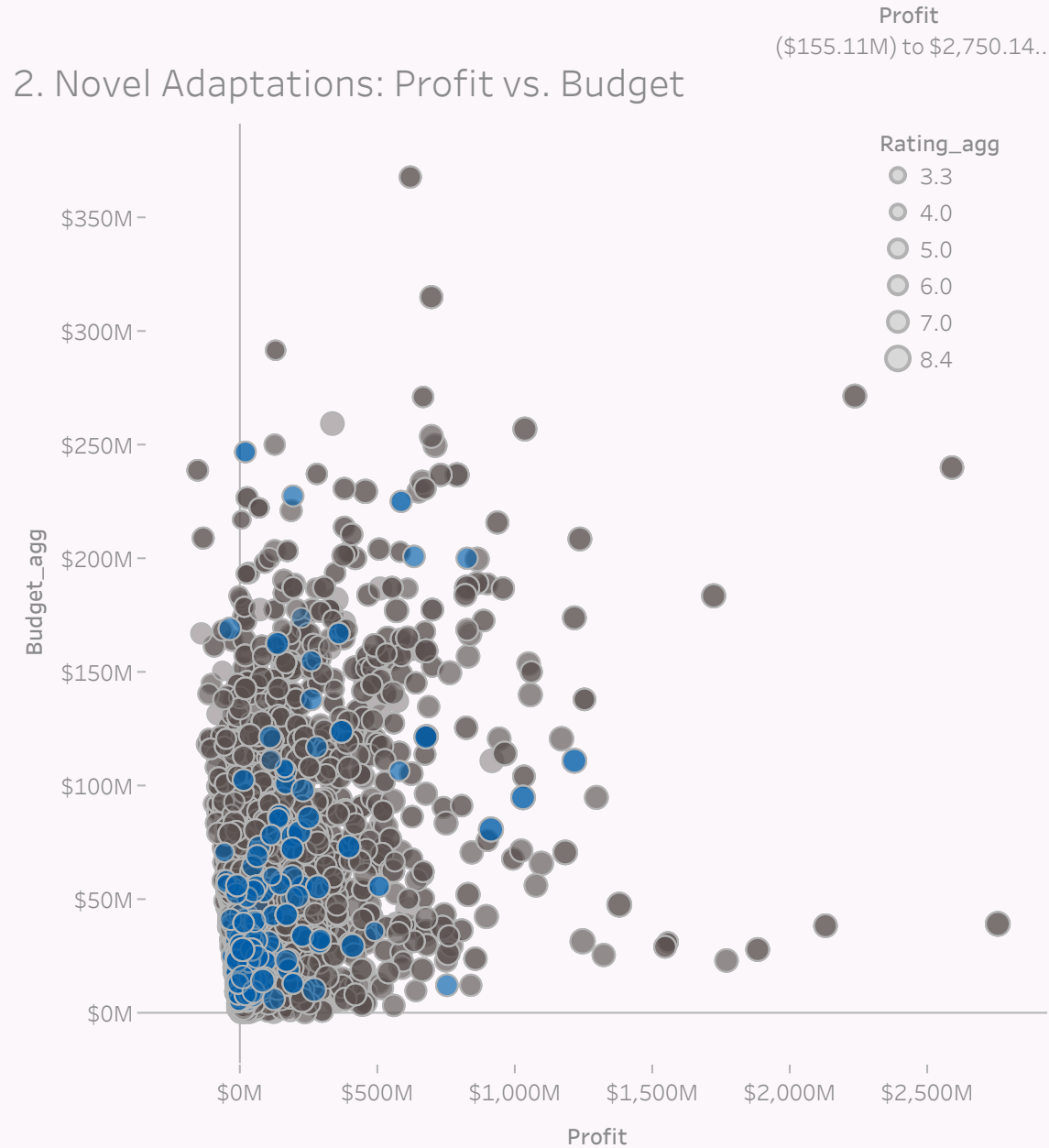
1. Novel Adaptations In Top Genres: Profit & Rating



| | In | Out |
|-----------------------------|------|------|
| Avg. Rating_agg | 6.5 | 6.2 |
| Percentage in Numbers | 5% | 100% |
| Percentage in Median Profit | 135% | 100% |

Summary:
1) Compared with movies not based on novel, although the number of movies adapted is only 5% of the latter, the average rating is 0.3 higher, and the median profit is actually 31% higher than the latter.
2) In almost all genres, except for comedies, the movies adapted from novels have a higher median profit.

2. Novel Adaptations: Profit vs. Budget



Q4: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

| | | | | |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|
| Overview: Key Indicators | Correlation Coefficient | Classified by profits and ratings | Comparison of profits and numbers | Top movies based on novel released by studios (Sorted by profit) |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|

3,561

Overall Movie Number

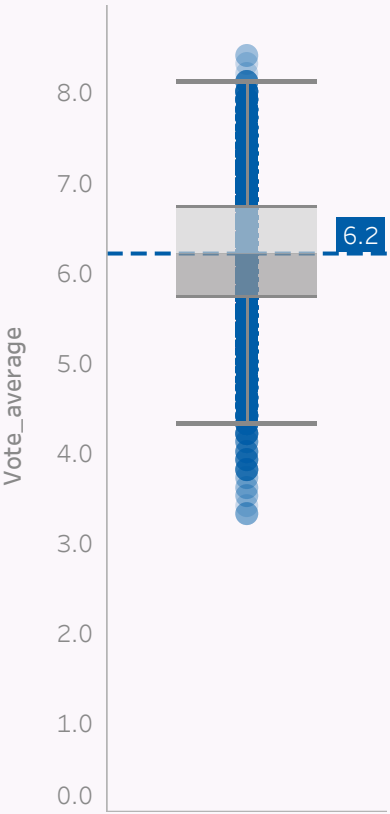
6.2

The median and average rating

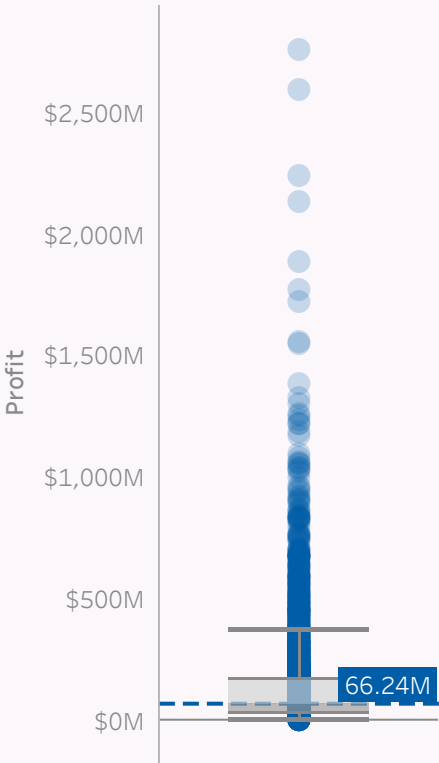
\$66.24m

The median profit of profitable movies

Rating Distribution

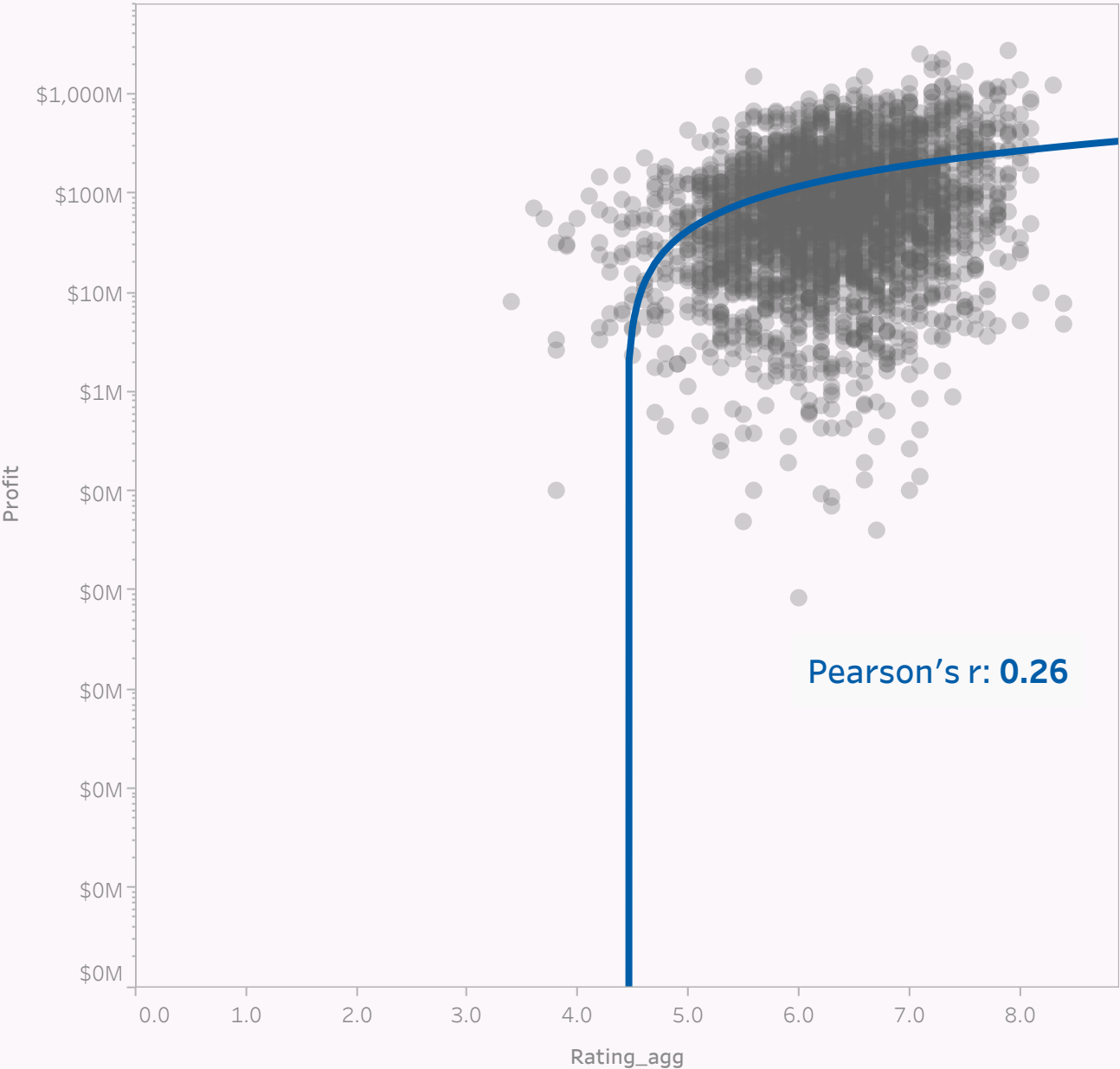


Profit Distribution of Profitable Movies



Q4: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

| | | | | |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|
| Overview: Key Indicators | Correlation Coefficient | Classified by profits and ratings | Comparison of profits and numbers | Top movies based on novel released by studios (Sorted by profit) |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|

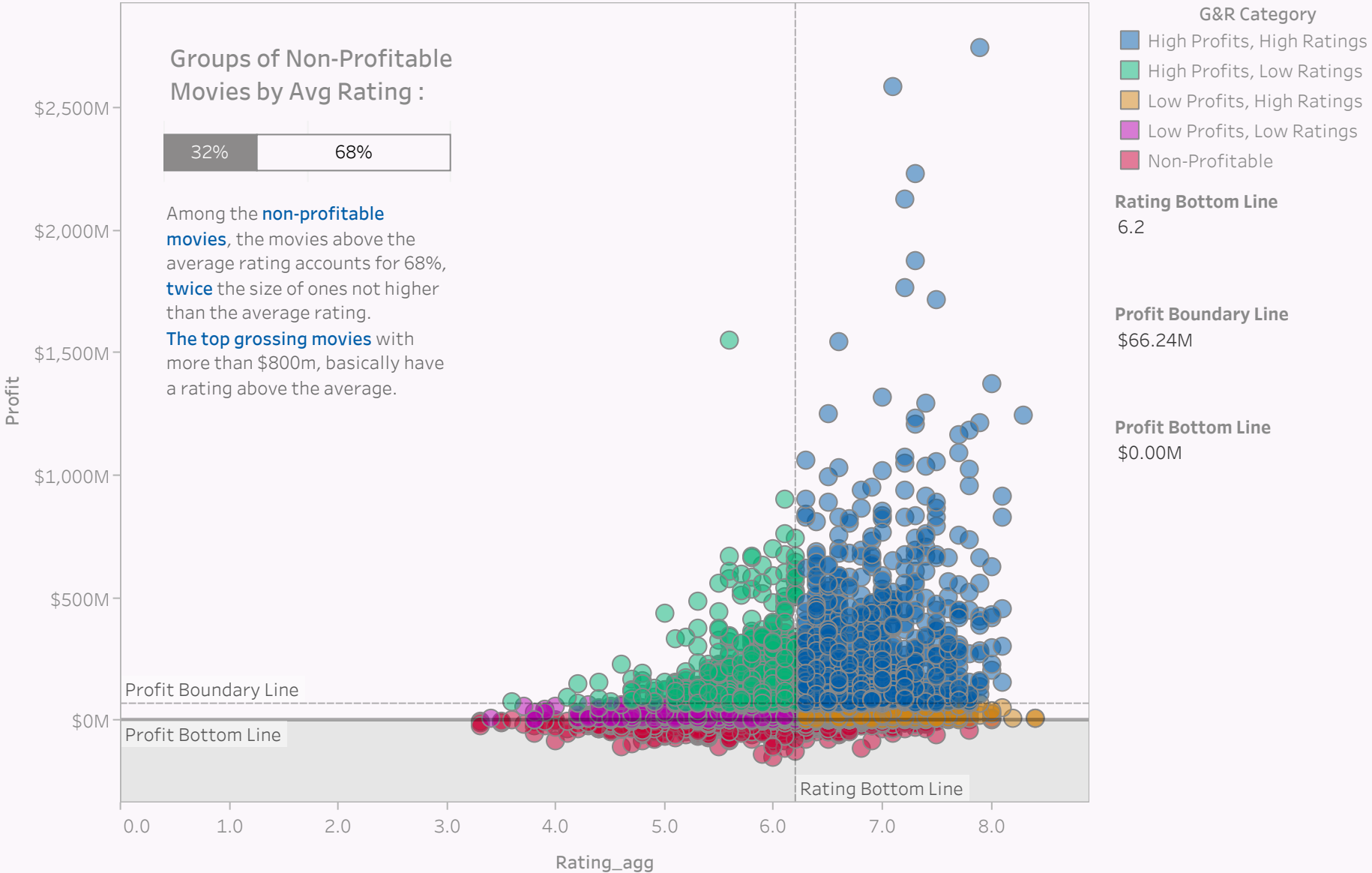


The Pearson correlation between the ratings and profits is **0.26**, implying **a weak positive linear relationship** between the two variables.

Q4: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

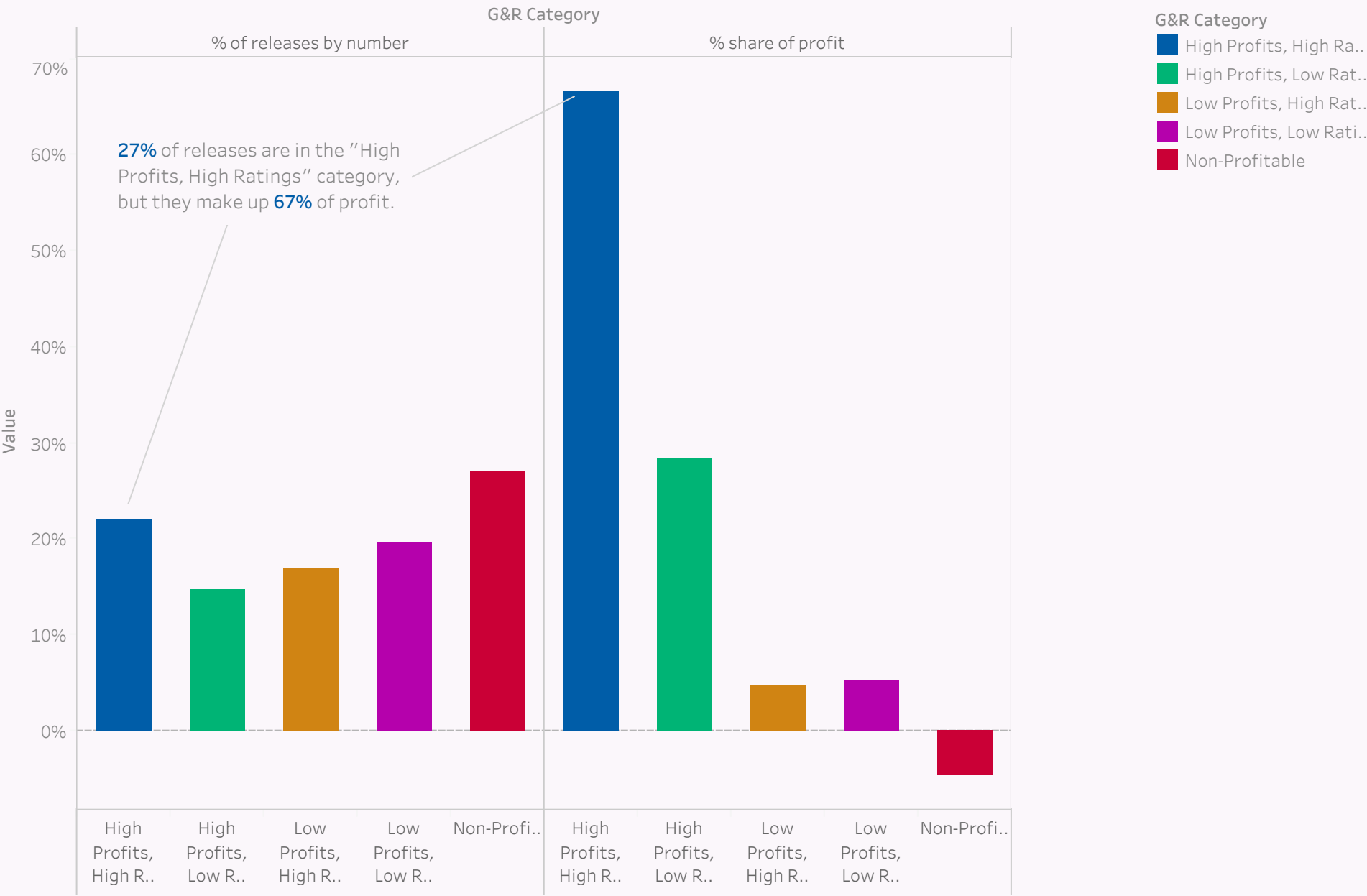
| | | | | |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|
| Overview: Key Indicators | Correlation Coefficient | Classified by profits and ratings | Comparison of profits and numbers | Top movies based on novel released by studios (Sorted by profit) |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|

Profitability & Rating: Quadrants



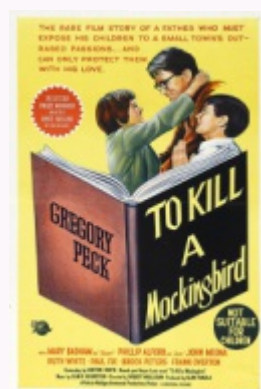
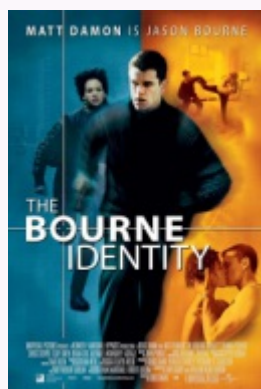
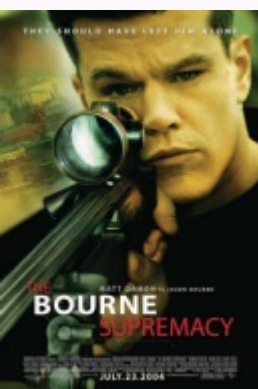
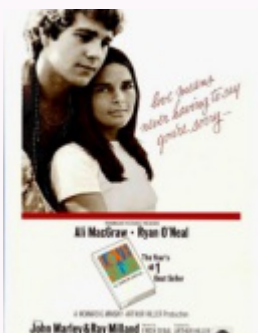
Q4: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

| | | | | |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|
| Overview: Key Indicators | Correlation Coefficient | Classified by profits and ratings | Comparison of profits and numbers | Top movies based on novel released by studios (Sorted by profit) |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|



Q4: What's The Relationship Between Profitability And Ratings? (Budgeted over \$100m)

| | | | | |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|
| Overview: Key Indicators | Correlation Coefficient | Classified by profits and ratings | Comparison of profits and numbers | Top movies based on novel released by studios (Sorted by profit) |
|--------------------------|-------------------------|-----------------------------------|-----------------------------------|--|



The top grossing movie series adapted from novels in history are <The Lord of Rings> and <Bourne>, of which the latter is released by Universal Picture. The corresponding movie from Paramount Picture is <Forrest Gump>.

The Lord of Rings

Data Source

