Does increasing movie budget lead to an increase in audience ratings?

Team A3

Meet the Team


Kushal Murthy


Irene Varsou


Yao Cheng


Steve Feng


Gihyeon Kwon


Theo Bravos


Raymond Moy

Does a Bigger Budget Lead to Better Audience Ratings?

Problem Statement

Studios invest heavily in production, but the relationship between budget and audience ratings remains unclear. Does higher spending directly lead to better reviews, or do factors like genre, runtime, and budget allocation play a bigger role?

Research Questions

- What is the causal relationship between a movie's budget and its audience ratings?
- What additional factors (e.g., genre, cast, marketing) moderate or confound this relationship?

Does a Bigger Budget Lead to Better Audience Ratings?

X

Budget Expense ↑

Y

Audience Ratings ↑

?

Runtime

Number of Votes

Cast

Genre

Director

Why it matters

Studios often assume that bigger budgets lead to higher ratings and greater revenue. But is that always the case?

Key Considerations:

- Could overspending result in diminishing returns?
- How can studios strategically allocate resources for success?

Possible Outcome Scenarios

1

High Budget
Good Quality

High Rating | Low Rating

2

High Budget
Poor Quality

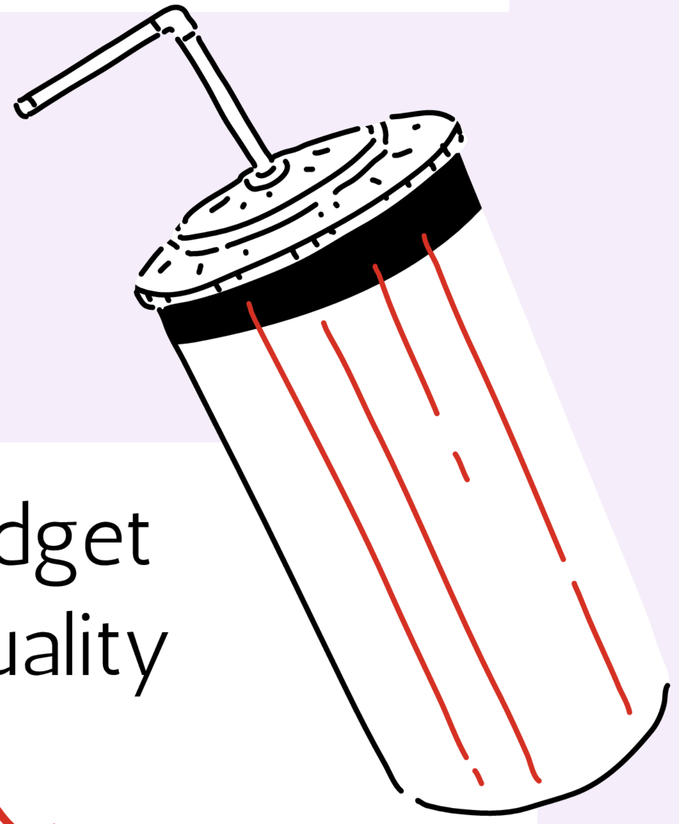
High Rating | Low Rating

5

Budget

No Relationship

Ratings



3

Low Budget
Good Quality

High Rating | Low Rating

4

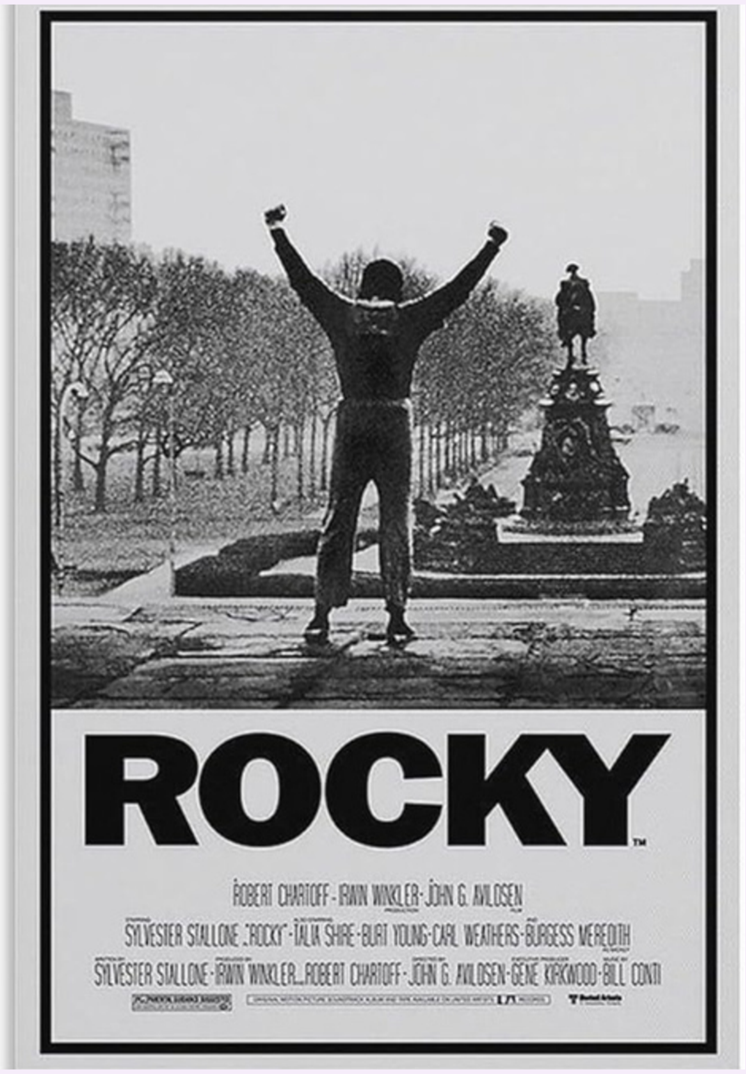
Low Budget
Poor Quality

High Rating | Low Rating

Examples

Low Budget High Rating

Budget: 1.1 million USD



93% Tomatometer

IMDb RATING 8.1/10 654K

High Budget High Rating

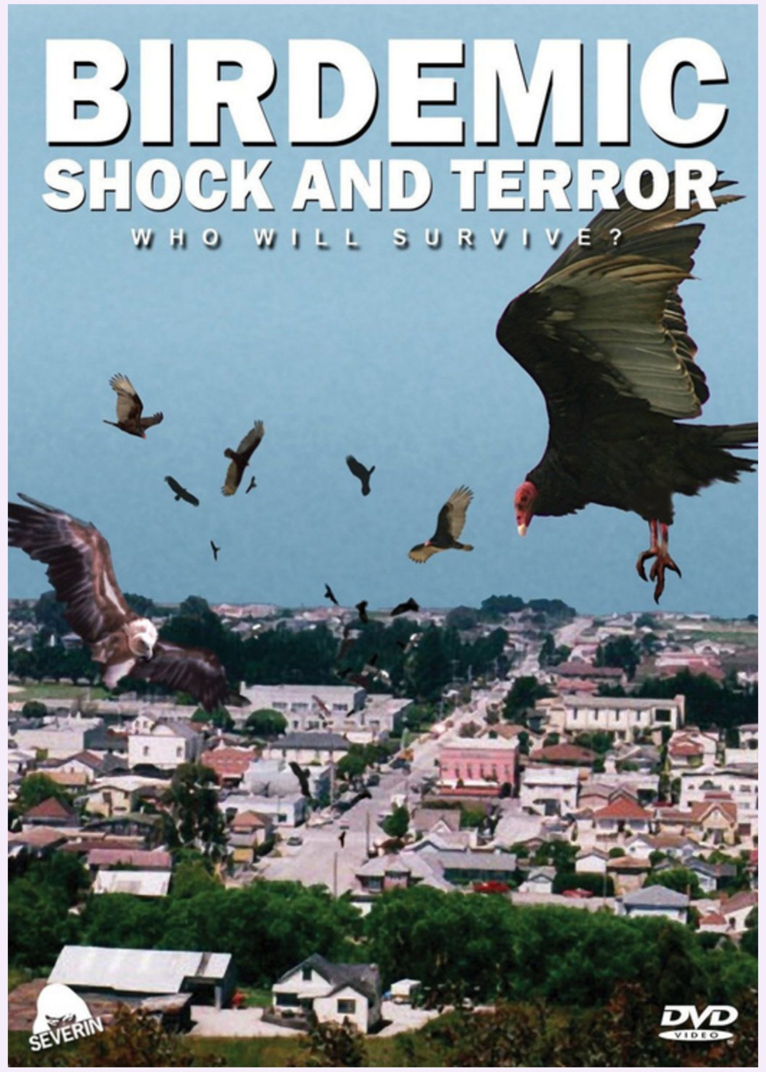
Budget: 356 million USD

94% Tomatometer

IMDb RATING 8.4/10 1.3M

Low Budget Low Rating

Budget: 10,000 USD



19% Tomatometer

IMDb RATING 1.7/10 26K

High Budget Low Rating

Budget: 146 million USD

35% Tomatometer

IMDb RATING 5.2/10 106K

Dataset Overview

Data Source: *TMDB Movies Dataset* (Kaggle) & IMDB Movies (IMDB.com)

01 Categorical Variables

- Title
- Original Language
- Original Title
- **Genres**
- Production Companies
- Production Countries

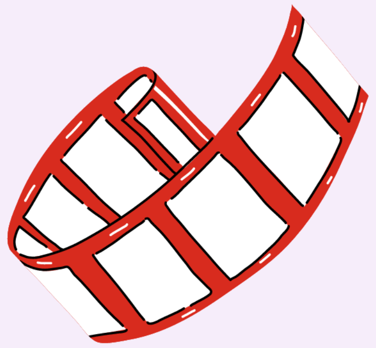
02 Numerical Variables

- ID
- Availability
- **Average Rating**
- **Budget**
- Revenue
- **Release Date**
- **Runtime**
- IMDB ID
- Popularity
- **Number of Votes**
- Year

Data Filtering

Date Range: Filtered movie data from 2000 to 2025 for analysis
-Rating websites became increasingly accessible in the digital age

Budget Outliers: Included movies with budgets between $8,000 and $300M to filter extreme cases



Average Rating Scale: Kept movies with an average rating from 1 to 10

Runtime Filter: Only movies with a minimum runtime of 30 minutes were included

Ideal Experiment

Randomized Controlled Trial (RTC) with Control Variables

- Independent Variable (IV): Movie budget allocation
  - Low: < $5M, Medium: $5M - $50M, High: > $50M
  - Control Group: Low Budget Allocation Films
- Dependent Variable (DV): Audience Ratings (IMDb, Rotten Tomatoes, etc.

Control Variables

To isolate budget effects, we need identical conditions across groups:

- Cast Quality: Assign actors with similar levels of fame across budgets
- Marketing Spend: Keep marketing budget proportional or fixed
- Genre: Ensure films within each budget group have comparable genres
- Platform Distribution: Ensure all movies get equal exposure (e.g., theatrical + streaming)

Alt: Longitudinal Observational Study

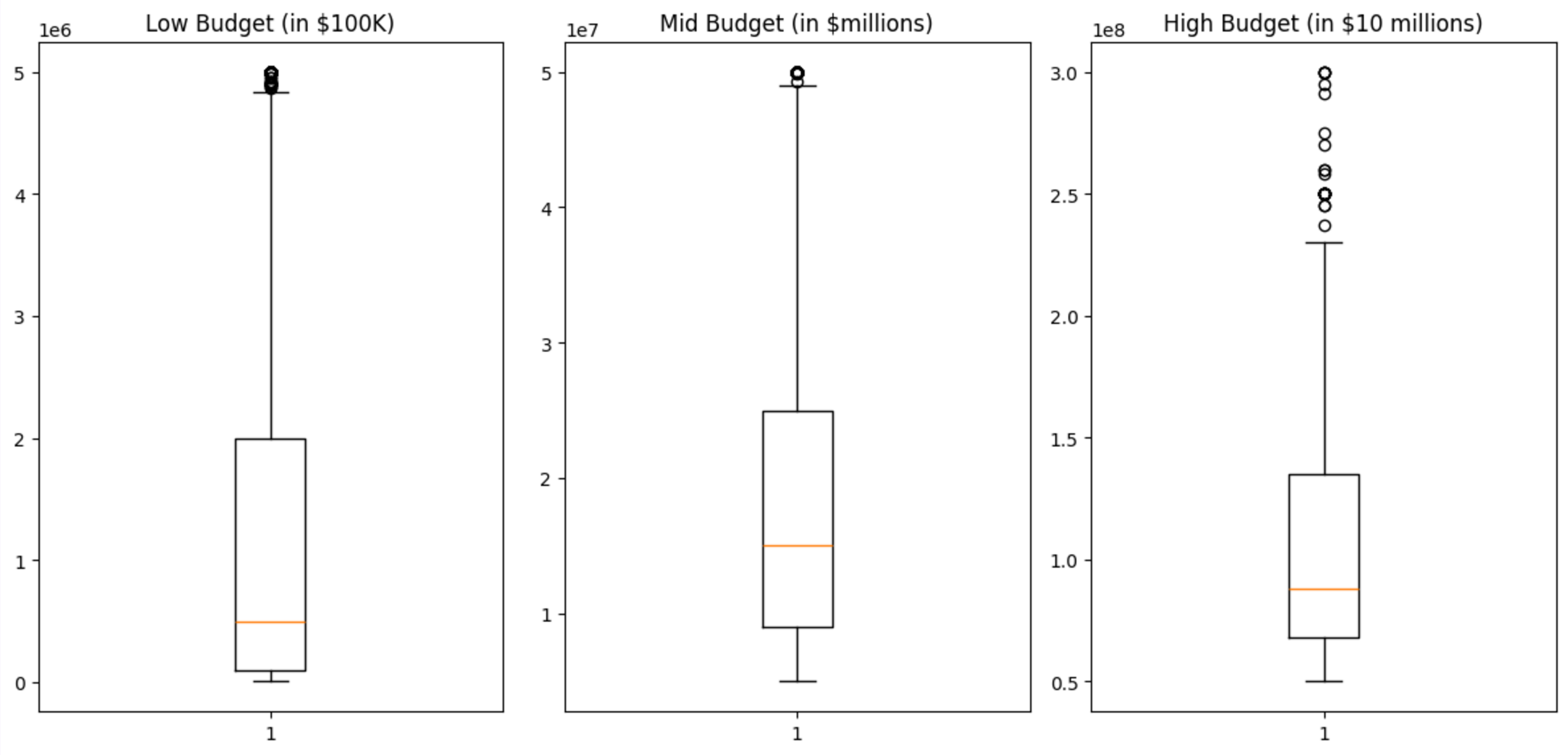
- Difference-in-Differences (DiD): Compare sequels with different budgets (e.g., Marvel films with varying production costs).
- Instrumental Variables (IV): Use external shocks (e.g., tax incentives, pandemic effects) to determine how budget influences ratings

Challenges

Not feasible due to real-world constraints

- Studios do not randomly assign budgets
- Genre, cast, and marketing impact are not present in the current dataset

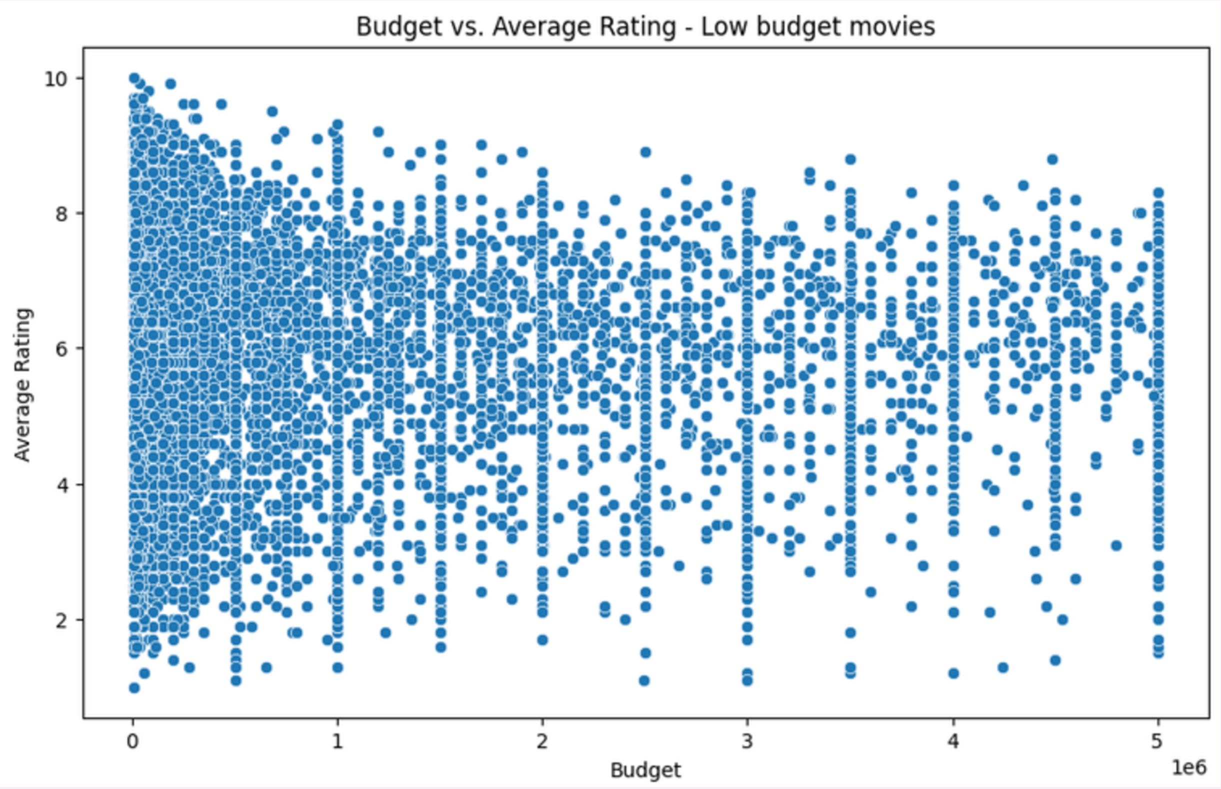
Descriptive Statistics



- Low Budget <= $5M
- Mid Budget between $5M and $50M
- High Budget > 50M
- Movie budgets are right-skewed, with most films having lower budgets and some blockbusters with extremely high budgets

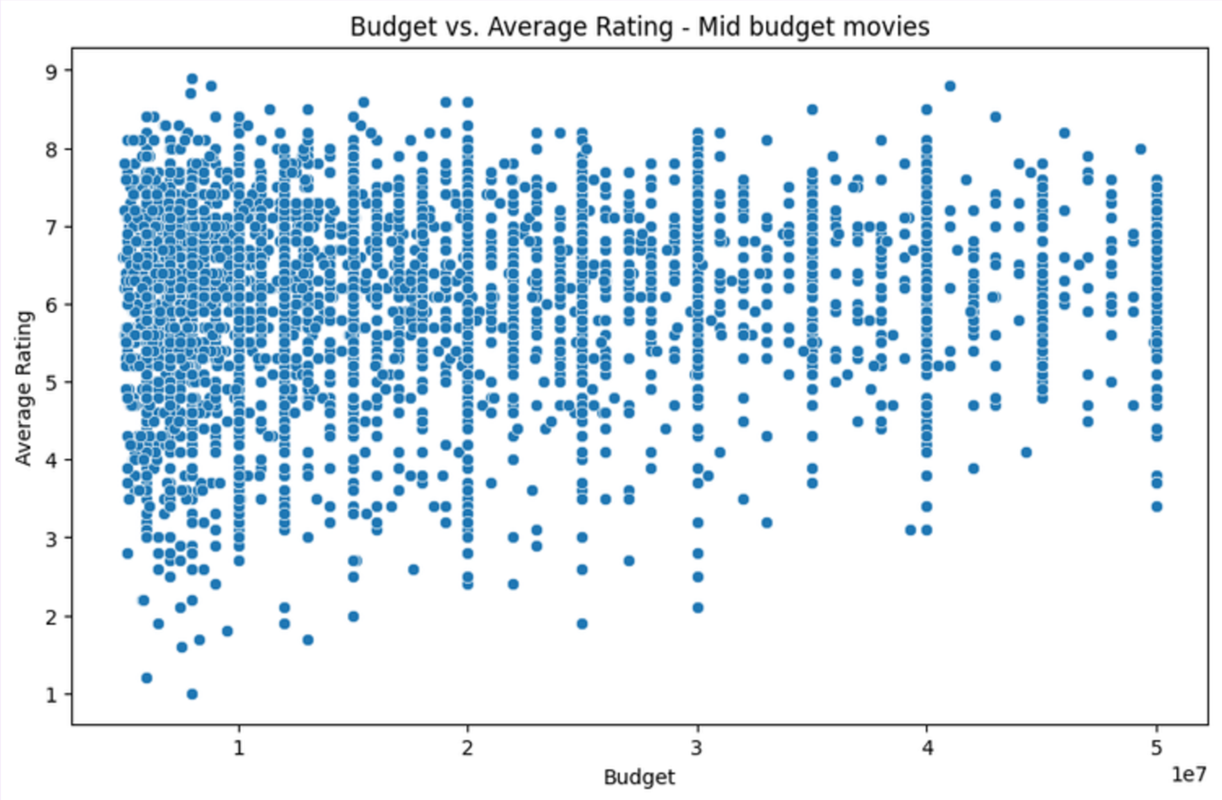
Descriptive Statistics

Low Budget Movies



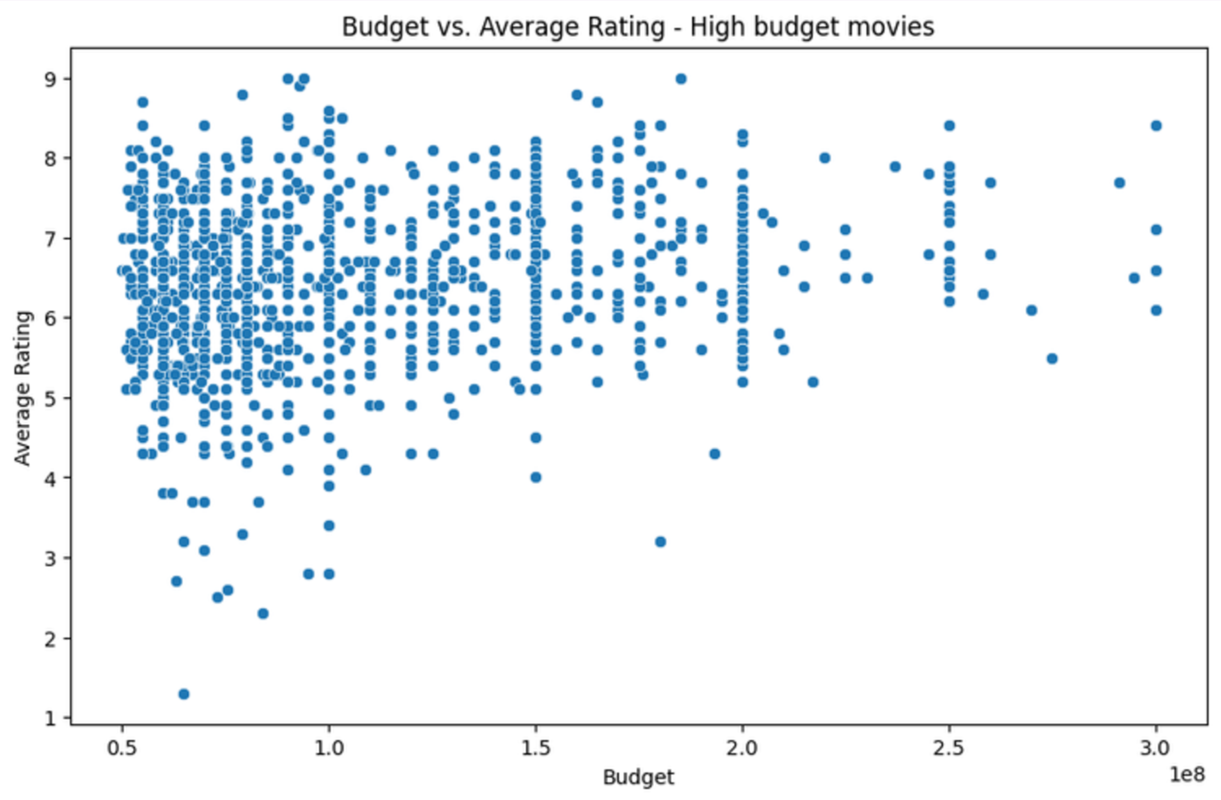
- There is a high concentration of movies with lower budgets and a wide spread in ratings
- Some low-budget movies achieve high ratings, but others receive very low ratings, leading to high variability

Mid Budget Movies



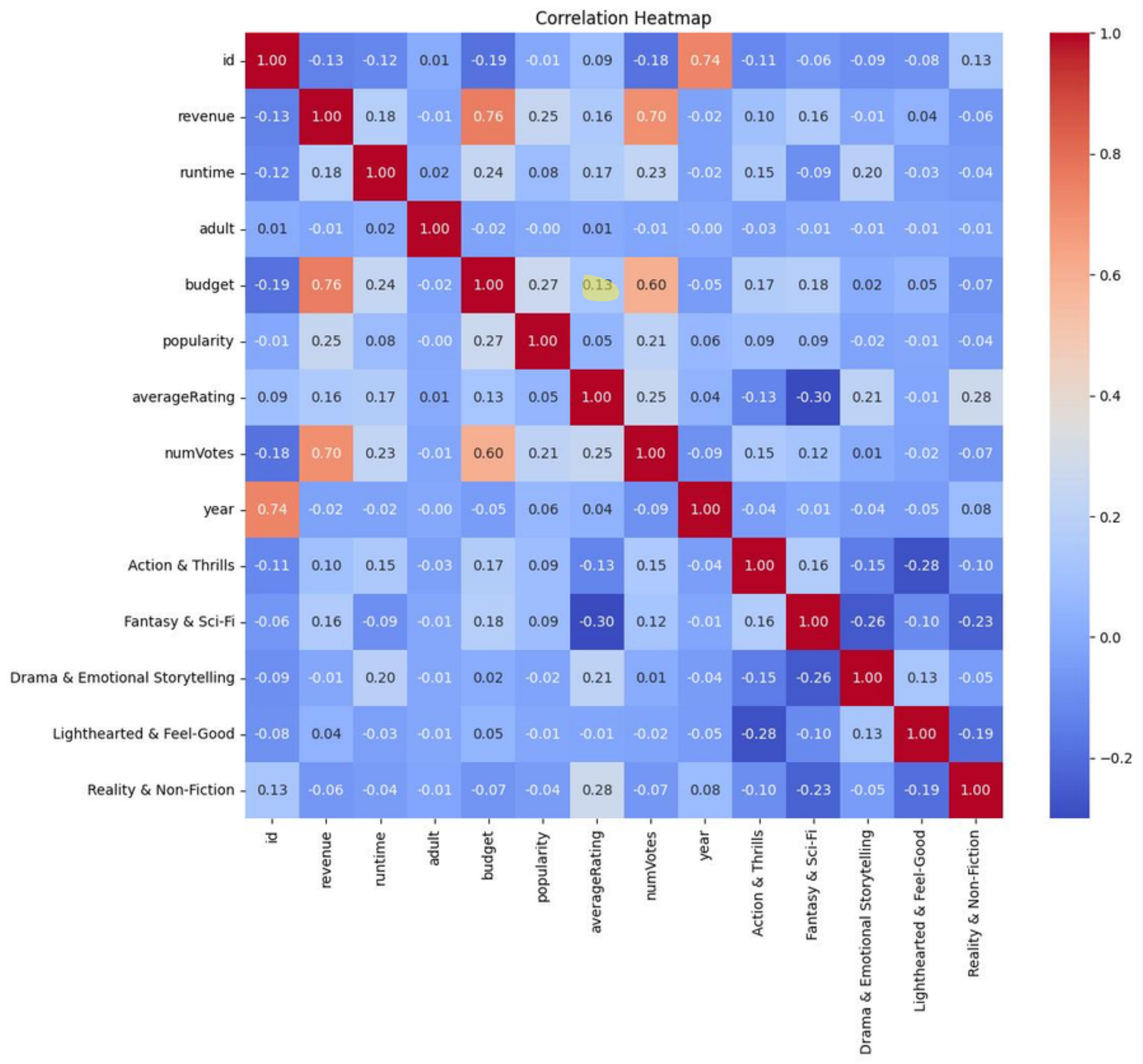
- Ratings appear more stable compared to low-budget movies, with a less extreme spread
- There is no strong trend showing that increasing the budget significantly improves ratings

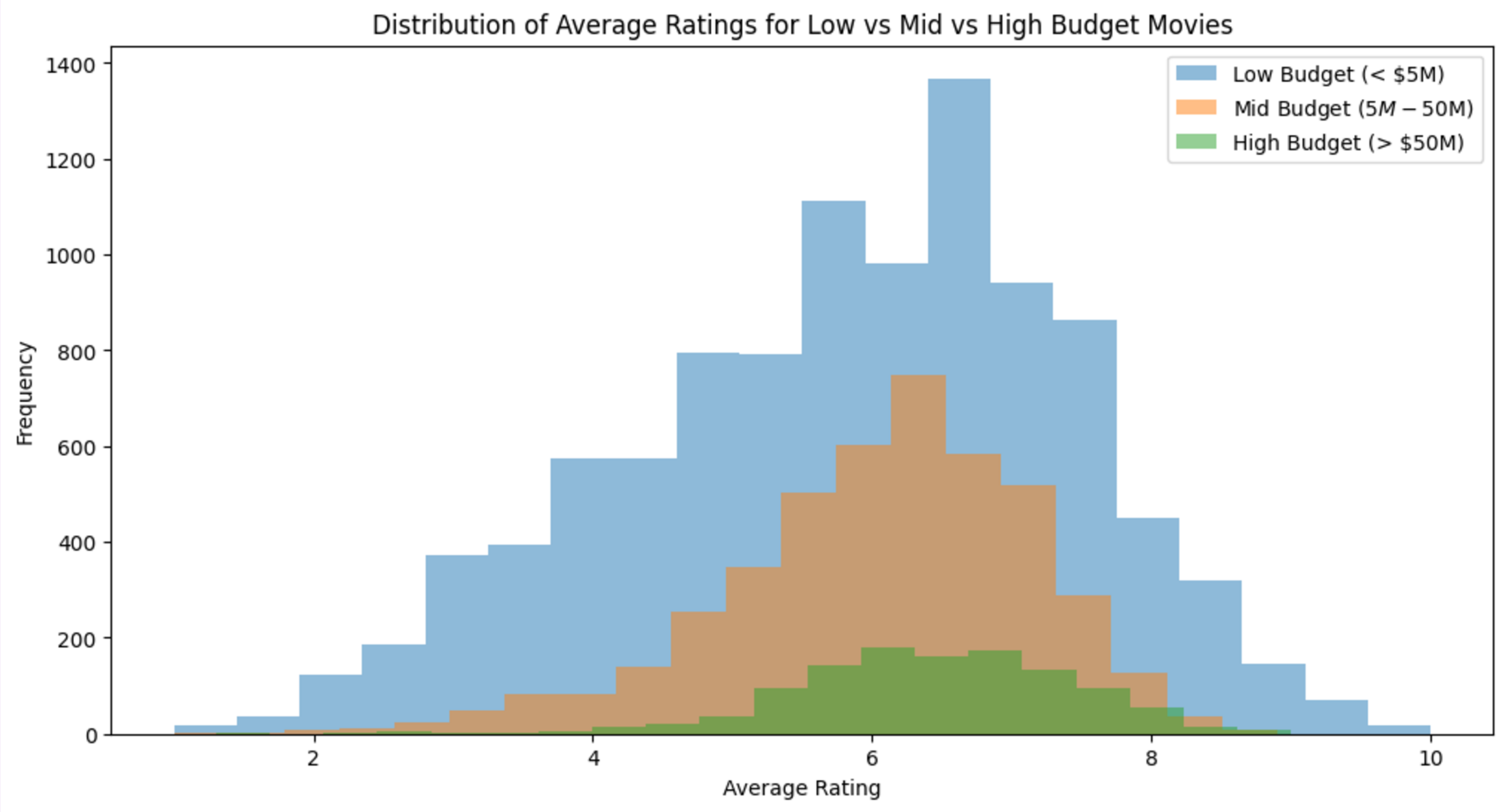
High Budget Movies



- Most high-budget movies have moderate to high ratings, but there are still some low-rated films
- The budget range is much larger, but ratings appear to plateau—increasing the budget further does not guarantee higher ratings

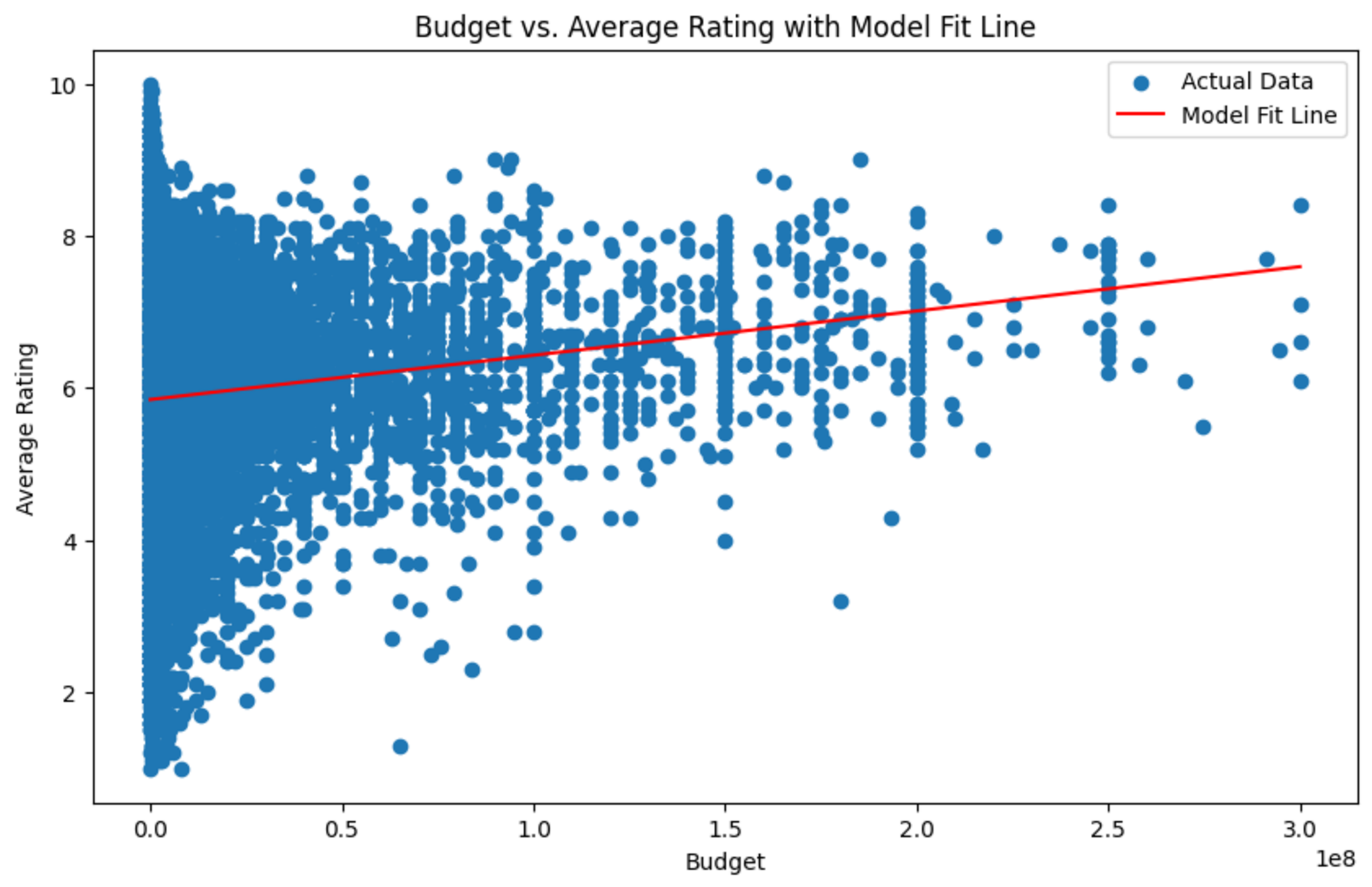
Descriptive Statistics: Bivariate





Simple Regression Model

Model ⋯⋯⋯⋯→ Average Rating = β0 + β1 x Budget




```
                            OLS Regression Results
==============================================================================
Dep. Variable:          averageRating   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.016
Method:                 Least Squares   F-statistic:                     249.3
Date:                Thu, 27 Feb 2025   Prob (F-statistic):           9.80e-56
Time:                        07:49:37   Log-Likelihood:                -27902.
No. Observations:               15692   AIC:                         5.581e+04
Df Residuals:                   15690   BIC:                         5.582e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          5.8491      0.013    467.644      0.000       5.825       5.874
budget      5.822e-09   3.69e-10     15.790      0.000     5.1e-09    6.55e-09
==============================================================================
Omnibus:                      400.911   Durbin-Watson:                   1.786
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              432.107
Skew:                          -0.406   Prob(JB):                     1.48e-94
Kurtosis:                       2.999   Cond. No.                     3.71e+07
==============================================================================
```

Our Approach

To accurately analyze the relationship between budget and audience ratings, we go beyond a simple linear model and consider alternative modeling approaches to capture underlying complexities in the data

- **Multivariate Regression Analysis:** Accounts for potential confounding variables such as number of votes, genre, runtime, year it was released
- **P-value Analysis:** Determines the statistical significance of budget's impact on ratings

Multivariate Regression Model

Null:
Budget has no effect on ratings (β Budget = 0)

Alternative:
Budget has a significant effect on ratings (β Budget != 0)

$$\widehat{\text{averageRating}} = \beta_0 + \beta_1 \cdot \text{budget} + \beta_2 \cdot \text{numVotes} + \beta_3 \cdot \text{yearsSinceRelease}$$
$$+\beta_4 \cdot \text{runtime} + \beta_5 \cdot \text{ActionThrills} + \beta_6 \cdot \text{FantasySciFi} + \beta_7 \cdot \text{DramaEmotional} + \beta_8 \cdot \text{LightheartedFeelGood} + \beta_9 \cdot \text{RealityNonFiction} + \varepsilon$$

Low budget group <= $5m

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          averageRating   R-squared:                       0.278
Model:                            OLS   Adj. R-squared:                  0.277
Method:                 Least Squares   F-statistic:                     432.7
Date:                Thu, 27 Feb 2025   Prob (F-statistic):               0.00
Time:                        20:09:34   Log-Likelihood:                -17417.
No. Observations:               10134   AIC:                         3.485e+04
Df Residuals:                   10124   BIC:                         3.493e+04
Df Model:                           9
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
const                            5.7114      0.062     92.107      0.000       5.590       5.833
budget                       -6.628e-08   1.01e-08     -6.546      0.000   -8.61e-08   -4.64e-08
numVotes                      8.636e-06    4.4e-07     19.616      0.000    7.77e-06     9.5e-06
years_since_release             -0.0127      0.002     -5.629      0.000      -0.017      -0.008
runtime                          0.0034      0.001      6.380      0.000       0.002       0.004
Action & Thrills                -0.3485      0.031    -11.273      0.000      -0.409      -0.288
Fantasy & Sci-Fi                -0.9765      0.033    -29.980      0.000      -1.040      -0.913
Drama & Emotional Storytelling   0.4098      0.029     13.902      0.000       0.352       0.468
Lighthearted & Feel-Good        -0.0464      0.031     -1.497      0.134      -0.107       0.014
Reality & Non-Fiction            1.1316      0.040     28.265      0.000       1.053       1.210
==============================================================================
Omnibus:                       93.764   Durbin-Watson:                   1.784
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               97.486
Skew:                          -0.223   Prob(JB):                     6.78e-22
Kurtosis:                       3.179   Cond. No.                     9.23e+06
==============================================================================
```

$50m <= Mid budget group > $5m

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          averageRating   R-squared:                       0.336
Model:                            OLS   Adj. R-squared:                  0.335
Method:                 Least Squares   F-statistic:                     248.0
Date:                Thu, 27 Feb 2025   Prob (F-statistic):               0.00
Time:                        20:09:34   Log-Likelihood:                -5842.2
No. Observations:                4416   AIC:                         1.170e+04
Df Residuals:                    4406   BIC:                         1.177e+04
Df Model:                           9
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
const                            4.5536      0.090     50.455      0.000       4.377       4.731
budget                       -1.209e-09    1.2e-09     -1.009      0.313   -3.56e-09    1.14e-09
numVotes                      3.716e-06   1.19e-07     31.226      0.000    3.48e-06    3.95e-06
years_since_release             -0.0041      0.002     -1.821      0.069      -0.009       0.000
runtime                          0.0111      0.001     16.815      0.000       0.010       0.012
Action & Thrills                -0.1501      0.032     -4.717      0.000      -0.212      -0.088
Fantasy & Sci-Fi                -0.1829      0.032     -5.758      0.000      -0.245      -0.121
Drama & Emotional Storytelling   0.4420      0.032     14.022      0.000       0.380       0.504
Lighthearted & Feel-Good        -0.1296      0.032     -4.055      0.000      -0.192      -0.067
Reality & Non-Fiction            0.3620      0.048      7.571      0.000       0.268       0.456
==============================================================================
Omnibus:                      736.203   Durbin-Watson:                   1.724
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1587.606
Skew:                          -0.977   Prob(JB):                         0.00
Kurtosis:                       5.193   Cond. No.                     1.49e+08
==============================================================================
```

High budget group > $50m

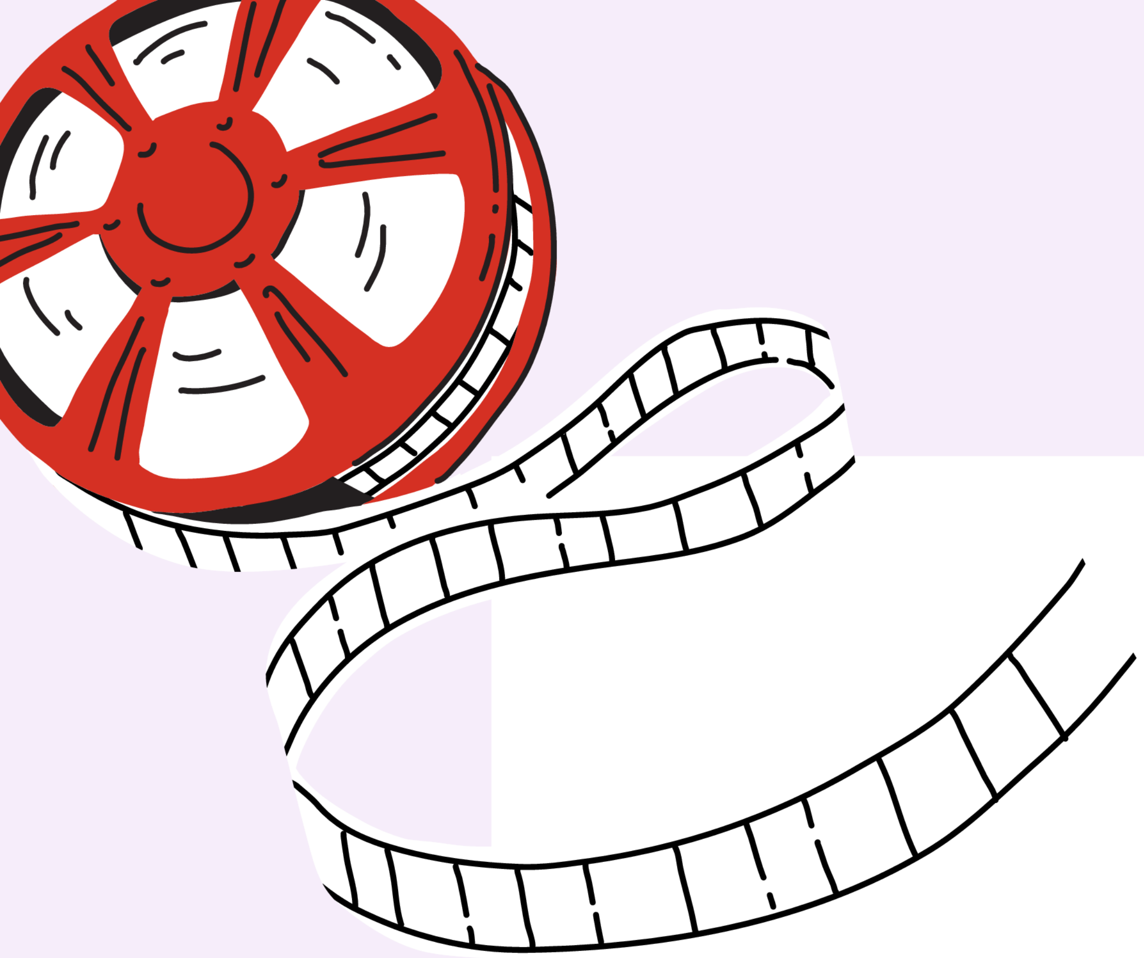
```
                            OLS Regression Results
==============================================================================
Dep. Variable:          averageRating   R-squared:                       0.442
Model:                            OLS   Adj. R-squared:                  0.438
Method:                 Least Squares   F-statistic:                     99.67
Date:                Thu, 27 Feb 2025   Prob (F-statistic):          7.74e-137
Time:                        20:09:35   Log-Likelihood:                -1281.5
No. Observations:                1142   AIC:                             2583.
Df Residuals:                    1132   BIC:                             2633.
Df Model:                           9
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
const                            5.7057      0.134     42.572      0.000       5.443       5.969
budget                       -2.228e-10   5.06e-10     -0.441      0.659   -1.21e-09    7.69e-10
numVotes                       2.02e-06   8.13e-08     24.830      0.000    1.86e-06    2.18e-06
years_since_release             -0.0120      0.003     -3.463      0.001      -0.019      -0.005
runtime                          0.0029      0.001      4.431      0.000       0.002       0.004
Action & Thrills                -0.1274      0.059     -2.155      0.031      -0.243      -0.011
Fantasy & Sci-Fi                -0.0686      0.051     -1.353      0.176      -0.168       0.031
Drama & Emotional Storytelling   0.2573      0.054      4.752      0.000       0.151       0.364
Lighthearted & Feel-Good        -0.0579      0.059     -0.975      0.330      -0.174       0.059
Reality & Non-Fiction            0.3984      0.091      4.391      0.000       0.220       0.576
==============================================================================
Omnibus:                      294.656   Durbin-Watson:                   1.902
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1086.936
Skew:                          -1.207   Prob(JB):                    9.44e-237
Kurtosis:                       7.125   Cond. No.                     7.54e+08
==============================================================================
```

Model Limitations

- **Budget Allocation Uncertainty:** A large portion of a movie's budget goes to marketing rather than production, making our financial impact analysis unclear
- **Star Power:** The model does not account for starring actors, which can significantly impact ratings
- **Review Bombing:** Audience ratings may be skewed due to intentional mass ratings (positive or negative)
- The current regression model assumes **a linear relationship.** However, diminishing returns is a possible factor

Conclusion

- **Low-budget films: Reject** the null
- **Mid-budget films: Fail to reject** the null
- **High-budget films: Fail to reject** the null
- **Key takeaway:** Simply increasing the budget **does not guarantee** better ratings: strategic allocation matters.
- **Next steps:**
  - Prioritizing **storytelling** and **production**
  - Allocate budget efficiently between **marketing** and **talent**
  - Conduct deeper **analysis** on **audience** preferences



The End