How does a movie's budget affect its box office earnings?

Team South Beach

Ada Guan, Kolby Devery, Nan Li, Heather Rodney, Matt Pitz

Berkeley
UNIVERSITY OF CALIFORNIA

Questions

How does the change in budget affect box office earnings from 1990 to 2015?

Sub research questions:

1. Does genre affect the box office for movies?
2. Does the month of release really matter?
3. What are the effects of duration and maturity rating on box office earnings?

Introduction

Variables

DV

Global Box Office Sales - The amount of money (US Dollars) made after release of movie

IV

Budget - The amount of money spent on creating the movie

Genre - The movie category

MetaCritic Rating - Movie critique of the movie prior to its release

Movie Release Date - The time point in the year that the movie was released

Awards - Binary flag for whether the movie received awards

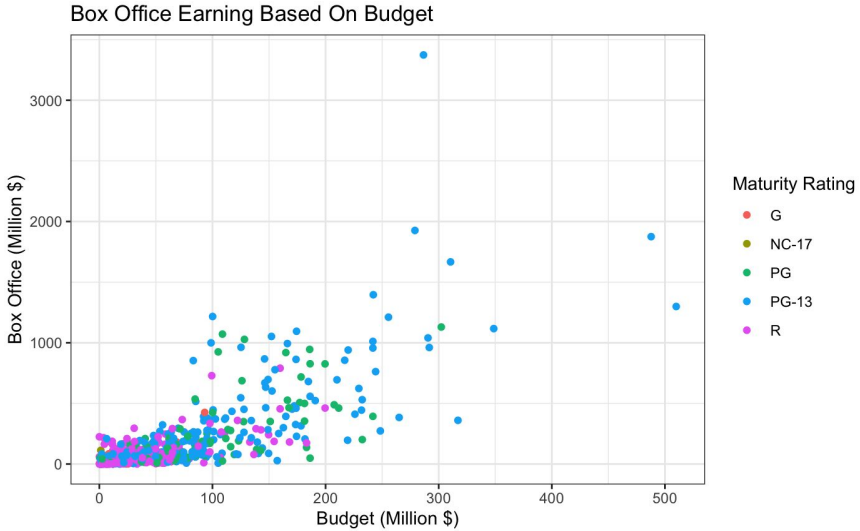
Maturity Rating - Movie content rating

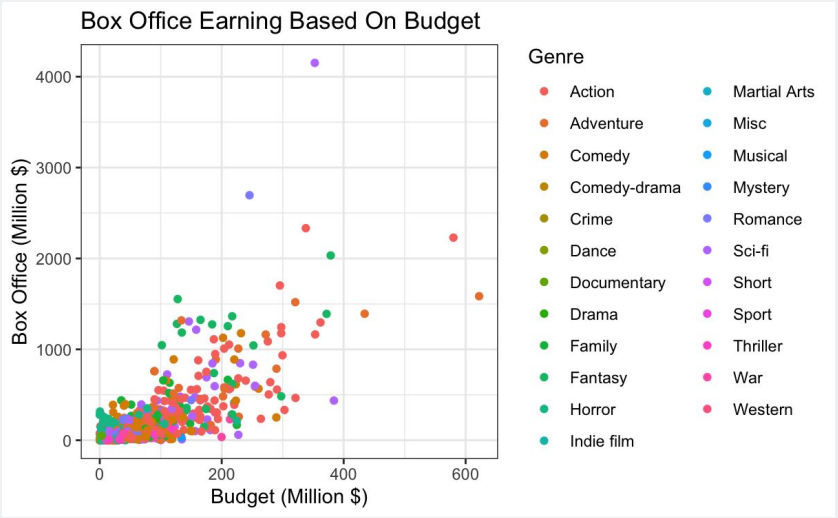
Duration - The length of the movie in minutes

Data Types

- Numerical and Categorical Data

Null Hypothesis: Spending has no effect on box office outcome





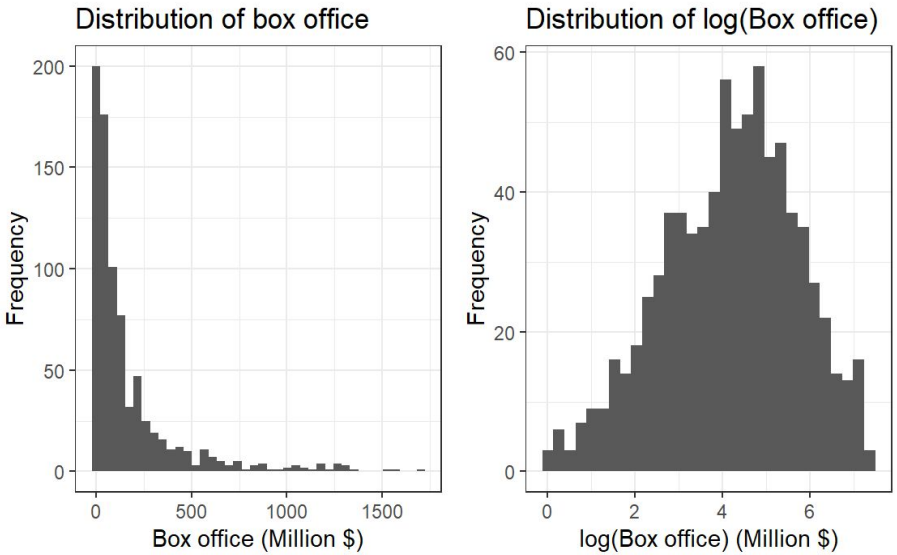
The results indicated that that movies with PG-13 rating with higher budgets likely had higher outcomes, compared to those with R ratings. We do not see much information for G (n=12) and NC-17 (n=1) ratings, but it is likely that the budgets were lower and likely did not have higher outcome.

The results indicated that those with lower budgets tend to have similar outcomes after release. Additionally, it appears that for some movies the genre can also affect the budget subsequently have a higher outcome compared to others, however those appear to be outliers.

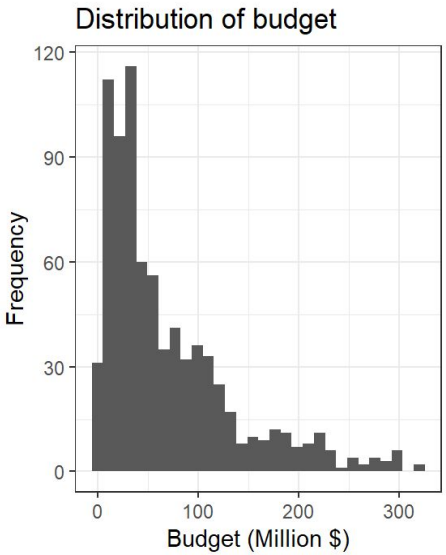
Introduction

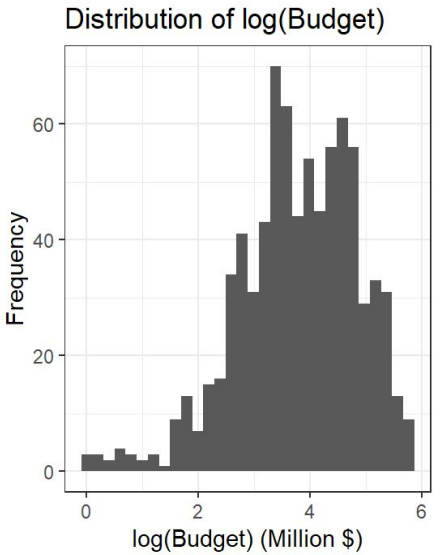
Variable transformation

Outcome variable



Independent variable

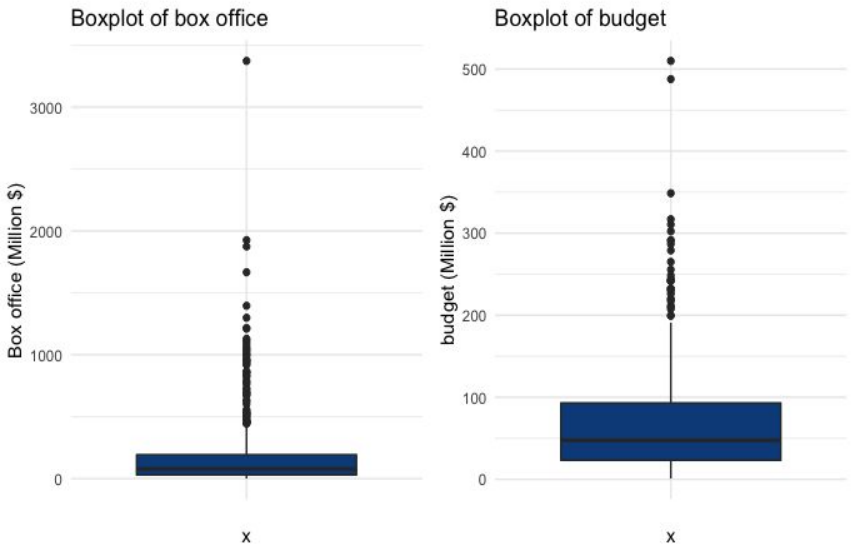


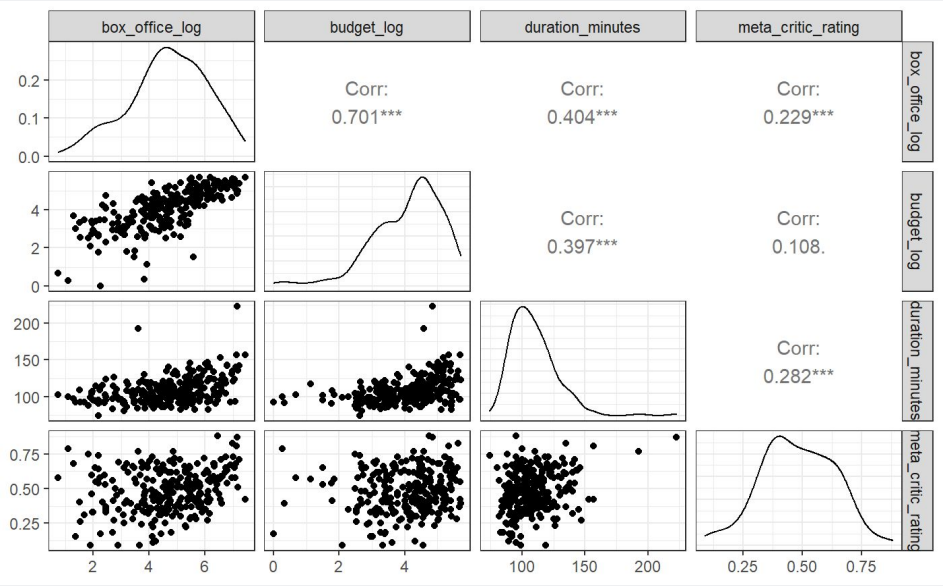


Log transformation applied to outcome and independent variable for a normalized distribution

Research Design

General EDA





Research Design

Boxplots were used to identify extreme outliers; Correlation matrix was used to identify meaningful relationship and potential collinearity between variables.

Data

- List of All American Films (1950-2020) from *Harvard Dataverse*
  - 794 movies with both budget and box office data
- Assumption of non-time series data
  - Used inflation data from the BLS to adjust box office and budget values
  - Treated 1990-2015 as single ‘snapshot’ in time
- Performed data transformation on the variables:
  - Release date to binary variables (summer and thankxmas),
  - Genre
  - Film awards to binary variable

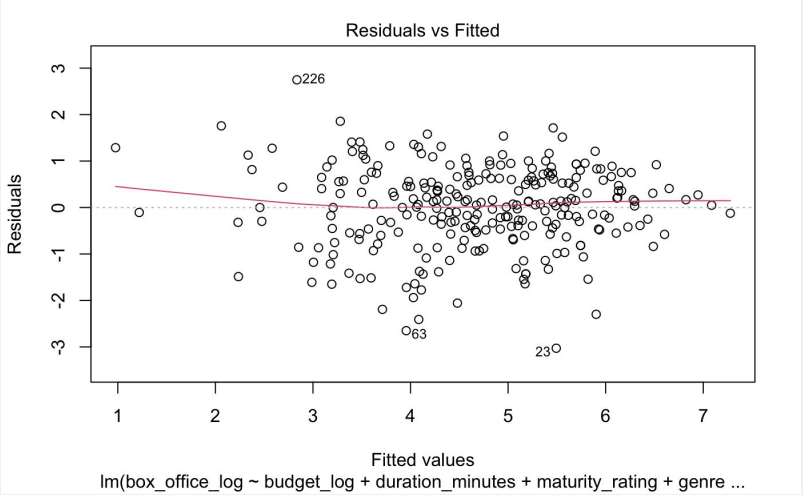
Research Design

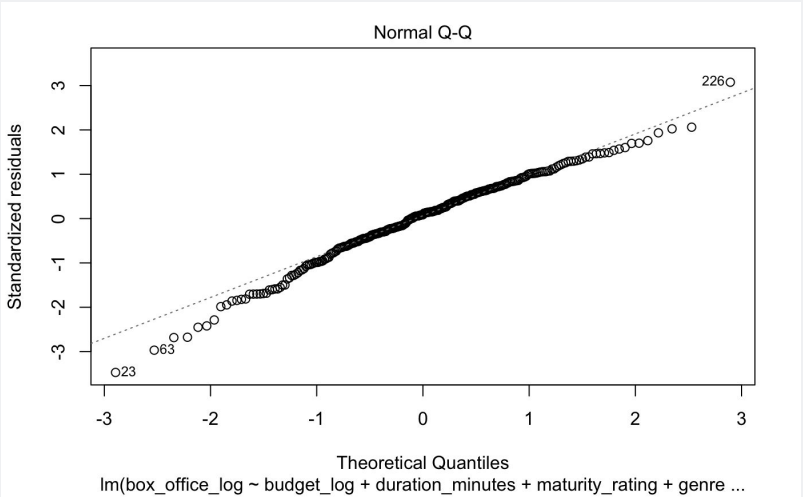
1. How does budget affect box office revenue?:
   - Log of budget vs. log of box office revenue
2. Study relationship between box office revenue and other explanatory variables
   - MetaCritic Ratings/Awards treated as marker of quality *before* movie’s release
3. Perform tests to determine which models:
   - Were statistically significant from base model
   - Reduce RSE
   - Maximize R2

Research Design

Classic Linear Modeling Assumptions

1. IID Data - **SATISFIED**
2. No Perfect Collinearity - **SATISFIED**
   - VIF < 2
3. Normality of Errors - **SATISFIED?**
   - Histogram of residuals looks normal but Q-Q plot is questionable
4. Homoskedasticity - **SATISFIED**
   - Breusch Pagan Test (p-value 0.105)
5. Linear Conditional Expectation - **SATISFIED**





Research Design

Model

Model 1: Baseline model (only including one independent variable):

$$\log(\text{Box office}) = \beta_0 + \beta_1 * \log(\text{Budget})$$

Model 2: Model including independent variables which affect both budget and box office:

$$\log(\text{Box office}) = \beta_0 + \beta_1 * \log(\text{Budget}) + \beta_2 * \text{duration} + \beta_3 * \text{maturity\_rating} + \beta_4 * \text{genre\_primary}$$

Model 3: Model including more independent variables of interest:

$$\log(\text{Box office}) = \beta_0 + \beta_1 * \log(\text{Budget}) + \beta_2 * \text{duration} + \beta_3 * \text{maturity\_rating} + \beta_4 * \text{genre\_primary} + \beta_5 * \text{meta\_critic\_rating}$$

Model Design

Model 4: Model including binary variables (thanksxmas, summer & received award)

$$\log(\text{Box office}) = \beta_0 + \beta_1 * \log(\text{Budget}) + \beta_2 * \text{duration} + \beta_3 * \text{maturity\_rating} + \beta_4 * \text{genre\_primary} + \beta_5 * \text{meta\_critic\_rating} + \beta_6 * \text{thanksxmas} + \beta_7 * \text{summer} + \beta_8 * \text{award}$$

Based on the F tests that were conducted, the binary variables did not have additional explanatory power, and conclude that model 3 is the optimal model.

Results

- Best performance: Model 3
- Release date was *not* statistically significant

Variable	Estimate
Budget	.82***
Duration	.008*
Maturity Rating: R	-.9*
MetaCritic Rating	1.31**
Genre: Documentary	1.38***
Genre: Horror	.512*
Genre: Fantasy	.42*
Genre: Drama	-.83*
Genre: Sport	-1.23***
Genre: Mystery	-1.67***

Results and Interpretation

Interpretation

- Budget has strong positive correlation with box office revenue
- Given a fixed budget, a movie's box office may perform better if it is:
  - *High quality*
  - *Documentary/Horror/Fantasy?*
  - *Not Rated R*
  - *Longer runtime?*
- Given a fixed budget, a movie's box office may perform worse if it is:
  - *Low quality*
  - *Mystery/Sport/Drama?*
  - *Rated R*
  - *Shorter runtime?*

Data Limitations

- Spending aggregated
  - No information on components (e.g. salary, special effects, etc)
- Panel data assumption
  - Time series analysis may be more appropriate even though release year was not significant
- Substantial loss of data from missing budget and box office
  - Potentially unreliable for indie films
- No control for limited geographic releases
- Does not consider impacts from increasingly important foreign markets

Structural Limitations

- Omitted variables:
  - Actor/Director/Studio specific effects
  - Previous award for content creators
  - Sequel (IID)
  - Novel adaptation (IID)
- Moderate model fit
  - Suggests moderate explanatory power
- Multiple tests: Use a wide range of categorical variables
- Model form may not be ideal to account for impact of genre
- Potential interdependence of RHS variables in longer models

Conclusion

- Central question:
  - We reject the null hypothesis that budget has no effect on box office
- Sub Questions:
  - Genre likely has an effect
  - Duration has low effect
  - Maturity Rating of R may have a negative effect
- What information do the models provide?
  - May provide insight into which factors may have an effect on box office performance

Closing thoughts or discussion of potential future analysis

- Does the timing of the release affect the box office results?
  - Perform a time-series analysis on the movies to further explore the data and patterns
- Find data on spending components
- Does talent recognition and movie trends(sequels, remakes, book adaptations, etc.) impact box office results?

Q&A