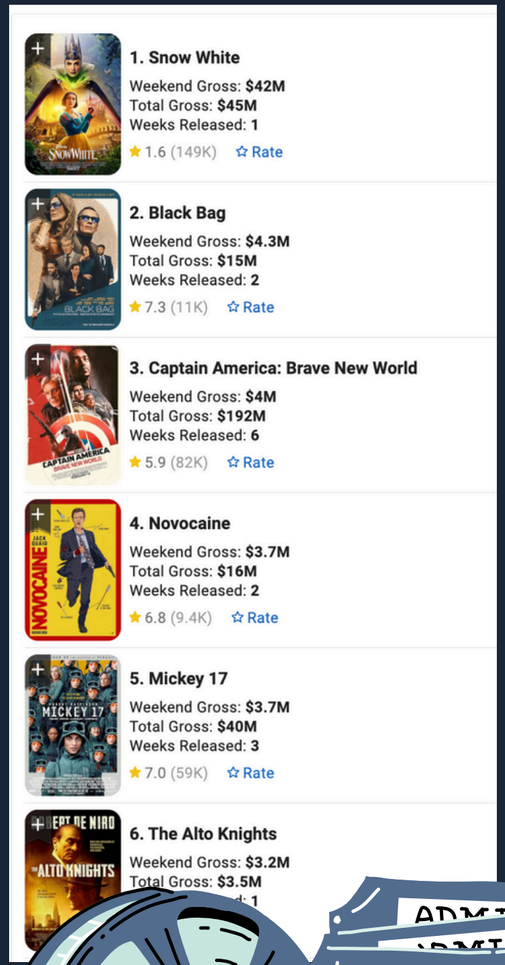
IMDB Movie Success: An Analysis of Movie Release Factors and Box Revenue

by Four-mula

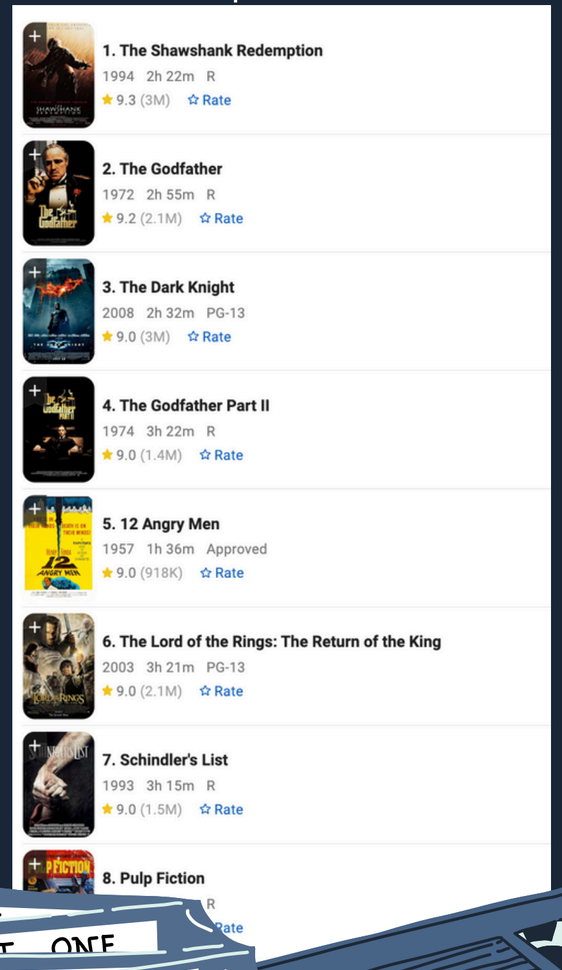
Background

- **Subject:**
  - IMDb Movies
- **Motivation:**
  - What drives a movie's box office success
  - How to better predict movie’s box office performance
- **Research Question:**
  - What production and release factors have the greatest impact on a movie's total revenue?

IMDb Top Box Office



IMDb Top 250 Movies



Dataset

- **Origin**
  - Kaggle
  - Collected by Anand Shaw from the IMDb website
- **Columns**
  - 903263 unique values
  - 21 columns on characteristics of a specific movie (basic information, ratings, monetary values)

`title`: Movie name

`vote_average`: Average rating from 0 to 10

`vote_count`: The total number of votes

`status`: The current state of the movie

`release_date`: The date the movie was officially released

`revenue`: The total earnings (usually in USD)

`runtime`: The duration of the movie in minutes

`adult`: Whether classified as adult content

`budget`: Total cost of production (usually in USD)

`imdb_id`: The unique identifier for the movie on IMDb

`original_language`: Original language produced

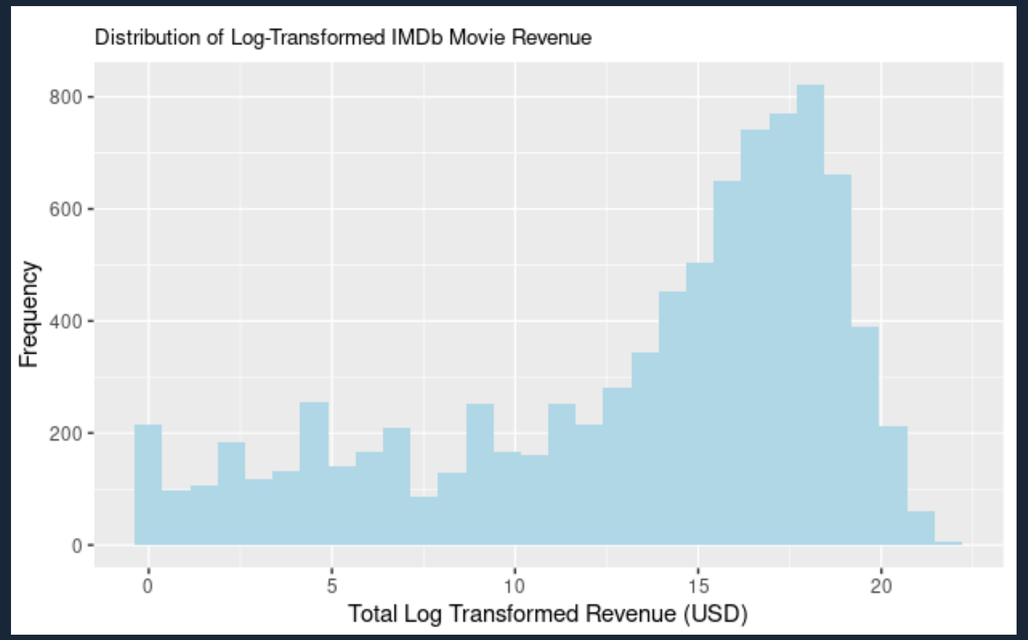
`popularity`: Popularity of the movie based on views or ratings

`genres`: The categories or genres the movie belongs to

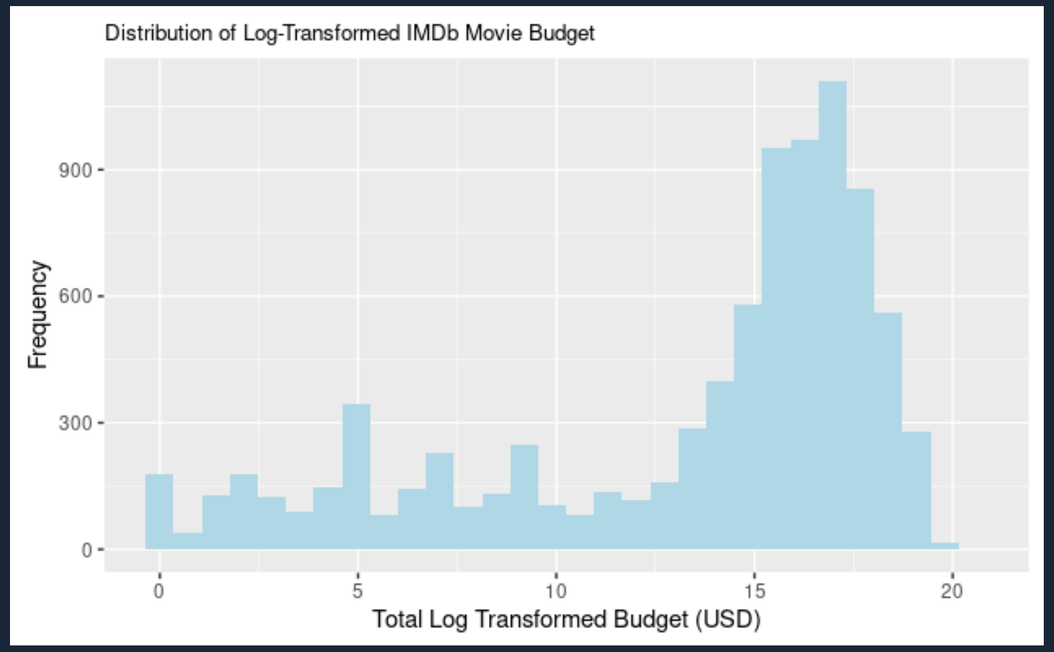
`production_countries`: Production countries

`spoken_languages`: The languages spoken in the movie

Univariate EDA

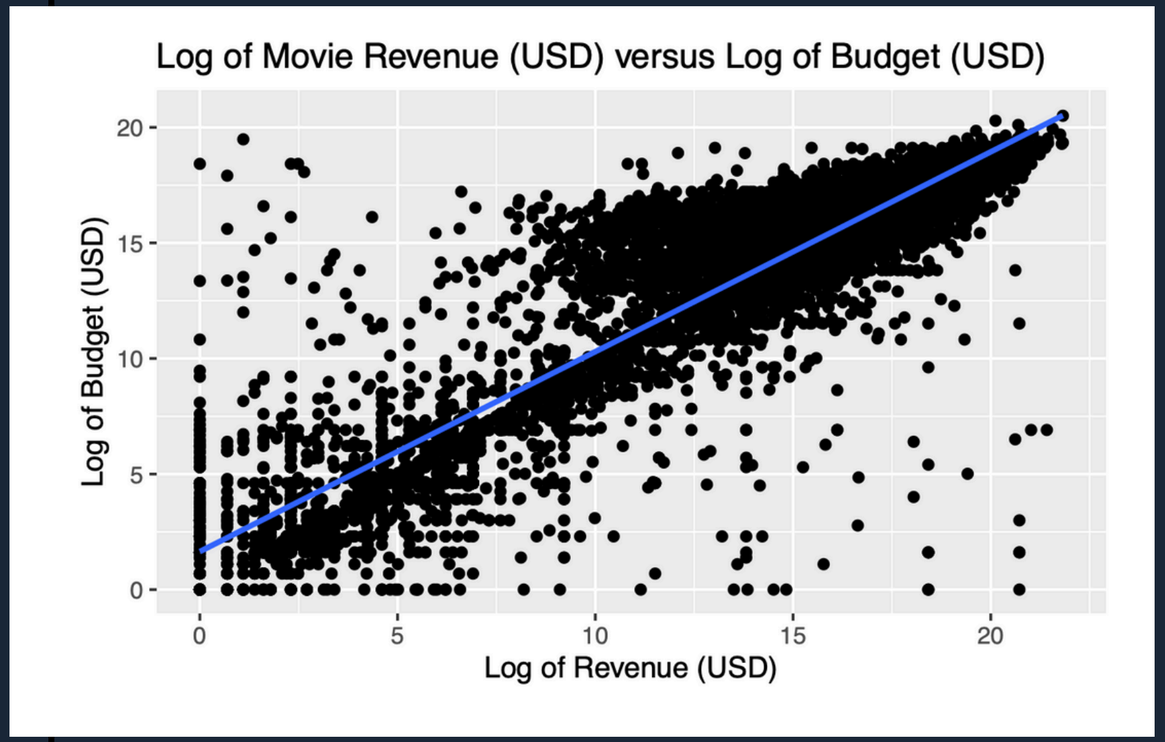


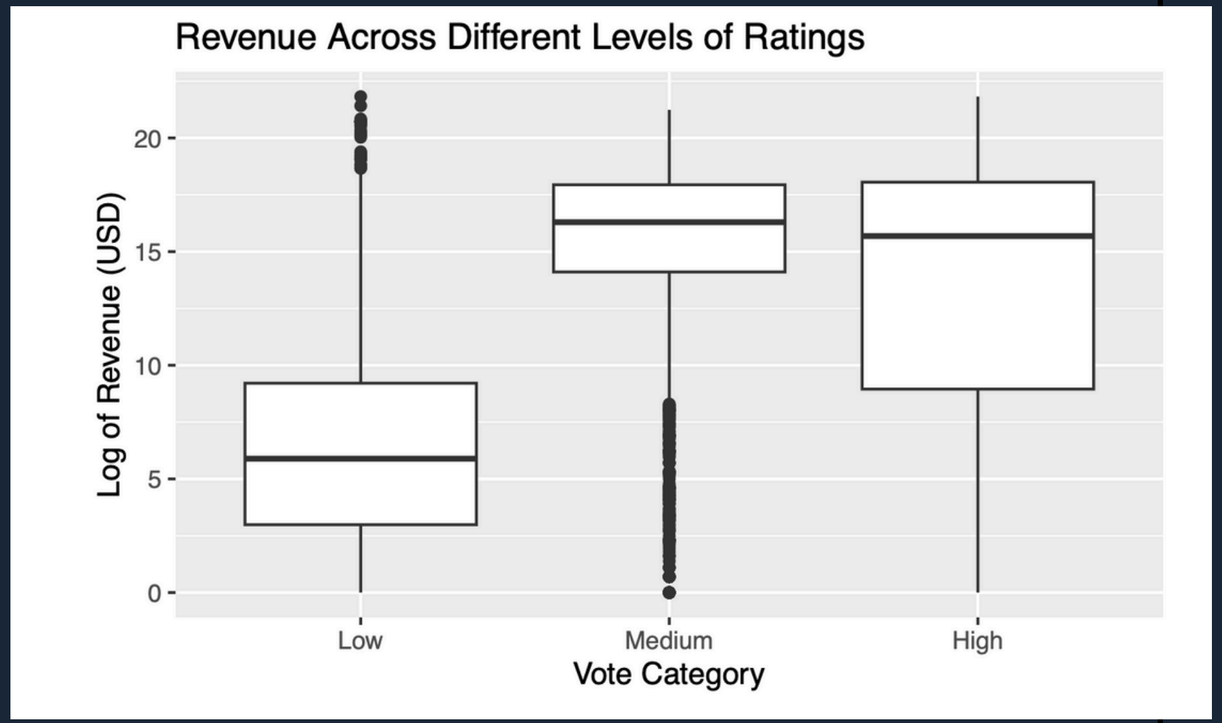
- Left-skewed, unimodal
- Center of approximately $1,800,000 USD
- Spread of approximately $18,600,000 USD



- Left-skewed, unimodal
- Center of approximately $5,000,000 USD
- Spread of approximately $25,000,000 USD

Bivariate EDA





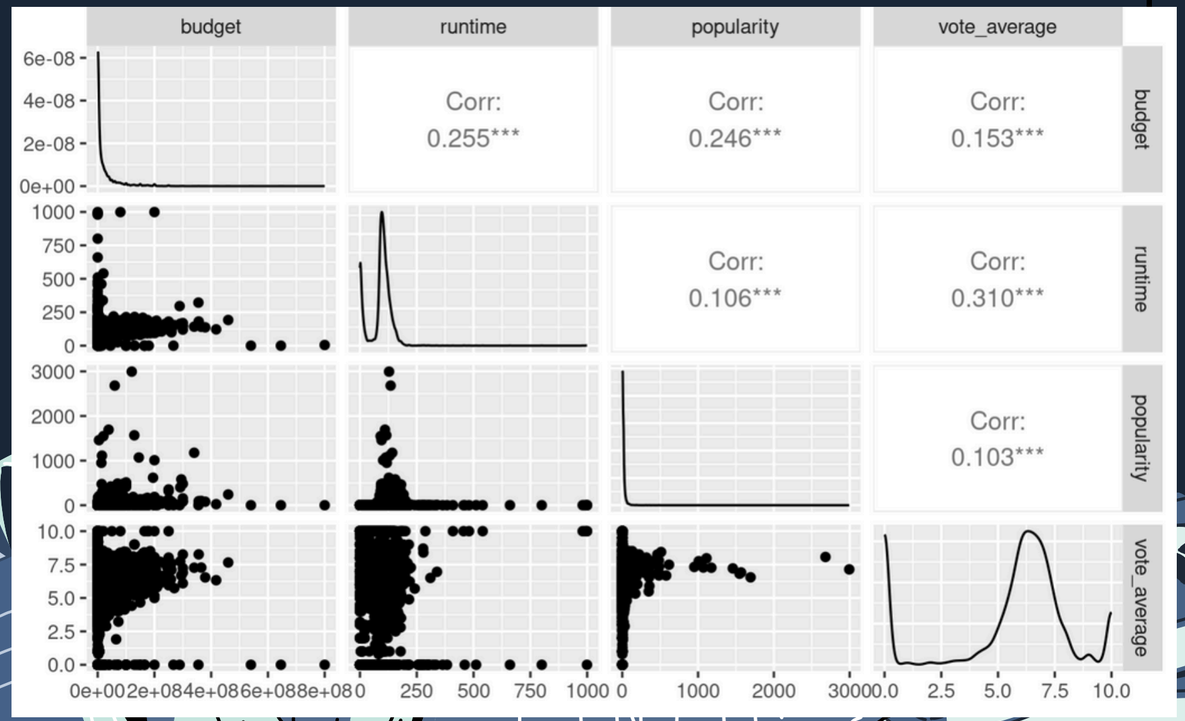
Log of revenue versus log of budget

- Positive, linear correlation
- Many potential outliers
- High concentration of movies with high budgets and high revenue

Modeling Strategies/Results

- Potential interaction effects (from conceptual evidence)
  - budget & popularity - not significant
  - vote average & vote count - significant
  - action movies & runtime - significant

- Modeling strategies:
  - logistic model
  - select predictors
  - log transformations



Next Steps

- Further analyze collinearity with categorical predictors
- Interpret the model and test the research question/prediction (higher budgets, higher average ratings)
- Reflect on limitations and applications from our findings