

Analysis of IMDb Movie Revenue

The Four-mula: Hellen Han, Lila Rogers, Amy Duan, Eva Aggarwal

2025-03-20

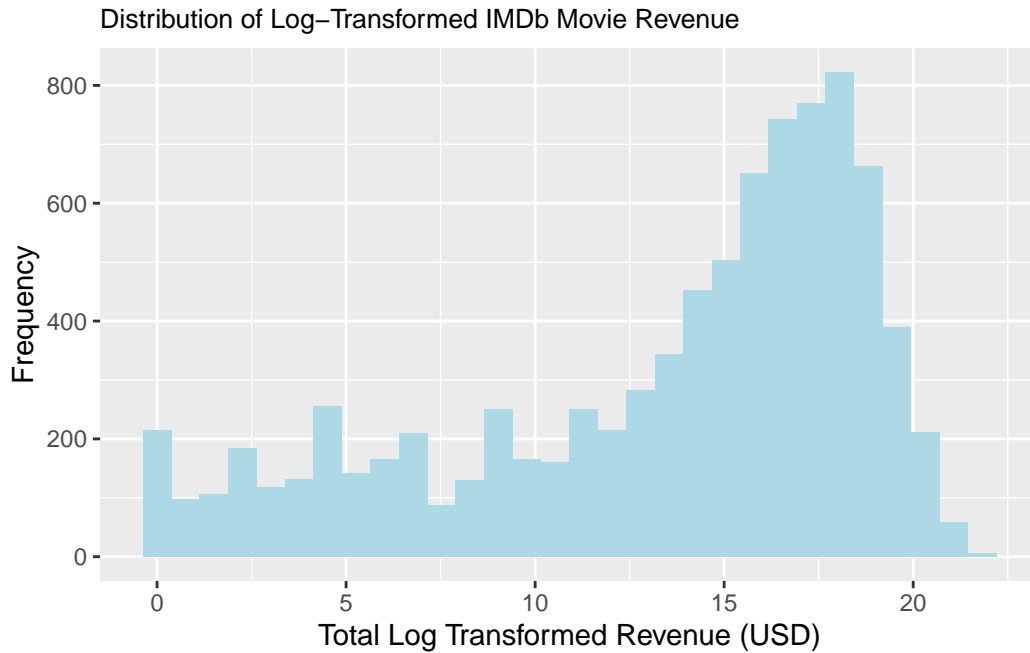
Introduction:

Exploratory Data Analysis:

To obtain informative interpretations of the distribution of the revenue response variable, we analyze IMDb movies released from 2000 and beyond that have a revenue greater than 0 (non-missing values in this dataset). Furthermore, we log transform revenue, as the revenue generated across the IMDb movies is heavily right-skewed. (hellen could you include this in the description of the variables section, maybe write a short blurb for log transformation of log transformation for budget too).

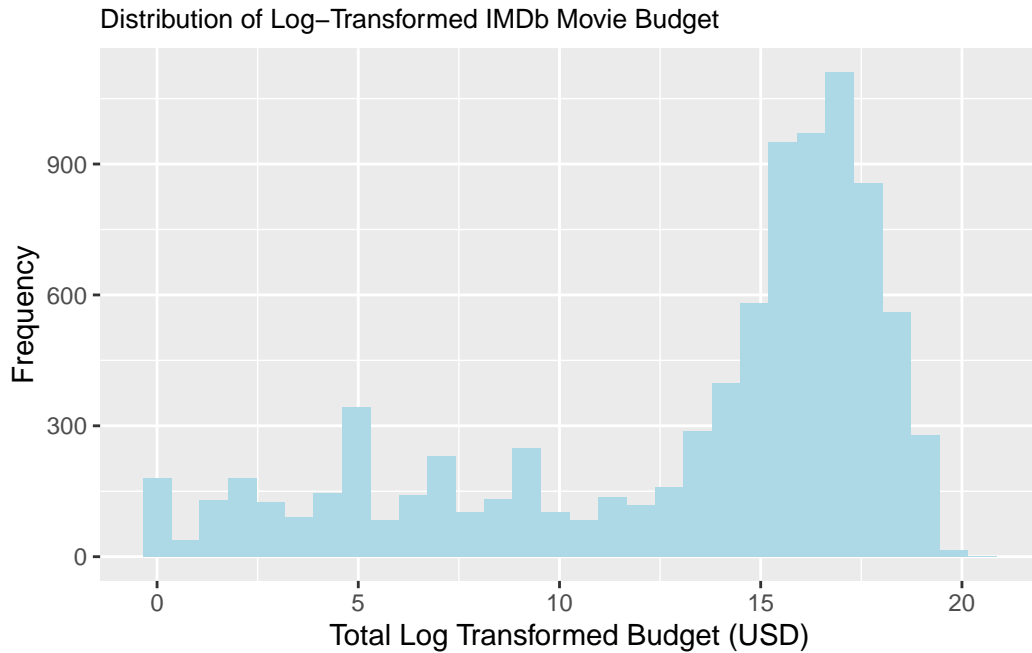
Univariate EDA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	9.933	15.323	13.387	17.646	21.822



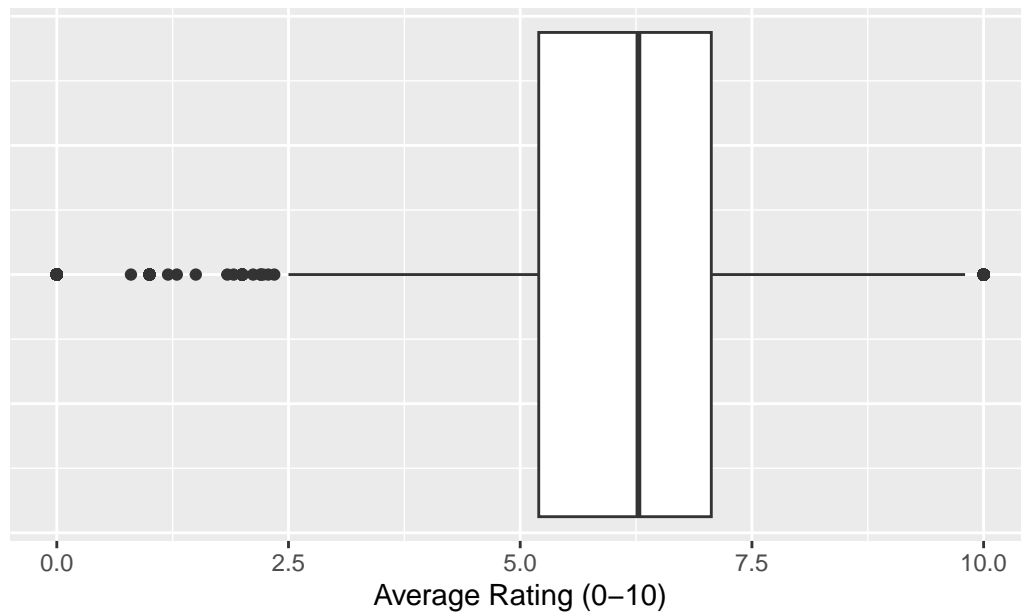
The histogram shows the distribution of the log transformed IMDb movie revenues response variable, which is left-skewed and unimodal, indicating that on average, movies listed on IMDb tend to generate higher revenues. Furthermore, the distribution of the log transformed IMDb movie revenues has a center of approximately 1,800,000 US dollars, described by the median, and a spread of approximately 18,600,000 US dollars, described by the interquartile range. The majority of IMDb movies generate between approximately 24,100 and 18,600,000 US dollars.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	9.629	15.425	13.222	17.034	20.500



The histogram shows the distribution of the log transformed IMDb movie budget predictor variable, which is left-skewed and unimodal, indicating that although IMDb movie budget varies greatly, on average, movies listed on IMDb tend to have higher budgets. Furthermore, the distribution of the log transformed IMDb movie budget has a center of approximately 5,000,000 US dollars, described by the median, and a spread of approximately 25,000,000 US dollars, described by the interquartile range.

Distribution of IMDb Average Rating



The majority of IMDb movies' average ratings are concentrated between a rating of 4.0 to 8.0. A few IMDb movies have a low rating (average rating below 2.5), contributing to the left-skewness of the distribution of IMDb movies' average rating predictor variable. Furthermore, a few IMDb movies have an extremely high rating (average rating of 10.0).

Bivariate EDA

Potential Interaction Effects

! Important

Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML.