

2018: A Statistical Odyssey of Movies

Eric Bengyak
Christopher Goode
Grayson Niehaus
Francis Vu





Agenda

- Questions, Data, and Sources
- Overall Movie Industry Trends Across Revenue, Budget, and Score
- Key Takeaways
- Future Considerations In Movie Industry



What we were looking for:

- Across total movies and total revenue, how has the movie industry changed over the past two decades?
- A comparison of revenue, budget, and viewer score and the relationship between the three factors



Dataset and Filters

- Kaggle – collective data set of revenue, budget, cast, crew, language of movie, and country of release
- Key Parameters:
 - Revenue: domestic revenue only
 - Budget: total budget
 - Language: English only
 - Country: United States
 - Distribution: Theatres only (sorry, Netflix)
 - Timeframe: 1995 – 2016
 - 2017 not fully complete

Data Filtering: Porn, Parsing, Price Index, and Potential Issues



- **Porn:** our original dataset from a separate database included many,many,many porn titles, leading us to abandon the original dataset. New one found on Kaggle
- **Parsing:** New data was very dirty, had to write functions that would convert our strings as well as extract the year only
- **Price Index:** All numbers are adjusted for inflation to 2017 numbers using the Consumer Price Index from the U.S Bureau of Labor Statistics
- **Potential Issues:** we used total budget, which may include overseas marketing spend, but only analyzed domestic revenue



KPIs for analysis:

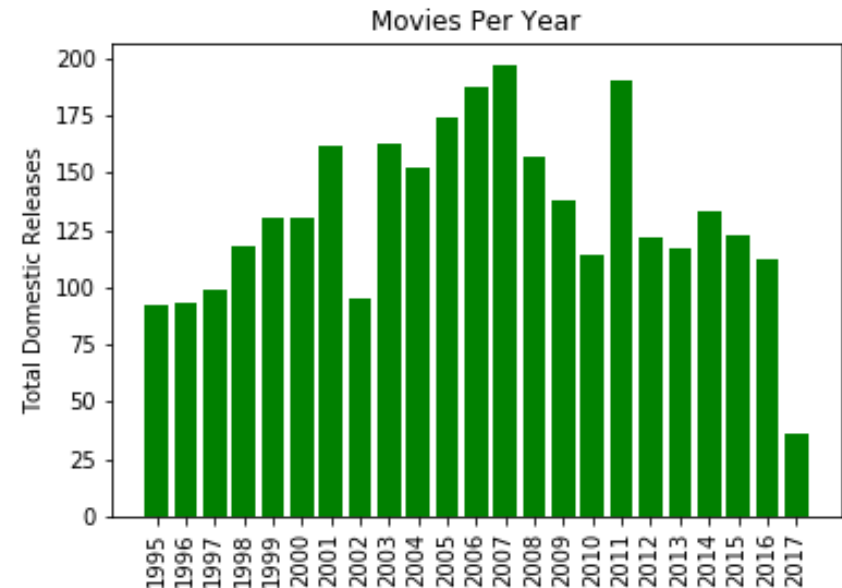
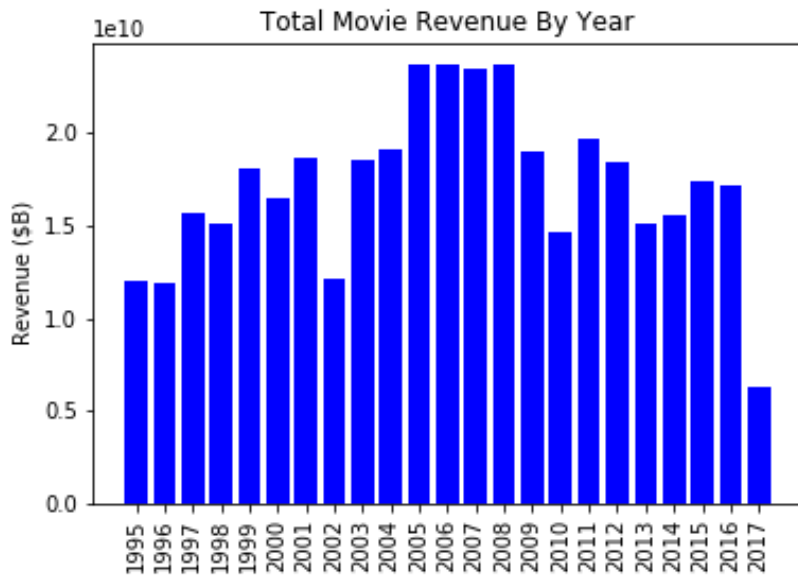
- Revenue – everyone needs to make money!
- Rating – viewer score based on votes from IMDB
- Budget – need to spend money to make money!
- The relationship between these three factors determined our analysis

Average Movie Revenue
\$130M

Average Movie Budget
\$48M

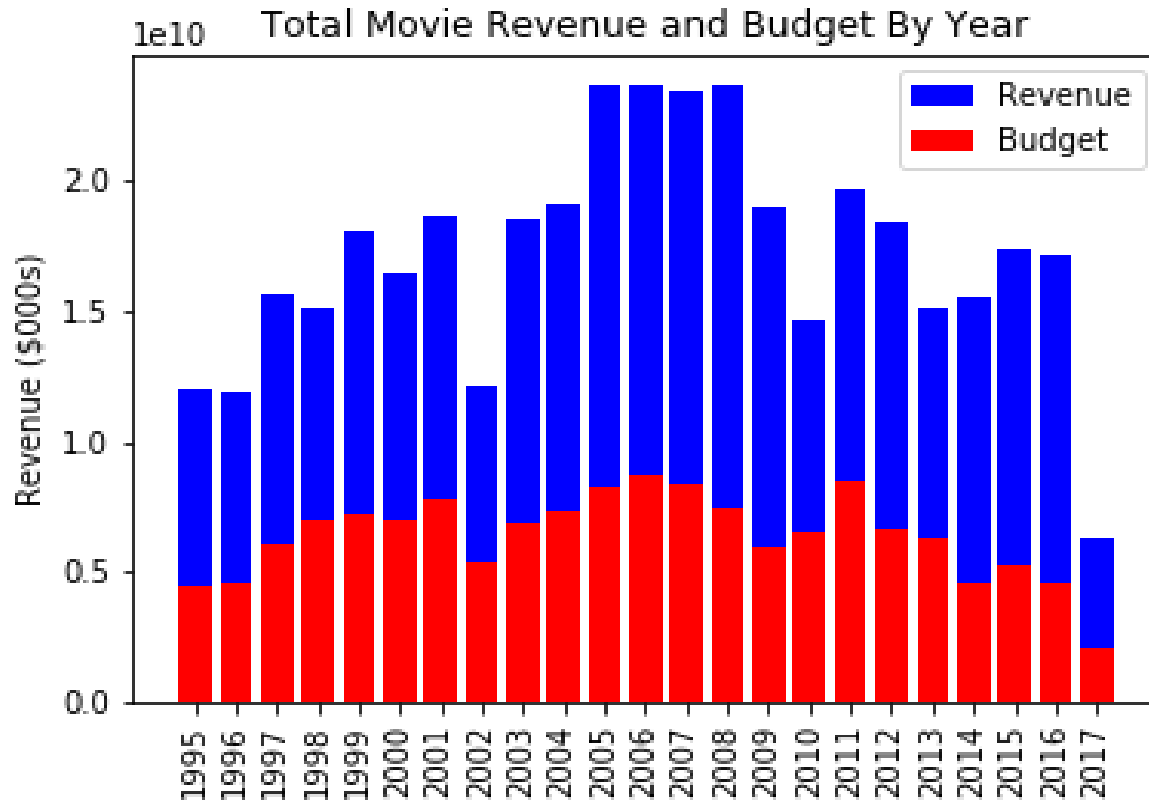
Average Movie Rating
6.2

The movie industry hit its peak in 2008, but has since faltered



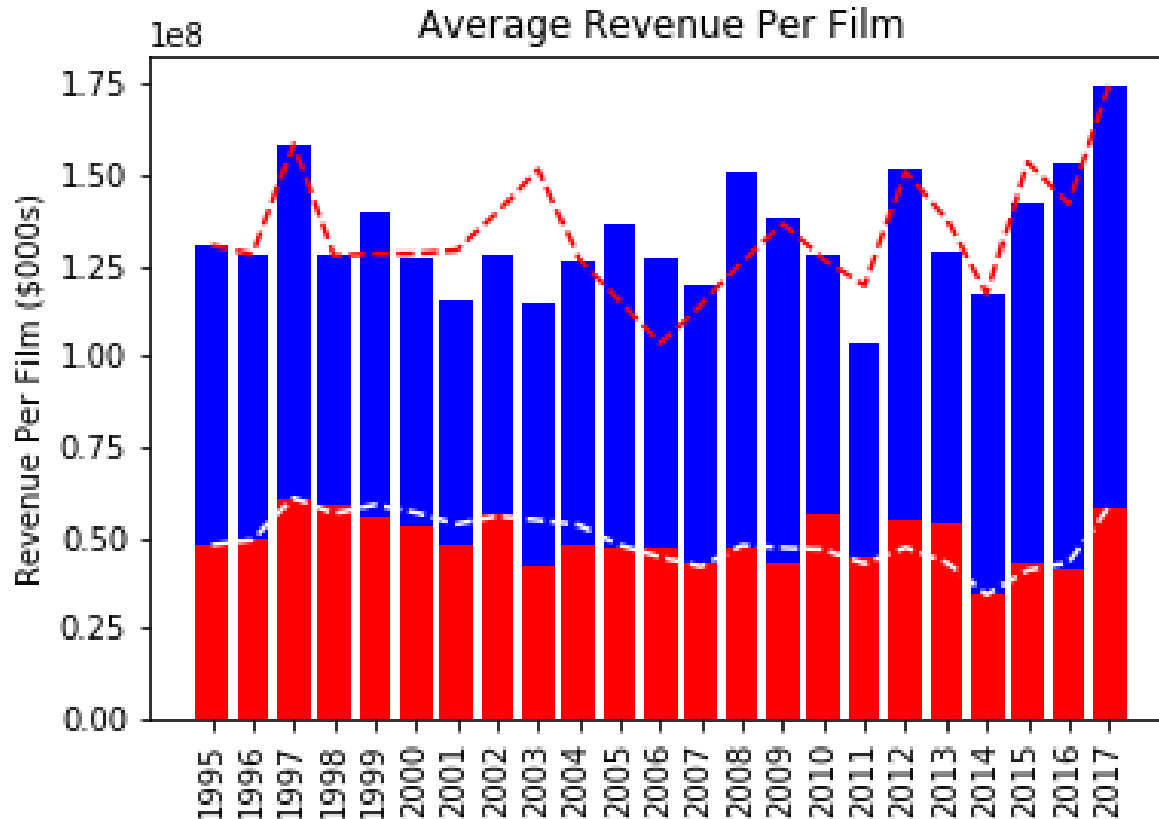
- Domestic movie revenue and total releases hit their peak in 2008, but both have been declining back to similar numbers seen in the 1990s
- Current revenue and movie releases have remained relatively similar over the last 6 years except 2011
- Key Considerations: Writer's strike in 2008, economic downturn, Netflix publishing digital platform in 2007

However, annual movie budgets have begun to decrease



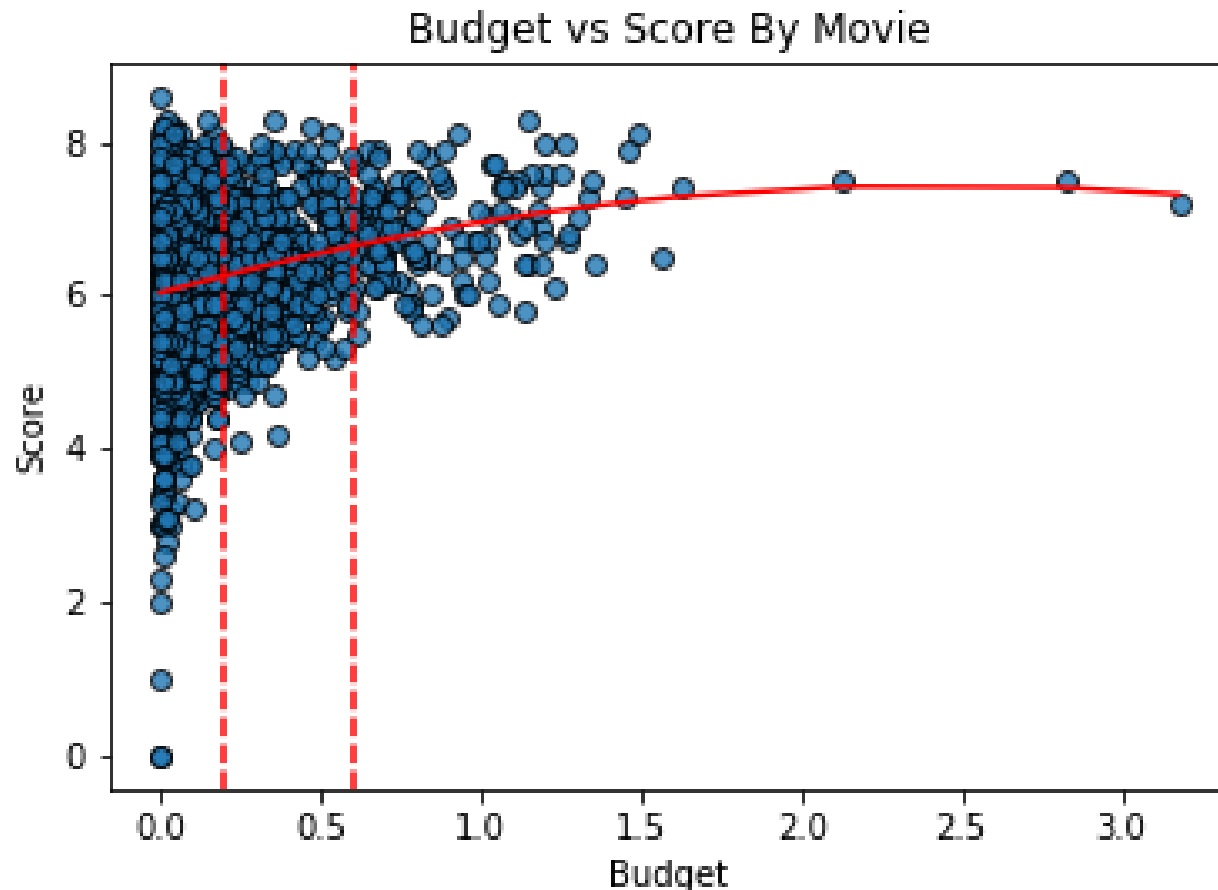
- As a percentage of revenue, budget has decrease from it's peak of 43% in 2006 to only 30% in 2016

On a per-film basis, revenue has increased while budget stagnated

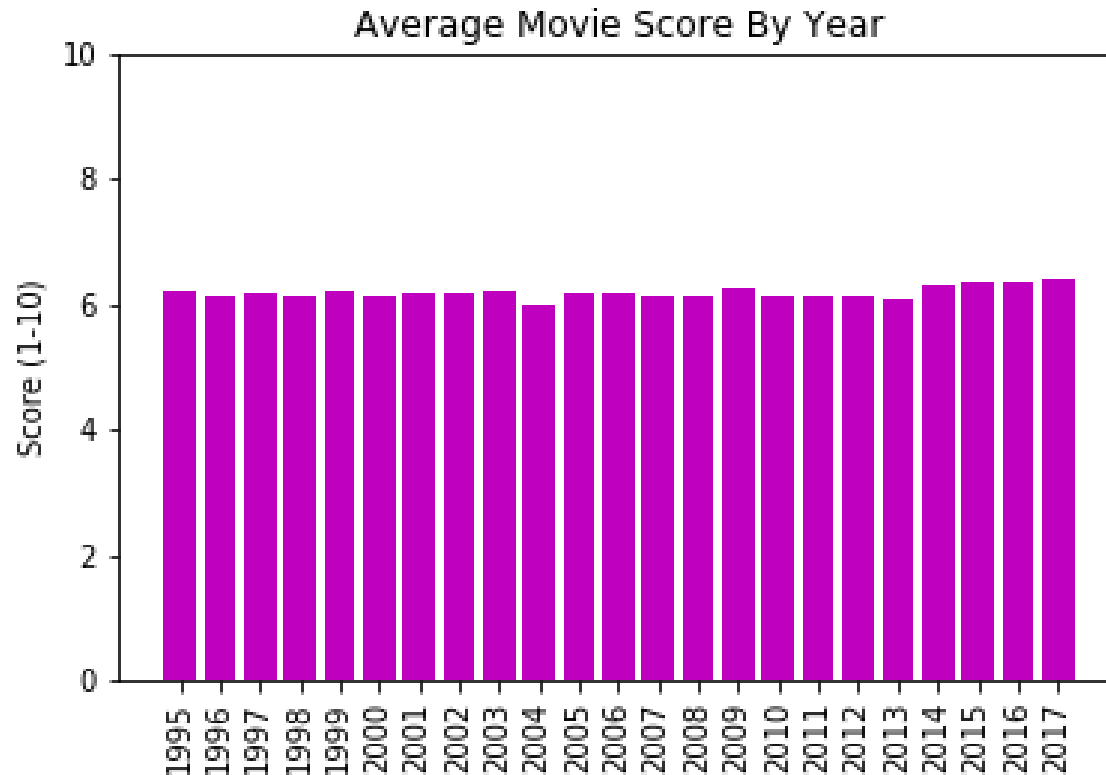


- Revenue per film has hit an all time high of \$170M in 2017, while budgets have remained stagnant
- In 2011, there was a drastic increase in number of films, however revenue had it's worse year

And production budget has had relatively little affect on score

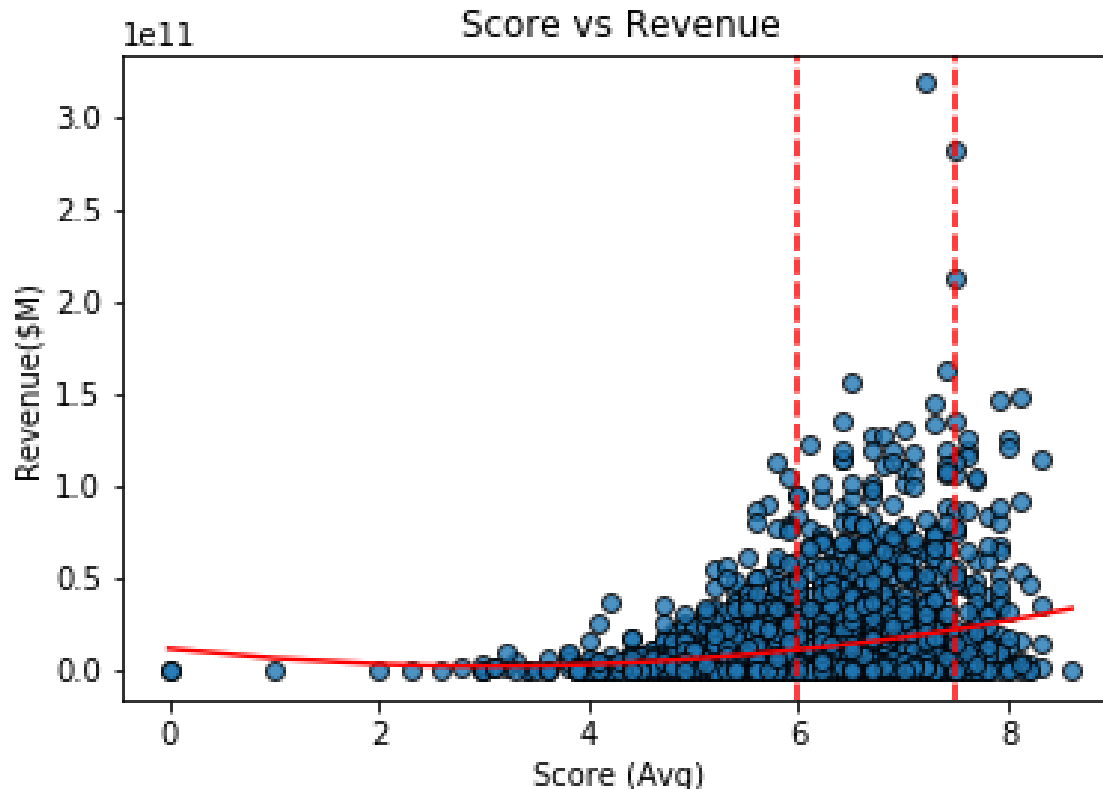


Overall Movie Ratings Haven't Changed Drastically



- The slight uptick shows a slight change in overall score – but the data doesn't offer an explanation as to why

However, once films exceed an average rating, revenue increases



- Rating does have an effect on revenue, as movies scoring above 6 (also about the average) tend to generate more revenue



Project Key Takeaways

- Sourcing data is a key component of data analysis
- Cleaning data is...hard
- The movie industry has become more productive, but isn't actually making better movies
- Production budget has relatively no influence on movie rating
- Having an above average rating does increase revenue

The movie industry looking forward



- Low budget and high ROI
 - Get Out
 - Paranormal Activities
- Data Analytics influencing creation of film
 - Kevin Spacey and Netflix
- New ways to account for revenue
 - Subscription services, high touch theatres
 - Disney/Hulu ownership model