

Introduction

Hello there, This analysis notebook aims to answer some questions I find interesting about the numbers & stats behind popular movies. Wish you find the results of the analysis intersting & hopefully get a better understanding of the mainstream movie industry as a whole.

Here are the stuff we are inspecting in this notebook.

- **Q1.** Profitable producers
- **Q2.** Profitable genres
- **Q3.** Pobular movies era
- **Q4.** Genres quality trends
- **Q5.** Sequel rating change
- **BONUS.** Descriptions wordclouds

NOTE: the movies used in the analysis are the top 2000 movies in vote count on TMDB, in other words the movies here are the **most voted** on TMDB.

Setting up

Technichal info about the cleaned dataframe used in this analysis

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2463 entries, 0 to 2462
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            2463 non-null   int64
1   id                    2463 non-null   int64
2   imdb_id              2463 non-null   object
3   release_date         2463 non-null   object
4   title                2463 non-null   object
5   collection           991 non-null    object
6   overview             2463 non-null   object
7   tagline              2387 non-null   object
8   language             2463 non-null   object
9   runtime              2463 non-null   int64
10  revenue              2360 non-null   float64
11  budget               2368 non-null   float64
12  vote_avg             2463 non-null   float64
13  vote_count           2463 non-null   int64
14  genres               2463 non-null   object
15  production_countries 2463 non-null   object
16  production_companies 2462 non-null   object
17  revenue_z_score      2360 non-null   float64
18  budget_z_score       2368 non-null   float64
dtypes: float64(5), int64(4), object(10)
memory usage: 365.7+ KB
```

The analysis

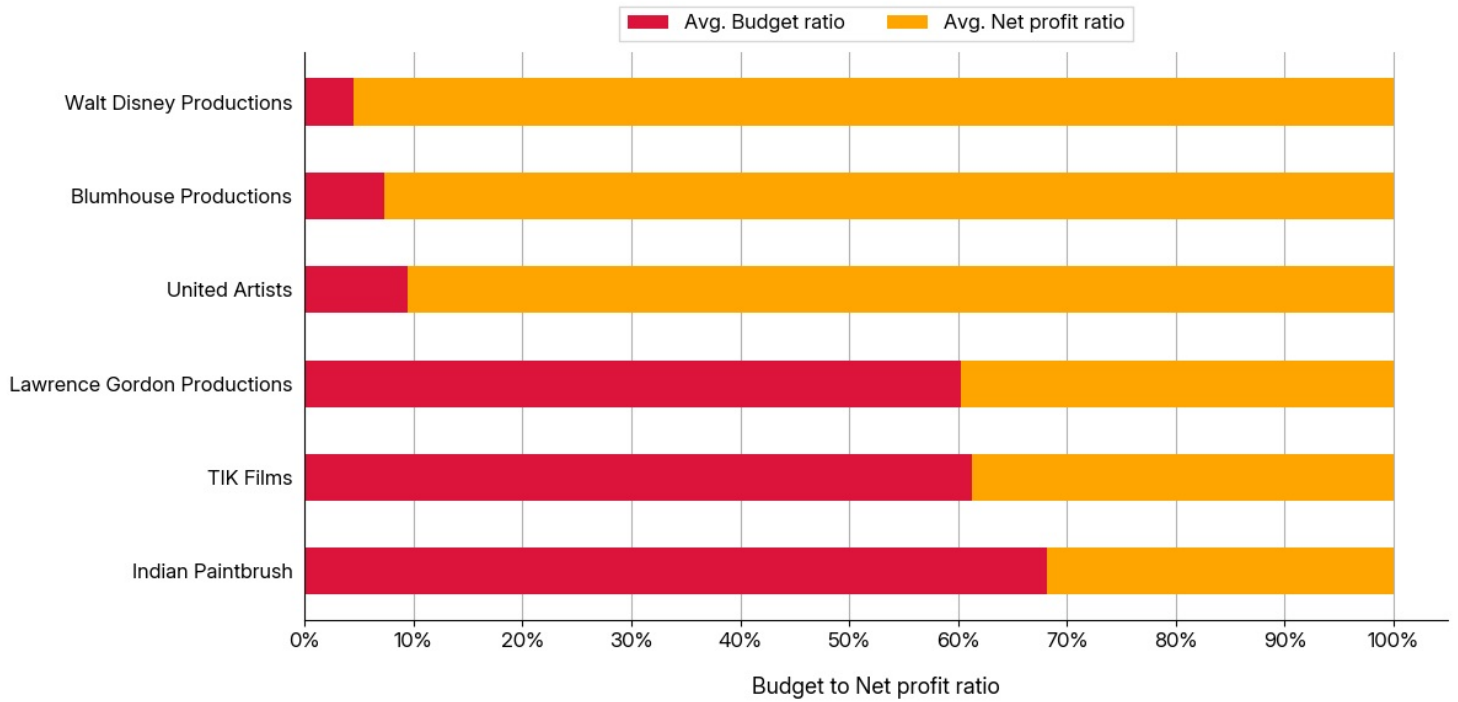
Q1. Profitable producers

Let's start by looking at best & worst production companies at generating large net-profit with small budget which might help anyone looking to invest in any any of those production companies.

This is simply caculated by represnting the ratio between average budget & average box office net-profit for the movies the production company has made.

NOTE: DON'T use any output of this analysis for decesions regarding investments, or money in general, You have been warned.

Average "Budget to Net-profit" per Producer



I think that people who watch the said production companies within the chart can relate to what it's saying. I just can't explain it more without a long paragraph which I won't do.

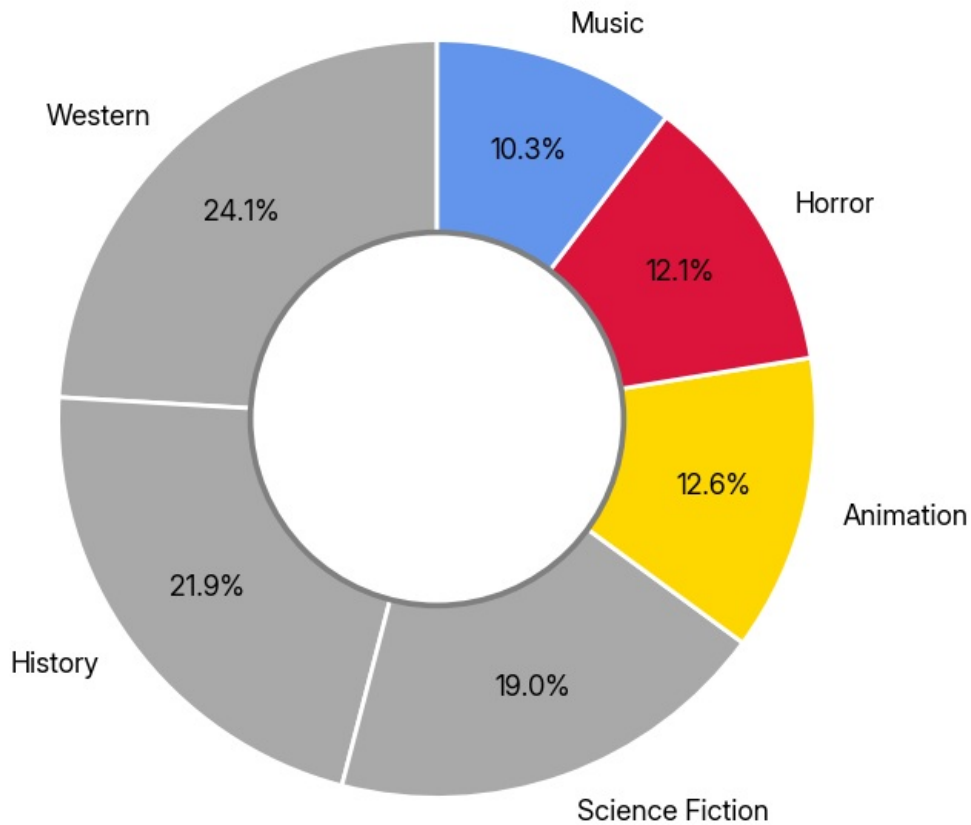
Q2. Profitable genres

Now we'll look at which genres generate high box office net-profit with the least budget. By showing thier budget to net-profit percentage.

This question will help us understand more about the film-making industry.

Budget to Net-profit per Genre

Top 3 & Least 3



Wow, this chart actually makes a lot of sense; History, Science Fiction & Western movies can be really expensive to produce which increases their *"budget to revenue"*. Where Horror & Music movies are relatively cheaper & able to produce a large revenue.

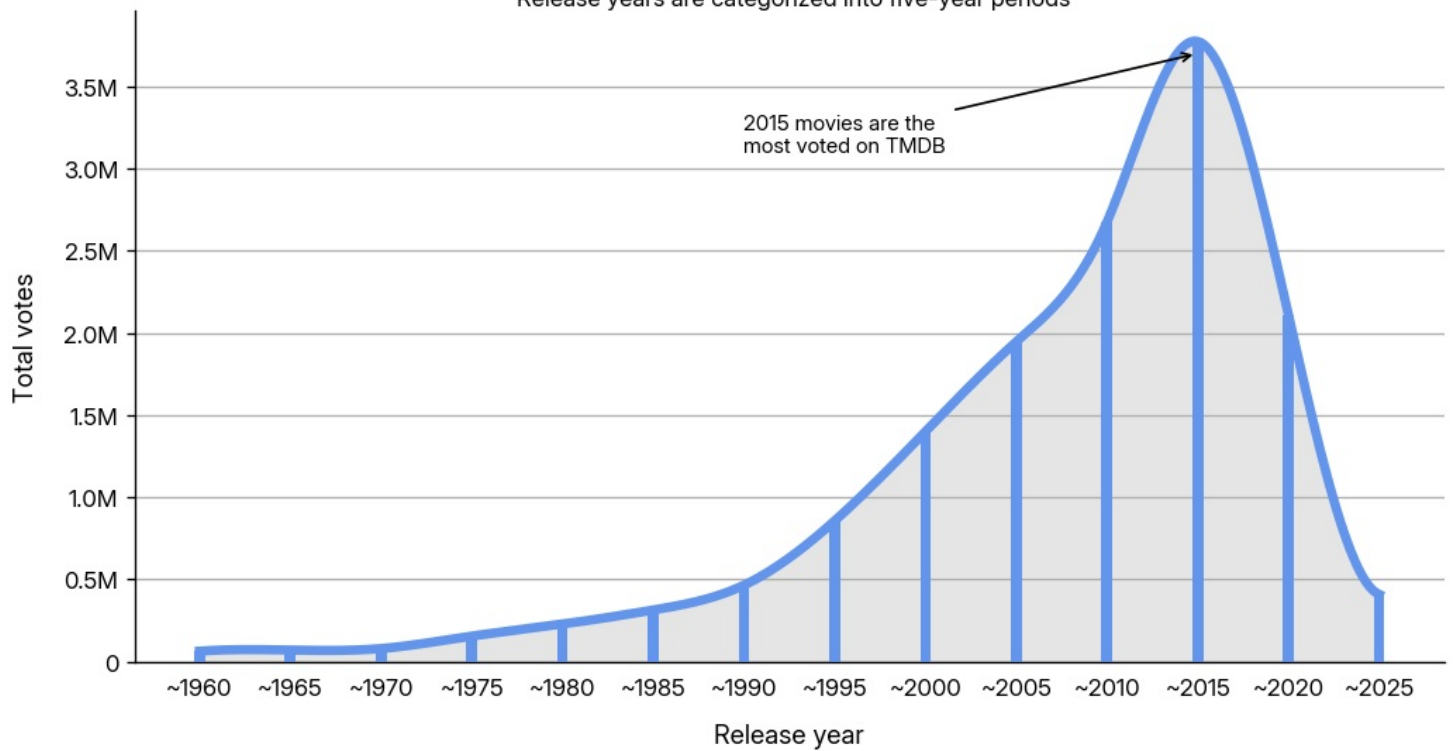
The unexpected is that Animation genre has one of the lowest *"budget to revenue"* though known to be one of the most expensive genres to produce, my guess of why that is true, is because production companies do really good marketing and won't risk making an animation movie if not 100% it'll succeed.

Q3. Popular movies era

In this chart I aim to find the movies era with the most total votes of all time on TMDB, in other words which movies era did reviewers review the most. Indicating us what era produced the popular movies.

Total Movie Votes by Release Year

Release years are categorized into five-year periods



We can see a few nice trends here are the most noticeable ones:

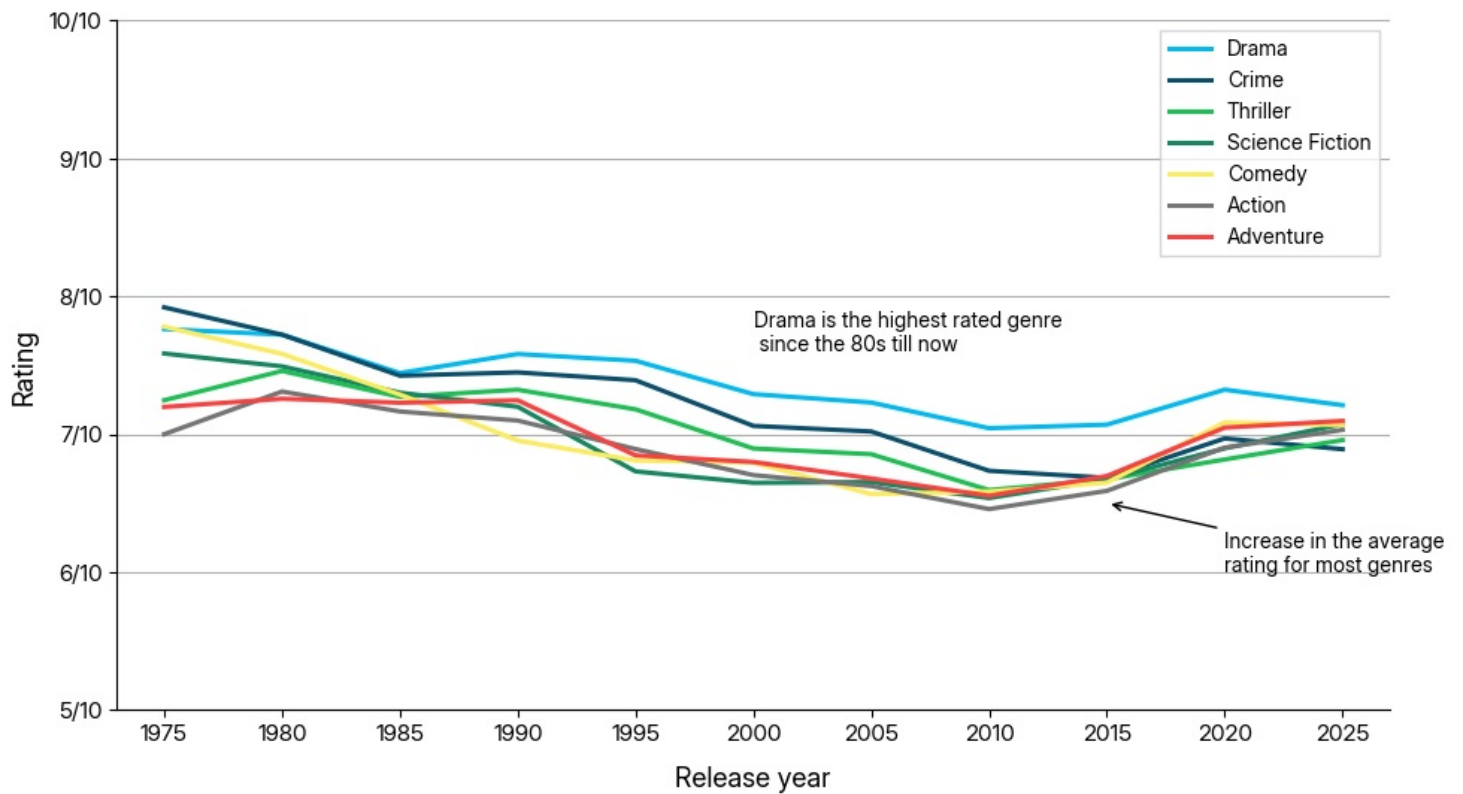
- 2013~2017 is the peak of total votes
- Votes fell drastically around 2020 & 2025
- Votes for movies sky rocketed after ~1995

Q4. Genres quality trends

This is the most "cliché" question here questioning: Are movies across different genres still the same quality *-indicated by rating-* as the good old days? or have it change to worse?

The latter for some reason is usually the answer.

Average genre rating over time



There are 4 notable things we can infer with ease from this chart:

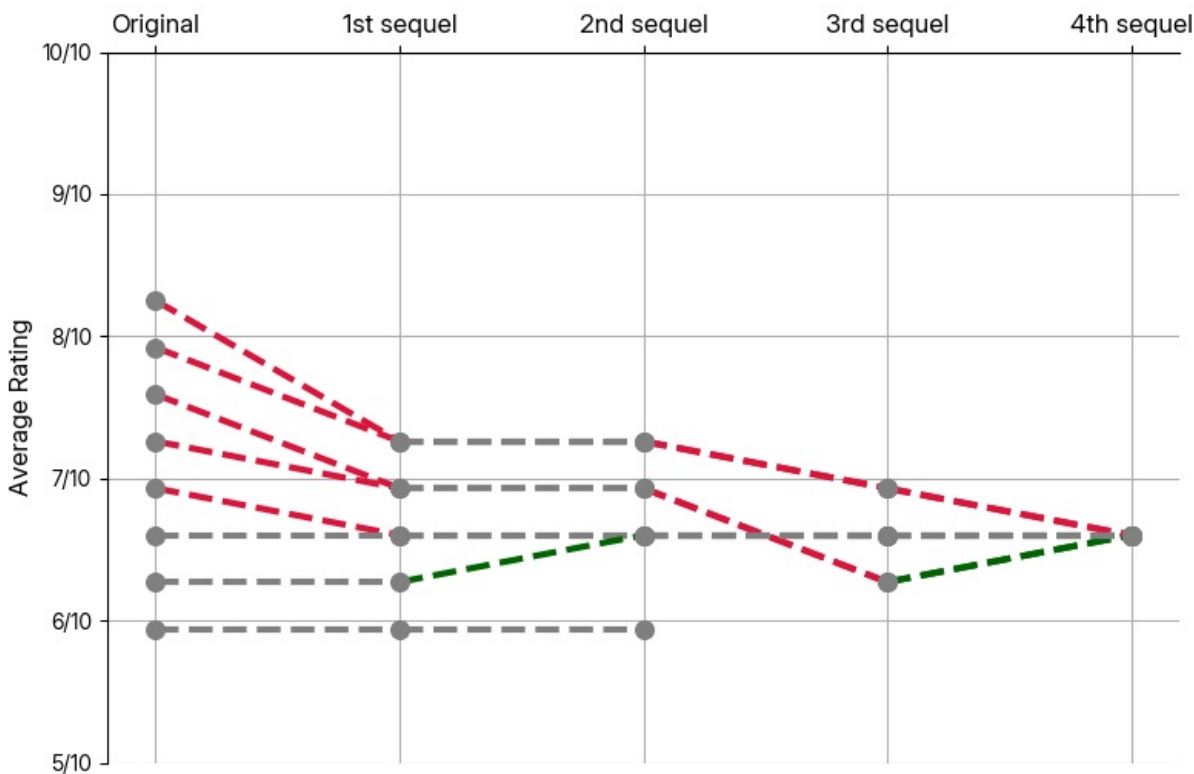
1. There was drop in rating across every genre around 1985
2. Drama followed by Crime & Thriller were the most to survive the 1985 rating drop
3. Drama genre is the highest rated genre overall with a decent positive edge
4. Unexpectedly there was an increase in the rating across all genres around 2015

Q5. Sequel rating change

How much does the rating change across a movie sequel in average ?

I think we all know the result of this beforehand, but let's see what the data has to say.

Average Rating change across Film Series

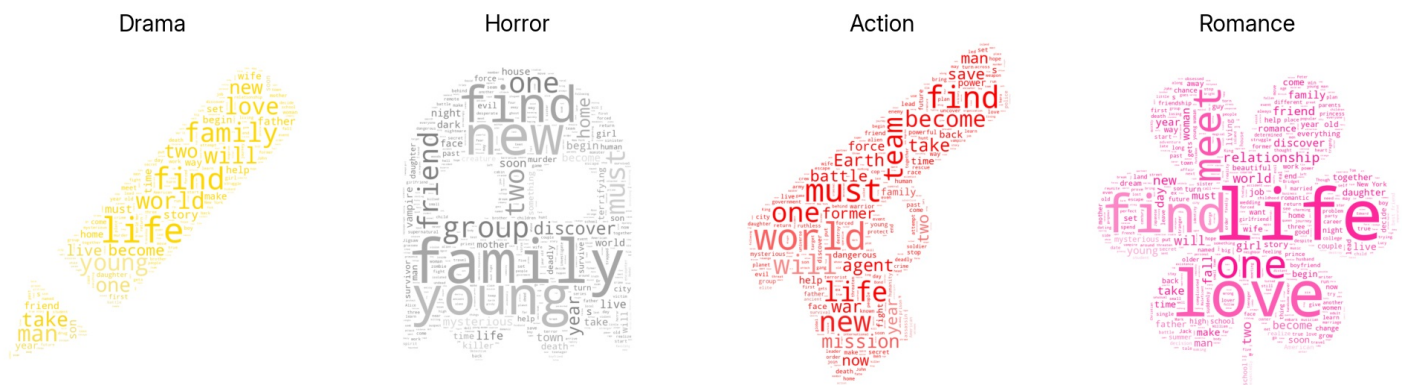


That was expected, movies usually get worse in quality throw the movies series leading to worse rating, what's kind of notable is how the transition between the second & third sequel in average doesn't change the rating.

Bonus. Descriptions wordcloud

Let's finish off with a little fun visual showing us frequent words used in movies movies descriptions across different genres.

Film genres Wordclouds



This chart has **0** practical uses but it has an eye-pleasant & fun look which is enough.

Summary

If you don't want to read all of the above here is a summary of the findings ;)

- Q2: Profitable genres:
History, Science Fiction & Western movies can be really expensive to produce which lowers their *"budget to revenue"*; Where Horror & Music movies are relatively cheaper & able to produce a large revenue.
- Q3: Popular movies era:

Around 2015 was when most voted movies -on TMDB- were released but fell drastically after that.

- Q4: Genres quality trends:
There was a significant drop in rating across every genre around 1985 that continued till around 2015 where it was starting to get better again.
- Q5: Sequel rating change:
Film sequels tend to diverge to worse & most of them diverged after 4 sequels to rating of about **~6.5**