Going to the Movies with Popcorn, Soda, and Python

Jack Pfeiffer

ABSTRACT

Movies today make up a major part of American Culture. This is not anything new. Movies were the most popular and impactful form of media during the beginning of the twentieth century[1]. During that time though there were no where near as many forms of media as there are today, and yet movies are still very popular. This paper is going to show how it has changed in last half-century. As the movie industry has been growing, along with the cost of production and the return in revenue. All the money moving around the industry it is interesting to understand why it has been and continues to grow along with how it all started. Why do we spend so much money on movies? With how many movies are coming out each year no one can see all that come out, no should they. A common filter the general public uses to determine if a movie is worth seeing is the critics'reviews. Do these critic reviews actually have that strong an impact on the success of a movie?

1 INTRODUCTION

Understanding that movies are not for everyone, there is a lot of time and money that goes into both seeing them and creating them. As previously mentioned movies have been around for a long time and continue to be one of the most popular forms of media. In recent decades with the advancements in technology and media platforms it is easier for studios to create several movies, and as a result more and more have been coming out each year over the last half-century. It is now starting to plateau, but still with how many releases there are it is hard to stand out as a studio. They need to put a lot of money into their movies to get recognizable actors, special effects, and movie sets.

As mentioned one of the ways we try to measure the success of a movie is through their reviews. Do these reviews really match up with the success of the movie financially, or are reviews just subjective? Could it possibly be a combination of the budget of a movie and its reviews? Could this be why we spend so much money on making movies? Maybe it is because movies have become more popular and to have your movie noticed you need to put more money in to making it. The budget, revenue, and reviews are all public knowledge. We can look at how these are related to start making inferences.

After we have found a data set that contains movies with these metrics (budget, revenue, and critics reviews) we can then import our data set into a Jupyter Notebook and use several python libraries to both clean up and present our data. Data it is not always perfect, not talking about for proving a point, but for usability or even makes sense. With our data set we had some movies that were listed in the past with a budget of 0, which doesn't make much sense. There were also some inconsistencies making it harder to present. Some of the movies release dates were formatted differently so to make
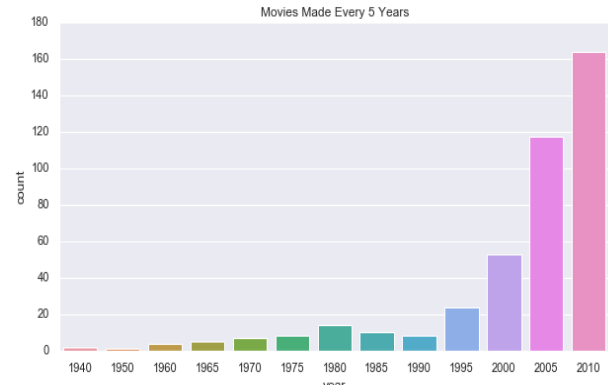
---

[1]SKLAR, ROBERT. 1994. Movie-Made America. Libary of Congress Cataloging in Publication Data

Figure 1: Number of Movies Made Every 5 Years

it easier we consolidated the dates. When were able to make the data easier to use, we are then better equipped to present our data better.

Once the data set is cleaned up and diagrams are created involving our variables of interest in a presentable manner, we are able to inference and hypothesize why we may see what we see, or why our preconceived ideas were not supported. After our inferences are supported by our data we can research historic and analytic texts to determine why it is the way it is. Once were able to backup our claims we can identify how these trends may impact the movie industry or related industries.

2 LITERATURE REVIEW

It quite interesting to research the film industry as text and film are quite different mediums themselves. There is a lot out there on the topic though, especially in terms of international movies. In order to narrow our research and data to parallel findings we needed to only look at American films.

The first article, *Movie-Made America: A Cultural History of American Movies* by Robert Sklar takes an almost anthropological perspective to the movie industry looking at the history and economics of how movies became such a significant part of human culture and society. The text touches on the beginning of the industry and its development through the uprising of mass mediums. It talks about how through the nine-tenth century and well into the beginning of the twentieth century movies went unnoticed.

In *Does film criticism affect box office earnings? Evidence fro movies release in the U.S. in 2003* is a more of a statistical analysis report on the correlation between critics reviews and box office earnings. The paper points out that their is a decent correlation and even a few examples with almost absolute correlation. The article explains why this correlation is not causal.

The final article, *Is love of nature in the US becoming love of electronic media? 16-year downtrend in national parts visits explained*

by watching movies, playing video games, internet use, and oil prices also talks about correlation. The correlation between national park visits and not just movies, but other forms of electronic media. The article brings up how it does not mean it is a causation, but does bring up a potential reason is, "that all these factors are competing for our limited time" [2]

I found these three articles after wondering after looking at my data visualizations. Do critic reviews actually affect revenue? When and why did movies become the main stream media they are today? And finally, Why do we spend so much on movies

3 METHODS

The data set used in the paper was found on Kaggle[3]. There was more information included in the data set that was not needed. The first thing done was to remove any information that would not be used. The columns of interest was the budget, title, genres, release date, revenue, runtime, vote average and vote count. The title of the movie does not make a difference in the analysis, but it was used to query any outliers and anomalies during the process. There was not much we could do with a JSON list of genres so we changed the column to be the first listed genre. For an unknown reason there were movies listed with a budget of zero. So we filtered the data to only look at movies with a positive, nonzero budget. To group the movies when looking at when they came out we did not care about the month nor date of the release so we changed the column to only list the year. In order to do this we needed to parse two different formats of date YYYY-MM-DD and MM-DD-YYYY. After cutting down the genres, there were some genres with only one movie. It is difficult to talk about movies in a genre or a genre as a whole with only one data point.

After we were able to clean up the data significantly, it was not time to develop visualizations to observe, infer, and support claims. A common method of looking at movies spanned over several genres was using small multiples to split up the relationship between two variables across different genres. By doing this we could see where a specific variable may have a bigger impact on another in one genre more than another. For example one may believe that the budget of an action movie may play a bigger role in the success of a movie because the movie creators are able to pay for more visually appealing actions scenes and special effects. This relationship may not be as strong in Comedy where writing is a bigger factor than visual effects. With the small multiples I decided to use scatter plots I best represent to spread of all movies in each genre. Another form of data visualization was the bar chart both a for frequency diagrams and to represent the sum of all values with a common value. The frequency bar chart was used to show how many movies of each genre are present in the data set and the another was used to show how many movies were made each year in the data set. The other bar charts grouped movies by every five years and plots the average budget of movies in each five year span and another for average revenue every five years.[4]

[2] R.W.Pergams O, A.Zaradic P. 2006. Is love of nature in the US becoming love of electronic media? 16-year downtrend in national park visits explained by watching movies, playing video games, internet use, and oil prices. Journal of Environmental Management. https://doi.org/10.1016/j.jenvman.2006.02.001

[3] https://www.kaggle.com/rounakbanik/the-movies-dataset$movies_metadata.csv$

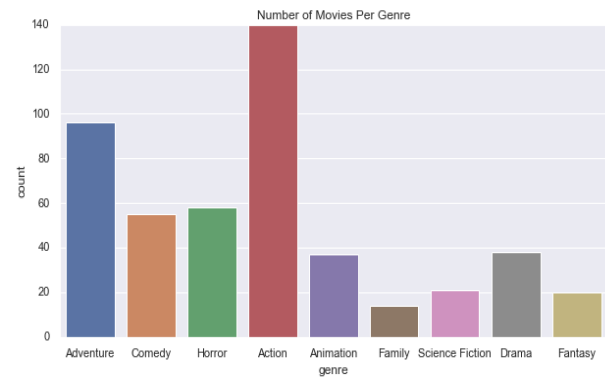[4] https://github.com/Jl5her/IntroDataScience-Fa18/tree/master/Project1

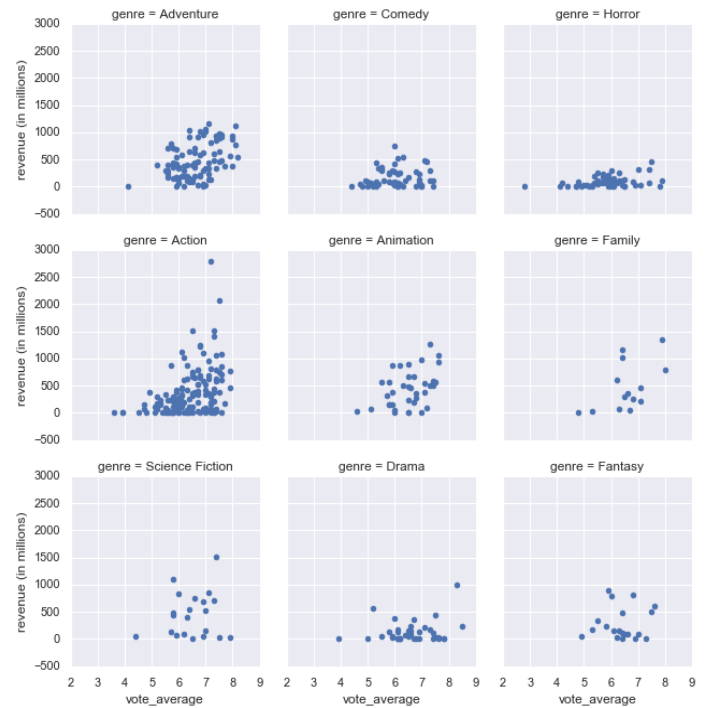**Figure 2: Number of Movies in Each Genre**

Figure 3: Vote Average vs Revenue (in millions)

4 RESULTS

The first research questions examines the explores the unconfirmed correlation between overall critics vote average and the revenue of a movie. With critics reviews as a misconceived measure of the success of a movie against the financial success of a movie being revenue, it does not take a lot to presume that these would correlate and maybe even be causal. Our findings plotting a small multiple scatter plot showing how vote average and revenue relate across different genres. The graphics in Figure 3 do not have a similar shape nor a correlation. Several of the charts do though have a flat bottom that would suggest movies with varying vote averages can
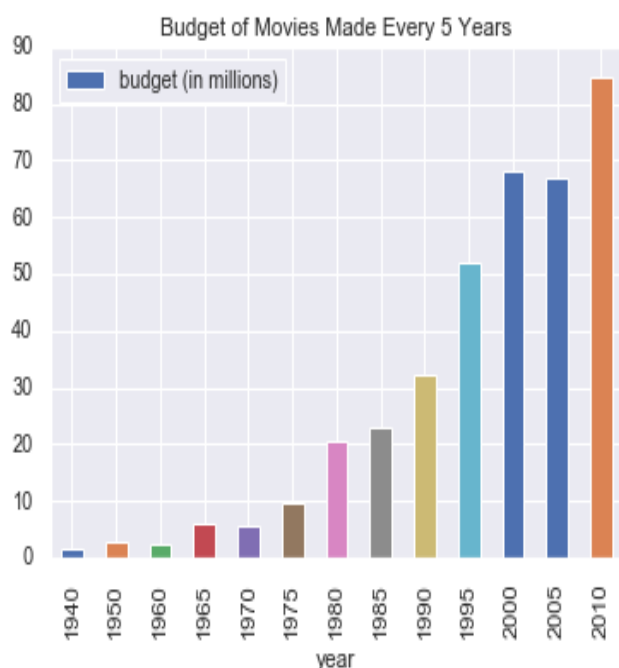
Figure 4: Budget (in millions) Per Movie Every 5 Years

perform the same financially. This means that revenue is in no way dependent on critics reviews.

Another question proposed is when and why did movies become what they are today. Figure 1 shows that the amount of movies made now is compared to 1940s is astronomically larger, but why is that and when did it shift. Looking back at figure 1 starting around the 1990s the amount of movies being made each drastically accelerates. If you are able to focus more in on the diagram between 1940s-1980s there is another increase in movies being made, not as much as later on but it is still worth observing

My last question was why do we spend so much on movies. Specifically production costs. Figure 4 shows the average million dollars spend on a movie over the span of five years since 1940s. Compared to Figure 1, the increase in movies being made every five years, the budget begins to rise a whole lot earlier on. During the years between 2005-2009 the budget of product seems to plateau and settle but not until 2010 when it rises again.

5 DISCUSSION

Does film criticism affect box office earnings finds similar results to Figure 3 and talks about how critics reviews can play a stronger role in initial earnings during the opening weekend. Although this data set cannot support his claim I would believe that this could be the case and the correlation breaks once a substantial amount of audience members response to the movie is different from the reviews, since the general public most likely has a stronger impact on box office earnings. Could there potential by a implanted bias set in our minds from the critics reviews when going to the movie that leads us to observe the faults higher than movies best moments. This selection bias formed through the growing number of critics

empowered through the advancements in social media can span further than movies. These biases can also cross over to the medical fields. A patients preconceived ideas of a medication developed from shallow research on the internet and others could cause the body to reject the effects. If were able to find the underlying psychological reason behind how a movie going experience can be affected the same way a patients treatment we could better understand the human mind and provide doctors with the tools to help patients better accept their treatment.

Is love of nature in the US becoming love of electronic media does not exactly provide an explanation, but bring up interesting factors that may have ultimately had an impact on the movie industry. The paper brings up the idea of these two activities are competing with our limited 24 hours in a day, 7 days a week. Could the be more today trying to compete with our time. Do we have too many things we want to get done in a day, month, year that we may push as side important things like education away when we find something we'd rather do. The paper continues to mention that "The average person went from spending 0 h/year on the internet in 1987 to spending 174 h/year on the internet in 2003"[5]. That's just till 2003, fifteen years later and it could be quadruple. Society has already started to explore this but with the internet having a large lead in the competition on our time, maybe we need to move education and other things we are being distracted from to the internet.

This competition on our time can also play into the last question. Why do we spend so much on movies. There is a lot of competition out there not just on what to do with our time, but inside the activities themselves. In our case movies, which is a strong contender in the competition of the average person. The movie industry as showing in figure 1, has plenty of movies coming out each year, for a studio to get there movie to stand out they have to spend a lot producing their movie, and that can be expensive. There are some studios out there that have the money to compete like Disney, and Warner Brothers. Smaller studios struggle to even play the game. It leads to the consolidation of studios, for example how Disney now owns Fox, and Marvel two very successful studios. This ruthless competition is not only in the industry. Just as other industry have to compete for our time and money, something we earn through our time at work, companies in those other industries need to compete as well. Industries just like the movie industry consolidate just as well. Maybe we need to give a little credit to the independent, indie companies to keep the competition alive before the same company owns every studio, restaurant, or electronic store.

The two interesting parts of the Figure 1 is in 1940s and 1990s when the amount of movies being made start their increase. During the 1990s and through today the advancements in technology have been exponential. Just glancing the Wikipedia[6] page for 1990s in science and technology the amount of advancements in Video/Audio alone can contribute to the number of movies a studio can roll out a year and also contribute to the increase in quality. Just because the technology is there doesn't mean it is cheap, that is probably why the budget of the average movie has also grown significantly

[5]R.W.Pergams O, A.Zaradic P. 2006. Is love of nature in the US becoming love ofelectronic media? 16-year downtrend in national park visits explained by watchingmovies, playing video games, internet use, and oil prices. Journal of EnvironmentalManagement. https://doi.org/10.1016/j.jenvman.2006.02.001
[6]https://en.wikipedia.org/wiki/1990s$_ins_science_andt_echnology$

through the 1990s. The other time period around the 1940s may be a factor of the post World War II era when the American family started to settle down into suburbs and spend more time in leisurely activities including going to the movies.

6 CONCLUSION

While everyone may not be the most enthusiastic about movies as some people there are still some parallels to other phenomena we see anywhere else. Movies no matter what one may say is apart of American Culture. Movies have been with us along the way over for more than the last century. The provide just as much a documentary to our progression as do books and can show us societal constructs in a way that books cannot. When you going to the movies today there will be something similar to how we as a society act today and as time goes on we may change but the movie will always be able to show us how we were, how we thought back then. Along the way as our computers, phones, and transportation has advanced so has the way we make and watch movies. With how many ways you can watch a movie they better do it right. We may have a better understanding of why and when movies started to increase in popularity and quantity, but there is no telling when it will stop growing. I bet the popularity of YouTube and the power of and individual through internet will not slow it down.