

How To Break The Box Office

Making a movie could be a daunting task. It requires one to employ the right decision in different sectors like what genre to focus on, which actress to cast, which producer to hire, and a bunch of different other elements including the ability to listen to your gut feeling. No one knows the future but what we do know is history. We have a history of decades of movie data collected by IMDB, TMDB, Rotten Tomatoes, Box Office Mojo. Humans never really changed, sure we do change in our outward appearance over a long course of time but our key characteristics and behavior are hardwired in our DNAs and this will take a large amount of self-awareness to decode. and let's be realistic no one starts thinking introspectively when deciding on a movie to watch. It's more of an impulse decision, like when I decided to watch the movie, *Dangerous Lies*, at 11:00 PM when I was definitely self-aware that I need to be up at 6:00 AM. What I'm trying to say here is if there was one place where data science could be applicable it will be here, in the movie industry. We have large amounts of data from viewers behaviors which we can use to predict their future behaviors.

Take Netflix for instance, I never went to the Netflix search box and typed in *Dangerous Lies*, it just appeared on my homepage as a recommendation so I just impulsively clicked on it. Regardless, it came of no surprise to me that I actually enjoyed watching it a lot. Netflix has thousands if not millions of movies and TV shows and in order to help customers sieve through their ever-changing catalog, they recommend movies and TV shows to watch based on a viewer watching habit. Netflix understands they can't offer up their entire catalog customers so they make use of algorithms which are computer-controlled database sets of rules. Netflix even said publicly that the Netflix experience is controlled by a number of machine learning algorithms such as personalized ranking, search similarity, watch history, ratings, and more. They understand they must curate what to present customers based on what the algorithm feels the viewer would like if not one could spend minutes if not hours searching through the database for the perfect movie.

Netflix even goes as far as tailoring the cover art off movies to fit their viewers taste. According to internal studies by Netflix, they found out that viewers spend an average of 1.8 seconds considering a title before moving to the next and they believe they only have 90 seconds to capture attention before you moved to another activity. They use Aesthetic Visual Analysis, AVA, to capture the perfect personalized thumbnail for each customer. AVA uses tools and algorithms that search Netflix videos for the best images to capture in order to make thumbnails out of them. This way they present to you, after multiple tests, the thumbnail they feel will grab your attention the most.

Because there are so many human decisions, Netflix understands its extremely useful to tag movie titles with very rich information so that customers can gain more insight into a movie, without reading much of the description. Netflix uses a very intricate tagging system to form these tags, there's even a job at Netflix called a tagger, these people watch every single movie and TV show on Netflix and tag them consistently so that Netflix can have consistent ways to do things anywhere where there's the human to machine interface such as members trying to look at their homepage and be able to see organized titles with labels in the rows. The tags also help Netflix predict the potential popularity of a title that hasn't launched yet on the platform. They use this to plot a graph showing the predicted popularity of a title against time. They split the time axis into four segments: pitch phase, development & production phase, pre-Launch phase, and launch phase. They use this graph to make decisions at the

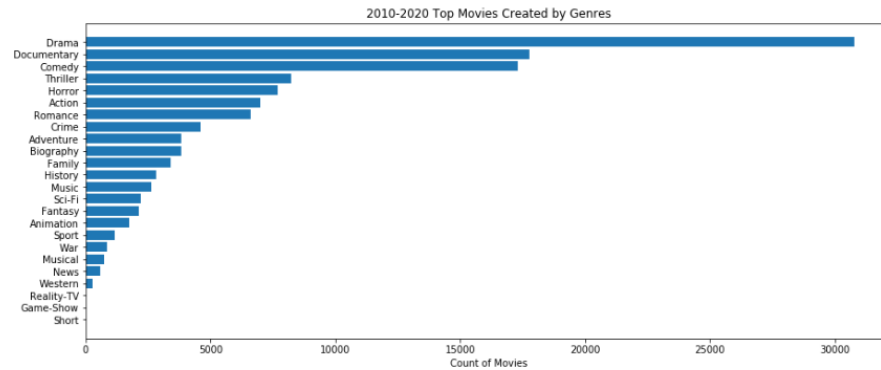
end of each of the four segments of time about if a title has the green light or if they'll be pulling the plug on it.

Even Legendary Entertainment, known for King Kong, Dark Knight, etc, uses their own type of AI to predict what type of impression to show to prospective customers in order to lure them into either watching a trailer or buying a movie ticket. If you notice here, the successful companies in the movie industry maintain this success with an eagerness to understand people, and how best can you understand people? Data. It is no more show me your friend and I'll tell you who you are. Rather, show me your data and I'll tell you who you are, and even maybe, who you will be. They use the data of people they have for analysis in order to create an audience and make individual predictions from it.

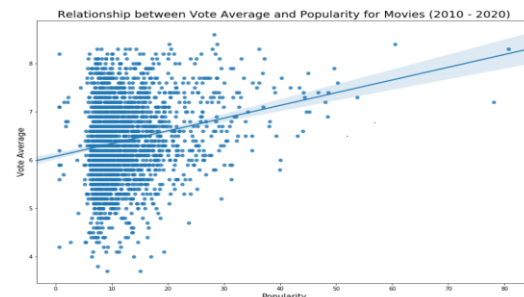
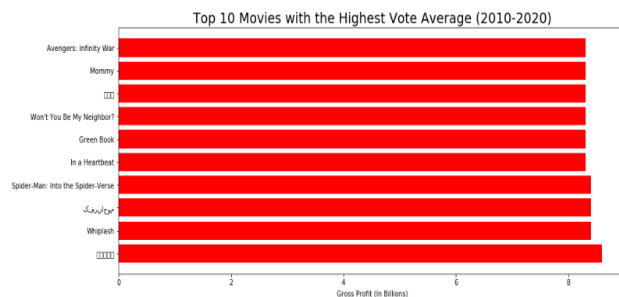
They use this insight to hone in on who they think their ideal audience would be. Using this they create three categories of audiences, the givens, who are avid long-time fans of the movie and will watch it no matter, for example, me for James Bond, the Nevers, who are people that have never watched it and have no intention on watching it, for example, me for Keeping Up With The Kardashians, and in the middle, they have the Persuadables, who are people that are convincible with the right impression at the right moment. Most of their targeting ads and impressions will be geared towards the Persuadables, they have no interest pursuing the Nevers and there's no use of selling something that is already sold so there's no need to focus on the givens. This is why most of their marketing in time and ads are spent trying to convince the Persuadables. They take this Persuadables and scored them individually from 0, meaning unlikely, to 100, meaning very likely. They then take a small group and carry out various tests using various marketing campaign strategies targeted to this small group. They then use the successful marketing campaigns and apply them to a broader group of the Persuadables, then as the release date approaches, they downsize this group to make for a more targeted audience.

In order to get a better understanding of the movie industry and what type of films succeed I and a couple of colleagues took on a project to analyze data extracted from websites with a huge amount of movie data such as IMDB, TMDB, The Box Office Mojo, etc. The goal of this was to answer several questions about changes in humans watching behavior and questions related to the success of a movie holistically.

The first question we answered was; what is the distribution for movies in each genre. we wanted to find the top three individual genres that film studios create the most. After a bunch of data cleaning, and merging we then plotted a horizontal histogram to analyze their mounted movies created for each genre. By doing this we concluded that most studios prefer to create drama films, followed by documentaries.

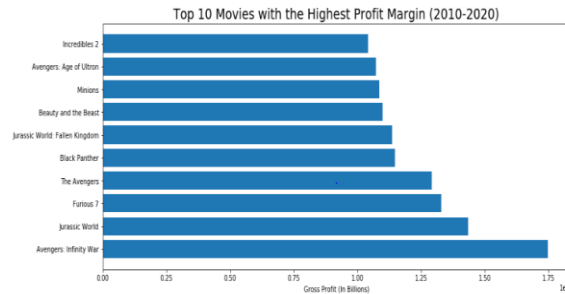


We then went on to investigate the 10 most popular and highest rated films and also if there was a relationship between the popularity of a movie and the movie's rating. Before the same initial steps for the first question but for this we filtered the data for movies made between the last decade that have an average vote count higher than 190. We only used movies made in the last decade because we found out that movies made in the previous decade had really low box office numbers and this might have dragged the average box office numbers down. I believe this is mainly because of The Internet that lets us have a lot more advertising in this decade in comparison to previous decades. After finding the voting average we plotted a horizontal histogram to analyze our results.

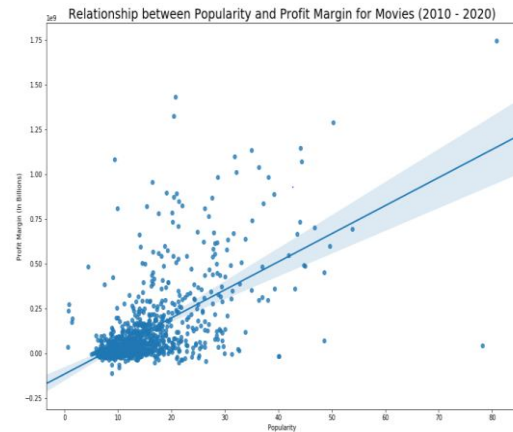
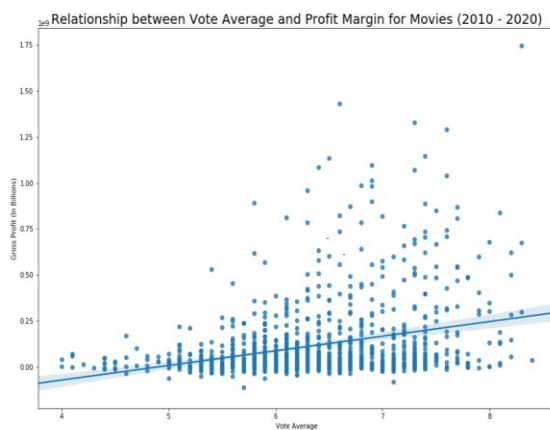


From the histogram we can see the top 10 movies in the last decade with the highest voting average and in the second graph we can see the scatterplot showing the relationship between the voting average and popularity of a movie. From this we can conclude that action and adventure movies are the most popular movies by far and foreign and drama films had the highest ratings, but this could also mean that not a lot of people knew about them so not a lot of people would not like them, hope you understand the logic. When we looked at the correlation between the voting average and popularity, we found the correlation is 0.22 which means the correlation is negligible. So, in other words, a popular movie doesn't exactly mean the movie will have a high rating, but it's plausible.

Next, we found out the top 10 movies with the highest gross profit margin and also the relationship between the popularity and the profit margin. What did this by plotting yet another horizontal histogram, getting sick of them yet? Sorry.

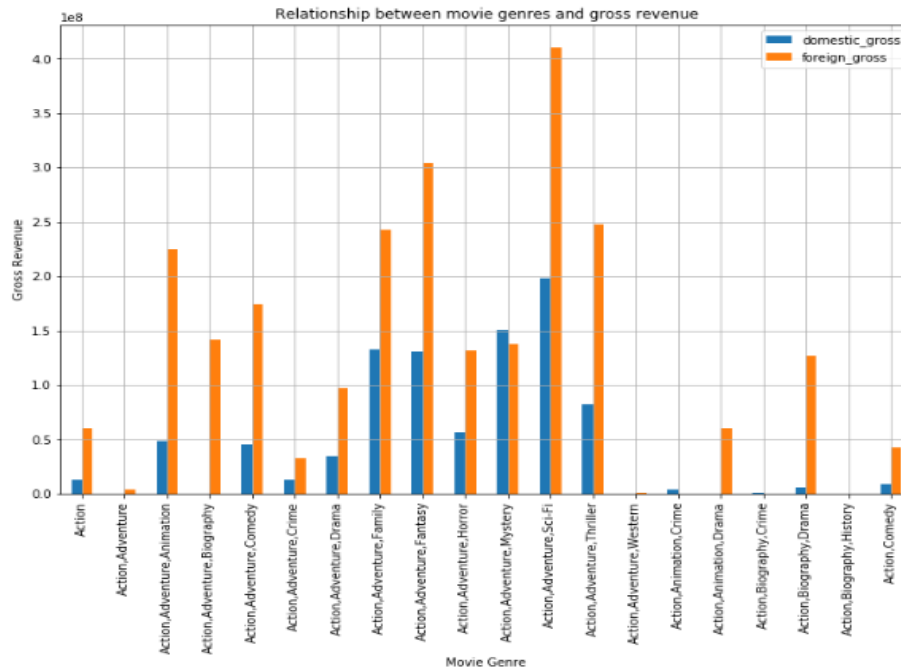


From this, we can see the top-earning movie within the last decade is The Avengers Infinity war, even though it's only been out for about a year from me writing this. We also plotted scatterplots to help us understand the relationship between the voting average and the profit margin of movies within the last decade and also the relationship between the popularity and the profit margin of a movie. From analyzing the scatter plots we could conclude that both have a positive relationship, which means the higher the voting average, the higher the profit margin, and the higher the popularity, the higher the profit margin as well so movies should definitely focus on marketing.



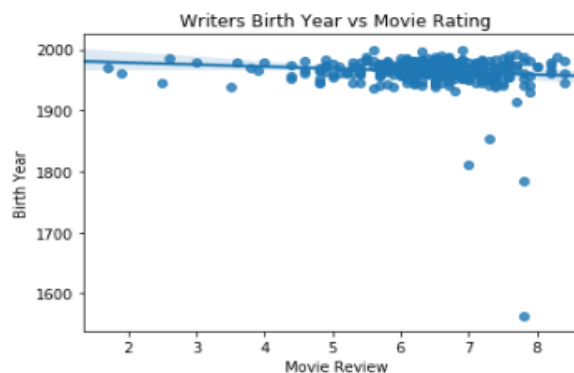
From this analysis we can see people's tastes have changed a little bit over the decades. the 40s and 50s people who or more inclined to watch a movie that incorporated heroism and freedom in the storyline. these days people just want a bunch of drama and gun-slinging, laser beaming action, and there's nothing wrong with that, I love those too. Although we could look into a relationship between the amount of action movies with huge amounts of violence and the average amount of homicides over the last two or three decades, but that's for another project for another story.

Furthermore, we dug a little deeper to find the most lucrative genre combination since movies, these days especially, usually are compromised off multiple genres. I know I like my action with thrillers the same way I like my tea with biscuit.



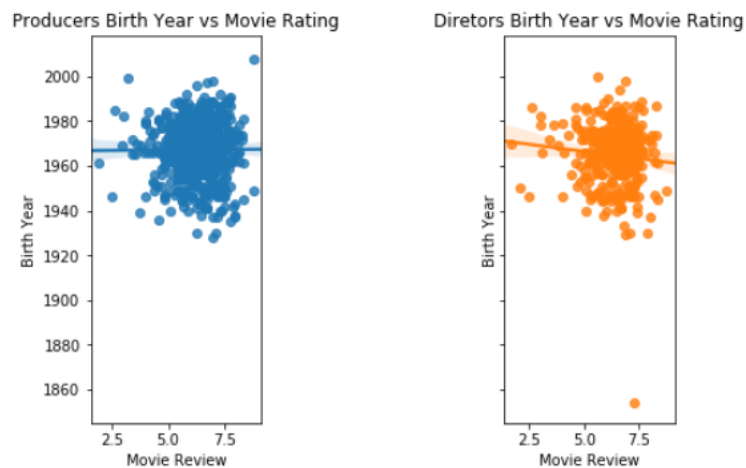
We were able to plot a histogram to show the relationship. From this we see action, adventure, and fantasy is the genre with the third-highest average profit margin, action, adventure, and thriller is the genre with the second-highest average profit margin, wink, wink. As we can see the action, adventure and sci-fi genre has the highest average profit margin among all the genre combinations, so I believe the smart decision for movie studios is to try and include these genres in the movies they make, in as much as I personally don't care for sci-fi movies at all.

We also try to understand the relationship between the age of or writer, producer, and director of a movie and the publics affinity towards that movie. I decided not to bore you with all the numbers, a rather torture you with the graphs instead. As we can see below, a writer's age has a negative relationship with the movie rating.

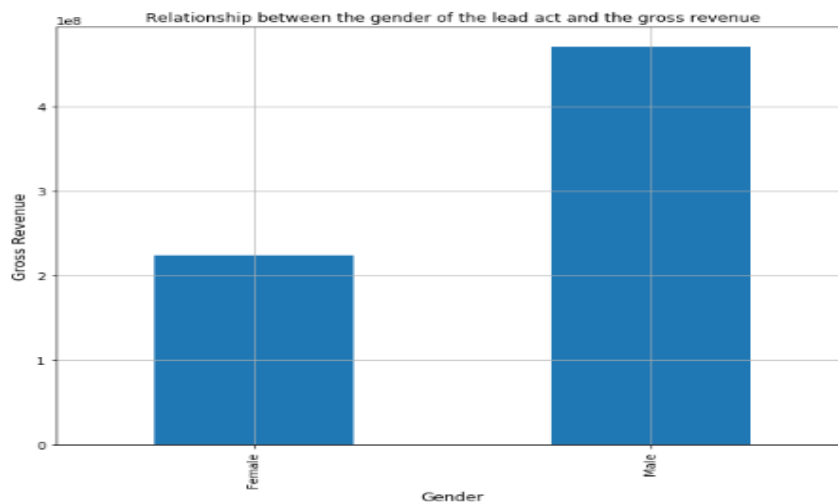


This means the older the writer, the better the storyline, which leads to higher ratings. We can also see a producer's age has a positive relationship with the movie rating. In other words, the younger the producer the higher the movie ratings. Finally, we analyzed their relationship that's really directors

age and movie ratings. From this we see there's a negative relationship which means proves that older directors are better at directing movies than younger directors.

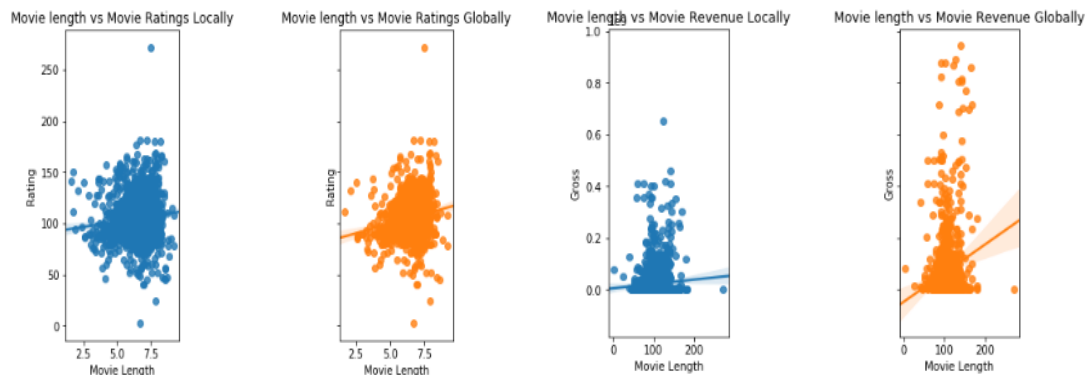


Another relationship we sought to you investigates was the relationship between the gender of the protagonist in an action movie and the profit margin of that movie. To do this we singled out the lead acts for action movies alone and grouped them by their gender. Furthermore, we plotted a histogram of the average gross revenue against the gender and it showed action movies starring males as the protagonist makes approximately double in gross revenue as movies starting females as the protagonist. This analysis was done using about six decades of movies and I know there weren't a lot of action movies starring female protagonists back in the 1900s so I'll be doing deeper analysis on this to see the relationship within the last decade or the last two decades.

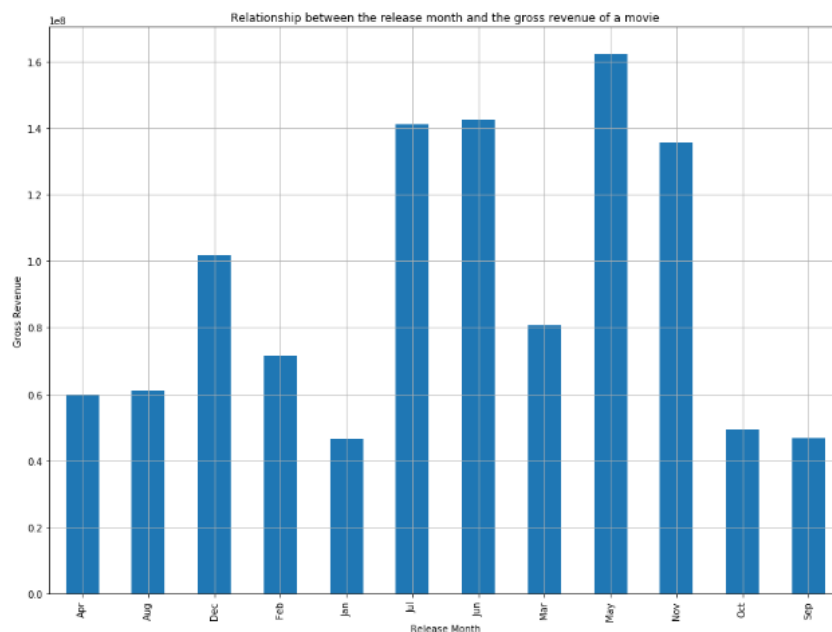


Next, we analyzed the data to find the relationship between the length of the movie and the gross revenue and rating of their movie. I know there's tons of research on this already but I wanted to conduct my own personal analysis so I can better understand the industry. we investigated using both data from just the United States, as well as the global data. We plotted some scattered graphs as you can see below. The first couple are using movie ratings against the movie length and the last two use the gross revenue and the movie length. As we can see, all have a positive relationship which means

regardless of just in America or globally, the longer the movie, the higher the profit margin and rating of that movie.



We could do it hundreds of analyses on the data we have but for now I decided to and understanding the relationship between the month a movie was released and how well that movie does. I grouped the movies by their release months and found to average profit margin for each month, then I used this to plot a histogram. From the histogram below, we can see that May is the best month to release a movie, and this confirms other studies I saw on the internet.



To summarize the results of what I analyzed, I'll recommend movie studios to focus on making movies with a combination of the action, adventure and sci-fi genres since these have proved to be the most profitable general combinations. I dug a little deeper and I created a fantasy team based on the highest-grossing movies globally of what an ideal cast would be for each of the highest-grossing individual genres. Thank me later when your movie breaks the box office.

Genre	Producers	Directors	Writers
Action	Chris Hemsworth	Stanley Tucci	John Logan
Adventure	Jennifer Lee	Ian McShane	John Logan
Sci-Fi	Don Murphy	Irrfan Khan	Steven Spielberg

I'll also recommend hiring a combination of a younger producer, with an older more experienced writer and director for the best creative experience. Also recommend making action movies with a male protagonist if you want to be on the safe side, but given that these this niche is saturated, it might be profitable to take a risk and star a female instead, the movie might do well since the competition is very low in this niche. Other recommendations out gave are make you longer movies because apparently people I don't really enjoy short movies and releasing movies during the summer months, especially May.

Bibliography

Netflix Technology Blog. (2018, February 7). AVA: The Art and Science of Image Discovery at Netflix. Retrieved from <https://netflixtechblog.com/ava-the-art-and-science-of-image-discovery-at-netflix-a442f163af6?gi=f0453d40208c>

How Netflix's Recommendations System Works. (n.d.). Retrieved from <https://help.netflix.com/en/node/100639>

How Netflix's Recommendations System Works. (n.d.). Retrieved from <https://help.netflix.com/en/node/100639>