Fresh Over Rotten: The Value of a Critic's Pick in Hollywood

Sergio E. Betancourt
Master of Science in Statistics, University of Toronto
STA2453 - Data Science Consultation

2018-11-13

Abstract

An analysis of the movie business based on New York Times and OMDB data, in the form of sentiment analysis of critics' reviews, factors that drive box office revenue, and factors that influence award nominations (and possibly subsequent wins).

Motivation

A report by the firm PwC (available in https://www.statista.com) estimated global film revenue to be approximately USD **$38.3 billion** in the year 2016, with forecasted growth for the foreseeable future. In the US alone, film revenue in the year 2016 was estimated to be USD $11.38 billion, about 30% of the World's revenue. With such a sizable market it is important to understand the factors influencing film earnings, as well as film prestige in the shape of critics' reviews and reputable awards.

In this report, I consider movie data from the NYT Movie Reviews (https://developer.nytimes.com/movie_reviews_v2.json) and OMDB (http://www.omdbapi.com/) APIs the for the years 2010 to 2017 inclusive. Although rich, the data procured is not entirely complete, which poses a major challenge when carrying out the intended rigurous analysis. Nonetheless, below I discuss the extent to which I use, modify, or throw out any of these exceptions.

I shall focus my analysis around the below questions:

1. What is the marginal financial contribution of a New York Times (NYT) critic's pick to a movie's revenue?
2. Is there a relationship between critics' picks and earning a nomination to a prestigious award?
3. What are the factors that contribute to earning a nomination and what is the role of review sentiment?
   - Does gender play a role in a movie's likelihood of earning a nomination to an award?

To address the above questions effectively it is necessary to clarify that a critic may not pick a movie with the sole intention of determining whether it shall be a blockbuster or not–critics will often look for originality, timeliness, and quality, and there are many blockbusters out there that they do not deem worthy of their pick (e.g., Marvel's The Avengers series). Moreover, although not discussed in detail here, the timing of a critics' review may or may not play a part in box office performance and/or nomination for a prestigious award. Some critics may get to review the movie before it is released to the public, whereas others may not have the chance to review if no pre-screening was done.


Analysis at a Glance

The below boxplots (excluding movies with NA values in box office earnings) outline summary statistics for movie earnings, comparing (1) those picked by critics versus those that were not, (2) earnings across genre for the entire time period. For the sake of cleanliness in the following visualizations, I limited maximum earnings in all of the below boxplots to USD $300 million. I also excluded genres with less than 5 entries in plot (2). I eliminated these filters and considered the complete dataset in my numerical analysis, except where indicated otherwise.

Note the large dispersion of boxoffice revenue in the below plots, with a sizable number of movies with revenue above the plots' whiskers. To make a fair comparison, I sampled randomly a number of non-picked equal to the number of picked movies I pulled from the APIs. Some movies tend to do extremely well compared to others in both categories, and this data might give us a rough idea of the why. Moreover, there appears to be a larger number of relative outperformers in the critics' pick category, even though median revenue for the non-picks appears slightly higher.

Genre-wise, the variability of revenue accross genres is evident, with some exhibiting a higher median but also higher dispersion (Action and Animation), while others show neither (Documentary and Crime). This might be due difference in sizes of each genre level.

Boxplot of Revenue by Pick vs Non in 2010 – 2017
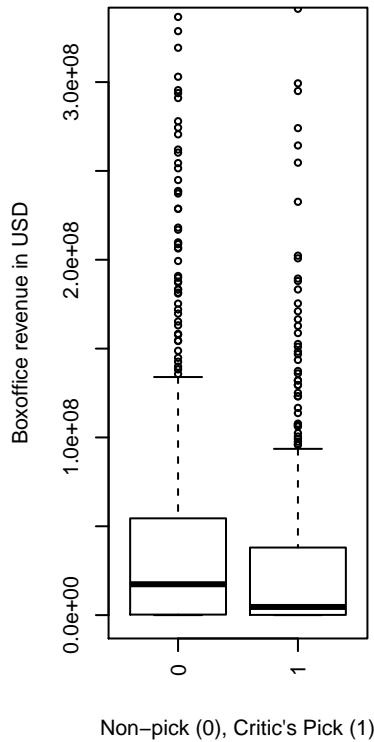
Boxplot of Movie Revenue by Genre in 2010 – 2017
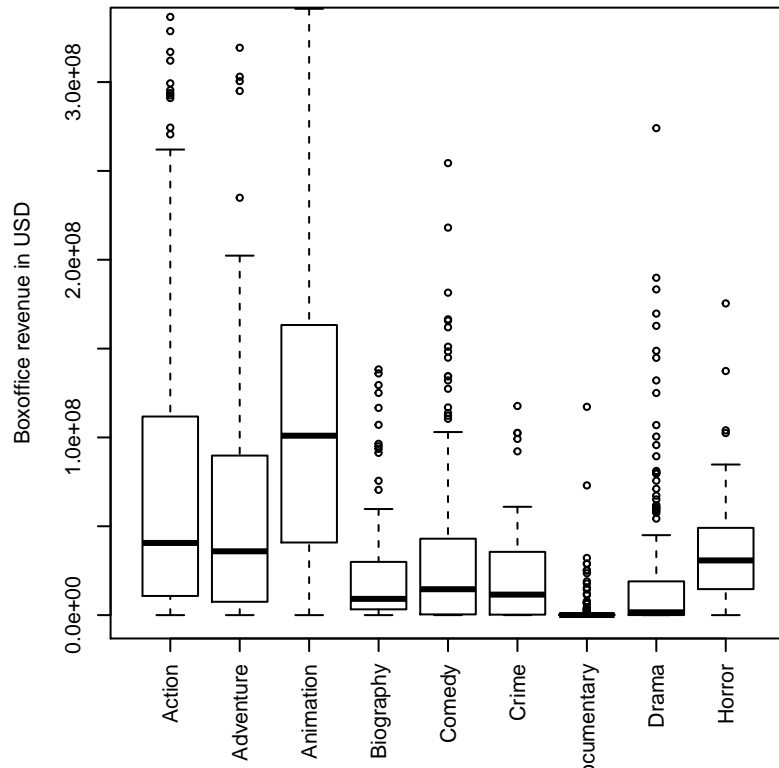
The selective nature of the critics' picks means the overall number of movies deemed critics' picks is substantially smaller than that of movies undeserving of this title. Of the **3245** movies collected for this analysis, only **659** were picked by a New York Times's critic.

Below is a more detailed breakdown of picks by genre:

	Not Picked	Picked	Total	Proportion Picked (%)
Action	248	31	279	11
Adventure	47	26	73	36
Animation	51	25	76	33
Biography	47	34	81	42
Comedy	233	57	290	20
Crime	41	12	53	23
Documentary	106	65	171	38
Drama	180	97	277	35
Horror	31	11	42	26

Of importance are the extreme values in the above table for proportion of movies picked by NYT critics, grouped by genre. Biography, Documentary, Drama, Adventure, and Animation show the highest proportions, whereas Action and Comedy show the lowest. This is a further line of investigation that will not be addressed in this project directly but in a future effort. However, we shall incorporate genre as a factor in this analysis.

Analysis

1. What is the marginal financial contribution of a New York Times (NYT) critic's pick to a movie's revenue?

To quantify the added value of a critic's pick to a given movie, I can carry out a standard regression with revenue as dependent variable, and other information as explanatory variables. Adjustments were made on the basis of data features and the the usual diagnostics.

$$\log(\text{boxoffice}_i) = \alpha + \beta_1 \text{criticpick}_i + \beta_2 \text{genre}_i + \beta_3 \text{nomination}_i + \beta_4 \text{rating}_i + \epsilon$$
$$\text{where } \epsilon \sim N(0, \sigma^2) \text{ and } \alpha \text{ is the grand mean}$$

An ANOVA test performed on the regression model of all movies with the logarithm of boxoffice revenue as the dependent variable, shows the variables **genre**, **nomination**, and **MPAA rating** to be significant at the 5% level.

This regression model indicates statistically significant differences of the following indicators from the grand average of boxoffice revenue across the entire time period:

- **Genre**: Action, Documentary, and Drama (Multiplicative factors of approximately **3.10**, **0.17**, and **0.40**, respectively)
- **Nomination** (Multiplicative factor of approximately **7.15**)
- **Rating**: Not Rated and Rated R (Multiplicative factors of **0.011** and **0.206**, respectively)

Therefore, this model suggests there exists a significant positive contribution of a nomination to box office revenue. With regards to genre, everything else constant: The average action movie seems to do exceedingly well, especially when compared to documentaries and dramas, while the average R-rated movie performs poorly. This might be due to the age restriction, which curtails young individuals from watching some of these restricted films.

As per the diagnostics of this model, the QQ-plot of residuals does not offer strong evidence for their normality, which I sought to ameliorate by taking the logarithm of box office. It is possible for a generalized linear model to fit the data better. Other aspects of concern are the number of entries with missing data, in some cases due to spelling differences between the data APIs, while in other cases due to lack of revenue amount (international/restricted films).

2. Is there a relationship between critics' picks and earning a nomination to a prestigious award?

Now I consider the role of a critic's pick in a movie's award nomination (restricted to BAFTA, Oscars, and Golden Globes).

	no_nomi	nomi	rowtotal	prop_nomi
non-pick	2412	174	2586	0.067
pick	415	151	566	0.267

I obtained a 95% confidence interval for the relative risk: **(0.229, 0.275)**. This interval and its middlepoint estimate sugggest that, according to the data, failing to be picked by a New York Times critic results in an approximate reduction by **75%** of the chance of earning a nomination to an award, when compared to a movie that earned a pick. Therefore, there is a positive relationship between being picked by a critic an earning a nomination and possibly winning an award. However, this may not be the only factor at play.

3. What are the factors that contribute to earning a nomination and what is the role of review sentiment?

To quantify the relationship between critic's picks and other variables to the probability of earning a nomination to a prestigious award (Oscar, BAFTA, or Golden Globe), I can perform a logistic regression. I codified the genders of lead actor and director for all movies for which the data was available.

A logistic regression model with an indicator variable for nomination as the dependent variable ("1" if nominated, 0 otherwise), was fit against genre, gender of lead director, and gender of lead actor. I was able to encode a the lead actor and director gender variables of a large number of movies across both critics' and non-critics' picks by harnessing the

Gender API (https://gender-api.com/en/). This API takes a first name in lower case and returns its classification of "male" versus "female." Unfortunately, this API has a limit of 500 free calls for any registered user, so I had to do a great portion of the encoding by hand.

$$\text{logit}(\pi_i(1)) = \alpha + \beta_1 \text{genre}_i + \beta_2 \text{fem-dir}_i + \beta_3 \text{fem-lead}_i$$

This logistic model suggests that, for a given genre, having a female lead actor almost **doubles** the odds **(1.809)** of earning a nomination to any of the BAFTA, Oscar, or Golden Globe awards, when compared to a movie with a male actor. However, the genre of the director was not significant in determining the probability of earning a nomination. This can mean good news for female participation in lead roles as production companies may seek to increase their prestige by creating stories with female casts in mind.

The results could be further improved with more extensive codification and more data without missing information. Another interesting covariate that could be considered in this model is the IMDB score that is present in the OMDB API. These scores are determined by groups of critics, which means that the higher the score, the more positive the average critique of the movie.

As it pertains to sentiment, I harnessed Python's language processing capabilities to analyze the reviews' summaries available via the NYT API. I then proceeded to use the built-in, Naive Bayes Analyzer tool (already trained on movie reviews) to compare sentiments between movies that were critics' picks versus those not picked. This classifier is a supervised learning approach that works well with binary classification of text bodies. The classifier's outputs are the classification ("pos" or "neg" in this case), as well as probabilities of assignment to each of these two binary categories.

Note that the Naive Bayes Analyzer used in this analysis was pre-trained on Bo Pang and Lillian Lee's Movie Reviews Dataset, which was built from thousands of IMDB reviews. Therefore it performs extremely well when applied this type of review text.

	not_picked	picked	total	propnotpicked
neg	561	92	653	0.859
pos	2023	567	2590	0.781

A Wald 95% confidence interval for the risk ratio of a movie failing to be picked by a NYT critic conditional on review sentiment is **(1.06, 1.142)**. This indicates that a review with a negative sentiment classification is approximately **1.1** times more likely to correspond to a movie that was not picked by a critic versus one that carries a positive sentiment classification.

The above risk ratio is not very dramatic, which potentially underestimates the relationship between review sentiment and critic's pick. Nonetheless, this may be due to the fact that only a dozen or so of words corresponding to each review were available via the API. Perhaps with a larger portion of the review text the classification could be enhanced.

Further Considerations

The goal of this work was to implement some sound yet simple analysis to a topic of my interest. Films and television are very popular forms of entertainment that can educate and entertain us, and I was very much influenced by the hours I spent in front of a screen while growing up.

The scope of this project can be extended to a longer time period, even incorporating time into the modelling, as movie release dates may affect their performance. Moreover, it would be interesting to get hands on data on TV shows and digital-only productions (e.g., Netflix and Amazon exclusives, etc.) in order to look at some of the elements (qualitative and quantitative) that separate traditional from non-traditional media. Other explanatory variables of interest would be movie duration, weekly/periodic box office revenue, production cost, etc.

The sentiment analysis portion of this work can be greatly extended to incorporate full review text (through targeted scraping) and better analytics. More sophistication can lead to more interesting and controversial questions. Moreover, social media content could also allow the construction of a portrait of movies, and may be linked to box office revenue.

I do not make mention of a literature review in the construction of this report because there was none. My goal from the start of this project was to brandish curiosity as my only weapon. There are many interesting findings and methods applied to this exact dataset/situation, and the next steps in this project would be to review, confirm, and possibly extend some of these.

Appendix: Data Considerations

I collected the data via the NYT and OMDB APIs, limiting the scope of my search to the years 2010-2017.

NYT has the following important data available:

- Title
- MPAA Rating (missing for several entries)
- Critics' Pick (0 or 1)
- Short Summary
- Opening Date

OMDB has the following important data available:

- Title
- Box Office Earnings (missing for several entries)
- Genre
- Director
- Writer
- Actors
- Country
- Awards
- Various Critical Ratings

This is a random sample of ten movies from the final dataset I constructed for this analysis. This table should illustrate the data structure and information considered in this report.

	display_title	mpaa_rating	critics_pick	publication_date	opening_date	date_updated
448	Somewhere	R	1	2010-12-21	2011-01-21	2017-11-02 4:18
2748	The King's Speec	R	0	2010-11-25	2010-12-25	2017-11-02 4:18
27	A Summer's Tale	G	1	2014-06-19	2014-06-20	2017-11-02 4:18
972	Blue Valentine	R	0	2010-12-28	2011-01-28	2017-11-02 4:18
327	Megamind	PG	1	2010-11-04	2010-11-05	2017-11-02 4:18
246	Inherent Vice	R	1	2014-12-11	2015-01-09	2017-11-02 4:18
2244	Red	PG-13	0	2010-10-14	2010-10-15	2017-11-02 4:18
147	Eat That Questio	R	1	2016-06-23	2016-06-24	2017-11-02 4:18
284	Lady Bird	R	1	2017-10-31	2017-11-03	2017-12-01 17:44
1872	Man on a Ledge	PG-13	0	2012-01-26	2012-01-27	2017-11-02 4:18

	display_title	boxoffice	genre	director	actor	nomination
448	Somewhere	1768416	Comedy	1	0	0
2748	The King's Speec	138300000	Biography	0	0	1
27	A Summer's Tale	198126	Comedy	0	0	0
972	Blue Valentine	9600000	Drama	0	0	1
327	Megamind	148244405	Animation	0	0	0
246	Inherent Vice	6100613	Comedy	0	1	1
2244	Red	88900000	Action	0	0	1
147	Eat That Questio	354981	Documentary	0	0	0
284	Lady Bird	48285330	Comedy	1	1	1
1872	Man on a Ledge	18400000	Action	0	0	0

In order to build the core dataset, I queried the NYT API using a loop that covered 10-day periods from January 2010 to December 2017. Then I used the movie names to make requests to the OMDB API to get the above additional information. Specifically, I added box office, genre, director, writer, actors, and awards to the core NYT table. The OMDB API has a daily call limit of 1000 entries returned, which meant that the data enrichment process took a few days to complete. Moreover, several movie titles are spelled differently or contain different characters between these two data sources. However, I took the NYT title spelling as canon and obtained the extra data where there was an exact match. I treated missing data as NA in order to simplify the data procurement process, aware that with more careful consideration the data set could be further enriched.

To simplify the dataset I assume that the first name in the list of actors obtained from OMDB is the lead actor, and the first genre that appears in this same source is the main genre. I also converted the strings in the boxoffice field to numbers, disregarding the currency while simultaneously assuming that all numbers available are in USD. Moreover, I encoded the awards variable, indicating the value "1" for at least one nomination (including wins) to either a BAFTA, Oscar, or Golden Globe award. Movies with no nominations (or wins) to any of these prestigious awards were assigned a value of "0".