

TMDb Movie Profitability Analysis

Background

What can we say about the success of a movie before it is released? Are there certain companies (Pixar?) that have found a consistent formula? Given that major films costing over \$100 million to produce can still flop, this question is more important than ever to the industry. Film aficionados might have different interests. Can we predict which films will be highly rated, whether or not they are a commercial success?

This is a great place to start digging in to those questions, with data on the plot, cast, crew, budget, and revenues of several thousand films.

Click [here](#) to check out data source from Kaggle.

1 Summary Statistic

Below are statistics for quantitative variables and release dates.

##	release_date		budget		revenue		profitability	
##	Min.	:1916-09-04	Min.	: 0.00	Min.	: 0.00	Min.	:-165.71
##	1st Qu.:	1999-07-14	1st Qu.:	0.80	1st Qu.:	0.00	1st Qu.:	-0.80
##	Median	:2005-10-03	Median	: 15.00	Median	: 19.18	Median	: 2.53
##	Mean	:2002-12-27	Mean	: 29.05	Mean	: 82.28	Mean	: 53.23
##	3rd Qu.:	2011-02-16	3rd Qu.:	40.00	3rd Qu.:	92.92	3rd Qu.:	55.34
##	Max.	:2017-02-03	Max.	:380.00	Max.	:2787.97	Max.	:2550.97
##	runtime		popularity		vote_average		vote_count	
##	Min.	: 0.0	Min.	: 0.0004	Min.	: 0.000	Min.	: 0.0
##	1st Qu.:	94.0	1st Qu.:	4.6717	1st Qu.:	5.600	1st Qu.:	54.0
##	Median	:103.0	Median	: 12.9249	Median	: 6.200	Median	: 235.5
##	Mean	:106.9	Mean	: 21.4968	Mean	: 6.093	Mean	: 690.4
##	3rd Qu.:	117.8	3rd Qu.:	28.3320	3rd Qu.:	6.800	3rd Qu.:	737.0
##	Max.	:338.0	Max.	:875.5813	Max.	:10.000	Max.	:13752.0

The data sets include movies from Sept 1916 to Feb 2017. Highest budget is \$ 380 million, revenue \$ 2,788 million and the most profitable movie was \$ 2,551 million in profit.

2 Qualitative Analysis

2.1 Keyword / Idea / Story Line

For a starter, let's identify keywords for movies that recorded profit of \$ 50 million and above. 75% (3rd Quartile) of movies in the data set recorded profitability of \$ 55 million and above, hence I am doing deep dive into keywords of Top 25% of movies.

2.2 Genres

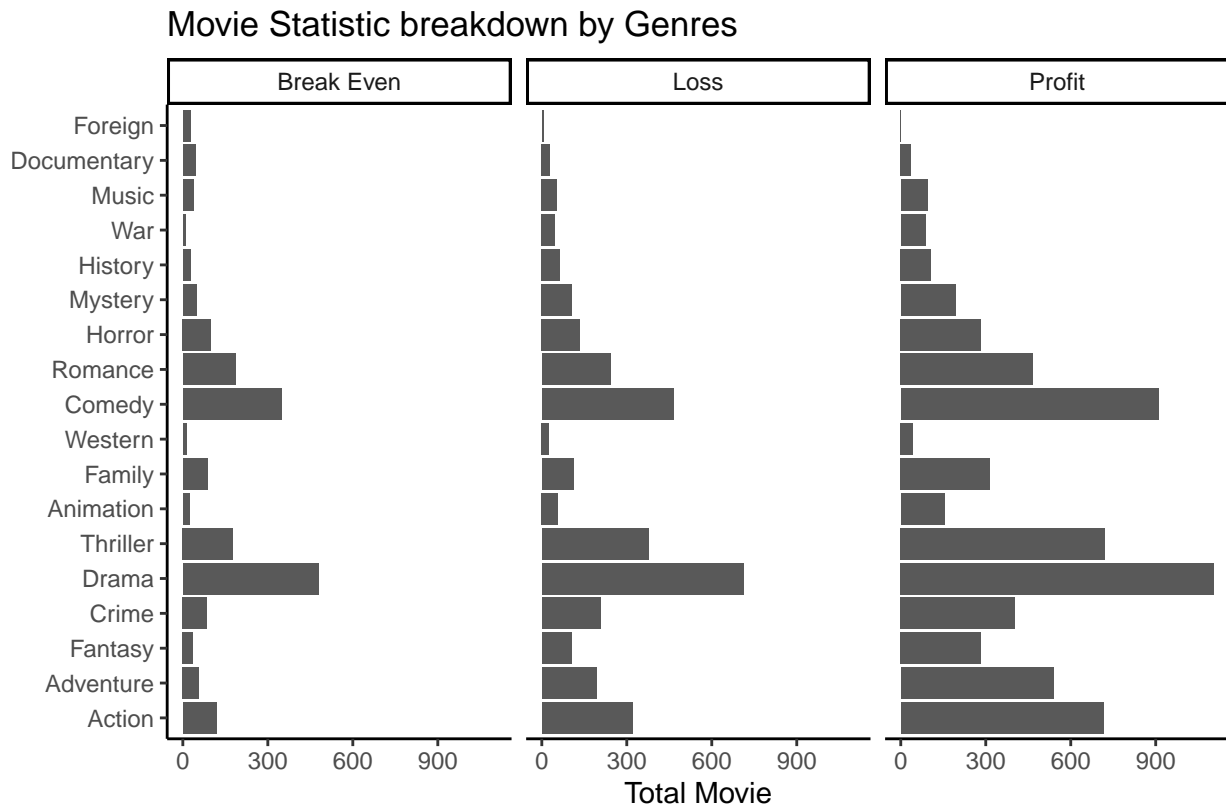
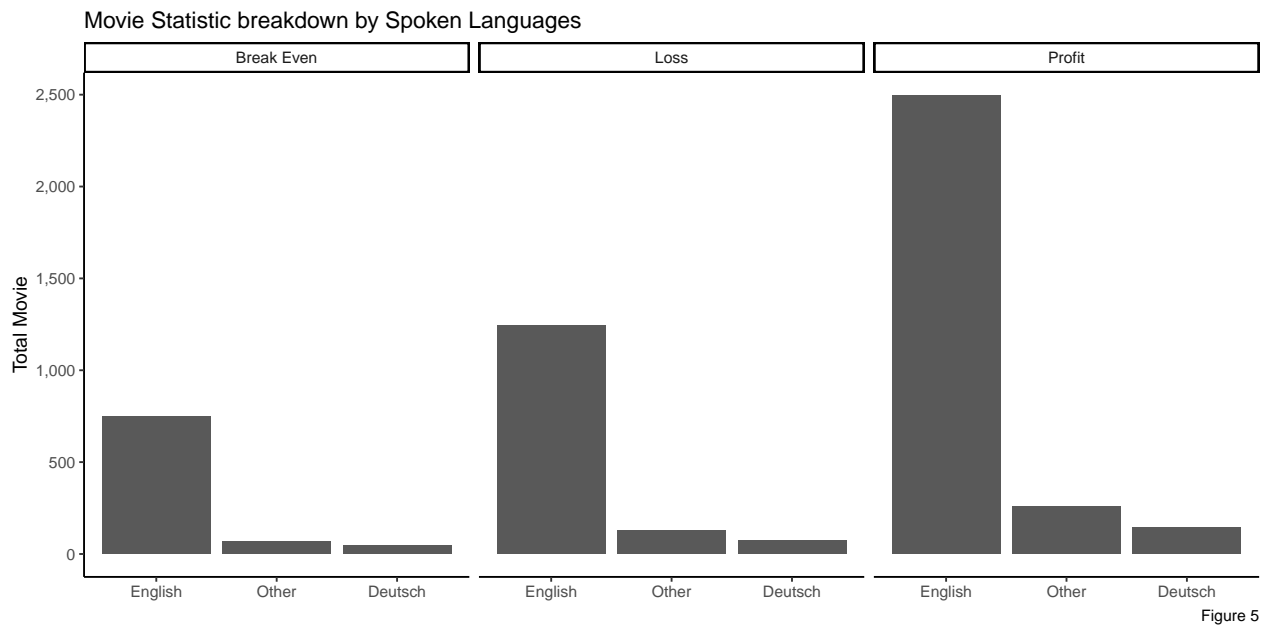


Figure 4

Based on Figure 4, it is clear that genres such as *Comedy*, *Romance*, *Horror*, *Crime*, *Thriller*, *Drama*, *Adventure* and *Action* top the list of profit-making movies. Movies that are loss-making / break-even also share the same genres.

2.3 Language



Most movies are English-oriented hence language doesn't matter for profitability either.

2.4 Production Companies

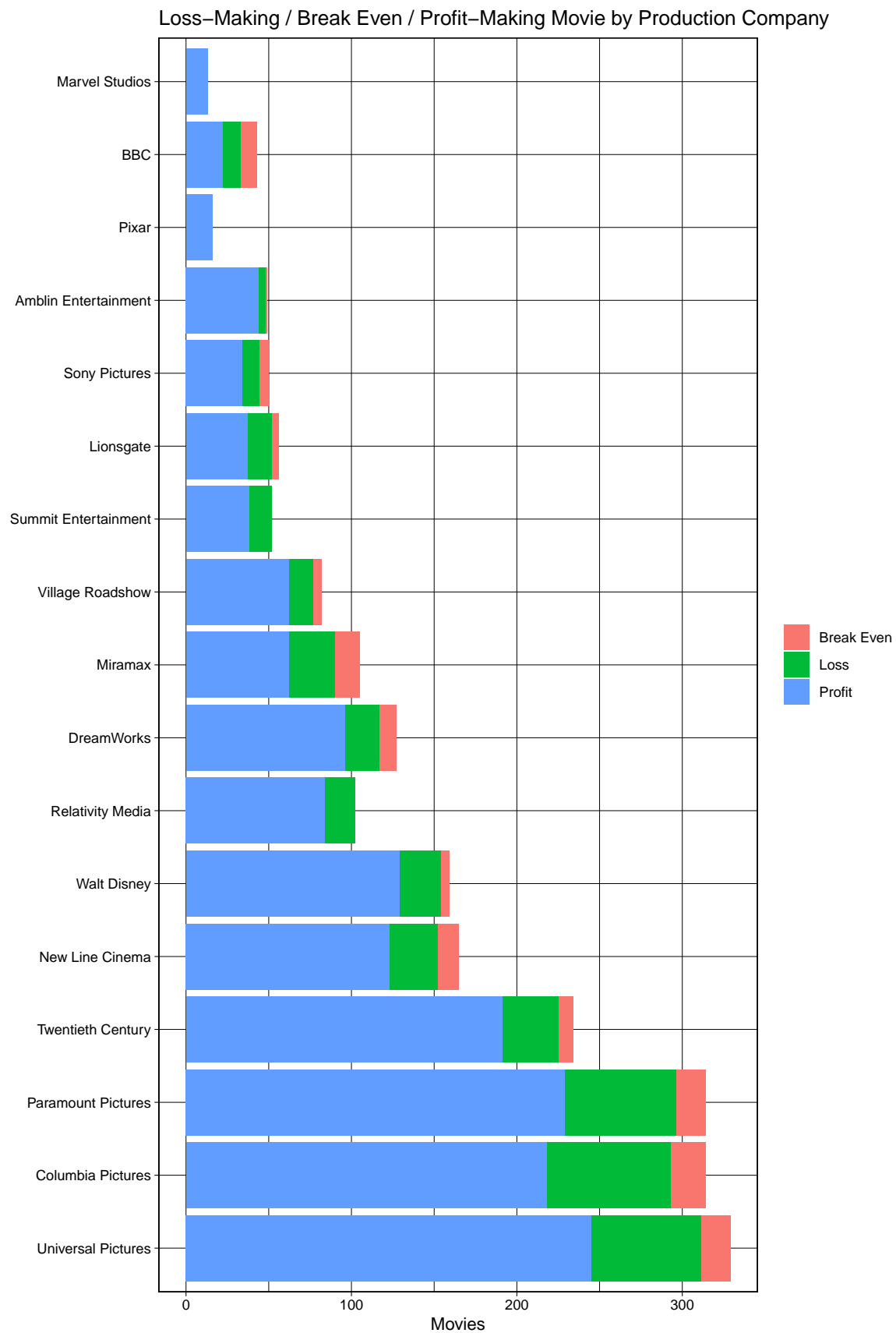


Figure 6

Figure 6 shows globally well-known production companies, for instance, *Universal Pictures*, *Warner Bros* and *Twentieth Century*. These are well established movie producers. Notice all of them produced more profitable movies than loss-making one. *Pixar* and *Marvel Studio* were pretty impressive as all movies produced are profitable, although they produced the least movies among all production companies.

Marvel Studios is well known for the Marvel Avenger series, Iron Man series and other super-hero movies. Pixar have been famous for animation such Wall-E, Toy Story series and Finding Nemo. It is fair to say that even the world's major producers can't guarantee profitability of a movie.

2.5 Casts

Given complexity of a movie, number of actors / actresses involved could be huge. With this in mind, my analysis will only focus on specific individual which has been part of some blockbuster movies, *Tom Cruise* and *Denzel Washington*. Let's find out some of their movies, loss-making or profitable, given their fame. Of course, I believe any actor / actress would have worked for some blockbuster, low quality and okay-okay movies.

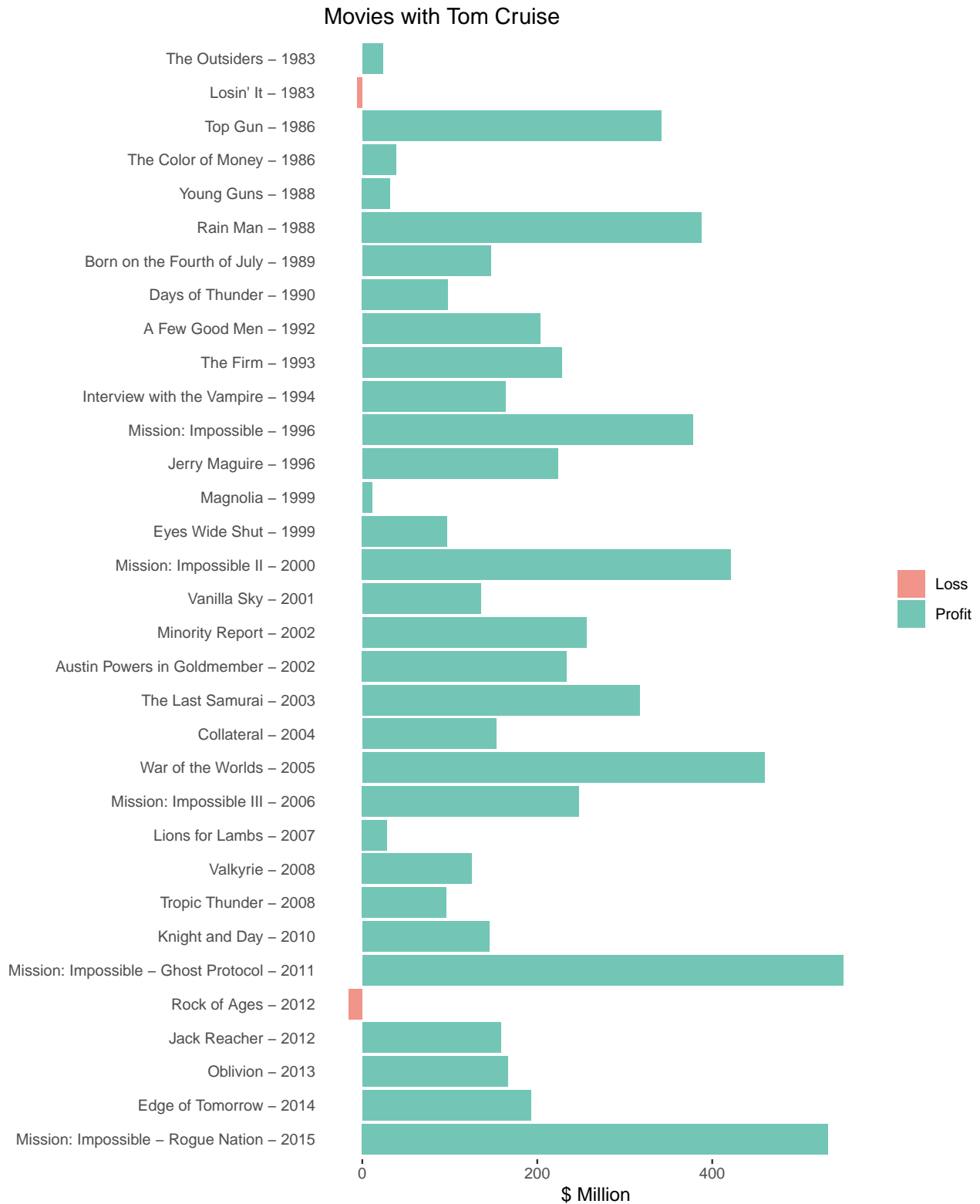


Figure 7

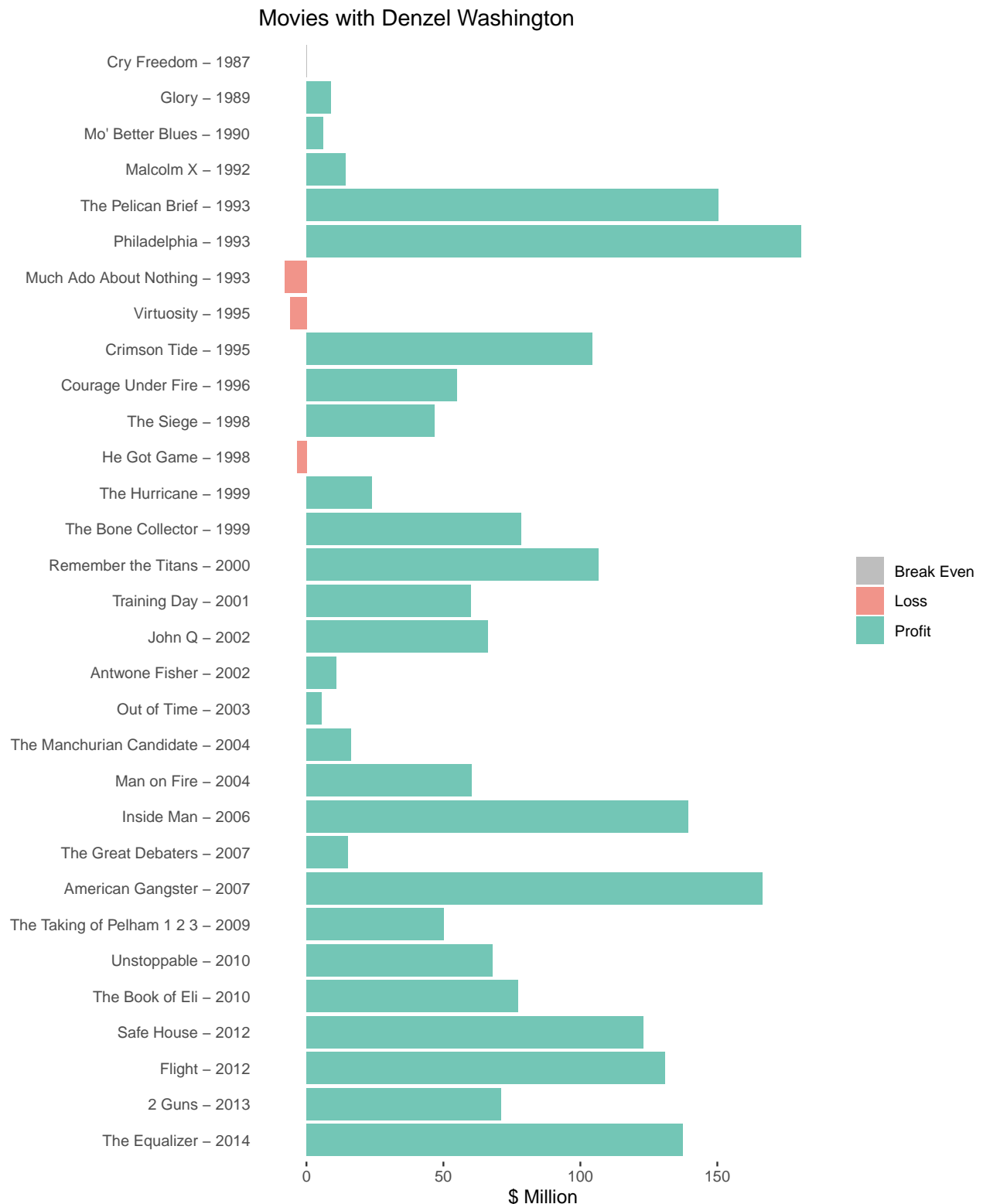


Figure 8

We can observe from Figure 7 (Tom Cruise) & Figure 8 (Denzel Washington) that popular celebrities such as Tom Cruise and Denzel Washington were part of blockbuster movies (Mission Impossible series & Equalizer) that are extremely profitable and also few loss-making movies (Rock of Ages & He Got Game) between 1980s and 2000s. Take Tom Cruise for example. He became a super star after the release of Mission Impossible

series in 1996, 2000 & 2007, but *Magnolia* (1996), *Lions of Lambs* (2007) don't share the same fate and *Rock of Ages* (2012) even suffered losses.

Denzel Washington didn't involved in any movie series like Tom Cruise did between 1980s and 2000s, however, he did deliver consistently profitable movies, except for a few loss-making movies in 1990s.

This mean fame of an actor / actress carry lots of weight but fame alone is not the only factor that gives glory to a movie. Speaking of fame / popularity, how does audience / director / production company evaluates popularity of an actor / actress worth the investment?

3 Quantitative Analysis

3.1 Correlation among key variables

Identify any relationship (positive / negative / none) among the key variables such as *budget*, *revenue*, *runtime*, *vote_average*, *vote_count* with *profitability*. How much each key variable contributes toward *profitability*, if any relationship is identified.

The result is known as **correlation coefficient (r)**. Example: $r = 0.78$ indicates **positive & strong** relationship.

Figure 9: Correlation among Key Variables

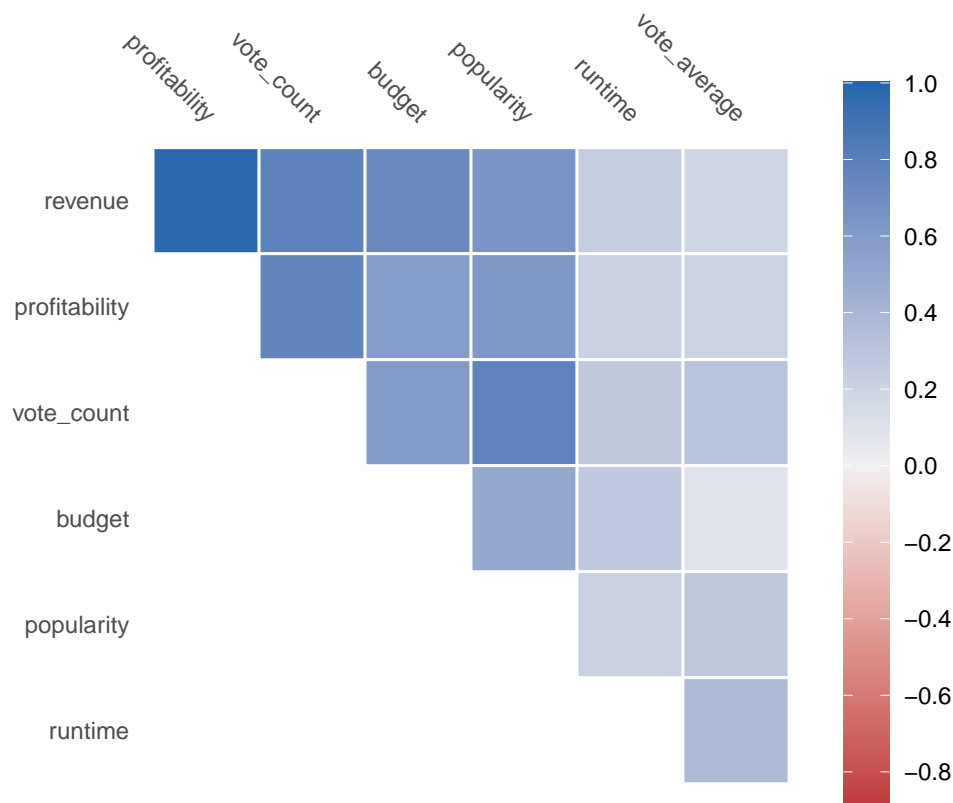


Figure 9 above shows budget, revenue, vote count and popularity are the main factors contribute significantly to profitability.

Logically, budget for movie is given & fixed and producers tend to determine if casts, story line and etc of the movie attract movie-goers. In other word, casts (ie: Tom Cruise) and unique idea / story line (ie: mission impossibles) are the lead factors that contribute to vote count and/or popularity and revenue, and

in turn makes a movie a huge success (measured in profitability). Vote count, popularity, revenue and profit are lag factors.

Therefore, I should only focus more on non-financial metrics - vote count & popularity.

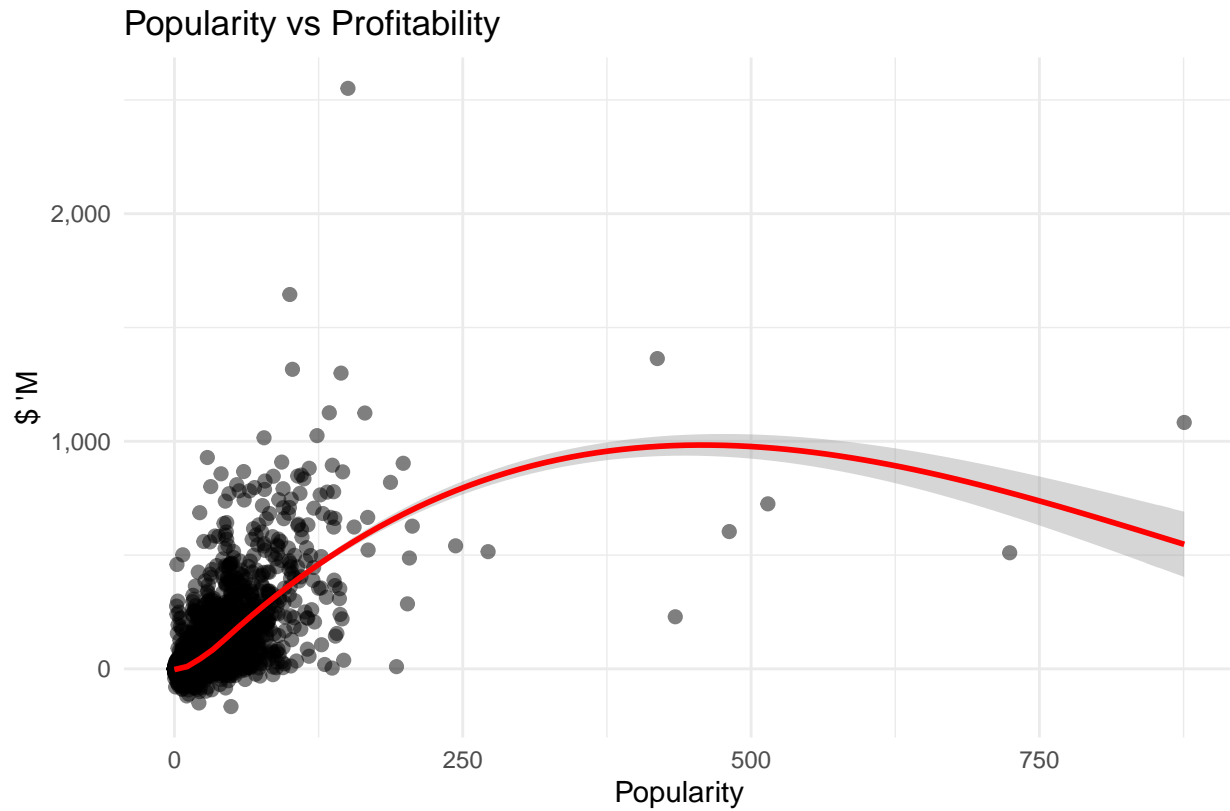


Figure 10

Figure 10 produces an interesting finding where popularity of a movie contribute to profitability only to certain extent. Among all movies, a few movies were extremely popular (popularity > 500) and only one hit \$ 1 billion profit. In contrary, there are also less popular movies (popularity < 250) and recorded \$ 1 billion and more in profit.

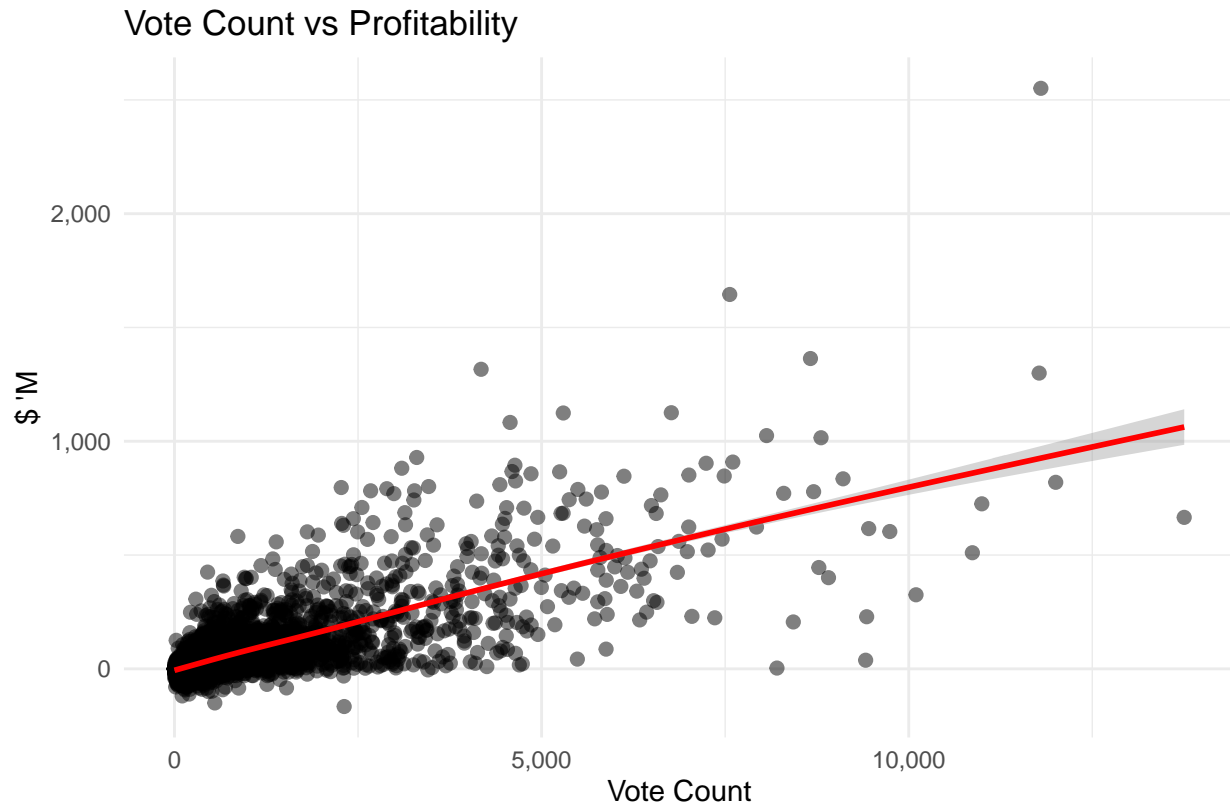


Figure 11

Refer to Figure 11, the trend is more obvious where higher vote count translates to higher profit. This is especially true where vote count is within the range of 7,500.

From quantitative perspective, vote count is better predictor than popularity.

3.2 Conclusion

Based on quantitative and qualitative reasons, can we jump into conclusion that success (measured in revenue & profit) of a movie is the magic combination of the actor / actress and story? Should we attribute the glory to director as well?