

# Will vs Denzel

## STAT 419: Final

**Kathleen Rivas**     *Washington State University*

---

Using data analysis, the question being asked is, “Who is the better actor, Will Smith or Denzel Washington?” The answer is derived using return on investment (ROI) and a graphing rank matrix of the top 2000 movies from the last 40 years as gathered from IMDB.

**Keywords:** summary statistics; boxplots, normalized data, feature scaling

---

December 14, 2020

### 1 Introduction

In the last 40 years, Will Smith and Denzel Washington are two of the most prominent African-American actors in the film industry. At least one, if not several movies of theirs shows up regularly in lists compiled by public users on IMDB and other websites. Currently, on Ranker, a website that uses public votes to rank a wide variety of topics, has Denzel Washington at #2 of 101 on the “The Best African-American Film Actors” list. Will Smith is #5. (ranker.com 2020)

The question of “Who is the better actor?” is highly subjective with no “right” or “one” answer. It really depends on what is important to the person. Do film ratings determine the best? Or how diverse the type of film which shows a willingness to experiment or a sign of an actor less concerned with film quality? Perhaps profit determines the best. Not only that, but what is more important – is the ability to potentially pull in 900% profit once in a while better than the ability to consistently pull in 100-200% profit? These are not the only “metrics” (inasmuch as metrics can be applied to a subjective subject) that can be used.

So how will I determine who is the better actor?

### 2 Defining "the best"

The film industry’s primary reason for existence is to generate profits. It’s a business. Ratings of films, rankings, lists, movie and actor fans wouldn’t exist if it wasn’t profitable. Putting myself in the mind of a film executive, my primary concern with who is the better actor is financial. Secondary to that is the reputation of the movie – how film critics and the populace think how good the film is – this would be important for reputation contests like the Academy Awards or the Oscars.

With this in mind, the factors used for profitability analysis are budget, box gross, total gross. With this data, I was able to see how much profit the movies generate, and more importantly to me, the return on investment. Think about it. What good does generating \$900million dollars if it took \$900 million dollars to produce? As a film executive, this means you have gotten a 0% return on investment – i.e. you broke even and made no money.

What is more important to me (if I put myself in the shoes of a film executive), is how consistently and how large of an ROI an actor’s movie can generate for the film company. My first metric is purely profit driven and looking at it as “two fish in a barrel”– Will and Denzel’s profit data, side-by-side comparison.

Second in importance is how well the movie is received, and its popularity. Popularity can help determine profit, and critical reception is good reputation for future movies from that actor/producer/etc. Data used for this is through a professor generated graphing rank matrix used on the top 50 movies for each year on the last 40 years. Using the fancy adjacency and eigenvector matrix algorithms, similar to how Google Page Rank algorithm works for their search engine results, the MovieRank dataset has each movie ranked (a “movieRank” number from 0-100) within the ecosystem of the top 50 movies over the last 40 years.

Therefore, my second metric uses this larger matrix of movies as Will and Denzel as “two fish in a pond”– Will and Denzel’s movieRank and ROI within the larger matrix of 1,753 “top50” movies.

## 2.1 Normalizing the Data

Will and Denzel’s movies are ranked within the ecosystem of a larger database of movies. However, their budget and ROI data are not (as far as I know). Therefore, I normalized the data so that Will and Denzel’s money data is normalized within the larger movie ranking ecosystem. This way, both movie rank and financial rank uses the same dataset and set within the same “pool”.

Normalization was done on a scale of 0 to 100. To elaborate, all 1,753 movies budget and gross information was calculated to give you a sense of how the movie did against all other movies within 0 to 100 – people can intuitively get a sense of the “big picture”. The larger the number, the “better” the score or rank. The formula used:

$$X^{scaled} = \frac{X_i}{\sum_{i=1}^{\infty} X_i} \quad (1)$$

## 3 Data Description

The size of this ecosystem is 1,998 movies. With all the NAs removed from budget and global gross columns, I ended up with 1,753 movies. The dataset was provided for me by the professor who used a graph rank matrix algorithm to calculate “movieRank”.

I calculated ROI for each movie, both on US profit, and total profit (worldwide and US combined). Then, I normalized three features – budget, ROI on US only, and ROI Total.

From this dataset, I then extracted the Will and Denzel’s movies. Each had 19 movies appear in an IMDB top 50 list over the course of 40 years. This was further reduced to 17 movies each for comparison. Some movies were removed due to lack of financial information (Ma Rainey’s Black Bottom (2020) [Denzel] and Fallen (1998) [Denzel]). One movie was removed due to an error – Will Smith was tagged with Austin Powers: The Spy Who Shagged Me (1999), but did not appear in the cast credits on the IMDB webpage. A William Smith is in the credits, but not as an actor. I removed it from Will Smith’s list of top movies, but kept the movie information within the larger dataset for normalization reasons.

The top movies that Will and Denzel made that were analyzed are as follows:

Will Smith	Denzel Washington
Independence Day (1996)	The Pelican Brief (1993)
Enemy of the State (1998)	American Gangster (2007)
Men in Black (1997)	Glory (1989)
Men in Black II (2002)	Training Day (2001)
Men in Black 3 (2012)	Inside Man (2006)
Bad Boys II (2003)	Deja Vu (2006)
Wild Wild West (1999)	The Book of Eli (2010)
I, Robot (2004)	Much Ado About Nothing (1993)
Hancock (2008)	The Magnificent Seven (2016)
Suicide Squad (2016)	Crimson Tide (1995)
Hitch (2005)	Man on Fire (2004)
The Pursuit of Happyness (2006)	Remember the Titans (2000)
Focus (2015)	Philadelphia (1993)
Bad Boys (1995)	The Equalizer (2014)
I am Legend (2007)	The Equalizer 2 (2018)
Aladdin (2019)	Malcolm X (1992)
The Karate Kid (2010)	Flight (2012)

### 3.1 Summary of Will and Denzel

In order to understand Will and Denzel’s movies in context to the larger “pool” of movies, understanding Will and Denzel’s individual data comes first. In order to do that, here is an overview of Will and Denzel’s

movies:

The total number of movies Will has made: 48 The total number of movies Denzel has made: 40 Note: I removed many pre-production and possible productions from the datasets. (Will had a large amount of these.) Summary Statistics on the Return on Investment: Will Smith

Min	Q1	Median	Mean	Q3	Max	NA's
-81.05%	22.66%	177.35%	221.21%	323.20%	989.87%	5

Denzel Washington

Min	Q1	Median	Mean	Q3	Max	NA's
-79.66%	48.41%	99.10%	129.58%	180.45%	694.92%	2

In exploring the datasets, I thought of several questions related to money. I noticed that there were many instances where the movies lost money in the US, yet still managed to profit in total due to the worldwide sales. It showed me, data-wise, how important the global market is in the film industry.

As an aside, this made me wonder if the recent push towards “diversity” is less altruistic and more financially driven – Matt Damon’s participation in *The Great Wall* (2017) and the recent influx of Asian-American based content (such as *Crazy Rich Asians* (2018)). The Internet has really made global content easily flexible which has caused interesting shifts such as the influence of K-pop through YouTube, etc. It makes sense to create content that would appeal to this demographic, albeit in a more niche sort of fashion.

For the following summary statistics, due to NA’s, Will’s total movies analyzed are 43 and Denzel’s are 38.

In any case, how many times did the global market save a movie from US profit loss?

For Will, that is 9 times out of 14 (64.29%). 17/43 movies did not break even in the US, yet only 8 movies

out of that did not break even total. However, three of those were not globally distributed, so out of 14 movies, the global market made enough to break even or more 9 times.

For Denzel, that is once (100%). Out of 6 movies that lost profit total, only 1 of these 6 profit-loss movies were distributed globally and lost. This sounds like Denzel loses more profit than Will, but that is not the case. Movies that did not break even in the US: Will: 17/43 (47.1%) Denzel: 13/38 (34.21%)

Movies that did not break even total: Will: 9/43 (20.93%) Denzel: 6/38 (15.79%)

Movies that made more than (an arbitrary) \$50 million in profit: Will: 25/43 (58.14%) Denzel: 20/38 (52.63%) At this point, I am not sure who is the better actor. In a sense, what type of film executive am I?

Am I the type of person who wants a consistent US profit only type of profit? Am I the type that wants to make as much profit as I can in the long run (i.e. with global profits)?

If I was only interested in consistency, Denzel is the better actor. His percent to return has a lower minimum loss, a better Q1 than Will. He will generally lose less money than Will will. However, if I was interested in the greater profits, Will is the better actor – his median score is better than Denzel’s, and his upper range at Q3 and Max exceed Denzel’s by 142.75% and 294.95% times.

### 3.2 Summary Statistics of Data

Looking at Will and Denzel as “two fish in a barrel” provided some perspective on their individual ROI performance. Will’s potential range is wider than Denzel’s – he not only can generate greater ROI, but also lower ROI.

However, critical reception is also another metric important to a film executive. There is prestige when a movie wins Best Picture at the Oscars or other film competitions. There may be long term profits associated with films that have attained Best Picture.

The following is a summary table, a side by side comparison of the movies that have made it to the top50 lists, and where it ranks within the larger dataset of movieRank.

Will's Movies	Movie Rank	ROI Rank	ROI Rank	Movie Rank	Denzel's Movies
Independence Day (1996)	47.49	.077	.011	58.66	The Pelican Brief (1993)
Enemy of the State (1998)	43.02	.014	.006	48.04	American Gangster (2007)
Men in Black (1997)	41.9	.043	.007	44.69	Glory (1989)
Men in Black II (2002)	41.34	.017	.007	43.58	Training Day (2001)
Men in Black 3 (2012)	40.78	.014	.013	43.58	Inside Man (2006)
Bad Boys II (2003)	34.64	.009	.004	39.66	Deja Vu (2006)
Wild Wild West (1999)	34.08	.002	.004	39.11	The Book of Eli (2010)
I, Robot (2004)	34.08	.015	.016	37.99	Much Ado About Nothing (1993)
Hancock (2008)	32.96	.025	.01	37.99	The Magnificent Seven (2016)
Suicide Squad (2016)	24.02	.025	.015	37.43	Crimson Tide (1995)
Hitch (2005)	22.35	.033	.019	35.2	Man on Fire (2004)
The Pursuit of Happyness (2006)	21.79	.036	.024	29.05	Remember the Titans (2000)
Focus (2015)	20.11	.017	.008	27.37	Philadelphia (1993)
Bad Boys (1995)	16.2	.05	.03	26.82	The Equalizer (2014)
I am Legend (2007)	12.85	.023	.054	20.67	The Equalizer 2 (2018)
Aladdin (2019)	12.85	.037	.033	20.11	Malcolm X (1992)
The Karate Kid (2010)	6.15	.062	.028	17.32	Flight (2012)

The key takeaway from this table is that you can see that Will's movies score higher in ROI rank. Inversely so, Denzel's movies all have a higher movieRank score than Will's.

Keep in mind that the scaled ROI had two extreme outliers – Paranormal Activity (2007) (ROI Rank: 100) and The Blair Witch Project (1999) (ROI Rank: 32.14). The third highest rank after these two is Clerks (1994) (ROI Rank: .89). It drops from 32.14 to .89! And subsequently lower down to the negatives (those that have lost money). So that needs to be kept in mind when looking at Will and Denzel's ROI rankings.

The movieRank scale is more evenly dispersed throughout so it's easier for the human brain to comprehend those movie rankings of "24.02(%)" than ".036(%)".

#### 4 Key Findings

So, my original premise was to create a type of indexed score that would represent both of these ranks by multiplying movieRank and ROI rank together. Then, I would evaluate this score, discuss the findings, and then construct a visualization of this indexed score. This indexed score, in a side by side comparison resulted in Will's "win". (9 Will wins and 8 Denzel losses)

I've counted the above summary table multiple ways. I've multiplied actorRank, movieRank, and ROI rank – Denzel wins in this case, because his higher actorRank score (35.99 vs Will's 32.01) helps boost his lower ROI score. I thought this was an unfair advantage of Denzel's because actor Rank is pretty subjective. So I removed actorRank from the scoring. I've also tried multiplying movieRank and their individual ROIs (instead of ROI rank), and while they may have adjusted the "wins and losses" a little, Will still "won".

The only time (other than using actorRank) that Denzel "won" was when I would tally who had the higher score, irrespective by how much higher. For example, if Will's movie 1 had a higher movieRank, he would get a 1, if Will's ROI was higher, another tally to 2, and a third metric was used in this case as a "tie-breaker", the ROI on US profits alone (the modern Hollywood strategy to break even in the US box office and profit in the world box office). If Will's movie 1 had a lower score than Denzel's movie 1, then the final score for the first side by side comparison was Will:2, Denzel:1 – Will wins the first round. And so on and so forth.

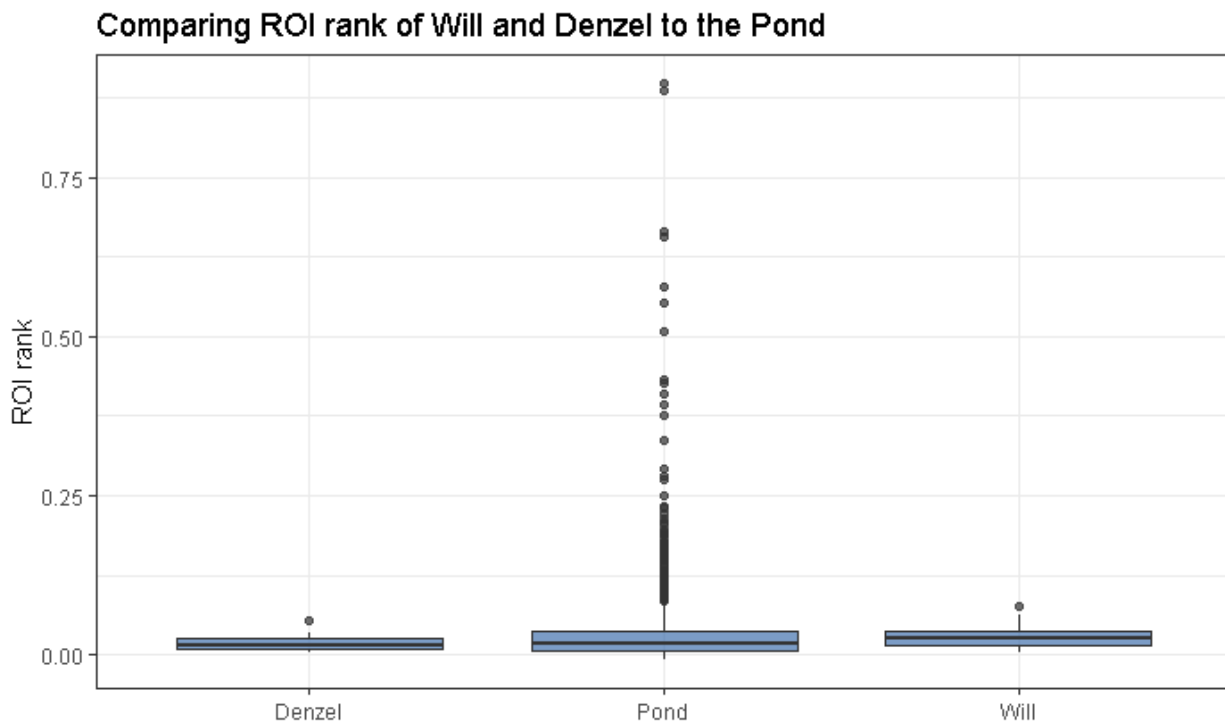
With this method, Denzel won.

But I decided that this wouldn't be fair to Will's range of profit potential. Can Denzel's ROI of 333.929% for *The Pelican Brief* to Will's *Independence Day* ROI of 989.868% really be compared equally to Denzel's ROI of 207.097% for *The Equalizer 2* to Will's ROI of 290.273% for *I Am Legend*? The difference between the first is huge, and the second is quite close.

The same would go for movieRank differences.

However, during my data analysis journey, I came across webpages that decried the use of an indexed score, saying that it reduced the complexity of a multivariate system and provided an oft-misleading convenient single number. (Anonymous 2016) And I am torn. Because when I read these opinions, intuitively I agree. But a single number would be convenient since my summary table of Will and Denzel's top movies ranked with the larger "pond" has opposite results – Denzel's movieRank is higher, and Will's ROI rank is higher, across the board. Combining the score allows me to rank them more conveniently.

The following is a visualization of three boxplots. The first is the boxplot of the "pond" ROI rank scores. This is to provide context to the following boxplots – Will's and Denzel's ROI rank scores. I did remove the two extreme outliers (*Paranormal Activity* and *The Blair Witch Project*) because they distorted the visualization too much to read Will's and Denzel's boxplots. Even with the outliers removed, Will and Denzel's ROI rank boxplots are small in perspective to the "pond".



If you look closely at the boxplot, you will see that the median ROI rank for Will is higher than the median ROI rank of "the Pond" and the median ROI rank score for Denzel is lower than "the Pond's".

In the end, I define "the best" as the one with the better ROI. The "best" actor is the one who had a higher ROI. The movieRank was placed as a secondary consideration. An "indexed" score would help but not looked upon favorably but not necessary. When placing Will's and Denzel's ROI rank within the framework of the "pond's" ROI rank, it is clear that Will's ROI is higher than Denzel's.

## 5 Conclusion

While Denzel's high movieRank scores mitigate his lower ROI rank score, it is not enough to overcome Will's higher ROI rank scores. The data analysis of the ROI of the two actors, as well as how they do compared

to many movies within the same time period shows this. As an imaginary film executive, I would pick Will Smith to be the actor over Denzel Washington.

However, movies are not created in isolation. It is a vast machinery, with many components. Even just focusing on ROI alone can be complicated. There are many interesting layers to answering this question that did not get explored. If there were a way to include actorRank, and the actorRanks of their co-stars, would be interesting. The popularity of other actors (and directors) play a part in the success of a film. I also wanted to look into profitability of movies over time. Perhaps Denzel's "star power" is fading over the years. Or perhaps Will's. When is their "peak" profitability or perhaps it's a uniform distribution.



**REFERENCES**

Anonymous. 2016. combining multiple variables into one "score". (Accessed: December 2020).  
<https://bit.ly/37kY4fo>.

ranker.com. 2020. The Best African-American Film Actors, ranker.com. (Accessed: December 2020).  
<https://bit.ly/2WhlcVH>.



---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Defining "the best"</b>	<b>1</b>
2.1	Normalizing the Data . . . . .	2
<b>3</b>	<b>Data Description</b>	<b>2</b>
3.1	Summary of Will and Denzel . . . . .	2
3.2	Summary Statistics of Data . . . . .	3
<b>4</b>	<b>Key Findings</b>	<b>4</b>
<b>5</b>	<b>Conclusion</b>	<b>5</b>