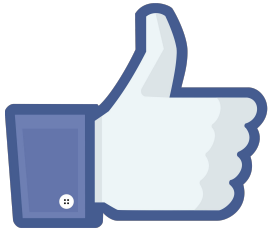
Movie Criticism vs. Elitism: A Correlation Analysis on Movie Reviews and Facebook Likes

Martin Santos
Michael Hissey
Jonathan Chan

Data Exploration & Analysis



Using the amount of Facebook Likes on a set of movies' Facebook Pages as a proxy for that movie's overall audience popularity, we compared this popularity with the critical reviews of those movies from three different online rating systems: IMDB, Rotten Tomatoes, and Metacritic.

The critical review data for each film was collected through an OMDB API call. The data on facebook likes was collected from a publicly available (data.world) CSV file (movie_metadata_csv). Movies were selected from the years ten movies each from the years 2015 and 2016 were randomly selected using the "rand" pandas function to ensure statistical validity.

Once the data was collected, cleaned, and sorted into a dataframe using Pandas, we performed a regression analysis to determine the correlation coefficient between (all) the movies' audience popularity and the critical score it received from each rating system. This correlation coefficient measures how closely the rating system came to reflecting the overall audience popularity.

Hypothesis:

For our hypothesis, we endeavored to show that the IMDB Rating System is the most reflective (highest Pearson R) of a movie's overall audience popularity, as determined by Facebook Likes, and that Rotten Tomatoes would be least reflective (lowest Pearson R).

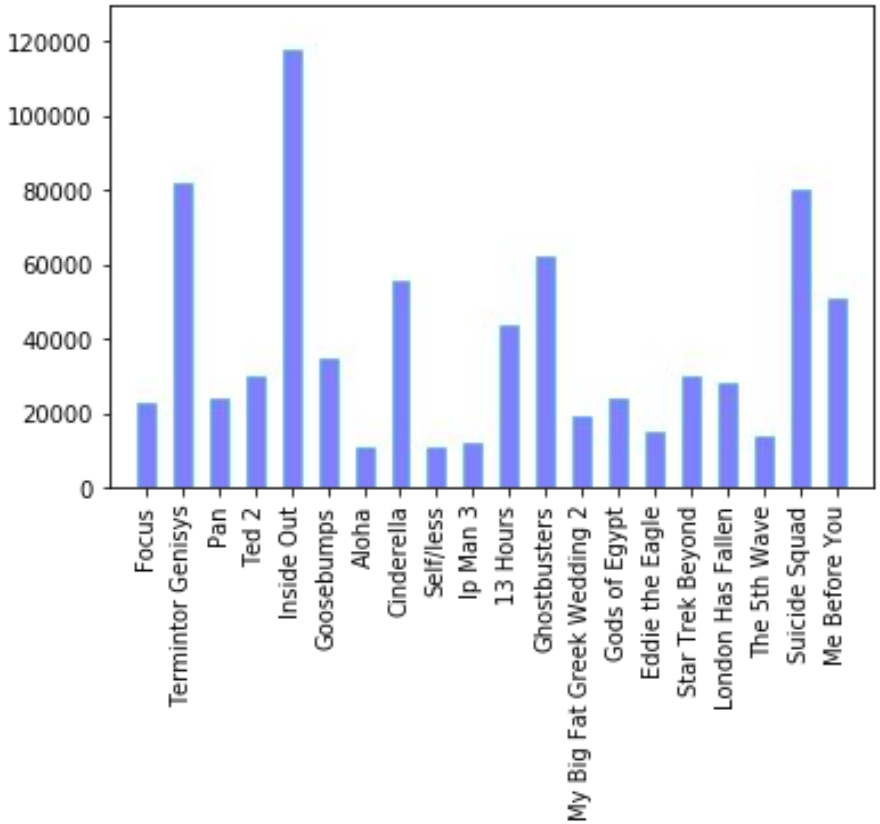
Movie Likes

Due to Facebook's worldwide popularity and availability, we determined that the amount of Facebook Likes a movie page received would be an acceptable metric to serve as a proxy for that film's general popularity. We wanted to discover how close each rating system came to this through determining the correlation coefficient.

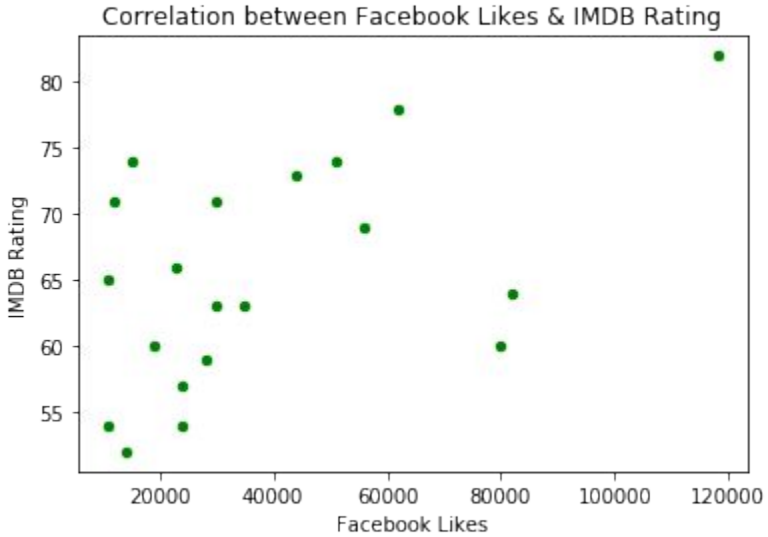
We wanted to know:
1) Which rating system came closest?

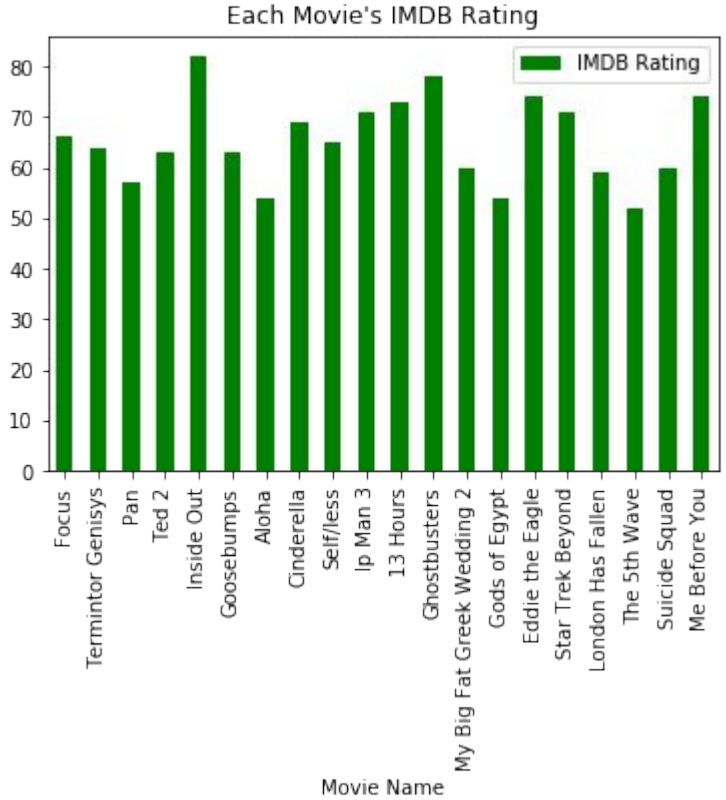
2) Was the result in line with our hypothesis?

3) What further questions can we ask about rating systems given our results?



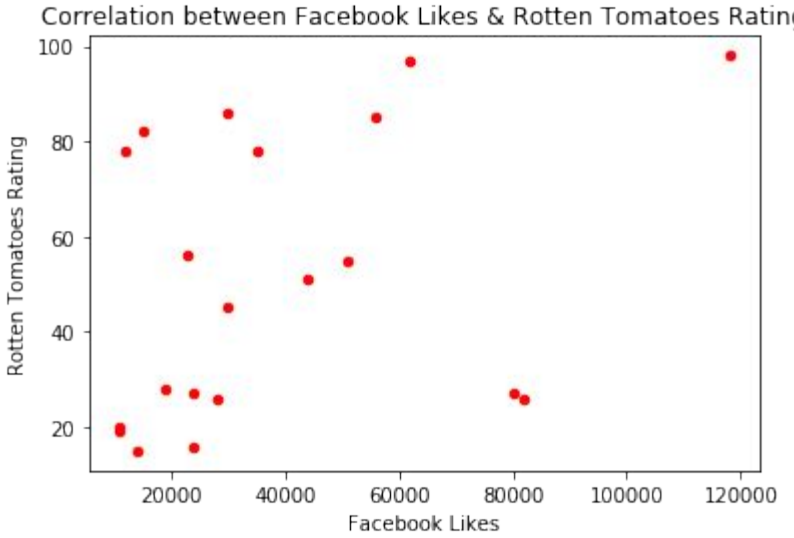
IMDb

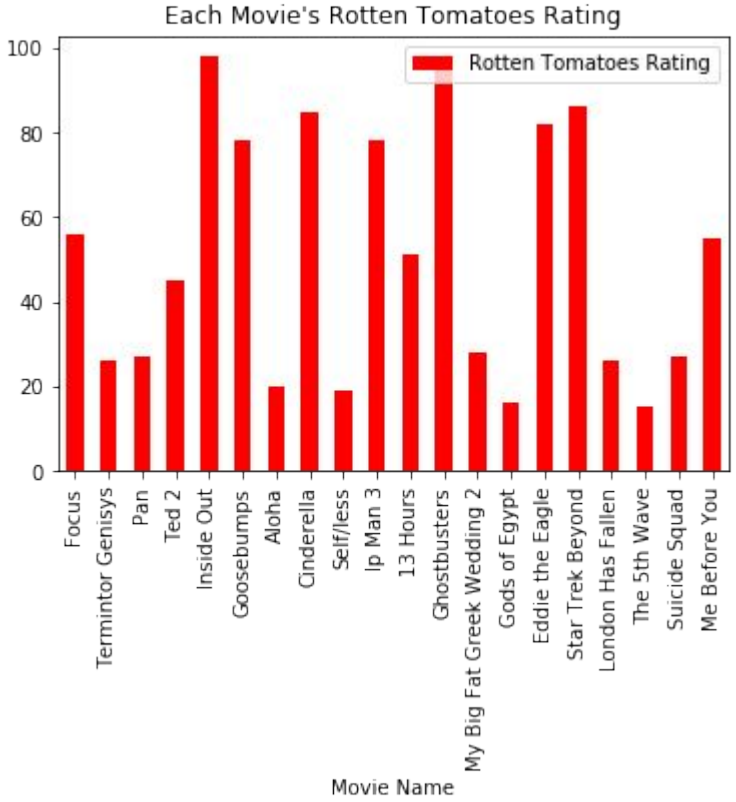

Correlation between Facebook Likes & IMDB Rating
IMDB Rating
Facebook Likes
80
75
70
65
60
55
20000
40000
60000
80000
100000
120000


Each Movie's IMDB Rating
IMDB Rating
Movie Name
80
70
60
50
40
30
20
10
0
Focus
Termintor Genisys
Pan
Ted 2
Inside Out
Goosebumps
Aloha
Cinderella
Self/less
Ip Man 3
13 Hours
Ghostbusters
My Big Fat Greek Wedding 2
Gods of Egypt
Eddie the Eagle
Star Trek Beyond
London Has Fallen
The 5th Wave
Suicide Squad
Me Before You

Facebook vs IMDB

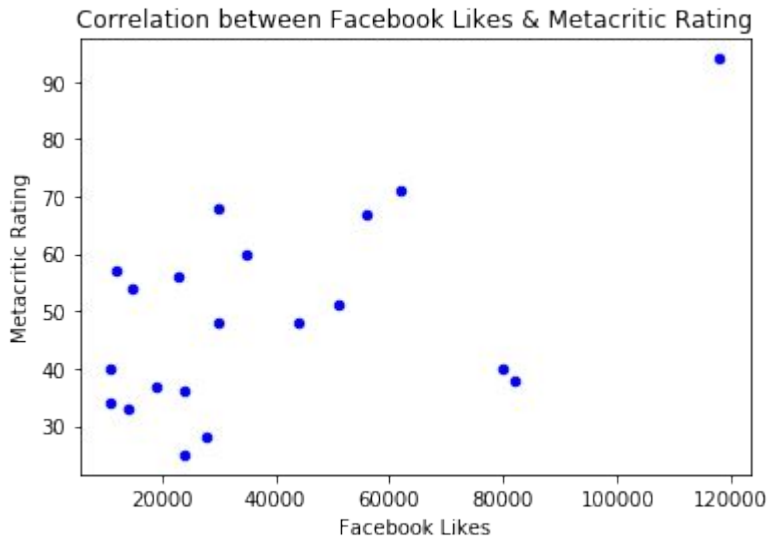
1) IMDB: The first graph is a scatter plot displaying each film as a datapoint, juxtaposing that Film's facebook likes against its IMDB rating. The correlation coefficient is 0.497, which is mildly positive, but is not strong enough to show that the IMDB rating scheme is reflective of the overall audience popularity of the films.
2) The second graph displays each Film's IMDB Rating. It is scaled from 0-100.

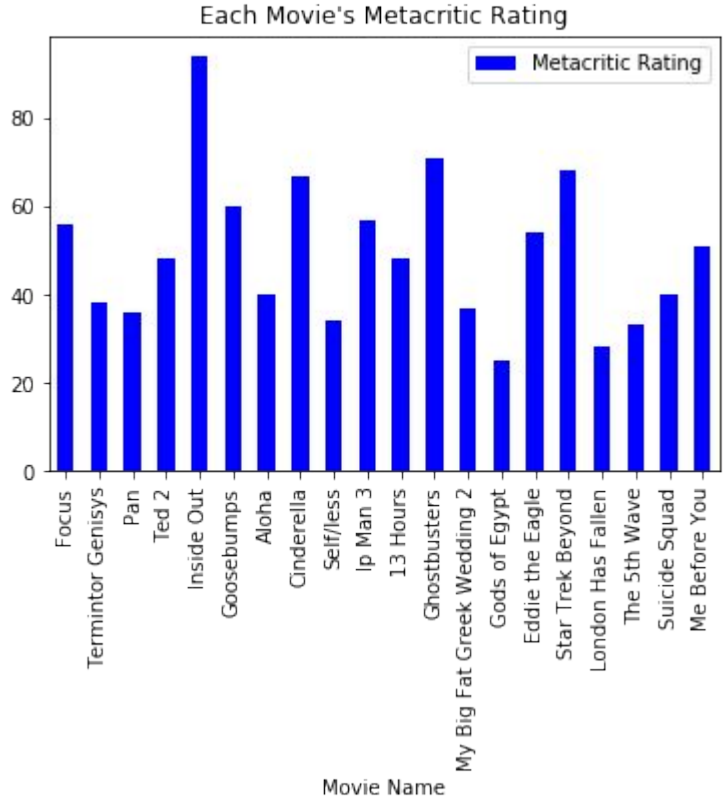

Correlation between Facebook Likes & Rotten Tomatoes Rating
Rotten Tomatoes Rating
Facebook Likes
100
80
60
40
20
20000
40000
60000
80000
100000
120000


Each Movie's Rotten Tomatoes Rating
Rotten Tomatoes Rating
100
80
60
40
20
0
Focus
Termintor Genisys
Pan
Ted 2
Inside Out
Goosebumps
Aloha
Cinderella
Self/less
Ip Man 3
13 Hours
Ghostbusters
My Big Fat Greek Wedding 2
Gods of Egypt
Eddie the Eagle
Star Trek Beyond
London Has Fallen
The 5th Wave
Suicide Squad
Me Before You
Movie Name

Facebook vs Rotten Tomatoes

1) Rotten Tomatoes: The first graph is a scatter plot displaying each film as a datapoint, juxtaposing that Film's facebook likes against its Rotten Tomatoes Rating. The correlation coefficient is 0.353, which is positive, but very weak, and does not show this rating scheme to be reflective of overall audience popularity of the films.
2) The second graph displays each Film's Rotten Tomatoes Rating. It is scaled from 0-100.


Correlation between Facebook Likes & Metacritic Rating
Metacritic Rating
Facebook Likes
90
80
70
60
50
40
30
20000
40000
60000
80000
100000
120000


Each Movie's Metacritic Rating
Metacritic Rating
80
60
40
20
0
Focus
Termintor Genisys
Pan
Ted 2
Inside Out
Goosebumps
Aloha
Cinderella
Self/less
Ip Man 3
13 Hours
Ghostbusters
My Big Fat Greek Wedding 2
Gods of Egypt
Eddie the Eagle
Star Trek Beyond
London Has Fallen
The 5th Wave
Suicide Squad
Me Before You
Movie Name

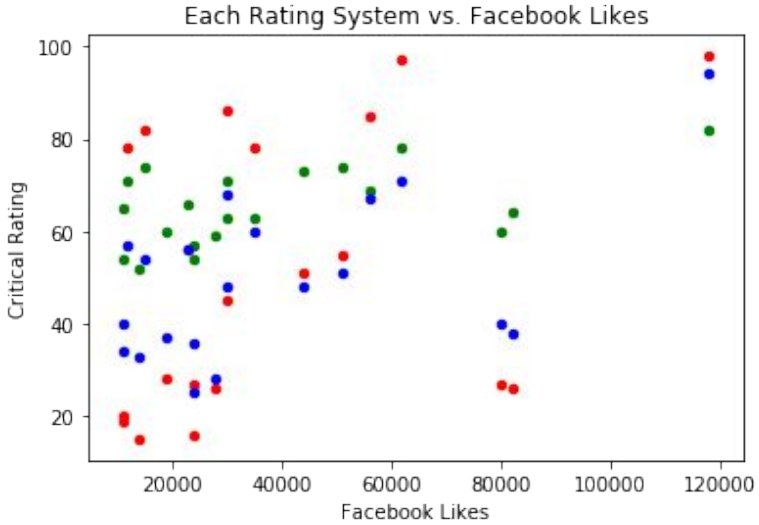
Facebook vs Metacritic

1) Metacritic: The first graph is a scatter plot displaying each film as a datapoint, juxtaposing that Film's facebook likes against its Metacritic Rating. The correlation coefficient is 0.535, which is positive, and the strongest of all three rating systems. Thus, according to our analysis, the Metacritic Rating is most reflective of the overall audience popularity of the films.
2) The second graph displays each Film's Metacritic Rating. It is scaled from 0-100.

Difficulties

Our primary difficulty came from finding a proper proxy for overall audience popularity. We initially considered twitter mentions, but ran into a barrier that came with paid API access. We then considered Youtube views of trailers, but then realized that not all of the movies we were looking to consider had trailers up on Youtube. Finally, we settled on Facebook Page likes, for the aforementioned stated reasons.

Conclusion



Our Hypothesis had two parts. We hypothesized that the IMDB Rating would be most reflective of overall audience popularity, and that Rotten Tomatoes Rating would be the least reflective. We based our hypothesis on the metrics that inform each rating scheme. IMDB aggregates viewer ratings from the general population, while Rotten Tomatoes only takes into account the reviews of respected critics.

We were incorrect about IMDB - it was not the most reflective. Metacritic was. However, we were correct about Rotten Tomatoes being the least reflective.

This poses **further questions** about the rating schemes that we can ask and endeavor to solve:

- How far does Metacritic go in its reach to aggregate reviews? How many reviews does it source, and from where?
- Can answers to the above question provide insights into why it had a stronger correlation coefficient with the overall audience popularity of the films?
- Is there a better proxy we can use, other than Facebook Likes, to get a sense of overall audience popularity?

Thank you for listening!