SENTIMENT

WARS:

Film Critics vs. Fandom

From research in my last project involving the movie industry (network analysis on three recently released movies), I had come across differing viewpoints on the importance of critics' reviews and what role they play in creating buzz for a movie (if any). To investigate this, I decided to track down datasets aggregating critic reviews and compare that with box office revenue. This should shed some light on the (un)importance of critics since ultimately only box office totals matter to the movie makers.

To explore this idea, I decided to use a movie franchise which I'm quite familiar with: Star Wars. As a boy I remember seeing *Return of the Jedi* in a theater in my small farming community and was quite enthralled. I had always heard that critics were less so when it came out.

To answer my questions, I needed two datasets: 1) one showing historical box office totals (adjusted for inflation if possible), and 2) one with critics' reviews. After several hours of searching for historical, detailed box office totals, I finally found a reputable site called the-numbers.com run by analytics firm Nash Information Services out of Beverly Hills (Nash Information Services, 2021). They've been used by firms like Deloitte and Insider.com for publishing data (Deloitte Touche Tohmatsu Limited, 2016; Guerrasio et al., 2020). Aside from accurate data, they also provided a user-friendly web app for comparing data.

Secondly, for the historical reviews, Rotten Tomatoes has become the go-to source for Hollywood critiques. Movie commercials even broadcast "fresh" scores. Fortunately for this analysis, I was able to find an aggregated list of over 1 million reviews on Kaggle updated in 2020 (Leone, 2020). From that, I isolated only the first three first canon movies:

Figure 1 – All three original trilogy, canon Star Wars *posters in order of release date:* Episode IV – A New Hope *(1977),* Episode V – The Empire Strikes Back*, and* Episode VI – Return of the Jedi *(Image source: starwars.com).*

Sentiment Analysis of Critic Reviews

To examine the overall *feeling* critics had for a movie, I decided to perform a text analysis technique called "sentiment analysis." Using the software title Semantria, a sophisticated text analysis engine, I analyzed all the available reviews from critics in the dataset for a given movie. I then compared that overall sentiment, or feeling, to the respective movie's box office performance. It should be noted this analysis isn't perfect because Rotten Tomatoes is relatively new, arriving some time after the internet itself; the first *Star Wars* movie (a.k.a *A New Hope*) was released in 1977. Some of these reviewers have the luxury of hindsight, meaning they already know *Star Wars* was one of the most enigmatic sci-fi movies ever released, so bias in later reviewers plays an important role.

The Rotten Tomatoes dataset has (among other fields) the critic's name, review text, and other metadata (review date, whether they are considered a "top critic," etc.). Here I should note that standard cleaning techniques were used to prepare the data for analysis. The three movies of choice were extracted from the larger dataset for performance reasons (drastically improving loading/analysis times). Next, duplicate rows were removed and blank reviews were replaced with neutral NAs.

For the sentiment analysis, I set Semantria to use the critic's name as the Identifier Column, and the review text for the analysis, or source of content. The rest of the columns were transferred over to investigate possible trends or insights with the pivot tables provided by Semantria. Semantria has a dictionary of about 200,000 sentiment-bearing words in English ([2021, Lexalytics](#)). It searches each text cell for these words and tallies them up. It is able to assign each word a positive or negative score depending on how it is used. For example, if a critic says *Return of the Jedi* was "not great," Semantria understands this and flags the word "great" as having a negative usage. I actually had to perform what is called "tuning" of the sentiment phrases. For example, many reviews used the words "evil" and "monsters" to describe the movies as having battles between good and evil but Semantria always recognized these as negative. Tuning involved manually adding these words as neutral (setting to value of zero) for these particular movies. Performing this analysis is important for both moviegoers and movie

makers because as we'll see, critical insights can be exposed. Now that I've explained what we are looking for, what datasets we're looking in, and how we're performing the analysis, let's see the findings.

Key Points – Critics' Sentiment Analysis

Figures 2-4 contain data visualizations of the results of the sentiment analysis: each sentiment phrase occurrence and whether it is used positively or negatively towards that movie. To compare with the respective box office performance, *Figure 5* contains a line graph of the box office performance over time along with the cumulative critics' sentiment totals. This tells us whether the critics – overall – liked or disliked the movie.
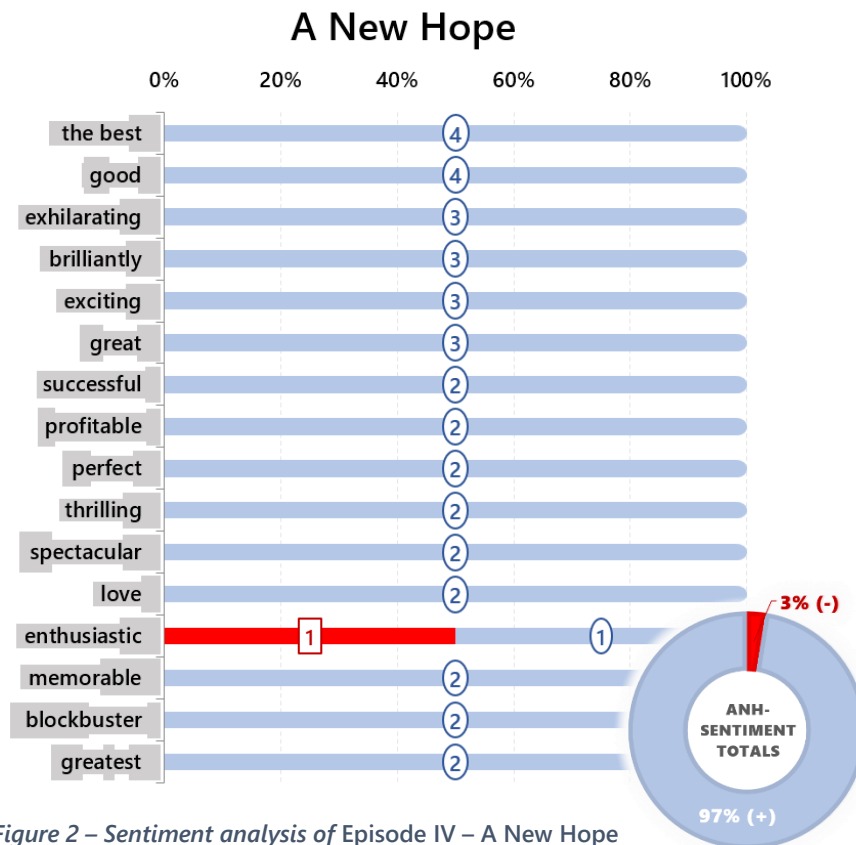

Figure 2 – Sentiment analysis of Episode IV – A New Hope *results showing only positive and negative words (neutrals removed), with inset donut chart showing total percentage of each. With 97% positive words, the critics really liked A New Hope.*

The Empire Strikes Back

the best

good

great

greatest

exciting

most successful

spectacular

powerful

thrilling

most mature

superior

happy

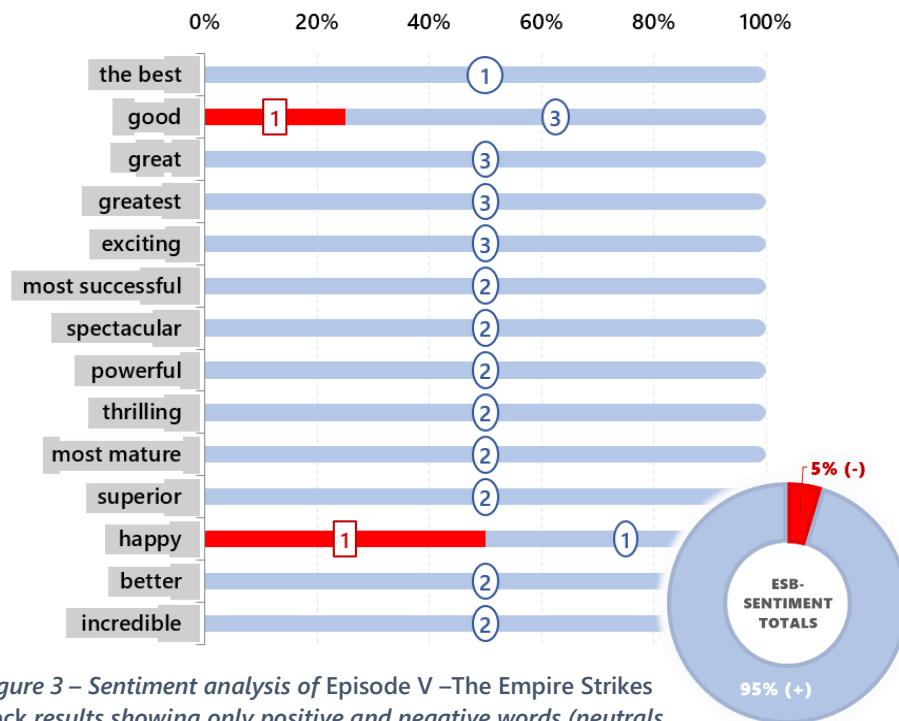better

incredible

5% (-)

ESB-
SENTIMENT
TOTALS

95% (+)

Figure 3 – Sentiment analysis of Episode V –The Empire Strikes Back results showing only positive and negative words (neutrals removed), with inset donut chart showing total percentage of each. Ninety-five percent positive is still very favorable.

Return of the Jedi

satisfying

great

the best

good

bad

enjoy

disappointment

magnificent

thrilling

spectacular

better

entertaining

captivating

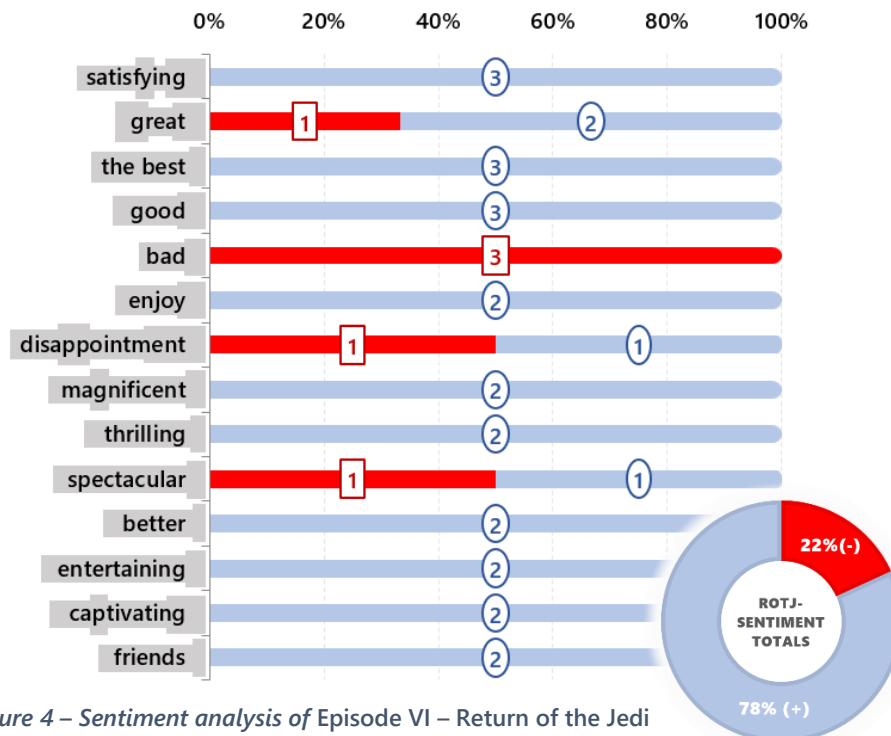friends

22%(-)

ROTJ-
SENTIMENT
TOTALS

78% (+)

Figure 4 – Sentiment analysis of Episode VI – Return of the Jedi results showing only positive and negative words (neutrals removed), with inset donut chart showing total percentage of each. At 78% positive, this movie was the critics' least favorite.

Discussion – Critics vs. Box Office Earnings

Now let's compare the findings from the sentiment analysis with the box office revenues from each movie and see what influence, if any, critics have on the franchise.
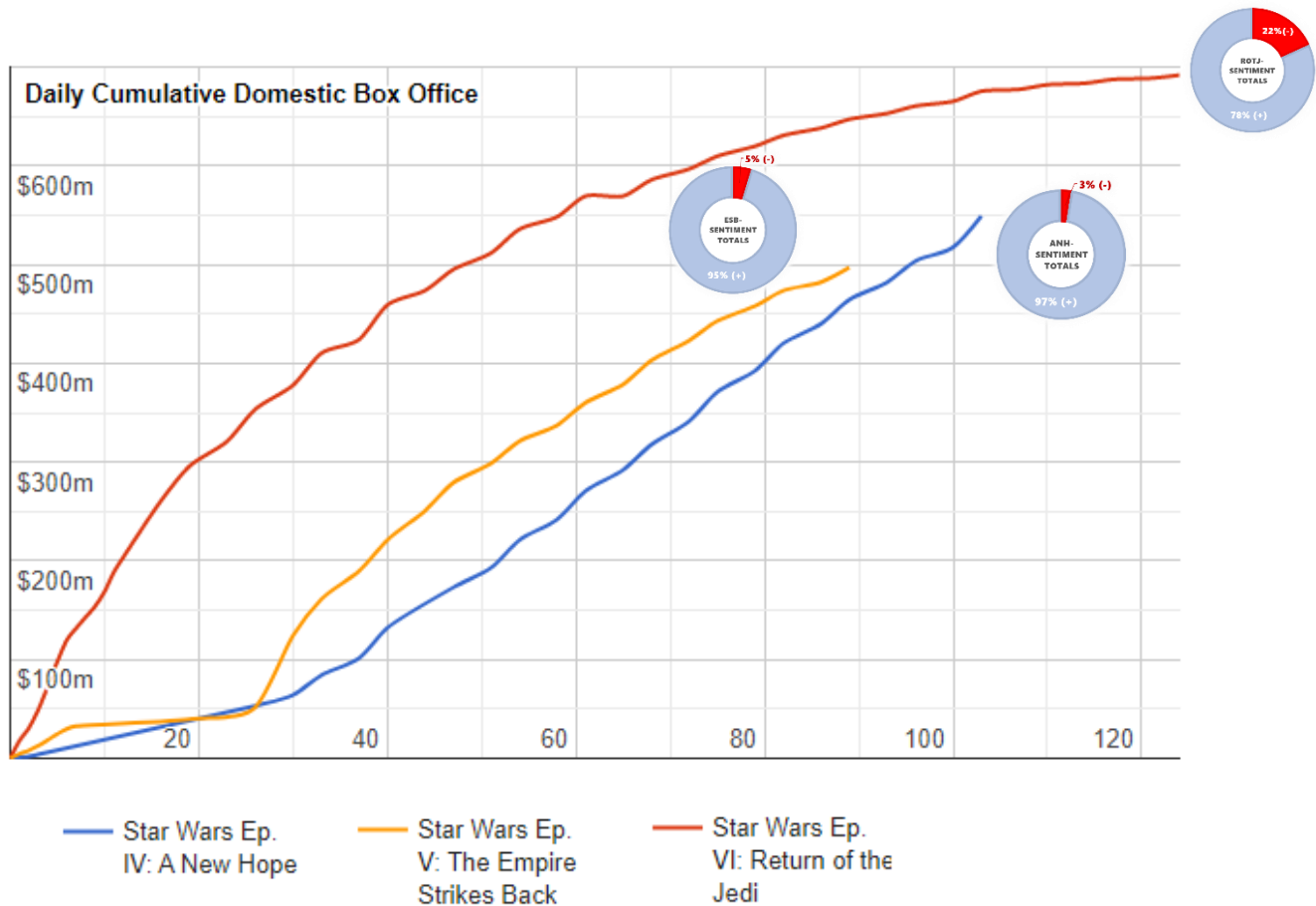

Figure 5 – Inflation-adjusted, domestic box office earnings for each Star Wars movie from the original trilogy as a function of days in theaters. Data was compiled using the-numbers.com movie comparison page.

Figure 5 shows each movie's earnings as a function of time (lines) as well as the same donut charts of cumulative sentiment totals from the previous three figures. Some interesting things can be seen in this comparison. First, the overall observation that jumps out is that *critics' reviews don't seem to matter*, at least in the Star Wars universe. In fact, there is almost a negative correlation between review positivity and box office earnings. For example, the worst received movie of the three, *Return of the Jedi*, was the highest earner. Also note the slow start *Empire Strikes Back* had (~ days 10 – 25). It could be that

moviegoers read the reviews and decided to stay home, but as more and more of their friends went to see it with positive feedback, they changed their minds.

Keep in mind, this analysis is only three movies from nearly 40 years ago BUT this initial comparison is intriguing. More analysis is needed. Luckily there are plenty of Star Wars movies released since and many more in the pipeline.

"Difficult to see, always in motion is the future."

- **Jedi Master Yoda (The Empire Strikes Back)**

References

Deloitte Touche Tohmatsu Limited. (2016). *The award for stable box office revenues in the face of digital media goes to…* Deloitte network. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/gx-tmt-prediction-box-office-revenue.pdf

Guerrasio, J., Shamsian, J., Loudenback, T. (2020). The biggest box-office movie that came out the year you were born. Insider Inc. https://www.insider.com/most-popular-movie-every-year-2017-8

Leone, S. (2020). *Rotten Tomatoes movies and critic reviews dataset*. Public Domain. Available from Kaggle.com: https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset

Lexalytics. (2021). *Semantria for Excel 5.3 – Document Detail*. 2:03. Semantria for Excel/Videos. https://youtu.be/nGg_e0CoxC4

Nash Information Services. (2021). the-numbers.com About webpage. https://www.the-numbers.com/about