MULTIVARIATE COMPARISON OF EXPERTS' VERSUS USERS' REVIEWS ONLINE

By

Mariana Repsold França

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the

JOSEPH WHARTON SCHOLARS

Faculty Advisor:

Professor Jehoshua Eliashberg

Sebastian S. Kresge Professor of Marketing

Professor of Operations, Information and Decisions

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2018

TABLE OF CONTENTS

ABSTRACT

This study aims to compare and contrast professional critics' movie reviews with those of general public users' to understand the differences and similarities between both and its effect. This will allow better informed decisions to be made highlighting when and for what ends the use of each type is more appropriate. This study analyzes length, ease of reading, sentiment analysis, main topics, emotional tones and the effect of time to evaluate the reviews. All the data used in this study pertains to the movie Avatar and was obtained from the IMDb movie database online. The research conducted found that while critics' and users' reviews may truly have some generally expected differences in fields like length or ease of reading, they may also present unexpected similarities in others that include, but are not limited to, the polarity, the main topics and even certain emotional tones of the reviews. Evaluating how users' reviews changed over time additionally allows for a comprehension of the change in sentiment with the gradual distancing from the movie release date.

Keywords:

Reviews, Movies, Ratings, Critics, IMDb, Text Analytics

INTRODUCTION

Motivation and Purpose of Thesis

Data has become a vital resource for the modern society and the increasingly digital and connected world. It is everywhere and is increasingly used to guide decisions ensuring that the probability of positive results is maximized. Recognizing the importance of interpreting data today has been a key factor in designing this study and topic. Being able to manipulate data to use it as a predictive and explanatory tool is distinctively valuable and applicable to a wide set of fields.

When exploring the field of data analytics, one of the newer fields that growing in importance fields is text analysis. It offers many data points and allows for a wide set of inferences to be drawn, making it an attractive opportunity for data scientists and enthusiasts. Given its fairly recent nature, there are still many areas to further develop research. One that is of particular interest, is review analysis as reviews have exponentially grown in the last two decades. They have become increasingly influential and important opinion and behavior drivers with the booming expansion of social media and the digital space. Today's society is an extremely social one where individuals like to share their experiences and give recommendations to others and thus online reviews are a very effective outlet for this trend. With the growth of social media, people base their purchase decisions more and more on what they read online from others.

Moreover, the accessibility and availability of the data are also attractive elements for this topic. Today, internet is easily accessible from almost anywhere you go and with a few clicks you may leave a review, read others' reviews, or find the answer to almost anything you want to know.

There are endless websites to connect with other users and exchange information in the form of reviews for places you have been, products you have tried or services you have used. Similarly, professional critics may now also directly distribute their own evaluations directly to the public considerably expanding their reach, instead of having to rely on print publications and a substantial level of effort from the audience to access and read them. This behavior and industry trend have led to new consumption movements and so, by analyzing reviews, we may uncover a valuable forecasting potential of future consumption patterns.

The research is focused on the movie industry as it is a theme that raises a lot of interest not only in society's day-to-day lives, but also online. There is a myriad of varied opinions and discussions and these are very useful data points that can provide substantial insight to highly profitable businesses. Additionally, it is an industry where it is possible to easily observe outcomes in the form of box office demand and revenues.

Method

This study utilized the IMDb database to extract nearly 4,000 reviews on the movie Avatar (2009) that were then vetted to obtain 322 expert critics' reviews and 322 general public users' reviews. Users' reviews were extracted using a python crawler while critic's reviews had to be pulled manually. Analyses were then done on Microsoft Excel, IBM Watson Analyzer, Flesch Reading Ease calculator and MeaningCloud (Excel add-in) to compare and contrast reviews' length, ease of reading, sentiment analysis, main topics, emotional tone and publication time lag between users' and critics' reviews. This thus enabled a better understanding of how the two sets of reviews differ or assimilate and what information can best be gathered from each.

Findings & Contribution of the Paper

The results of this study indicate that while users' and critic's reviews differ in terms of length, ease of reading and the presence of irony, they can be relatively similar in their polarity level and some emotional tones. Additionally, using users' review publication dates against the movie launch date to analyze the change on the reviews over time also yielded interesting results as in the beginning there appeared to be a balance between positive and negative reviews which stopped being the case with the passage of time as the number of negative reviews decreased. All analyzes were done solely on the text data and thus this study contributes to the field of text analytics and especially highlights and emphasizes how much information may be obtained from text.

RELEVANT LITERATURE

In today's fast-moving society, data and its interpretation are paramount. Being able to use data to predict future patterns and behaviors is of immense value. This study will look at how text can be analyzed to improve interpretation and gain insight on upcoming trends. With machine learning, a lot of work has been done in the field of text analytics for classification purposes and Jostein Gripsrud's 1995 book "The Dynasty Years: Hollywood Television and Critical Media Studies" is a common example of a study that places people in categories based on their responses to a television program. It is a research field however, that is still growing as text analyses become more precise and more valuable data can be extracted from them.

There are numerous existing literatures on text analytics to make forecasts as it is a subject that has gained traction and thus been considerably developed in the last decade increasing its preciseness and effectiveness at making estimates. Text is available everywhere and being able to classify and use it to make future predictions is a tool that is extremely powerful and valuable in any industry. Better understanding the power of text and drawing parallels between what different words can mean is highly applicable to a wide set of contexts. By taking on an angle that has not yet been extensively researched on, this paper plans to make the field of text analytics even more accessible and useful in its forecast potential.

Movie reviews offer interesting analyses and have not yet been so thoroughly examined that there is an excessive amount of data published. Although the movie industry and its revenue models have been thoroughly studied for years, the text analysis of reviews has attracted scholars' interests only more recently. The rest of the section will discuss some previous research on topics surrounding the questions this study intends to answer.

Expert critics' reviews impact on moviegoers

Professional movie critics' reviews are widespread and may often provide contrasting opinions. In the movie industry – and the entertainment industry in general – they play a significant role as they may directly impact the financial health of a film production. Interestingly however, critics' reviews appear to be better predictors of movie performance than influencers of moviegoers' behaviors and preferences (Eliashberg and Shugan 1997). This seems to be the case as critics' reviews appear to be in line with late box office performance as well as with accumulated box office profits, but not with early box office success, which is directly linked with the public's perception of the movie. This shows that when a movie is first released, critics' reviews had a very marginal effect on the public's decision to watch a movie. Movie goers that are looking forward to the release of a movie will go independent of how it is reviewed, but those who choose to watch it after the initial release buzz will tend to prefer making a more informed decision and thus their choice will be affected by the movie's ratings.

A succeeding study (Basuroy, Chatterjee and Ravid 2003) found that critics may play the dual role if influencers and predictors of box office revenue. The impact of negative reviews was found to diminish over time, however negative reviews tend to more negatively impact movies than positive reviews help during the first week a movie is exhibited at the theaters. Additionally, the researchers also found that having famous actors and actress as well as a large budget for the movie improve box office revenue more for movies that receive negative critical reviews than for those that were positively reviewed. Therefore, producers and studios should look at critics' reviews strategically in order to successfully maximize their box office revenues.

Movie awards and recognition effect on movies' box office

Movie producers, directors and actors seek recognition for their performance in particular movies through important well-known awards like Oscars and other Academy Awards. A study on the effect this recognition has on the financial aspect of the movies (Dodds and Holbrook 1988) has shown that there is a direct relationship between the box-office revenue of a movie and the awards it receives. For example, movies that received the most sought-after awards – best picture, best actor/actress or best director – experienced a boost in their box-office numbers, directly increasing their revenue stream. Moreover, another study (Terry, Butler and De'Arno De'Armond 2005) was actually able to quantify the dollar amount effect that these recognitions may yield and in 2005 an Academy Award nomination was actually worth more than $6 million dollars representing an extremely significant financial return. Similarly, a best picture nomination by the Academy will yield a 25% weekly increase in box office revenue while an Oscar win translates into a 50% increase (Ginsburgh, Gutierrez-Navratil and Prieto-Rodriguez 2014). There has been considerable research on how recognition through awards impacts actual box-office revenues and it became clear that receiving a distinctive honor from one of the most respected associations in the movie industry will directly impact a movie financially.

Social media influence on movie demand

Social media has become an integral part of our everyday lives and thus it has recently attracted researchers' interest resulting in close scrutiny of the activity within their forums and pages. Looking at the behavior people have on social media has appeared to be a great forecasting tool. A study on Twitter (Asur and Huberman 2010) that covered almost 3 million tweets showed that the amount of attention an upcoming movie receives in terms of volume of tweets actually directly impacts the ranking the movie will receive. Additionally, including a sentiment analysis of the tweets helped improve the accuracy of predictions once the movie was released. This shows that text comments – similar to user reviews – may directly impact the box-office revenues of the film. Text clearly plays an important and significant role on the future perception the public has of a movie.

Users' Internet search behavior impact on box office performance

Still within the sphere of consumers' digital behavior, researchers found that there is an interesting link between Internet searches and box office revenues (Lee, Cheon Cha and Kim 2016). The researchers of the study that explored this relationship found interesting and perhaps slightly unexpected results as Internet searches appear to be a bigger factor following high box-office revenues rather than leading to one. What this means is that the searches are serving more the purpose of information sharing and extending than as an informational channel. The study found that people who had a positive experience watching the movie were more likely to conduct searches about the movie and thus movies with the highest box-office numbers were also the ones with the highest search volume. These findings thus served as a springboard for the discussion of this paper as the information sharing described could and most frequently is translated into user reviews, which is one of the facets this study will explore.

Online reviews and the power of online word-of-mouth

When analyzing online reviews, many varied parallels may be drawn. Certain characteristics present in a review may shape customer behavior and thus are very relevant for this study. The study by Hua-Ning Chen and Chun-Yao Huang (2013) identified four main variables that play a key role in differentiating reviews from one another. These are as follows: average rating, average length of textual reviews, feedback from others, and whether the reviewer uses their real name or discloses other personal information. The study found a positive relationship between rating and frequency as well as continuity meaning that reviewers were more likely to leave a high rating when satisfied with the movie. The length of the review is also informative as longer reviews translate into a higher level of effort and commitment by the reviewer. The longer the review, the higher the frequency of reviews by that specific reviewer. Feedback from others was quantified as number of comments and average helpful votes. The more interactions the reviewers had with others – i.e. comments on their reviews – the more social benefits they enjoyed and thus the more motivated they were to post reviews. The helpfulness aspect also provides an interesting perspective as the higher the helpfulness vote, the more consumers will include it in their set of considerations when deciding whether to watch the movie. Additionally, for the reviewer a high helpfulness also motivates them to write more review as it functions as recognition mechanism. Moreover, the use of reviewers' real name instead of anonymity, leads to a smaller number of reviews to be published by them. This likely occurs as the disclosure of personal information results in the reviewer's own reputation to become open to scrutiny and thus in order to ensure a high reputation is maintained, he will write reviews more carefully and thoughtfully so as to not leave any space for uncertainty.

Similarly, it is possible to find a direct relationship between online word-of-mouth and review helpfulness to box office revenue. A recent study (Lee and Choeh 2018) has found that the higher helpfulness scores a review receives, the more it will impact the box office numbers. More specifically this influence comes from the number of reviews a movie has and the length of its reviews. Factors that directly influence a review's helpfulness score tend to be the length and depth of the review, the review rating and the helpfulness for the reviewer (based on his past reviews). As can be observed, quite a bit has been done on understanding the role of review helpfulness and how reviews are perceived and analyzed by the customers reading them. Reviews may affect individuals' decision to watch a movie and thus it is helpful to understand how different factor may affect public perception and opinion of them.

Contrasting online discussions of pre-released movies with critical experts' reviews

From the discussion this far, it is possible to perceive that a lot has been done in the field of expert critics' reviews and their influence, but not as much has been researched about the effect of users' or public reviews, and especially not on how the reviews from the two groups compare to each other. The study by Chakravarty et al. (2010) has perhaps a more similar objective and focus to this research paper itself, but instead of looking at experts versus users' reviews, they use users' discussions, which are not as directly comparable. The study looks at pre-released movies and thus the general public does not yet have access to them. Researchers found that negative word-of-mouth – or in this case the online users' discussions – had a more significant impact on the infrequent moviegoers than on frequent ones. The influence of negative users' reviews is still bigger for these infrequent moviegoers than positive critic reviews on the same movie. Contrastingly however, frequent moviegoers tend to rely and be more influenced by professional critics' reviews than users'. It is interesting thus to observe that there is a clear difference in the effect reviews play on consumers' response depending on their pre-existing behavior pattern towards movies and movie theater attendance.

Analyzing the previous existing relevant literature, it becomes clear that a lot more should be studied in the sphere of contrasting users' and expert critics' reviews as these two may impact moviegoers in significantly different ways. Therefore, the identified empty space in the existing literature prompted the development of this research topic and question in order to better address, analyze and understand the different impact and influence of the different kinds of reviewers and reviews in the motion picture industry.

Theoretical Framework

As previously discussed, the development of the research question came from an identified gap in the field of study. The research question of this paper is **"*How do expert and public online reviews of movies differ across different dimensions?"*** This question will contribute to the field and provide an interesting and useful new perspective by contrasting two different kinds of reviews – users' and expert critics' – which are often consulted by the same audience. In terms of the dimensions, the users' and critics' reviews will be compared and contrasted against five different aspects: Length (number of words), Ease of Reading (what age group and education level is the review suitable for), Sentiment Analysis (what feelings does the review project), Emotional Tone (positive or negative responses), and Main Theme (what is the central topic of the review). Additionally, the users' reviews will also be analyzed based on their Publication Time Lag (how much time has elapsed between the movie release and the publication date of the review) to find out if there is a different in the kind of responses a movie receives depending on how long has elapsed since its launch. Examining the reviews across all of these dimensions will allow for the construction of a complete image of each review and will allow the identification of the most important characteristics of each review source (users or critics) after the individual results are aggregated.

Moreover, critics' reviews and influences have grown in interest as they serve a wide variety of roles. Basuroy's 2003 paper discusses Cameron's definition of professional critics where "critics provide advertising and information (e.g., reviews of new films, books, and music provide valuable information), create reputations (e.g., critics often spot rising stars), construct a consumption experience (e.g., reviews are fun to read by themselves), and influence preference (e.g., reviews may validate consumers' self-image or promote consumption based on snob appeal)" (Cameron 1995). Critics across industries serve many different functions and may have a considerable effect on their field of specialization. Looking precisely at the movie space, one of the most widely accepted definitions of professional film critics is "persons usually employed by newspapers, television stations or other media who screen newly released movies and provide their subjective views and comments on the movie for the public's information" (Cones 1992, 120). This definition will be used as the accepted characterization of professional movie critics in this research paper.

Significance of Research

This research project is especially noteworthy and relevant today as reviews play an increasingly important role on individuals' decision making and their extent and influence may often be overlooked. The timing of the study is also very appropriate as the field has seen considerable growth and still offers significant opportunities to be explored through further studies. With the spread of easily accessible internet wherever you go, finding the answer to almost anything can be done online with just a few clicks. It also means that people share more openly their experiences and we have seen a booming growth of social media, which has quickly become a central aspect of our daily lives. It is hard to find someone that does not have a profile on or utilize Facebook, Instagram, Yelp, or TripAdvisor – or often the three –, all of which provide direct access to reviews. These are just some of the numerous platforms that now exist to allow individuals to connect with each other and discuss what they have been up to. With the increasing popularity of Big Data and Machine Learning, people have also been looking more closely at the data that online reviews can provide us with. More and more researchers have begun to delve into the topic but there is still a lot to be done to map the true value of reviews. Furthermore, better understanding the role reviews play in the movie industry and how experts' versus users' reviews can impact the industry distinctively is of great value as it may shift future strategies, budgets and goals in the film space.

There is a large interest potential from a substantial portion of the population and diverse audiences for this research. As a growing field of study for text analysis and data interpretation, many researchers and scholars would find it valuable to learn about the how the two different kinds of reviews contrast. It also a topic of interest to the management teams of movie industry businesses as online reviews provide insightful and continuous information regarding how their brand/business is doing as well as assist on the prediction of future performance. Additionally, it is helpful in analyzing the competitive landscape and is a useful source of product improvement ideas. Movie directors and screenplay writers would be able to better understand what their customers care about and look for, thus being able to produce movies that directly cater to such interests. Similarly, movie studio owners and investors are interested in the financial results of movies and by having a better comprehension of what drives demand and income, they will be able to make better informed investment decisions and thus improve their margins and the profit potential of movies they produce or fund. Another group that will benefit from this study are the professional movie critics as they will be more aware of how their personal opinions are in line with that of the general public and this could help them focus their future reviews to become more informative and attract more public interest. Last but not least, this study matters to moviegoers and to the rest of the population interested in entertainment. Understanding how the reviews contrast would help people save time and make better informed decisions when choosing what movies to watch. Given how it impacts a varied set of industry players, movies would likely improve in the dimensions the public cares the most about and thus would lead to “better” movies from the audience’s perspective, benefitting the general population.

This particular paper was thus fomented from an exploration of different models for marketing strategy that led to an interest on how reading reviews can influence individuals’ decision making as well as understanding the growing influence this experience sharing/feedback giving approach has had and continues to have on our modern society. The amount of comprehension and forecast potential that can be gained from the interpretation of text in the form of reviews is impressive. It may shift future trends, markets and strategies and is thus of major importance. This research paper strives to make a significant contribution to the field of data analytics and provide valuable insight into the movie industry.

RESEARCH QUESTION

The research question of this study ***"How do expert and public online reviews of movies differ across different dimensions?"*** calls for a multivariate comparison between critics' and users' reviews. In order to construct a full image of the reviews of each group, the dimensions discussed below had to be analyzed for each individual review, then aggregated to build the complete set for each review and eventually combined with that of the other group to allow the results to be directly contrasted.

Dimensions:

Length

The length of the review is the word count of each review. Longer reviews tend to be more detailed and offer more in depth as well as more detailed discussions of the movie.

Ease of reading

The ease of reading rates the level of difficulty in comprehending the text of the review and fully understanding it. It is based on the Flesch Reading Score which identifies how understandable the text is and what level of education an individual should likely have to fully understand what they are reading. This analysis will be discussed more extensively in the Methodology section.

Sentiment Analysis

Sentiment analysis will discuss the polarity and irony of the review. In terms of polarity, the study will look at how positive or negative the reviewer's response towards the movies was. For the irony aspect, the reviews will be classified as ironic or non-ironic according to the words used by the reviewer and the structure of their sentences. This will also be detailed later in the paper.

Main Theme

The main theme analysis will classify the text of the review under one of the pre-existing categories based on the central focus of each review. This approach and the software used will be discussed under Methodology.

Emotional Tone

For the emotional tone of the review, the IBM Watson Tone Analyzer will be applied and the reviews will be evaluated based on the categories of Emotion, Language and Social. A more in-depth explanation of this analyzer is included in the Methodology section.

Publication Time Lag

In order to look at whether the passage of time affects the result of the reviews, users' reviews for which there are consistent publication dates will also be analyzed for changes over time. The study aims to draw parallels between changes in review behavior based on the time elapsed between movie launch and review publication.

HYPOTHESES

1. Critics' reviews are longer than users'

Critics in most cases write reviews professionally and therefore take the activity seriously. They tend to include many details and often discuss technical aspects, acting, directing, amongst other features. Users on the other had may just leave a short review unless they are passionate about the movie (positively or negatively), as they have less motivation to write an extensive review.

2. Critics' reviews are harder to read than users' reviews

Many critics likely write more formally than most of the general public who mostly post short reviews sharing their thoughts and opinions at the moment. As critics will many times have their own pages or work for a publication, they write extensive reviews that cover technical aspect and thus their language will often be more formal or at least harder to comprehend than the general user's as they may also make use of specialized terms and descriptions.

3. Critics' reviews are polarized towards neutral and positive reviews while users' reviews are polarized towards the two extremes

Critics may desire to be more careful about being very radical in one direction or another as their reviews will often have high visibility and posting something that is too far in one extreme may have repercussions. Additionally, given the fact that Avatar was a movie that received extensive recognition, especially for its technical aspects, critics will often agree with that praise.

As many critics have a tendency to discuss the technicalities of movies, a film that receive high technical acclaim will likely be one that critics view positively. Meanwhile, users will often be motivated to write reviews when they are very passionate about the movie, have radical views about it or feel very strongly about what they watched. This means that many of the users who post reviews are those who have extreme opinions about the movie. Furthermore, most users will not face any repercussions for posting an extreme review and thus there is nothing discouraging them from doing so.

4. **Critics' reviews are non-ironic while users' reviews are more frequently ironic**

Critics tend to take the review writing activity more seriously and therefore are less likely to post ironic reviews as their reviews are very public and part of their image. Ironic reviews could be seen as unprofessional. Users' review on the other hand reflect individuals' mindset and opinions and thus are more likely to ironic as the reviewer attempts to be funny or make fun of something.

5. **Critics' reviews have a bigger focus on technology while users' reviews center more around social issues and the environment**

As Avatar was a significantly groundbreaking movie for its technical innovations and application of cutting edge technology in the movie space, critics' will most likely discuss these factors and the new technology as a whole. Users on the other hand may be more interested in topics they have more extensive knowledge on or care more deeply about as for example social issues – especially facing the military, war and second chances as is the case in this movie – and the environment – the movie has a well-known "green" message aspect to it that may attract those that especially care about eco-friendly initiatives and developments.

6. Watson Tones:

a) Emotional Tone is split between joy and anger for both critics and users

Critics and users alike have been enthusiastic about watching Avatar in some cases and disappointed in others. Thus, those who enjoy the experience will likely write reviews that reflect that happiness and thus the review will have a joy tone. Conversely, those who go to the movie with high expectations but do not enjoy the movie will likely leave frustrated and annoyed, and therefore their review would likely have an anger tone.

b) Critics' reviews have a more analytical language tone while users' reviews present a more tentative language tone

As critics often write reviews as part of their work and carry detailed and extensive analyses, they are much more likely to use a more analytical language tone which makes sense given their purpose with the reviews. Users on the other hand may be uncertain about how they feel about the movie, may not know how to express themselves or not at ease with posting something publicly. This in turn would make them more likely to use a tentative term.

c) **Critics' and users' reviews both have an openness social tone, users' also exhibit an extraversion tone**

The openness tone reflects being open to new experiences and willing to try new things. Given that Avatar was a revolutionary movie, both critics and users who watched it were most likely open to having a new experience in the movie theatre. Similarly, their review probably included discussions of these new innovations which would further reflect the openness tone. The extraversion tone in the other hand reflects finding stimulation from the company of others an as users often reply, up vote/down vote each other's' reviews, they most likely reflect this extraversion tone as they benefit and enjoy the interaction with others.

7. **Early reviews will show more variation, extremes and a higher tendency to be positive with negative and neutral reviews appearing more extensively later on**

Early in the launched movie's lifespan, especially in the case of Avatar, there was a lot of hype surrounding the film and thus it would be more likely that those early reviews reflected it and were mostly positive. However, as time passes and the initial anticipation and excitement for the movie decreases or end, reviews become more neutral or negative as people start to pay more attention to the negative aspect of the movie, especially if they have already watched the movie before.

8. Ironic reviews will be more frequent earlier on

As early on emotions and passion will tend to have a bigger effect especially as the earliest viewers were mostly those that had anticipated for a long time the viewing of the movie. If they are disappointed with it they are more likely to be ironic and make ironic remarks about it. With the passage of time people may feel less strongly or passionately about the movie and thus not care as much to post ironic comments about it.

METHODOLOGY

Why IMDb

IMDb is globally recognized as the leading database on movies, television and celebrities. It has very extensive and detailed content making it the reference site for information on these industries. It has 250 million unique monthly visitors in the consumer end and more than 250 million data items. Most of the experts and researchers in the movie industry use IMDb as their leading source of information and thus, as this study focuses on movie reviews, which IMDb offers access to, the site is a great starting point for the research and data collection. The central focus of the study will be user and critic reviews and thus when deciding what database or data sources to use, the preference was to find one source that offered both and extensive samples as this would make the reviews more comparable across the dataset.

Why Avatar

Avatar directed by James Cameron and released in the United States on December 18, 2009, is a movie that completely revolutionized the movie industry. Not only did it take 12 years to produce, have a 237 million dollars budget, won 3 Oscars (85 other award wins and 128 nominations), and made a record-breaking 2.788 billion dollars in box office, but it also completely changed the special effects sphere of movies going forward. The graphics and 3D technology used for the movie were much more advanced than anything that had been seen until that time. Given these accomplishments and characteristics, Avatar is a movie that has been thoroughly discussed. Additionally, it is a movie that has sparked contrasting opinions and responses from both the general public and the professional critics community. Moreover, the movie was released over 8 years ago which has enabled it to accumulate a lot of information over time and allows for time series analyses that explore how time may have affected results. Avatar has very extensive information available on it, especially in terms of reviews from across the world and from the most diverse sources. With this, Avatar seemed like an appropriate movie to base this initial research to investigate how users' and critics' compare and contrast as there is a lot of available and interesting data on it

Users' Reviews: Python

Python

In order to analyze the movie reviews from the general user base, users' review text in IMDb was extracted using a crawler in Python. This resulted in 3,132 reviews being copied to Excel alongside with their respective numerical scores, helpfulness votes and totals, as well as the date the review was posted. Although most reviews had numerical and helpfulness scores, these ratings were not available for every user review and especially not for the critics' reviews, therefore these two dimensions were not included for further analyses or on the clean dataset moving forward. Python was chosen as the program to extract the reviews as it has been previously used in research for similar purposes and thus it was possible to adapt an existing crawler for this specific dataset. The use of crawler facilitated the data extraction and allowed a large number of reviews to be extracted much faster than doing so manually.

Randomization

Although there were 3,132 user reviews available in IMDb, there were only 720 critic reviews out of which inky 322 turned out to be usable. Thus, in order to make the analyses more comparable between the two datasets and make it possible to contrast them, it was necessary to have the same number of reviews for both. To ensure a representative, balanced and unbiased sample set of reviews was chosen from the reviews available, 322 reviews were chosen from the 3,132 available ones. Given that one of the multiple analyses this study aims to explore is whether reviews are significantly affected based in when they are published and how much time has elapsed since the movie was launched, it was necessary to select a random sample that was equally distributed across the time period. To do so, first the median date of the reviews was calculated and found to be January 15, 2010. Then, using this date as the central point for the pool of reviews, 161 reviews were selected from before and up to the median date and 161 reviews were selected from the median date and after. To select those two sets of reviews, the RAND function was used in Excel to assign the reviews a random number greater than or equal to 0 and less than 1 evenly distributed. Once the two sets had a random number assigned to each review, the lowest 161 numbers of each were selected in order to have a random sample of 322 user reviews spread across the review publication dates.

Critics' Reviews: Manual

To extract critics' reviews, first a similar methodology to the users' reviews through the application of a Python crawler was used, however it was found that this was ineffective as the critics' reviews with text available on the IMDb page itself were those from Metacritic which only showed a small portion of the full reviews, which was problematic as a big differentiator between users' and critics' reviews is their length. Critics' reviews tend to be significantly more extensive and thus having only part of the review would also affect the result of other analyses. Furthermore, there were only a total of 35 critic reviews available which represented a sample size that was too small to do significant analyses on.

In order to extract a larger dataset of critics' reviews, the 720 external critic reviews were used instead. As these were each links to outside pages, it became necessary to open each link and copy & paste each review individually into Excel. Although time consuming, this also allowed the issue of different formatting in each website to be overcome and ensured that the reviews that were selected were appropriate for this study. The process thus enabled some immediate data vetting. Not all websites had publication dates or scores and thus these were not included for further analyses. Reviews were cleared of spaces, links, advertisements and other characters that were not part of the review itself.

Data Vetting

Users

In order to ensure that the data was appropriately formatted for the analyses and thus that the results would not become influenced, especially for the Flesch Reading Score analysis, the reviews were cleared of symbols, spacing as well as "[ ]" which Python added to the reviews when they were extracted. Reviews were also corrected for accentuation that is not part of the US keyboard and other error markings with the body of the review text. This required manually going through the 322 users' reviews sample and checking for any anomalies or non-core review aspects.

Critics

Foreign Reviews:

Given that not all analyses could be run across different languages, the foreign reviews that were not written in English were not included in the consideration set. These were manually removed when the reviews were being pulled from IMDb. For a complete list of foreign reviews not included in the sample please see Appendix A.

DVD & Blu-Ray Reviews:

In order to be more consistent across reviews, the study focuses on Avatar reviews that are about the motion picture itself. Given that the movie had subsequent DVD and Blu-ray releases at two later dates, there was a considerable number of reviews that focused on the DVD and Blu-ray versions. This was especially significant for this movie as it stood out for its technical innovations and effects, which are not the same in the movie theater and home versions. There was a lot of discussion about the technical aspects and performance of the discs, however the focus of this study was the film and the release date was included in the analyses, therefore the sample selection was restricted to reviews on the feature itself.

To ensure that the reviews being analyzed were centered on the movies, the reviews that were mostly or entirely about the disc versions were simply not included. As these reviews were manually input into an Excel spreadsheet, excluding them happened when each site was opened and before they were imported. Some reviews however, had sections only on the movie and sections only on the DVD or Blu-ray and for those the reviews were manually adjusted to include only the movie part of it. Adjustments were only made in the cases where the movie itself was the central focus of the review and that any discussion of the discs came at the beginning or end of the review so as to minimize any effect on results. For a complete list of reviews not included or adjusted for the sample due to a focus on DVD and Blu-ray, please refer to Appendix A and B respectively.

Comparison to Other Movies Reviews:

Additionally, some critics' movie reviews were not solely on Avatar and instead discussed multiple movies. There were two kinds of these reviews: those that directly compared and contrasted Avatar to other films and those that discussed more than one movie but in different sections thus only focusing on one movie at a time. The first type of reviews were not included in the sample in order to restrict the reviews solely to the Avatar movie and make them more directly comparable across all reviews as response to other movies could impact the results if critics' had different opinions on those movies. An example of this is the review on the website "Critic After Dark" published on April 11, 2010 by critic Noel Vera that is entitled "Altar vs. Avatar". The entire review moves back and forth between the movies Altar and Avatar to assess the two side-by-side. For the latter type, the section that discussed solely the movie Avatar was adjusted and included in the sample. “The History of the Academy Awards: Best Picture – 2009” review by critic Erik Beck, for example, has a section on each of the nine movies that were nominated for the 2009 Best Picture Oscar, of which Avatar was a nominee. The Avatar section was thus included but the sections on the movies “Inglourious Bastards”, “A Serious Man”, “An Education”, “Up”, “Up in the Air”, “District 9”, “Precious”, and “The Blind Side”, were deleted from the dataset entry. For a complete list and details on exclusions and adjustments for reviews that included other movies, please refer to Appendix A and B.

Error Reviews

Furthermore, given that Avatar was a movie that came out in December 2009, over 8 years have elapsed since a significant portion of the reviews came out. What this means for the data collection is that multiple websites are no longer live as the domains have ceased to exist or have been sold directing the link to a different page. Others, although the link still directs to the correct website, the page administrators have updated their archives and removed the post. In these cases, although the reviews are still listed on IMDb’s external critics’ reviews page, these are no longer accessible and thus were not included in the sample.

Analyzing the Data

Length

To calculate the length of the reviews, a word count formula was used on Excel. Excel does not have a built-in word count formula and thus the formula below was used:

```
+IF(LEN(TRIM(A5))=0,0,LEN(TRIM(A5))-LEN(SUBSTITUTE(A5," ",""))+1)
```

where A5 is the cell with the review text.

Publication Time Lag

To calculate the Publication Time Lag between the Users' reviews and the and Avatar's launch date, a simple difference formula was used to find how much time elapsed between the movie's official launch date in the United States (December 18th, 2009) and the date each individual review was posted on IMDb.

Flesch Reading Score: Ease of Reading

About Flesch Reading Ease

The Flesch Reading Ease Formula was developed in 1948 by Rudolph Flesch, a writer and writing consultant who first published the formula as part of his article "A New Readability Yardstick" in the Journal of Applied Psychology. It is a simple approach that allows the ideal grade-level of the reader to be identified for an English language text. It has now been adopted for many different ends and is often used by government agencies in the United States.

The specific formula is the following:

$$\text{Readability Ease} = 206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$$

where ASL stands for Average Sentence Length (number of words) and ASW stands for Average number of Syllables per Word.

Although the formula is straightforward, in order to minimize mistakes and yield more reliable results, an online Flesch calculator was used: http://www.readabilityformulas.com/free-readability-formula-tests.php. Each review had to be copied and pasted individually into the website which although time consuming, ensured that the results were more consistent. Python allows for the calculation of the Flesch Reading Ease Score, and at first it was used to calculate the Flesch score, however it was soon found that the results obtained using Python and those in the online calculator were not consistent – likely due to formatting of the reviews when they were extracted from the IMDb database using the Python crawler.

Analyzing the Data

The Flesch Reading Ease Formula yields results between 0 and 100 and the higher the value the easier it is to read the text. The following table details exhibits the meaning of the possible results:

Score	Difficulty
90-100	Very Easy – easily understood by 5th grader
80-89	Easy
70-79	Fairly Easy
60-69	Standard – easily understood by 8th/9th graders
50-59	Fairly Difficult
30-49	Difficult
0-29	Very Confusing – easily understood by college graduates

Sentiment Analysis

About MeaningCloud Sentiment Analysis

Sentiment Analysis was done using an Excel add-in for Text Analytics called MeaningCloud. It determines if the text has a positive, neutral or negative sentiment; subjective or objective expressions; any irony characteristics; as well as if its messages and opinions agree or disagree with one another. It reaches these conclusions by analyzing individual phrases and then evaluating the relationship between them in the entire body of text. The phrase level analysis is output as Topic Sentiment Analysis and is aggregated to build the Global Sentiment Analysis. For the purpose of this study, the results of the Global analysis were used with a specific focus on polarity and irony.

Analyzing the Data

Polarity

The review is classified from positive to negative based on the polarity of each of the sentences in it. The following table shows the possible polarity values the review may be classified as:

Possible Values	Meaning
P+	Strong Positive
P	Positive
NEU	Neutral
N	Negative
N+	Strong Negative
NONE	No Sentiment

Irony

The review is analyzed for any irony remarks and whether these make the general review have an ironic tone to it or not. It can be classified as IRONIC (text includes ironic marks) or NONIRONIC (text does not include ironic marks).

Main Theme Classification

About MeaningCloud Text Classification

In order to identify the main theme under which each of the reviews should be classified, the MeaningCloud add-in was used again but this time for its Text Classification feature. Text Classification assigns a category from a previously established set to the text through the model the user chooses. Some of the most common ones are the International Press Telecommunication Council (IPTC) Subject Codes, EuroVoc, Business Reputation, IAB Taxonomy and Social Media. For the purpose of this study, given that the nature of the movie reviews, the Social Media model will be used.

Analyzing the Data

The Social Media model applies a simple taxonomy to classify social media into seventeen possible categories. Each review was thus classified into one of these based on its main discussion topic as shown below:

Code	Description
01	art and culture
02	crime, law and justice
03	disaster and accident
04	economy and finances
05	education
06	environment, weather and energy
07	health
08	social issue
09	labor
10	tourism, travel and commuting
11	lifestyle and leisure
12	politics
13	religion and belief
14	science and technology
15	sport
16	unrest, conflicts and war
17	greetings and thanks

Watson Tone Analyzer

About Watson Tone Analyzer

IBM offers a service called the IBM Watson Tone Analyzer that detects the emotional and language tones in written text. It allows the user to understand how their text is being perceived by others and in the case of this paper is used to detect the tone of the movie reviews.

Analyzing the Data

The text was analyzed and classified under three types of tones: Emotional, Language and Social. Each review receives a score between 0 and 1 across all the tones and for the tones that a review had a score below 0.5 these were omitted as they were unlikely to be perceived and the tones that scored above 0.75 were considered as having a high chance of being identified in the text. Below there are descriptions of the categories within which the review text can fall under.

Emotional Tones[1]

Tone	Description
anger	Anger is evoked due to injustice, conflict, humiliation, negligence, or betrayal. If anger is active, the individual attacks the target, verbally or physically. If anger is passive, the person silently sulks and feels tension and hostility.
fear	Fear is a response to impending danger. It is a survival mechanism that is triggered as a reaction to some negative stimulus. Fear can be a mild caution or an extreme phobia.
joy	Joy (or happiness) has shades of enjoyment, satisfaction, and pleasure. Joy brings a sense of well-being, inner peace, love, safety, and contentment.
sadness	Sadness indicates a feeling of loss and disadvantage. When a person is quiet, less energetic, and withdrawn, it can be inferred that they feel sadness.

[1] Retrieved from IBM Watson Tone Analyzer documentation. Available at https://console.bluemix.net/docs/services/tone-analyzer/using-tone.html#using-the-general-purpose-endpoint.

Language Tones[1]

Tone	Description
analytical	An analytical tone indicates a person's reasoning and analytical attitude about things. An analytical person might be perceived as intellectual, rational, systematic, emotionless, or impersonal. (
confident	A confident tone indicates a person's degree of certainty. A confident person might be perceived as assured, collected, hopeful, or egotistical.
tentative	A tentative tone indicates a person's degree of inhibition. A tentative person might be perceived as questionable, doubtful, or debatable.

Social Tones[2]

Tone	Description
openness	Openness is the extent to which the presented text demonstrates openness to experience a variety of activities.
conscientiousness	Conscientiousness is a tendency to act in an organized or thoughtful way as expressed in the input text.
extraversion	Extraversion is a tendency to seek stimulation in the company of others.
agreeableness	Agreeableness is a tendency, expressed in writing, to be compassionate and cooperative towards others.

[2] Retrieved from IBM Watson Tone Analyzer developer blog. Available at https://www.ibm.com/blogs/watson/2016/02/293/.

DATA & RESULTS

Summary Statistics

In the following histograms, it will be possible to visualize the summary statistics for the critics' and the users' review. Critics reviews' means are in blue and users' in orange. Standard deviations are in the form of yellow dots. F-Tests were conducted to test the variance between the means of the critics' and users' reviews in order to then conduct the appropriate t-Tests to analyze the significance of the results obtained. All of these statistics were obtained using Microsoft Excel.

Additionally, all analyses that reflect the movie launch date assume **December 18th, 2009** as the movie release date as that was the official date for the release of the Avatar movie in theatres in the United States. Furthermore, the median date of the reviews is January 15, 2010, thus early reviews are those between the release date and the median while late reviews are those post the median date.

Length

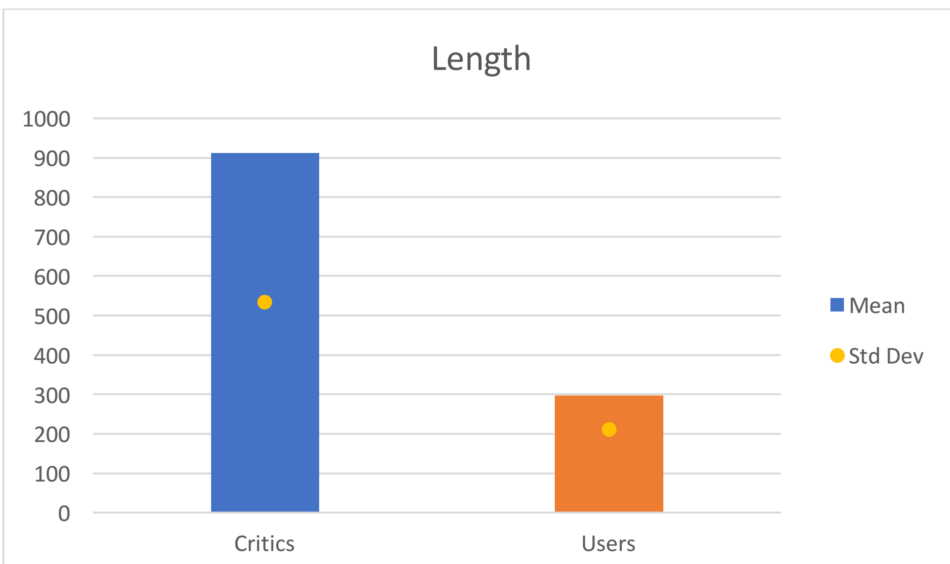
F-Test Two-Sample for Variances

Word Count	*Critics*	*Users*
Mean	911.6614907	296.5838509
Variance	284959.7199	44560.86055
Observations	322	322
df	321	321
F	6.394843287	
P(F<=f) one-tail	3.33598E-55	
F Critical one-tail	1.201878565	

t-Test: Two-Sample Assuming Unequal Variances

Word Count	*Critics*	*Users*
Mean	911.6614907	296.5838509
Variance	284959.7199	44560.86055
Observations	322	322
Hypothesized Mean Difference	0	
df	419	
t Stat	19.22722476	
P(T<=t) one-tail	8.04119E-60	
t Critical one-tail	1.64849841	
P(T<=t) two-tail	1.60824E-59	
t Critical two-tail	1.965641842	

The t-Test above shows that the difference between the length of the reviews by critics and by users is significant and thus the null hypothesis can be rejected.



From the graph above it is possible to perceive that the critics' reviews are on average about three times the length of the users' reviews. This is a significant difference and is in line with expectations given the nature and common motivations behind the two types of reviews. These results prove hypothesis 1: ***Critics' reviews are longer than users'***.

Ease of reading

F-Test Two-Sample for Variances

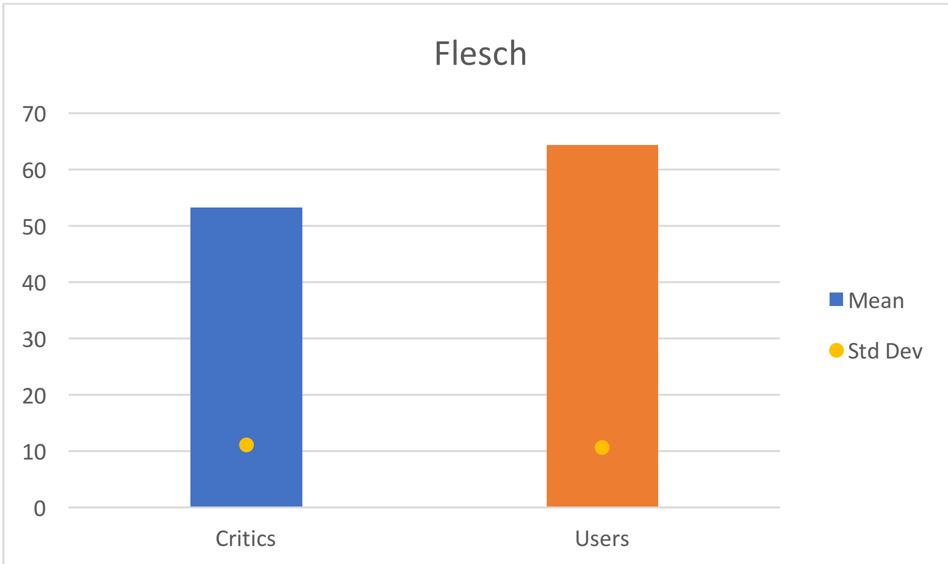
Flesch	*Critics*	*Users*
Mean	53.2784375	64.41785714
Variance	125.2501606	114.634957
Observations	320	308
df	319	307
F	1.092600057	
P(F<=f) one-tail	0.217238322	
F Critical one-tail	1.20497056	

t-Test: Two-Sample Assuming Unequal Variances

Flesch	*Critics*	*Users*
Mean	53.2784375	64.41785714
Variance	125.2501606	114.634957
Observations	320	308
Hypothesized Mean Difference	0	
df	626	
t Stat	-12.74764984	
P(T<=t) one-tail	1.48316E-33	
t Critical one-tail	1.647291391	
P(T<=t) two-tail	2.96631E-33	
t Critical two-tail	1.963760768	

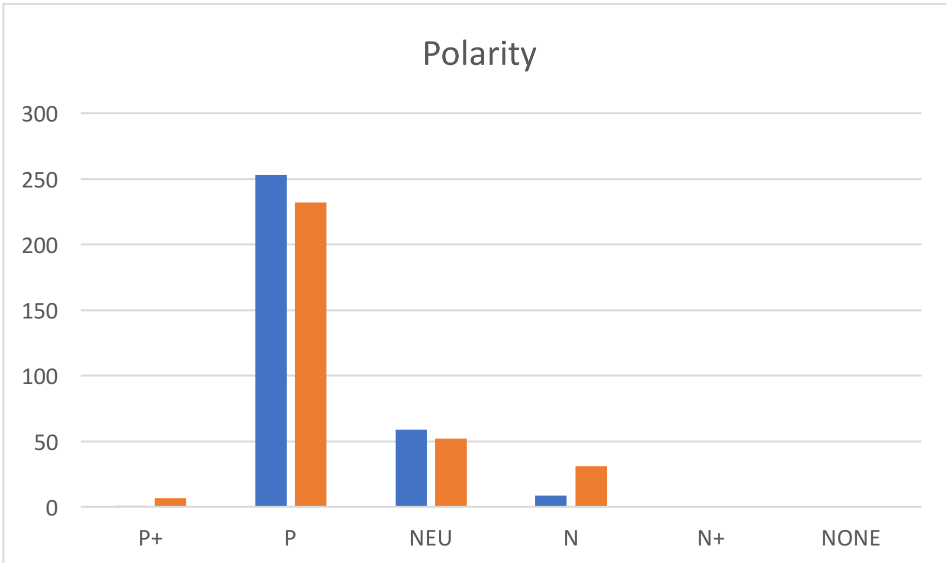
In order to be able to run the Flesch Reading Ease formula, reviews have to be at least 100 words long. The sample used for this study had two reviews within the critics' population and fourteen within users' that did not meet the minimum criteria and thus could not be analyzed using the Flesch Reading Ease scores. This explains why the number of observations for critics and users are 320 and 308 respectively, instead of 322 of each which is the full sample size being used throughout the study.

Nevertheless, the t-Test shows that the difference in reading ease between critics' and users' reviews is significant and thus the null hypothesis is rejected.

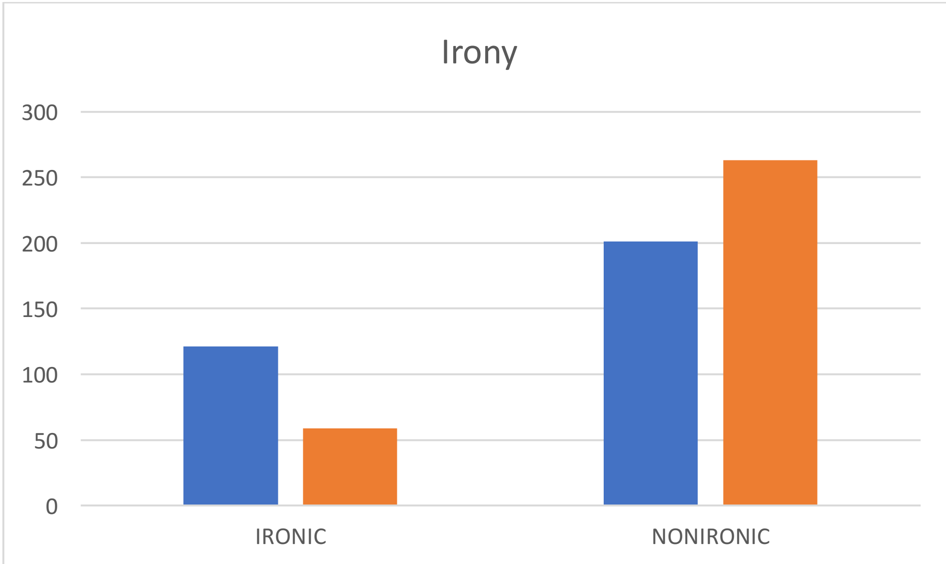


The higher the Flesch score, the easiest reviews are to read and thus the users' reviews are easier to read than critics'. Looking at the graph above it is possible to see that although the difference between the two sources of reviews was proved to be significant, this difference is not as sizable as had originally been anticipated. The results do however confirm hypothesis 2: ***Critics' reviews are harder to read than users' reviews***.

Sentiment Analysis

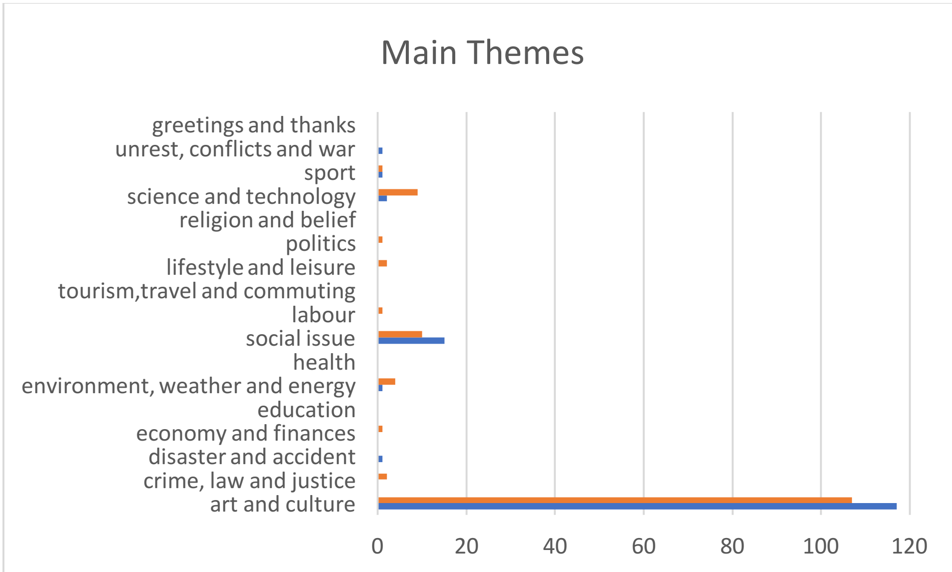


The graph above shows that critics' and users' reviews have quite similar polarity. Critics are more optimistic than users' as they have in general left more positive reviews. However, the numbers are close and thus we find that users' and critics' opinions are quite similar from this perspective. Given that Avatar received considerable recognition and awards, especially from technical aspects like special effects, was expected that critics' reviews to be more neutral and positive, which is mostly the case. However, there was also an expectation that users' reviews would be oriented more towards the two extremes, which apart from a couple of reviews classified as P+, was not the case. These results thus only prove hypothesis 3 partially: **Critics' reviews are polarized towards neutral and positive reviews**. However, it rejects the second half as **users' reviews are NOT polarized towards the two extremes**.



The graph above shows that critics' reviews tend to be more ironic than users' reviews which is contrary to expectations that critics' reviews would be more non-ironic and users' more ironic. These results thus reject hypothesis 4 as **Critics' reviews are IRONIC while users' reviews are more frequently NON-IRONIC.**

Main Theme



The graph above shows that the main themes present in the critics' and users' reviews are similar. It is not surprising that most reviews fell under the "art and culture" category given that all the texts analyzed were movie reviews, however there was the expectation of a more significant presence under the "science and technology" category given the movie storyline, science-fiction genre and acclaim for technical innovation. This is especially true for the critics' reviews as it was anticipated that they would focus their discussions more on this theme for its breakthrough technology. Additionally, there was also the expectation of a bigger presence of user reviews under the theme of "environment, weather and energy" as the movie has a "green" message as well as under "social issues" as it addresses the topics of war and overcoming adversities. Therefore, the results reject hypothesis 5 as ***Critics' reviews DO NOT have a bigger focus on technology while users' reviews DO NOT center more around social issues and the environment.***

Watson Tone Analyzer

Emotional Tone

F-Test Two-Sample for Variances

Watson - Emotional	*Critics*	*Users*
Mean	0.623198516	0.605815292
Variance	0.0026954	0.01774881
Observations	322	322
df	321	321
F	0.151863676	
P(F<=f) one-tail	0	
F Critical one-tail	0.832030813	

t-Test: Two-Sample Assuming Unequal Variances

Watson - Emotional	*Critics*	*Users*
Mean	0.623198516	0.605815292
Variance	0.0026954	0.01774881
Observations	322	322
Hypothesized Mean Difference	0	
df	416	
t Stat	2.181589755	
P(T<=t) one-tail	0.014849462	
t Critical one-tail	1.648524754	
P(T<=t) two-tail	0.029698925	
t Critical two-tail	1.965682905	

Looking at the results of the t-Test above, it is possible to conclude that the difference between the dominant emotional tone in the critics' and users' reviews is significant. It compares the most prominent emotional tone which had the highest score in each review between critics and users.

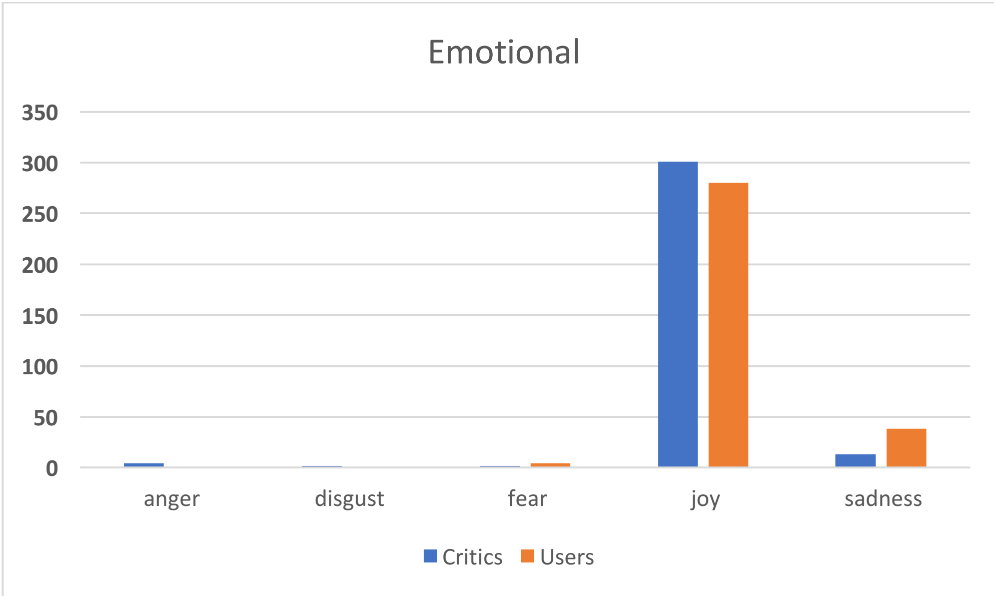
F-Test Two-Sample for Variances

Emotional - Joy	*Critics*	*Users*
Mean	0.62601088	0.619290764
Variance	0.002009376	0.014464057
Observations	301	280
df	300	279
F	0.138922035	
P(F<=f) one-tail	0	
F Critical one-tail	0.824155134	

t-Test: Two-Sample Assuming Unequal Variances

Emotional - Joy	*Critics*	*Users*
Mean	0.62601088	0.619290764
Variance	0.002009376	0.014464057
Observations	301	280
Hypothesized Mean Difference	0	
df	350	
t Stat	0.879872131	
P(T<=t) one-tail	0.189766152	
t Critical one-tail	1.64921887	
P(T<=t) two-tail	0.379532304	
t Critical two-tail	1.966765003	

The t-test above shows that although the difference between the dominant emotional tones of critics and users was significant, the difference for the joy emotional tone between the two is not. Joy was the dominant emotional tone for both critics and users, however the difference in its prominence between the two is not large enough to be significant. Therefore, although we may say that the reviews by critics and users have significantly different dominant emotional tones, we cannot conclude that the level of "joy" in their reviews significantly differs.



The graph shows that the most prevalent emotional tone amongst all of the critics' and users' is joy. There is a small presence in other emotional tones and this is especially visible for the users under the sadness tone. The results thus reject hypothesis 6a as **Emotional Tone is NOT split between joy and anger for both critics and users** as by a considerable majority, both critics and users exhibit almost entirely solely "joy" as the main emotional tone.

Language Tone

F-Test Two-Sample for Variances

Watson – Language	*Critics*	*Users*
Mean	0.594809689	0.57252782
Variance	0.021961833	0.057223608
Observations	322	322
df	321	321
F	0.383789725	
P(F<=f) one-tail	0	
F Critical one-tail	0.832030813	

t-Test: Two-Sample Assuming Unequal
Variances

Watson – Language	*Critics*	*Users*
Mean	0.594809689	0.57252782
Variance	0.021961833	0.057223608
Observations	322	322
Hypothesized Mean Difference	0	
df	536	
t Stat	1.420878345	
P(T<=t) one-tail	0.077966804	
t Critical one-tail	1.64770143	
P(T<=t) two-tail	0.155933607	
t Critical two-tail	1.964399705	

From the t-Test above, the null hypothesis cannot be rejected and there is not a significant difference between the language tone of the critics' and users' reviews.

F-Test Two-Sample for Variances

Language – Analytical	*Critics*	*Users*
Mean	0.584844909	0.523759123
Variance	0.020170285	0.067463646
Observations	198	73
df	197	72
F	0.298980059	
P(F<=f) one-tail	1.39877E-11	
F Critical one-tail	0.734831554	

t-Test: Two-Sample Assuming Unequal
Variances

Language – Analytical	*Critics*	*Users*
Mean	0.584844909	0.523759123
Variance	0.020170285	0.067463646
Observations	198	73
Hypothesized Mean Difference	0	
df	88	
t Stat	1.907041779	
P(T<=t) one-tail	0.029888788	
t Critical one-tail	1.662354029	
P(T<=t) two-tail	0.059777576	
t Critical two-tail	1.987289865	

This t-test shows that the difference in analytical language tone is not significant between users. The t Stat is close to the upper limit of the t critical two-tail but falls short of actually surpassing it to make the difference in analytical language tone significant. It has high prominence for both critics and users but is only dominant for critics as will be shown in a following histogram. However, the difference when comparing its presence level in both kinds of reviews is not large enough to be significant and thus it is not possible to conclude that the analytical language tone is significantly different between critics' and users' reviews.

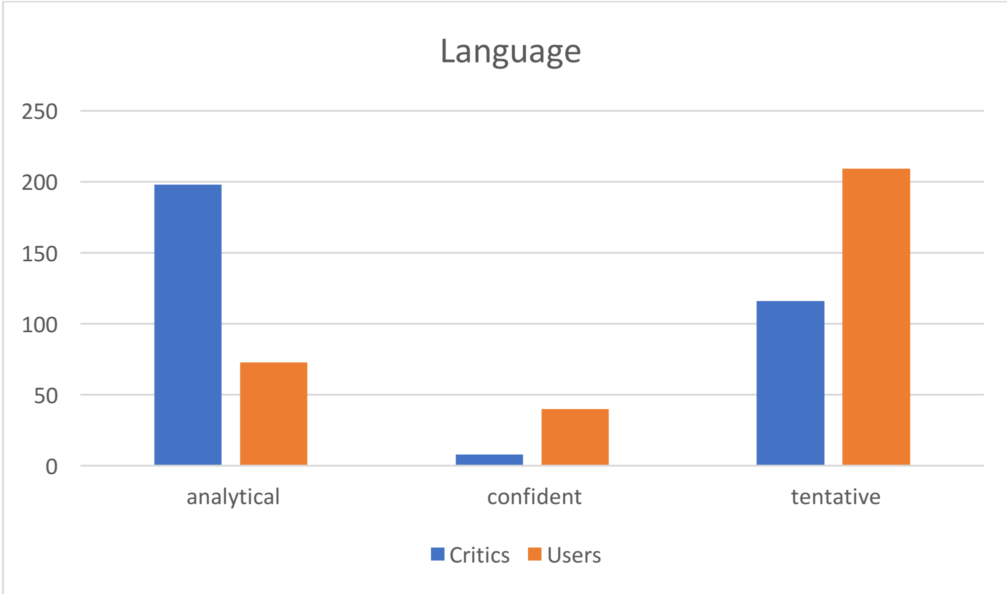
F-Test Two-Sample for Variances

Language – Tentative	*Critics*	*Users*
Mean	0.609813664	0.586812268
Variance	0.024504667	0.057316091
Observations	116	209
df	115	208
F	0.427535561	
P(F<=f) one-tail	5.16116E-07	
F Critical one-tail	0.757773138	

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	0.609813664	0.586812268
Variance	0.024504667	0.057316091
Observations	116	209
Hypothesized Mean Difference	0	
df	314	
t Stat	1.043915817	
P(T<=t) one-tail	0.148663672	
t Critical one-tail	1.649720831	
P(T<=t) two-tail	0.297327345	
t Critical two-tail	1.967547698	

Similarly, the tentative language tone t-test also shows that the difference between the analytical language tone between critics' and users' reviews is not significant. In this case the tentative tone was dominant only for users but the fact that it had a high level of presence in critics' reviews as well made the difference between the two not be significant.



Even though in the graph above there appears to be clear differences between the language tone of critics and users as the first exhibits preference for an analytical language tone while the latter falls mostly under the tentative language tone, the t-Test does not prove the difference is significant. The Language tone graph does support hypothesis 6b: **Critics' reviews have a more analytical language tone while users' reviews present a more tentative language tone**.

Social Tone

F-Test Two-Sample for Variances

Watson - Social	*Critics*	*Users*
Mean	0.881799118	0.840044255
Variance	0.004908466	0.014475552
Observations	322	322
df	321	321
F	0.339086618	
P(F<=f) one-tail	0	
F Critical one-tail	0.832030813	

t-Test: Two-Sample Assuming Unequal Variances

Watson - Social	*Critics*	*Users*
Mean	0.881799118	0.840044255
Variance	0.004908466	0.014475552
Observations	322	322
Hypothesized Mean Difference	0	
df	516	
t Stat	5.381620859	
P(T<=t) one-tail	5.61024E-08	
t Critical one-tail	1.647812009	
P(T<=t) two-tail	1.12205E-07	
t Critical two-tail	1.964572029	

The t-Test above proves that the difference between the social tones for critics and users is significant and thus the null hypothesis can be rejected.

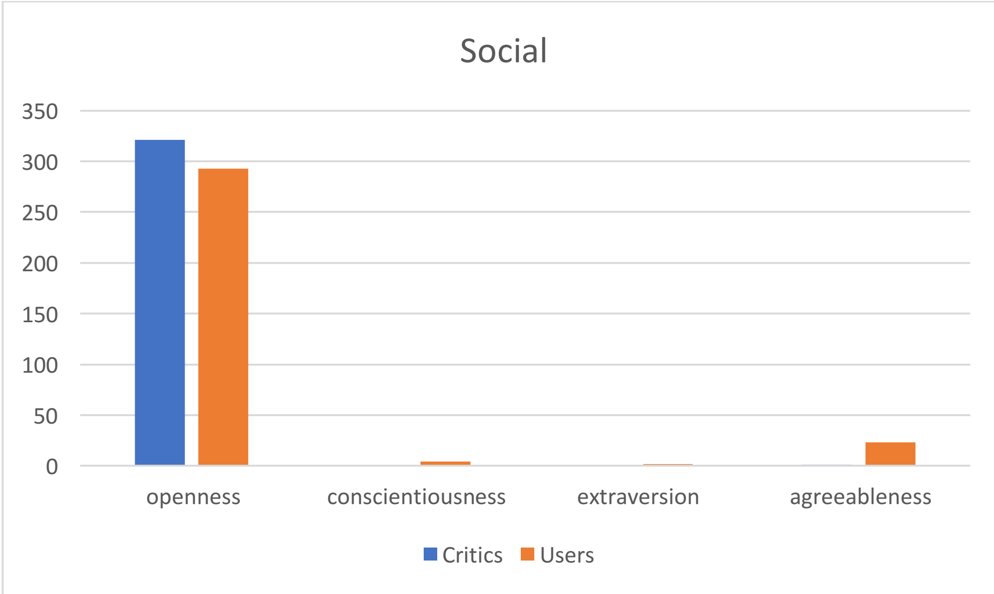
F-Test Two-Sample for Variances

Social - Openness	*Critics*	*Users*
Mean	0.882451442	0.851374478
Variance	0.004786357	0.012330959
Observations	321	293
df	320	292
F	0.388157743	
P(F<=f) one-tail	2.22045E-16	
F Critical one-tail	0.828548581	

t-Test: Two-Sample Assuming Unequal Variances

Social - Openness	*Critics*	*Users*
Mean	0.882451442	0.851374478
Variance	0.004786357	0.012330959
Observations	321	293
Hypothesized Mean Difference	0	
df	481	
t Stat	4.116390648	
P(T<=t) one-tail	2.26442E-05	
t Critical one-tail	1.648027693	
P(T<=t) two-tail	4.52884E-05	
t Critical two-tail	1.964908164	

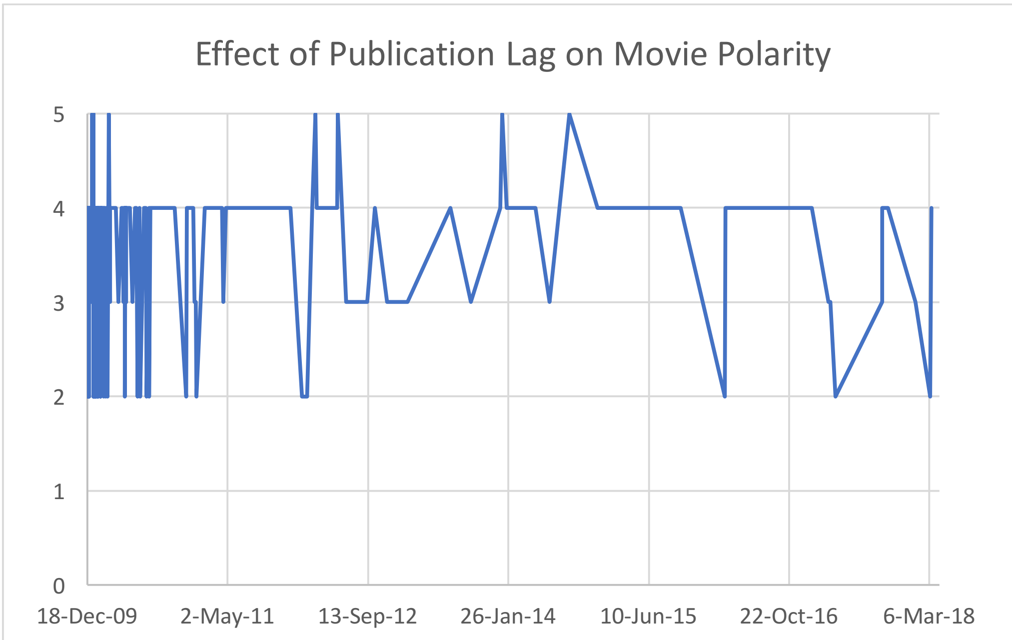
In the case of openness social tone, the difference between the level of openness in critics' and users' reviews is significant. This means that not only is the dominant social tone significantly different, but the difference between the two kinds of reviews within the overall dominant tone for social tones is also significant. This further highlights how the two review groups differ.



The social tone graph shows that critics and users again adopt a very similar tone in their reviews – both mainly exhibit openness. The t-Test shows that the difference between the reviewers is significant, but visually both reviewer groups appear to behave similarly. The social tone graph thus supports hypothesis 6c partially: **Critics' and users' reviews both have an openness social tone**, however it appears that **users' DO NOT exhibit an extraversion tone**. As the difference between the two groups is significant however, critics and users do not behave the same in terms of social tone.

Publication Time Lag

Polarity



5 = P+: strong positive

4 = P: positive

3 = NEU: neutral

2 = N: negative

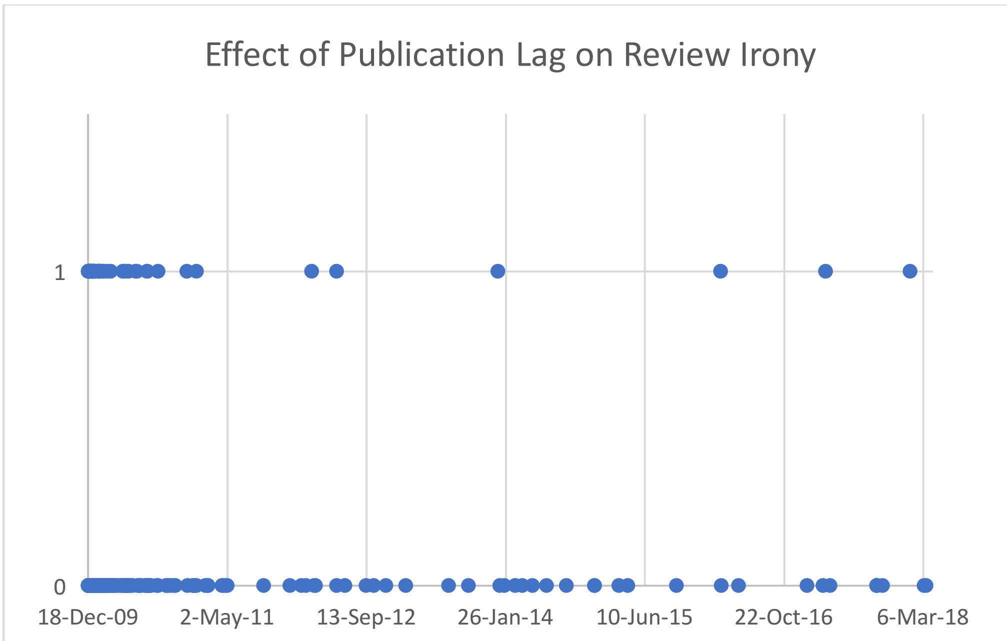
1 = N+: strong negative

0 = Neutral

The graph shows that the earliest reviews closest to the movie release on December 18, 2009 had more variation in polarity. Most of the negative reviews also appear to be concentrated in the first year of the movie release. Extremes in polarity were expected early on, however the fact that most of the negative reviews are present on the earlier end, contradicts part of hypothesis 7, thus **early reviews will show more variation and extremes, BUT not necessarily a higher**

tendency to be positive early on. Although **later reviews seem to be centered around neutrality, negative reviews DO NOT appear more extensively later on.**

Irony



1 = Ironic

0 = Non-ironic

The graph above shows that for the early reviews there were as many ironic as non-ironic reviews. Overall, substantially more non-ironic than ironic reviews which helps explains the very small number of ironic reviews after the first year following the movie release. The fact that ironic and non-ironic reviews are nearly balanced between the release date and the median reviews date rejects hypothesis 8: **Ironic reviews will NOT be more frequent earlier on.**

CHALLENGES

Data Scrapping

Perhaps the most challenging and time-consuming aspect of this study was collecting the data. Data scrapping was especially hard due to the fact that it was not as simple and direct to use codes and crawlers to extract the data. Not coming from a computer science background as well as the fact that IMDb is not a very crawler-friendly database made automatic data extraction hard and in multiple occasions manually copying and pasting data proved to be more reliable. For the critics' reviews on external sites linked from IMDb, the only option truly was going through all 720 links one by one. Additionally, the fact that reviews are formatted differently on every website required reformatting before analyses could be run and checking every row on Excel individually.

Limitations

This study was carried out on movie reviews of only one movie. Ideally some, if not most, of the findings from paper can be replicated for other movies and used as the general trend of the movie industry, however there has been no study done on its replicability and no other movies have been analyzed using the same methodology in order to compare and draw parallels. Avatar is a very unique movie due to a range of different factors that include but are not limited to being directed by one of the best well-known and acclaimed directors in the world, receiving various

Oscar nominations and awards, being recognized as a special effects and technological innovations breakthrough in the movie industry, having an extremely large budget, and taking 12 years to complete its production. To what extent these truly affect the results is hard to tell without running the same analyses for different movies. It would be interesting to see this study attempted to be replicated. Perhaps it could even be applicable outside of the movie industry with reviews on products or restaurants.

DISCUSSION & IMPLICATIONS

Significance

The significance of this study is understanding that critics' and users' reviews will differ in many of the dimensions they may discuss. Both offer extremely valuable insights and thus it is valuable to, as a movie consumer, to realize what is most important to you and what you care the most about. The two kinds of reviews will be able to offer more valuable insights in specific areas. For example, as the study showed, user reviews have an easier to understand language, at the same time critics' reviews are surprisingly more ironic than users' but are much more extensive and detailed. Nevertheless, the two sources of review are also similar in various aspects that sometimes would not be as expected, as for example the main themes and tones of the reviews.

This study thus hopes to better inform customers and professionals of the movie industry about how to interpret the different kinds of reviews. Very little has been studied about how professional and general public reviews compare and thus the research present in this paper may serve to bridge some of the existing gap and initiate further discussions on the topic which may prove beneficial for the entire industry.

Looking at the broader field of analytics, text is also an important medium that may not always receive as much critical attention as it should, given the extent of information that can be extracted from just a few lines of text. Text is an extremely valuable information source which can often be overlooked and thus this study is also significant in the sense that it highlights the importance and potential of text analytics.

Looking Ahead

Likely the most obvious next step in this research would be, as previously mentioned, test for the replicability of this study. How successful this will turn out to be is hard to forecast for the reasons previously mentioned, however if replicating this analysis to other movies or even diverse industries proves to be successful, this could improve forecasting and information sharing. Therefore, future research could be done on how to adapt this study to make its replicability possible. Additionally, further tests could be run to extend even further the understanding of how critics and users differ in opinion and how they present their views. It was interesting to find that a number of the original hypotheses once tested turned out not to be true and were instead rejected or partially rejected. There is no question that this is a growing field with a lot of potential for future research.

REFERENCES

Asur, Sitaram, e Bernardo A. Huberman. 2010. “Predicting the Future with Social Media.” *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.* Toronto, ON, Canada: IEEE. 492-499.

Basuroy, Suman, Subimal Chatterjee, e S. Abraham Ravid. 2003. “How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budg.” *Journal of Marketing* (American Marketing Association) 67 (4): 103-117.

Cameron, S. 1995. “On the Role of Critics in the Culture Industry.” *Journal of Cultural Economics* 19 (4): 321-331.

Chakravarty, Anindita, Yong Liu, e Tridib Mazumdar. 2010. “The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation.” *Journal of Interactive Marketing* 24 (3): 185-197.

Chen, Hua-Ning, e Chun-Yao Huang. 2013. “An investigation into online reviewers' behavior.” *European Journal of Marketing* 47 (10): 1773-1758.

Cones, John W. 1992. *Film Finance and Distribution: a Dictionary of Terms.* Silman-James Press.

Dodds, John C., e Morris B. Holbrook. 1988. “What’s an Oscar worth? An empirical estimation of the effects of nominations and awards on movie distribution and revenues.” *Current research in film: Audiences, economics and the law* 4: 72-87.

Eliashberg, Jehoshua, e Steven M Shugan. 1997. “Film critics: Influencers or predictors?” *Journal of Marketing* 61 (2): 68-78.

Ginsburgh, Victor, Fernanda Gutierrez-Navratil, e Juan Prieto-Rodriguez. 2014. “The Impact of Oscar Nominations and Wins on Box Office Revenues.” *ACEI 2014: The 18th International Conference on Cultural Economics.* MOntreal (Quebec), Canada: Association for Cultural Economics International.

Jostein, Gripsrud. 1995. *The Dynasty Years: Hollywood Television and Critical Media Studies.* London and New York: Routledge.

Lee, Sangjae, e Joon Yeon Choeh. 2018. “The interactive impact of online word-of-mouth and review helpfulness on box office revenue.” *Management Decision* 56 (4): 849-866.

Lee, Youseok, Kyong Cheon Cha, e Sang-Hoon Kim. 2016. “Internet Search Behavior and Box Office Performance.” *Korean Management Review* (Korean Academic Society Of Business Administration) 45 (5): 1501-1526.

Terry, Neil, Michael Butler, e De'Arno De'Armond. 2005. “The Determinants of Domestic Box Office Performance in the Motion Picture Industry.” *Southwestern Economic Review* 32 (1): 137-148.

s.d. *The Internet Movie Database (IMDb).* Accessed on May de 2018. https://www.imdb.com/title/tt0499549/.

APPENDIX

A. Reviews not included in Critics' reviews sample

Foreign Reviews:

2inthesoup (Greek)
Abus de ciné [François Rey] (French)
allesfilm.com [Reinhard Bradatsch] (German)
Alligotographe (French)
AllOfCinema [Evgeniy Nefyodov] (Russian)
Antepenultimo Mohicano [Emilio Luna] (Spanish)
Artecock [Rüdiger Suchsland] (German)
aVoir-aLire.com [Rüdiger Suchsland] (German)
Bakiniz.com [Suat Demirel] (Turkish)
Balbi.de - Micha's Filmboulevard (German)
Bestiarium [Diego Cabeza] (Spanish)
Betomovies (Spanish)
Bilimkurgukulubu.com (Turkish)
Boulevard do Crepúsculo [Renato Félix] (Portuguese)
C7nema.net [José Pedro Lopes] (Portuguese)
Cadependdesjours.com [Guilhem] (French)
callia.lt [Aleksandras T.] (Lithuanian)
CanalTCM.com [Sergi Sánchez] (Spanish)
CaPo [CaPo] (German)
Captain Charismas Filmblogg [Tom Kleppe] (Norwegian)
Cenas de Cinema [Cecilia Barroso] (Portuguese)
CervenyKoberec.cz [Eliska Bartlova] (Czech)
Christoph Hartung - Die besten Filmkritiken im Netz (German)
Cine Adanruiz10 [Adan Ruiz] (Spanish)
Cine para leer (Spanish)
Cine Planeta [Adan Ruiz] (Spanish)
Cine Players [Silvio Pilau] (Portuguese)
Cine y Letras. Revista de Cultura [Guzmán Urrero Peña] (Spanish)
Cine y Letras. Revista de Cultura [Guzmán Urrero Peña] (Spanish)
Cine.gr [Stavros Ganotis] (Greek)
Cineclub.de [Martin Wolkner] (German)
Cinefacts (German)
CineFile.biz [Alberto Cassani] (Italian)
Cinefreaks (German)
Cinema 2000 [João Lopes] (Portuguese)
Cinema em Cena [by Pablo Villaça] (Brazilian Portuguese)
Cinemafantastique [Damien] (French)
Cinemagazine [Wouter de Boer] (Dutch)
CinemaZone.dk (Danish)
Cinemovie.info [Marco Michele] (Italian)
Cinerama.no [Tor Arve Røssland] (Norwegian)
Cineycine.com [Alex Moraño] (Spanish)
Cineycine.com [Alex Moraño] (Spanish)
ComingSoon.it [Federico Gironi] (Italian)
Confraria de Cinema [Lucas Salgado] (Portuguese)
critic.de (German)
Criticos.com.br [Marcelo Janot] (Portuguese)
CriticsCinema [Jcr] (Spanish)
De Ultieme FilmBlog (Dutch)
Die Furche (German)
Die-besten-Horrorfilme.de [Marcus Littwin] (German)
digitalvd.de [Frank Brenner] (German)
Dipticos [Dario Lavia] (Spanish)
Dutch Movie Reviews [Basilios Mulder] (Dutch)
DVD-Headquarters.de [Mike Flinzner] (German)
DVD-opas.fi [Hannu Bjorkbacka] (Finnish)
e-media.ch (French)
Ekran [Janez Strehovec] (Slovenian)
El Bloc de Josep (Spanish)
El Criticón (Spanish)
El rincón de Carlos del Río [Carlos del Río] (Spanish)
Elitisti [Ilja Rautsi] (Finnish)
epd Film [Jan Distelmeyer] (German)
Escribiendo de cine [Mariana Mijares] (Spanish)
Expanded Cinemah [Roberto Matteucci] (Italian)
F.LM - Texte zum Film [Jochen Werner] (German)
Fanatisk Film [Tommy Söderberg] (Swedish)
film-zeit.de [Pressespiegel] (German)
Filmelskeren.no (Norwegian/Norsk)
Filmering.at [Michael Föls] (German)
Filmfenix.se [Pär Wirdfors] (Swedish)
Filmflash.nl [Evert van de Grift] (Dutch)
filmfuchs.de (German)
Filmkritik "Avatar - Aufbruch nach Pandora" (German)
Filmkritik (Swedish)
Filmliefhebber [Paul Hauer] (Dutch)
filmmusicjournal.com [Phil Blumenthal] (German)
Filmofiel.nl (Dutch)
Filmovie.it [Francesco Mangiò] (Italian)

Filmreporter.de [Andrea Niederfriniger] (German)
filmrezension.de [Tobias Vetter] (German)
filmsfantastiques.com (French)
Filmski Blog [H.Trobradovic] (Bosnian)
Filmstarts.de [Carsten Baumgardt] (German)
FiLmSTOP.NeT [Ozan Kanik] (Turkish)
FilmTotaal.nl [Thomas Hermssen] (Dutch)
filmzentrale [J. Distelmeyer] (German)
Franglaisreview [Monsieur D and Miss J] (French, English)
Fred Burle no Cinema.br [Fred Burle] (Brazil)
Freequency [Gabriele Guerra] (Italian)
GamersGlobal (German)
gutefilme.info (German)
Hartigan's World (German)
Hello Friki [M.C. Catalán] (Spanish)
Hideout.it [Antiniska Pozzi] (Italian)
gutefilme.info (German)
Hartigan's World (German)
Hello Friki [M.C. Catalán] (Spanish)
Hideout.it [Antiniska Pozzi] (Italian)
gutefilme.info (German)
Hartigan's World (German)
Hello Friki [M.C. Catalán] (Spanish)
Hideout.it [Antiniska Pozzi] (Italian)
Il Cancello (Italian)
Il cinema secondo me [Il cinefilo incolto] (Italian)
indyfilmblog [indy] (German)
JUICED.de [Daniel Höly] (German)
Kisisel Depresyon Anlari Film (Turkish)
Kulthit.de - Filmkritik (German)
L'Encyclopedie du Cinema Fantastique (French)
L'Internaute::Cinéma [Rédaction L'Internaute] (French)
L'Occhio Movie [Morena Mancinelli] (Italiano)
LaButaca.net [Joaquín R. Fernández] (Spanish)
LaButaca.net [Jordi Revert] (Spanish)
LaButaca.net [José Arce] (Spanish)
LaButaca.net [Julio R. Chico] (Spanish)
LaButaca.net [Miguel A. Delgado] (Spanish)
Las Horas Perdidas.com [Rafa Martín] (Spanish)
Le Blog Du Cinéma (French)
LeBuzz.Info [Isabelle Hontebeyrie] (French Canada)
Leer Cine (Sebastián Nuñez) (Spanish)
Les Ingoruptibles (French)
Manifest - Das Filmmagazin [Björn Lahrmann] (German)
mannbeisstfilm.de [Asokan Nirmalarajah] (German)
KinoGallery.com [Kornev Alexander] (Russian)
Kritiken [Tim Gieselmann] (German)
Moviemaster [Martin Günther] (German)
MovieMaze.de (German)
Movienerd.de [David Rams] (German)
Movieplayer.it [Adriano Aiello] (Italian)
Moviereporter.net (German)
Movies Ltd. [zerVo] (Greek)
MovieScene [Tom Rosens] (Dutch)
Mr. Karimi (Persian)
myFILM.gr [Jim Papamichos] (Greek)
MYmovies.it [Giancarlo Zappoli] (Italian)
myrating.dk [Thomas Ardal] (Danish)
noiseFromAmeriKa.org [Fabio Scacciavillani] (Italian)
Oh My Gore!! [Lan] (French)
OtrosCines.com [Diego Batlle] (Spanish)
OutNow.CH - DVD (German)
OutNow.CH - Kino (German)
Persönliche Buch- und Filmtipps [Dieter Wunderlich] (German)
Philm.dk [Tobias Lynge Herler] (Danish)
Portal de Cinema [Wikerson Landim] (Portuguese)
Público/Ipsilon [Jorge Mourinha] (Portuguese)
Quilombo (Spanish)
Quinlan.it [Raffaele Meale] (Italian)
Raúl Reis (Público) (Portuguese)
Remember it for later [Oliver Nöding] (German)
revue24Images.com [Philippe Gajan] (French)
sanatlog.com [Hakan Bilge] (Turkey)
Schnitt Online (German)
Schokkend Nieuws (Dutch)
screen/write [Thomas Lenz] (German)
Screeningmovies (German)
Screenshot-online.com [Sascha Koebner] (German)
simifilm.ch [Simon Spiegel] (German)
STIV [Christoph Stachowetz, Nina Tatschl] (German)
STIV [Christoph Stachowetz, Nina Tatschl] (German) → 2nd
Tempi Moderni [Diego Altobelli] (Italian)
Tetkam.Net (Russian)
The Cult of Ghoul [Dejan Ognjanovic] (Serbian)
The Director's Cup [Annarita Vitrugno] (Italian)
TVClassik [Franck Suzanne] (French)
Un Mundo de Cine (Spanish)
unnikath.de - Queer Film Blog (German)
Virtual DVD Magazine (German)
Virtual Illusion [Nelson Zagalo] (Portuguese)
Yolandamarin.net [Yolanda Marín] (Spanish)
zoom-Cinema.fr [Hervé Troccaz] (French)

Blu-ray and DVD focused reviews:

Blu-ray.com 3D [Martin Liebman]
Blu-ray.com Extended [Casey Broadwater]
Cagey Films [kgeorge]
ComingSoon.net - Blu-ray [Scott Chitwood]
ComingSoon.net - Extended Collector's Edition [Scott Chitwood]
DoBlu.com - Extended Collector's Edition Blu-ray [Matt Paprocki]
Entertainment Weekly [Jeff Labrecque]
High-Def Digest [Joshua Zyber] – extended version one
Home Theater Info DVD [Douglas MacLean] 3D
Movie Metropolis - Blu-ray [Dean Winkelspecht]
Movie Metropolis - Blu-ray [James Plath]
MovieWeb - Blu-ray [Paolo Sardinas]
NonModern [Jason Dietz] – 2nd one
Paste Magazine [Kristen Callihan]
Paste Magazine [Lindsey Lee]
PopMatters [Bill Gibron]
PopMatters [Jesse Hassenger]
Reel Reviews - Blu-ray [Loron Hays]
Review Maze (Blu-ray 3D) [John Moscow]
The Guardian [Ben Child]
The Independent [Tim Walker]
The Sci-Fi Movie Page - Blu-ray [Rob Vaux]
Theater Thoughts [John Carpenter]
UpcomingDiscs.com - Blu-ray [Gino Sassani]
UpcomingDiscs.com [Gino Sassani]
ViewLondon - Special Edition [Matthew Turner]

Comparison between movies:

Aisle Seat [Andre Dursin]
Critic After Dark [Noel Vera]

Non-reviews:

Podcasts:

Steady Diet of Film podcast [Erin Donovan and Mendi Menefee]
The Film Talk Podcast [Jett Loe and Gareth Higgins]

Videos:

Needcoffee.com Video [Widgett Walls]

Databases:

metacritic.com

Interviews:

Moving Image Source
Moving Image Source [David Schwartz]

Articles/Discussions on a different topic:

Moving Image Source [Tom McCormack]
Newsblaze [Prairie Miller]
The Sci-Fi Block [Robert Ring]
TotalFilm [Andy Lowe]

Non-critic:

The BigScreen Cinema Guide - Reader Reviews (x3)

Others:

movieshrink.com [Derek Dorris] – Cannot select text
Where the Mind is Without Fear . . . [Upamanyu] – Very long, does not fit into Excel cell

B. Critics' reviews that were adjusted to be included in the sample

Removed Blu-ray or DVD sections:

Review	*Adjustment*
411mania.com	Removed sections at the end
Blu-ray DVD / DVD talk (Brian Orndorf)	Removed section at the end
Blu-ray.com 3D [Martin Liebman] and Blu-ray.com Extended [Casey Broadwater]	Only included "Movie" section which was the same for both reviews so only included once; removed DVD sections
Bullz-eye.com	Removed extended version paragraph at the end
DVD Savant [Glenn Erickson]	Removed DVD/Blu-ray parts
DVD Talk - Blu-ray [Ian Jane]	Removed DVD/Blu-ray parts
DVD Talk - Blu-ray [Ryan Keefer]	Removed DVD/Blu-ray parts
DVD Talk - Extended Collector's Edition [William Harrison]	Removed DVD/Blu-ray parts
DVDActive [Gabriel Powers]	Removed DVD/Blu-ray parts
DVDActive [Marcus Doidge]	Removed DVD/Blu-ray parts
DVDcompare.net - Blu-ray [Jeremiah Chin]	Removed DVD/Blu-ray parts

DVDizzy.com - Blu-ray [Kelvin Cedeno] and DVDizzy.com - Extended Collector's Edition [Kelvin Cedeno]	Only included section on movie section which was the same for both reviews so only included once
Fulvue Drive-in - Blu-ray [Michael P. Dougherty II]	Removed two paragraphs on DVD
High-Def Digest [Joshua Zyber]	Removed second half
Just Press Play [Lex Walker]	Removed DVD/Blu-ray parts
MovieJuice [Mark Ramsey]	Removed DVD/Blu-ray parts
Movieman's Guide to the Movies - Blu-ray [Tyler Thomas] and Movieman's Guide to the Movies [Tyler Thomas]	Only included "Film" section which was the same for both reviews so only included once
New York Magazine [David Edelstein]	Removed DVD/Blu-ray parts
Popdose.com [Robert Cashill]	Removed DVD/Blu-ray parts
Slant Magazine Blu-ray [Ed Gonzalez]	Removed DVD/Blu-ray parts
Cagey Films [kgeorge]	Removed DVD/Blu-ray parts
Cinema Blend - Blu-ray [David Wharton]	Removed DVD/Blu-ray parts
DoBlu.com - Blu-ray [Matt Paprocki]	Removed DVD/Blu-ray parts

Disregarded reviews on other movies and included only Avatar portion:

ReelTalk [John P. McCarthy]
The History of the Academy Awards: Best Picture - 2009 [Erik Beck]
The New Yorker [David Denby]
The NYC Movie Guru [Avi Offer]

Other adjustments:

Review	***Adjustment***
DVD Movie Guide - Blu-ray [Colin Jacobson] and DVD Movie Guide [Colin Jacobson]	Only included 1 as same review
DVD talk (x2), briandorf.com and FilmJerk.com [Brian Orndorf]	Only included 1 as same review, 2 had an extra paragraph at the end which was included
Films in Review [Victoria Alexander]	Repeated so only included once
rogerebert.com [Roger Ebert] and DVDBeaver.com [Gary Tooze]	Only included 1 as same review