INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Critics vs. Audiences: The different opinions on the Marvel Cinematic Universe

João Pedro Gaspar Milheiro

Master's in Business Analytics

Supervisor:
Nuno Santos,
Professor Auxiliar Convidado,
ISCTE-IUL

October, 2023

BUSINESS
SCHOOL

Métodos Quantitativos para Gestão e Economia Department

Critics vs. Audiences: The different opinions on the Marvel Cinematic Universe

João Pedro Gaspar Milheiro

Master's in Business Analytics

Supervisor:
Nuno Santos,
Professor Auxiliar Convidado,
ISCTE-IUL

October, 2023

I dedicate this dissertation to my blonde and my parents,

I love you 3000.

Acknowledgements

I would like to thank all the people who made it possible to finish this dissertation, to my girlfriend who understanded that sometimes plans had to be cancelled, to my parents for all the support they gave me during all my academic journey, to all my family and friends for all the events that I had to miss but sill encouraged me to continue this thesis and lastly to teacher Nuno Santos for all the support and advice that he gave that made possible to finish this work.

Abstract

This thesis delves into the persistent disparity between critics and audiences within the movie and TV industry. Focused on the "Marvel Cinematic Universe" Phase 4, comprising seventeen projects, including movies, TV shows, and special presentations, the study sought answers to three critical questions: Do audiences and critics exhibit varying opinions for distinct projects? Does audience sentiment shift across different platforms? What aspects do both groups prioritize when sharing their opinions? Extensive data collection and analysis revealed significant differences between audiences and critics in opinions, particularly concerning heroes, feelings, supporting characters, and miscellaneous aspects of the projects. Whereas the most that audiences disagreed with each other was in opinions regarding heroes. Moreover, the study identified distinct dimensions of disagreement, with "WandaVision" emerging as the project with the highest level of discrepancy either when comparing critics with audiences or audiences against each other. Regarding what each group considers more important when sharing their opinion, it is discovered that critics are more restrictive than audiences.

Resumo

Esta tese investiga a disparidade persistente entre críticos e público na indústria do cinema e da televisão e na indústria televisiva. Centrado na Fase 4 do "Universo Cinematográfico Marvel", que inclui dezassete projetos, incluindo filmes, programas de televisão e apresentações especiais, o estudo procurou responder a três questões críticas: O público e os críticos apresentam opiniões diferentes para projetos distintos? O sentimento do público muda consoante as diferentes plataformas? Que aspetos é que ambos os grupos dão prioridade quando partilham as suas opiniões? A recolha e análise exaustivas de dados revelaram revelou diferenças significativas entre as opiniões do público e dos críticos, particularmente no que respeita heróis, sentimentos, personagens de apoio e aspetos diversos dos projetos. Enquanto os que os públicos mais discordaram entre si foi nas opiniões relativas aos heróis. Para além disso, o estudo identificou dimensões distintas de desacordo, com "WandaVision" a emergir como o projeto com o maior nível de discrepância, quer quando se comparam os críticos com o público ou o público entre si. Relativamente ao que cada grupo considera mais importante quando partilhar a sua opinião, verifica-se que os críticos são mais restritivos do que as audiências.

Index

1. Chapter 1: Introduction

1.1 Motivation and Context

The growing appearance in user-generated content (Beaudouin & Pasquier, 2017), particularly in the form of user reviews for various products, has led to significant disparities between user and critic assessments, as it can be seen within the recent studies concerning the divergence of perspectives between critics and audiences, the following findings were documented:"... it becomes evident that there exists a great deal of discrepancy between which characteristics professional critics value and which ones the audience favor"(Wallentin, 2016, p.80). This difference is corroborated by Basuroy et al. (2020) when they identify that user ratings (7.08) are, on average, higher than professional critics' ratings (5.82) when talking about the movie industry, by converting them into percentages and subtracting the ratings of the audiences by those of the critics we get the evidence that there is a difference of 1.26 percentage points (pp) between the two scores.

Rank	Title	Lifetime Gross	Year
1	Avatar	$2,923,706,026	2009
2	Avengers: Endgame	$2,799,439,100	2019
3	Avatar: The Way of Water	$2,320,250,281	2022
4	Titanic	$2,264,743,305	1997
5	Star Wars: Episode VII - The Force Awakens	$2,071,310,218	2015
6	Avengers: Infinity War	$2,052,415,039	2018
7	Spider-Man: No Way Home	$1,921,847,111	2021
8	Jurassic World	$1,671,537,444	2015
9	The Lion King	$1,663,075,401	2019
10	The Avengers	$1,520,538,536	2012

Figure 1.1 - Top 10 box office lifetime gross profit of all time

The decision to investigate the film industry was made due to the abundance of online platforms featuring user and critics' reviews. To determine how to study the film industry, it was chosen to examine the top 10 highest-grossing movies (as shown in Figure 1.1). It is noticeable that there is one movie considered independent and nine movies that belong to a cinematic universe and are considered blockbusters. Of these nine movies, four belong to the Marvel Cinematic Universe (MCU), and by searching the ratings attributed to MCU projects, we find that these discrepancies also happen for the MCU (figure 1). Basuroy et al. (2020) mention in their study that blockbusters are not "critics proof" and that they are very influenced by critical reviews. With this information, it seemed plausible that studying the reviews of MCU projects would help understand the different opinions of audiences and critics. Deng (2020) also talks about the difference between critics' and audience's opinions; he found that users say what they like about the movie, whereas critics comment on the more artistic parts of the movie.

Existing studies are related to the effect of reviews on the box-office or the influence of critics on consumers.

1.2 MCU

As Deng (2020), Legoux et al. (2016) and Verboord (2014) mention that the platform Rotten Tomatoes (RT) was used in their studies, with this it was decided to do a small analysis to verify how the opinion of critics and audiences regarding the MCU has evolved. For this, it is necessary to understand that the MCU is divided into four phases as it is explained by Sandwell and Longridge (2023), with the 1st phase having 6 movies and it started in 2008. The 4th phase ended in December 2022. In this phase, there are 6 movies, but here, the MCU decided to diversify and experiment with other types of content like TV Shows and Special Presentations; this means that in total, there will be 17 projects, which will total 41 complete projects on all the MCU phases combined. By averaging the past scores attributed to the MCU projects in RT by phase, differences are noticeable between the 1st phase ratings, where the difference is 1,7 pp between critics and audiences, and by doing the same calculation for the 4th phase, the difference is 10,7 pp. As we can see, over the years, a division is beginning to be created regarding what critics and audiences like, and it is close to the values of 1.26 pp previously mentioned as the average difference between each group for a movie.

By looking at Figure 1.2 and analyzing the scores provided by critics and audiences on RT, it is possible to notice the lowest score given by audiences (32 “She-Hulk: Attorney at Law”) and for the highest score, there is a three-way tie between two audiences scores (98 “Shang-Chi and the Legends of the Ten Rings” and 98 “Spider-Man: No way Home”) and one of critics scores (98 “Ms. Marvel”). It is also possible to see a few projects where the point difference is superior to ten, and in two cases, this difference should be very noticeable (“Eternals” 30 point difference and “She-Hulk: Attorney at Law” 48 point difference).

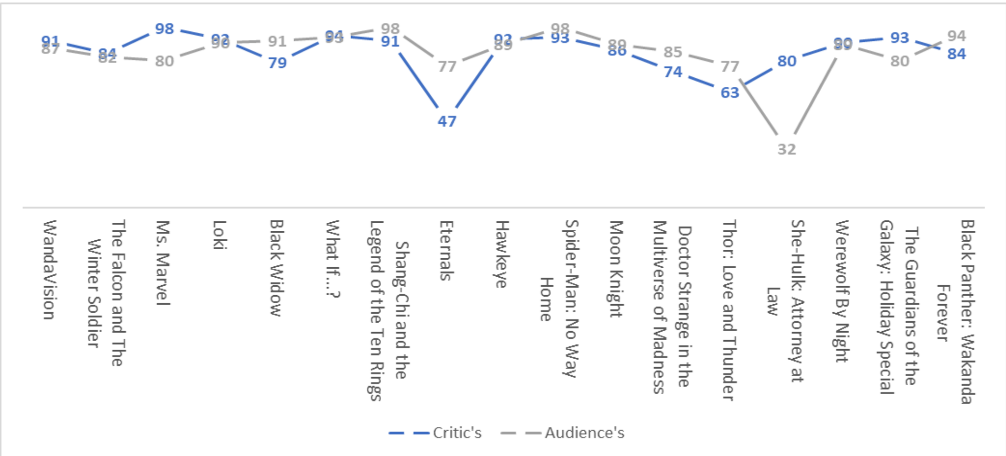

Figure 1.2 - Difference between critic's and Audience's RT scores

1.3 Research Questions and Hypothesis

To study the difference between critics and audience's opinions, research questions were created:

- IQ.1 - What are the main points of disagreement between critics and audiences, for the different projects?

To answer this question, first, we need to check if the opinions of both groups are related to one another; for that, it was created the following hypothesis: H1 – "Critics and Audiences have similar opinions for different projects".

- IQ.2 - Do the opinions of critics and audiences change depending on the platform/site where they are shared?

In order to answer IQ.2, we will create a hypothesis like it was done for the previous question. H2 will be "Critics or Audiences have the same opinions on different platforms".

If we assume that the null hypothesis of H1 or H2 was rejected, we will search for how their opinions differ.

- IQ.3 - What do critics and audiences consider most important when sharing their opinions? - To answer this question, we created a decision tree and analyzed the rules to find what each group considers important.

1.4 Objectives and Contributions

To be able to answer the research questions, there are a few objectives that will need to be achieved. They are the following:

1. Understand how reviews are scattered over the internet.
   - Quantitative analysis of the number of reviews for each selected platform.
2. Scrapping the various platforms for critics and audience's reviews.
   - Through applications and Python, retrieve reviews from different platforms.
3. Discovering what content is more mentioned in each group and attributing a feeling to the content.
   - Using natural language processing (NLP) methods, identify the most important terms and feelings attached to them.
4. Analyze the data retrieved from the text mining.
   - Using descriptive and inferential statistics and other machine learning techniques like decision trees.

After these objectives are completed, it will be possible to gain a new understanding related to reviews of the projects and platforms related to this study and how they differ from each other. This will allow the audience to find the best spaces to share their opinion and where to search for the opinions of others, whether this is an opinion that originated from a critic or an audience member. This is important because Tsao (2014, p.574) says that "...potential moviegoers attach greater importance to consumer reviews than to critic reviews..." and "...indicate that the positive influence of critic ratings on movie evaluations is more pronounced among individuals with low expectations.".

1.5 Document Structure

This thesis is structured as follows: in the first chapter, the theme of critics and audience reviews, as well as the MCU and the investigation questions that will help get answers about the differences between each group, are presented, and the contributions this thesis will provide. The second chapter comprises a systematic review of the literature (SRL), where 19 articles are analyzed, and the accomplishments of previous studies are presented. Chapter three explains the methodology, and it explains how the data was collected, how it was treated and how it was analyzed. In chapter four, we have an analysis of the hypothesis’s tests and the decision tree, and in chapter five, we present the conclusions of this study and some limitations encountered.

2. Chapter 2: Literature review

2.1 Protocol for the systematic review of literature

This section is an SRL, this type of document can offer a synthesis of the state of knowledge in a field and with this research priorities can be identified (Page, et al., 2021). Regarding this specific SRL it will be in the field of the analysis of opinions of critics and audiences. This way the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) methodology (Page, et al., 2021) will be used in this thesis. The objective of this SRL is to elaborate the state-of-the-art in a way to support the thesis "Critics Vs Audiences: The different opinions on the Marvel Cinematic Universe". The focus is on understanding the platforms where opinions are shared on the internet, how to collect these opinions, and the methods used for their analysis. To reach the objective, it is necessary to respond to a research question, which is: "How are the opinions of critics and audiences collected, analyzed and what knowledge was gained on the different platforms on the internet."

The research question was split into the following five specific questions in a way that permits a better understanding of the knowledge:

i) What is the main essence of the study?
ii) Does it describe how the data was gathered and what it contains?
iii) What methodology and techniques were used in the study?
iv) What were the main topics and conclusions found in both groups opinions?
v) What conclusions were described in the study?

The scientific studies found on this a SRL document were automatically retrieved from a scientific publication database. The method for this retrieval was based on keywords and inclusion/exclusion criteria like it was done by Caldas et al. (2017). The scientific database selected was "Web of Science", this was based on databases used in previous scientific studies (Caldas et al., 2017; Laureano & Santos, 2021). To retrieve the necessary articles to perform the SRL a query was executed on the database to extract papers related to cinema reviews, to do this the query was applied to the topic in the database, this means that a match like search was executed to match the query with words and expressions in the fields title, abstract and keywords. The query executed was the following: (Review* OR Opinion* OR Comment*) AND (Movie* OR Cinema OR Film*) AND (Critics OR Consumer* OR Audience*) AND (Compar* OR Different* OR Diverg* OR Evaluat*) AND ("Text-mining" OR "Sentiment Analysis" OR Model).

Table 2.1 - Inclusion and Exclusion criteria for SRL

Inclusion Criteria	Studies related to cinema reviews
	Studies related to text analysis of cinema reviews
Exclusion Criteria	Studies not published in peer-reviewed journals
	Studies before 2013
	Studies not in English or Portuguese
	Studies without access
	Duplicated studies

After applying the exclusion criteria by analyzing the articles characteristics, there were 119 articles where it was needed to verify the inclusion criteria, this was done by reading the abstract of the remaining articles. Upon the verification of the inclusion criteria, there were 19 articles left to carry out the SRL, these scientific studies are presented in table 2.2.

Table 2.2 - Articles analyzed by the SRL

ID	Year	Title	Jounal	Authors	Citations
1	2022	The Identification and Dissemination of Creative Elements of New Media Original Film and Television Works Based on Review Text Mining and Machine Learning	MATHEMATICAL PROBLEMS IN ENGINEERING	Yu, XA	0
2	2022	Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning	COMPUTATIONAL INTELLIGENCE AND NEUROSCIENCE	Chen, C; Xu, B; Yang, JH; Liu, M	0
3	2022	An integrative model of new product evaluation: A systematic investigation of perceived novelty and product evaluation in the movie industry	PLOS ONE	Luan, YY; Kim, YJ	2
4	2022	The cultural influences of narrative content on consumers' perceptions of helpfulness	INTERNATIONAL JOURNAL OF MARKET RESEARCH	Fu, N	0
5	2021	Typical opinions mining based on Douban film comments in animated movies	ENTERTAINMENT COMPUTING	Wu, T; Hao, F; Kim, M	1
6	2021	SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination	INFORMATION PROCESSING & MANAGEMENT	Sharma, SS; Dutta, G	15
7	2020	Investigating the effects of textual reviews from consumers and critics on movie sales	ONLINE INFORMATION REVIEW	Deng, TJ	6
8	2020	How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis	NEURAL COMPUTING & APPLICATIONS	Li, L; Goh, TT; Jin, DW	59
9	2020	What Is Important When We Evaluate Movies? Insights from Computational Analysis of Online Reviews	MEDIA AND COMMUNICATION	Schneider, FM; Domahidi, E; Dietrich, F	2

ID	Year	Title	Jounal	Authors	Citations
10	2020	Exploring contextual factors from consumer reviews affecting movie sales: an opinion mining approach	ELECTRONIC COMMERCE RESEARCH	Cheng, LC; Huang, CL	12
11	2018	Does Twitter chatter matter? Online reviews and box office revenues	APPLIED ECONOMICS	Vujic, S; Zhang, XY	7
12	2017	Forms of contribution and contributors' profiles: An automated textual analysis of amateur online film critics	NEW MEDIA & SOCIETY	Beaudouin, V; Pasquier, D	9
13	2017	Word of mouth quality classification based on contextual sentiment lexicons	INFORMATION PROCESSING & MANAGEMENT	Hung, CL	37
14	2016	Demand for cinema and diverging tastes of critics and audiences	JOURNAL OF RETAILING AND CONSUMER SERVICES	Wallentin, E	5
15	2016	Bidirectional Causality for Word of Mouth and the Movie Box Office: An Empirical Investigation of Panel Data	JOURNAL OF MEDIA ECONOMICS	Hsu, YL; Jane, WJ	4
16	2016	The effect of critical reviews on exhibitors' decisions: Do reviews affect the survival of a movie on screen?	INTERNATIONAL JOURNAL OF RESEARCH IN MARKETING	Legoux, R; Larocque, D; Laporte, S; Belmati, S; Boquet, T	22
17	2015	Everyone's a critic: The power of expert and consumer reviews to shape readers' post-viewing motion picture evaluations	POETICS	Jacobs, RS; Heuvelman, A; Ben Allouch, S; Peters, O	11
18	2014	The impact of peer-produced criticism on cultural evaluation: A multilevel analysis of discourse employment in online and offline film reviews	NEW MEDIA & SOCIETY	Verboord, M	48
19	2014	Feature-based opinion mining through ontologies	EXPERT SYSTEMS WITH APPLICATIONS	Penalver-Martinez, I; Garcia-Sanchez, F; Valencia-Garcia, R; Rodriguez-Garcia, MA; Moreno, V; Fraga, A; Sanchez-Cervantes, JL	106

The process executed during this SRL is synthesized in the following Figure 2.1.

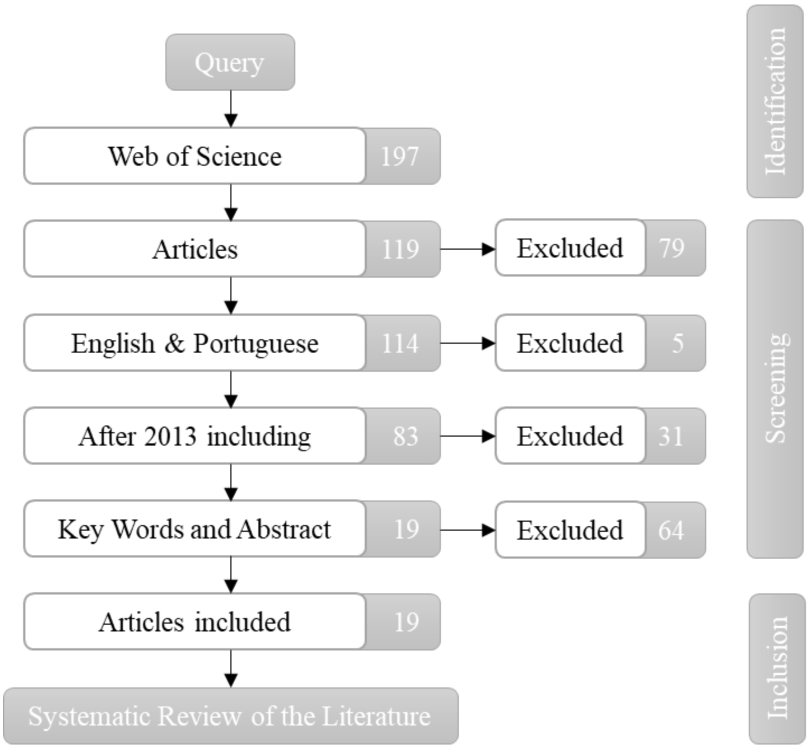

Figure 2.1 - SRL process

The last part of the SRL is the quality assessment of the remaining 19 articles, for this it was performed a critical evaluation of articles adapting the quality assessment used by Caldas et al. (2017) to this SRL. To do this, the articles were evaluated based on a checklist of questions created to respond to the specific research questions, presented in Table 2.3. The questions on the checklist were graded according to the following rules: YES (Y) = 1, PARTIALLY (P) = 0,5 and NO (N) = 0.

Table 2.3 - Quality assessment criteria

Study understanding	Q1.1	Is the scope of the study understandable?
	Q1.2	Is the objective of the study described?
	Q1.3	Does the study mention the data source and the collection time?
Data understanding	Q2.1	Is the source platform of the data mentioned?
	Q2.2	Is the data described?
Methodology	Q3.1	Is the process of data treatment described?
	Q3.2	The techniques/tools used during the study are mentioned?
	Q3.3	Is the methodology of the study explained?
Opinions	Q4.1	What topic/emotions were found in the study?
	Q4.2	Is it described the importance of a platform?
	Q4.3	Is it described the importance of an opinion?
	Q4.4	Do audiences follow critics, vice versa?
Conclusion	Q5.1	Were any limitations identified in the study?
	Q5.2	Were the contributions of the study identified?
	Q5.3	Were clues for future studies identified?

By analyzing Table 2.4, where the articles are presented, it is noticeable that exists 5 types of scope “Demand of Cinema”, “Model Testing”, “Opinion Mining”, “Financial Benefit” and “Text Quality”. The more prevalent scope is “Opinion Mining”, with the most recent three articles being about model testing. Looking at where the data was obtained previously it is noticeable that in most studies (13 articles) the data was obtained by extracting it from websites, whereas the remaining six studies used data from previous studies and one article didn’t mention how they collected their data. Concerning the period for the data collection, most studies collected data between two fixed points in time, but Deng (2020) has a different approach because he only wanted to study the impact of WOM in movie ticket sales, so he only considered the first 10 week for each movie that he decided to research.

Table 2.4 - Study Understanding

ID	Scope	Objective	Origin of Data	Period of the Data
1	Model Testing	Improve the performance of text sentiment analysis	N.S.	N.S.
2	Model Testing	ML sentiment analysis	Extracted from websites	2015-2021
3	Model Testing	Creation of a integrative theoretical model for product evaluation	Extracted from websites	2016
4	Opinion Mining	The helpfulness of User Generated Content	Extracted from websites	2013 - 2016
5	Opinion Mining	What are the problems with animation movies	Extracted from websites	2015
6	Model Testing	Increase accuracy of sentiment classification	Previous Study Extracted from websites	2012 - 2018 (IMDb)
7	Financial Benefit	Rating and reviews impact on sales	Extracted from websites	The first 10 weeks of a movie release
8	Text Quality	Impact of textual quality features on sentiment classification performance	Previous Study	N.S.
9	Opinion Mining	Comparing topics with SMEC	Open source	N.S.
10	Financial Benefit	The effect of reviews over time in box-office sales	Extracted from websites	2013 - 2014
11	Financial Benefit	Does tweets explain box office revenues	Previous paper	03/02/2012 - 07/03/2012

ID	Scope	Objective	Origin of Data	Period of the Data
12	Opinion Mining	How user talk in their reviews	Extracted from websites	2011
13	Text Quality	Identify high quality WOM (high and low reviews) and low quality WOM (medium reviews)	Previous Study	N.S.
14	Demand of Cinema	Analyze the demand for cinema of critics and general audiences	Extracted from websites	1999 - 2011
15	Financial Benefit	WOM importance for success of a movie	Extracted from websites	01/12/2010 to 30/04/2013
16	Financial Benefit	How critic reviews influence exhibitor's decisions	Extracted from websites	2002 - 2011
17	Opinion Mining	Influence that reviews have	Extracted from websites	N.S.
18	Opinion Mining	Different discourse of amateur critics and critics	Extracted from websites	First 3 months of 2010
19	Opinion Mining	Developing a new feature-based opinion mining approach	Previous Study	N.S.

2.2 How and what data was gathered

One of the limitations identified in 11 of the 19 analyzed studies is the dataset size, as they considered it was small. For instance, Li et al. (2020) worked with a dataset comprising 50,000 reviews, while Wu et al. (2021) utilized a dataset containing approximately 5,500 reviews. In contrast, Peñalver-Martinez et al. (2014) had access to a dataset consisting of 200 reviews, and both Chen et al. (2022) and Wallentin (2016) did not specify the number of reviews included in their analysis. Another limitation lies in the fact that the data originated from only one or two distinct sources or countries, as indicated in Table 2.5 (Beaudouin & Pasquier, 2017; Cheng & Huang, 2020; Deng, 2020; Fu, 2022; Luan & Kim, 2022; Verboord, 2014).

Table 2.5 - Data Understanding

ID	Source Of Reviews	Source of other Data	Nº Movies	Nº Series	Nº Critic Reviews	Nº User Reviews	Other Data	Country of Origin
1	N.S.	0	N.S.	N.S.	0	N.S.	0	China
2	N.S.	No other data source	N.S.	0	0	N.S.	No other data	China
3	IMDb	TMDb Box Office Mojo	147 (more than 100 reviews per movie)	0	0	49 835	Rating Movie Characteristics	U.S.
4	IMDb and Douban	Boxofficemojo.com	167 American Movies	0	0	111 857 (57 762 from IMDb, 54095 from Douban)	Information about the movies Helpful votes Country of origin (CH: 1 - China, 0 - USA) Age of the review	USA and China
5	Douban	Douban	1	0	0	500 short 4970 long	Rating Movie Characteristics	China
6	LMRD and CMRD IMDb	GitHub	1000 (500 from Hollywood and Bollywood, IMDb)	0	0	No more than 50 reviews per movie (IMDb)	Reviews of Kitchen, Yelp, DVD, Books and Electronics	N.S.
7	Rotten Tomatoes	IMDb, Wikipedia, OMDb, Ad$pender, Number.com	90	0	12 848	182 338	Movie Characteristics Weekly advertising Weekly box-office performance	
8	IMDb	No other data source	N.S.	0	0	50 000	No other data	N.S.
9	IMDb	No other data source	Max 30 reviews per movie	0	0	100 000 (25 000 positive, 25 000 negative, 50 000 unidentified)	No other data	
10	IMDb	IMDb Amazon Box Office Mojo	N.S.	0	0	18 131 (movies had to have more than 100 reviews)	Number of comments and Stars Other attribute information Box-office earning per week	N.S.
11	Twitter	Box Office Mojo	5	0	0	1,77 million tweets	Revenue data	N.S.
12	French Platform	French Platform	140	0	2300	40 000	User and Movie data	France

ID	Source Of Reviews	Source of other Data	Nº Movies	Nº Series	Nº Critic Reviews	Nº User Reviews	Other Data	Country of Origin
13	IMDb	IMDb	N.S.	0	0	27 886 (2000 have a pre-assigned sentiment tag)	Hotel Review Data Set Movie category and rating	N.S.
14	Swedish webiste similar to Metacritic	Swedish Film Institute (SFI) IMDb Metacritic	3212	0	0	N.S.	Movies Characteristics Ratings	Sweden
15	Yahoo! Movies eyny Cityyalk @movies	Cinema Yearbook in Republic of China IMDb Yahoo! Movies taiwancinema.com atmovies.com truemovie.com	769	0	0	401 000	Movie Characteristics	China
16	Rottent Tomatoes Mediafilm	CINEAC IMDb	788	0	N.S.	0	Movie Characteristics Weekly box-office performance Weekly advertising	Canada
17	filmkrant.nl moviemeter.nl	IMDb	1	0	40	38	10 control reviews Ratings of movies	Netherlands
18	IMDb	IMDb Rotten Tomatoes	109	0	N.S.	624	Media attention Background information on the movies	USA
19	N.S.	N.S.	1000	0	0	200 (100 positive, 100 negative)	No other data	N.S.

When it comes to data sources, IMDb stands out as the primary source utilized in 8 studies, notably being the most prominent. Following IMDb, two other platforms, RT and Douban (a Chinese cinema platform), are employed in two or more studies. However, Douban is not considered for this study due to the opinions being in Chinese. Other platforms are either not mentioned at all or are referenced in just one study, and therefore, they may not warrant further discussion. As such, IMDb and RT emerge as two potential data sources for this study, and we will delve into their usage in more detail in the methodology section. Looking at the data used in each movie it is possible that most studies collected data from different movies, some even created criteria relatively to the selection of the movies, like Cheng and Huang

(2020) that only used movies that had more than one hundred reviews. Others put the restrictions on the reviews such as no more than 30 (Schneider et al., 2020), 50 (Sharma & Dutta, 2021) or 100 reviews per movie (Luan & Kim, 2022). No studies were found that also analyzed TV Shows/Series.

Only one article does not use user generated reviews (Wallentin, 2016), since this one focus on the impact of critic reviews on ticket sales. Regarding studies related to critic reviews there are 5, with this it is noticeable that user reviews are studied more in the past. Another data point used in various studies is movie characteristics like actors, directors, budget, etc.

2.3 Methodology of previous articles

By analyzing the following table 2.6 it is possible to understand how previous studies were conducted. Looking at the tools used it is easily noticeable that python and R are the two predominant technologies for this type of research. Starting with the data treatment it is possible to identify a few procedures common between some studies.

Table 2.6 - Methodologies

ID	Data Treatment	Tools	Package	Techniques Used	Methodology
1	Filtering illegal characters Word segmentation Stop words removal Feature selection Feature weighting Vectorization	N.S.	N.S.	TF-IDF Epoch Parameter Comparison Experiment Batchsize Parameter Comparison Experiment Naive Bayes SVM	Construction of data set Preprocessing Classifier
2	Deduplicated Denoised Word segmentation	Python Word2Vec	SnowNLP - package	Multidimensional visualization LSTM (RNN)	Data Collection Data Preparation Modeling Analyze results
3	Remove emoji, URLs, stop words and punctuations Excluded reviews not in English, reviews less 10 words and without ratings	R	N.S.	HLM	Hypothesis development Data Preparation Quantify perceived novelty
4	Spell-checking Stop words removal Stemming Lemmatization	Python based web-scrapping TextMind - chinese reviews	N.S.	LWIC Negative binomial model with random effect Descriptive statistics	Hypothesis development Data Collection LWIC Statistical analysis
5	Deleting numbers, English text and null text Word segmentation Keyword extraction Stop words removal Vectorization	Python Octopus ROSTCM6 Word2Vec SDK - BOSON	Jieba	N_Gram K-means	Data Collection Data Preparation Modeling Analyze results

ID	Data Treatment	Tools	Package	Techniques Used	Methodology
6	Tagging Tokenizing Stop word removal Stemming Feature Selection	Stranford Log-Linear Tagger	SentiWordNet SentiDomain SentiPosNeg SentiDraw	SPLM	Data Preparation Sentiment Lexicon
7	Deleted records for lack of data.	N.S.	N.S.	Heuristic-systematic Rule-based approach with VADER Stepwise OLS regression	Data Preparation Compare sales with various types of reviews.
8	Tokenizing Specific words removal Lemmatization	Python v3.5.2	Natural Language Toolkit v3.2.5 Tensorflow v1.3.0 Keras v2.0.8 Gensim v3.1.0 Pandas v0.21.1 Scikit-learn v0.19.0	CNN SRN LSTM ARI CLI T-test OLS WLS	Dataset definition Data preparation Data quality measures DL - based sentiment classification Evaluation Statistical analysis
9	Exclude duplicates. Remove text like HTML tags, etc. Removed words with low tf-idf.	R	textclean topicmodels Idatunin 1.0.0 package spacyr	TF-IDF Topic modeling	Focus on positive and negative reviews. Data Preparation Estimated a correlated topic model. Group topic as categories Topics with at least 600 reviews
10	Tagging Feature extraction techniques	R	N.S.	Temporal Abstraction	Data Collection Data Preparation Data attribute selection module Temporal abstraction Module Association Rule Mining Rule Base Report
11	Only used tweets with the classification relevant	N.S.	N.S.	Descriptive statistics Panel Regression	Data Collection Data Understanding Modeling
12	First 400 reviews per movie Lemmatization Tagging Name entities and low frequency words removal	Alceste in Iramuteq	N.S.	Logistic regression Descriptive statistics	Inductive textual analysis Statistical analysis
13	Removal HTML tags Lemmatization Stop words removal Vectorization	Python	Natural Language Toolkit WordNetLemmatizer WEKA	Contextual Lexicons SVM (Linear) SVM(RBF) Decision Tree-J48 Naive Bayes	Data preparation Modeling Analysis
14	Movies with less than 10 visitors and 2 reviews were deleted Exclude duplicates	N.S.	N.S.	Ordered logit OLS	Review score model Box office model

ID	Data Treatment	Tools	Package	Techniques Used	Methodology
15	Indexed by date Word segmentation Tagging	The Chinese Knowledge and Information Processing System (TCKIPS) - Academia Sinica	N.S.	PGC test	Data Collection Data preparation Sentiment calculation
16	N.S.	N.S.	N.S.	Discrete-time survival model - DTPO	N.S.
17	N.S.	PASW Statistics 18	N.S.	Douma readability formula Inferential Statistics	Creation of hypothesis Selection of participants Show a movie Respond to a questionnaire Analyze results
18	N.S.	MLWin 2.22	N.S.	Multilevel analysis	Sampling of movies and reviews Coding the discourse in reviews Identifying information on context
19	Tokenizing Sentence Splitter Tagging Lemmatization	Movie Ontology	leftwords-wsj-0-18.tagger SentiWordNet 3.0	N_Gram Before, N_Gram After, N_Gram, All Phrase Euclidean vector	NLP Ontology-based feature identification Polarity Identification Vector Analysis

Two authors emphasized the importance of identifying reviews in English as a part of data preparation. Wu et al. (2021) carried out this step to eliminate non-English reviews, while Luan and Kim (2022) undertook it to remove reviews that were not in English. In this study, the identification of English reviews serves as a selection criterion for the reviews. Another procedure that holds significance for this thesis is the removal of numbers, special characters, punctuation, and HTML tags. This data cleansing method was implemented by several authors, including Hung (2017), Luan and Kim (2022), Schneider et al. (2020), and Wu et al. (2021).

The procedure of tokenizing a text is used by three authors (Li et al., 2020; Peñalver-Martinez et al., 2014; Sharma & Dutta, 2021). OLi et al. (2020) provides the most comprehensive explanation of this process. According to Li, tokenizing involves breaking down a text into individual tokens. The algorithm accomplishes this by identifying commas, single quotes followed by whitespace, and periods occurring at the end of a line as points for separation. Additionally, most punctuation marks are treated as separate tokens. To illustrate, one example highlights how the word "doesn't" is transformed into the tokens "does" and "n't."

Some studies also touch upon the concept of word segmentation, which is a similar process tailored for Chinese characters, given their distinct rules in writing.

The most frequently data preparation process among the studies (Fu, 2022; Hung, 2017; Luan & Kim, 2022; Sharma & Dutta, 2021; Wu et al., 2021; Yu, 2022) is stop word removal. Yu (2022) is the one that explains this process the best, he defines that stop words are auxiliary and function words and high frequency words and that they have insignificant meaning for sentiment classification, the best way to proceed is to filter them.

In multiple studies, the techniques of lemmatization have been discussed by Beaudouin and Pasquier (2017), Fu (2022), Hung (2017), Li et al. (2020) and Peñalver-Martinez et al. (2014). Similarly, stemming has been explored in different research papers, including Fu (2022), Li et al. (2020) and Sharma and Dutta (2021). Li et al. (2020) highlights the distinction between these word reduction methods. Stemming involves transforming a full word like "revival" into its base form, "reviv," while lemmatization deals with words like "drove" or "driving," reducing them to the base term "drive". The author further notes a preference for lemmatization over stemming, a choice also adopted in this thesis. The rationale behind this preference is the desire to comprehensively understand the text, and using full words is deemed more suitable for achieving that objective.

Tagging, as elucidated by Peñalver-Martinez et al. (2014), is the practice of attributing a part of speech, such as a noun, verb, or adjective, to a word. This technique, which was discussed in several studies (Beaudouin & Pasquier, 2017; Cheng & Huang, 2020; Hsu & Jane, 2016; Hung, 2017; Peñalver-Martinez et al., 2014; Sharma & Dutta, 2021), enables the creation of grammatical structures within phrases.

Regarding the techniques employed for data modeling, the following methods held particular significance. Support Vector Machines and Naïve Bayes, as indicated by Hung (2017) and Yu (2022), are classification algorithms frequently utilized for generating sentiment scores in various machine learning applications (Hung, 2017). Regressions, as explored by Beaudouin and Pasquier (2017), Deng (2020), Li et al. (2020), Vujić & Zhang (2018) and Wallentin (2016), play a crucial role in establishing relationships between reviews and other data. Although these methods will not be applied in this research, it is noteworthy to acknowledge the techniques employed by prior authors.

Term frequency-inverse document frequency (TF-IDF), as elucidated by Schneider et al. (2020) and Yu (2022), constitutes a technique wherein each term within a document is assigned a weight indicative of its relevance. This method allows for the assessment of the frequency of term occurrence in each document relative to its occurrence across other documents (Schneider et al., 2020). The previous researcher also made the choice of excluding terms with TF-IDF scores lower than 0,05. In the present study, TF-IDF is going to serve as the principal approach for the identification of significant terms.

The application of descriptive statistics, as demonstrated in prior studies (Beaudouin & Pasquier, 2017; Chen et al., 2022; Fu, 2022; Jacobs et al., 2015; Vujić & Zhang, 2018), will be integrated into the research methodology. These statistics will serve a dual function: initially aiding in data comprehension and subsequently facilitating the comparative analysis of critic and audience reviews.

In accordance with established conventions found in prior research, the fundamental structure depicted in figure 2.2 is a common thread among various methodologies. This thesis similarly adheres to this foundational framework, and the subsequent chapter will elucidate the methodology employed.

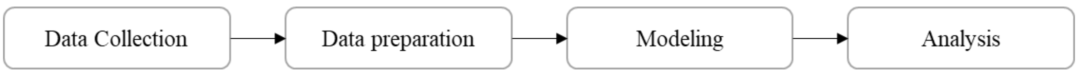

Figure 2.2 - Base methodology from previous studies

2.4 What is known about reviews from previous studies

In table 2.7 it is possible to understand that the two topics discussed the least are “Topic/Emotions Importance in Platforms” and “Audiences Follow Critics”.

Table 2.7 - What is in reviews

ID	Topics/Emotions (T/E)	T/E Importance in Platforms	Importance of Opinions	Audiences Follow Critics
1	N.S.	N.S.	N.S.	N.S.
2	N.S.	N.S.	N.S.	N.S.
3	N.S.	N.S.	A moderate level of perceived novelty in a movie leads to highest product evaluation of that movie. More likely to penalize novel movies by high reputation directors than low reputation directors Product novelty was unrelated to perceived novelty	N.S.
4	Four LWIC categories: - I - social - past-focus - future-focus	Websites managers should choose the most indicated reviews to provide to customers based on their cultural background	Individualists perceive reviews with a high frequency of "I" more helpful Collectivist perceive reviews with more social-related words more helpful Collectivist are more likely to vote for helpfulness for reviews that are more "past-focused", but less likely for "future-focused"	N.S.

ID	Topics/Emotions (T/E)	T/E Importance in Platforms	Importance of Opinions	Audiences Follow Critics
5	Domestic Conscience works Plot is weak Special effect are also good	Same movie in different platforms can have huge differences in rating scoring	N.S.	N.S.
6	N.S.	N.S.	N.S.	N.S.
7	User positive reviews. Critics balanced reviews. Critics lower ratings and less positive sentiment in reviews.	N.S.	User rating positive association with weekend sales. Sentiment in reviews of Critics positive association with weekend sales.	No, equal for reviews and ratings.
8	N.S.	N.S.	Word count has a statistically significant effect on the performance on sentiment analysis. Reviews with ARI or CLI lower than 10 are considered highly readable reviews	N.S.
9	10 categories of topics for 35 136 reviews. Main categories Acting, Comedy, Fun and Story & Plot.	N.S.	HE category reflects audience reactions. AP and NA most important in reviews.	N.S.
10	5 Categories of aspects 2 Categories of opinions 1 Implicit feature	N.S.	Good WOM means a stable decrease in box office instead a rapid decrease. "Customers use ratings to screen potential items and use text reviews to evaluate the limited set of screened items." Reviews of experience goods are richer and more dynamic than reviews of physical products.	N.S.
11	N.S.	N.S.	There is a positive relationship between total volume of tweets and box office revenue Tweets posted before watching a movie have a significant positive effect on the box office Negative tweets decrease cinema box office revenues	N.S.
12	Two types of reviews: Film-centered reviews (about the content and form of the film) Reception-centered review (about the effect that the film had on the viewer)	N.S.	The 2 types are equally distributed (52/48) "…the more a critic writes reviews, the lower their ratings and the longer their reviews tend to be..." The more a user write reviews they tend to become more film-centered	N.S.
13	N.S.	N.S.	N.S.	N.S.
14	N.S.	N.S.	The number of reviews that a movie has received helps to capture general recognition US and Swedish movies odds of getting a higher score from critics is 50% lower than other movies	Significantly differ for 11 out 18 parameters

ID	Topics/Emotions (T/E)	T/E Importance in Platforms	Importance of Opinions	Audiences Follow Critics
15	N.S.	N.S.	WOM is a outcome of sales but it also a driving force in consumer purchases WOM valence shows that causality direction for positive and negative critics with box office is asymmetric	N.S.
16	N.S.	N.S.	"… only a positivity bias of exhibitors such an excellent review allows a movie to stay longer on screen while negative reviews do not shorten a film´s run." " ... negative reviews hurt the performance more than positive reviews help performance. However, this negative effect on box office was only observed during the first week, which may explain why the same pattern does not apply to survival in theaters."	N.S.
17	N.S.	N.S.	Decreased readability and greater text complexity of professional reviews Confirmed the hypotheses that readers of negative reviews, evaluate product lower than readers of positive reviews	Rejected the hypothesis that says that readers of professional reviews would conform to their evaluations to professional evaluations
18	N.S.	N.S.	"On average, the reviewers use at least three high art and three popular aesthetic criteria in an inside review." "… institutionalized medium types appear to feature more high art and less popular aesthetic criteria…"	N.S.
19	N.S.	N.S.	N.S.	N.S.

Only two studies talked about the importance of the platforms where reviews are presented, Wu et al. (2021) said that the same movie in different platforms can have huge differences in terms of the ratting score and Fu (2022) mentioned the importance of website managers to monitor reviews in accordance with the platform cultural background.

When trying to answer if audiences follow critics and vice versa, there are 3 opinions, Deng (2020) says that they don't follow each other, and Wallentin (2016) confirms this by showing that they differ in 11 of 18 parameters used in that study. Jacobs et al. (2015) explains that their hypothesis "professional reviews influence post-viewing evaluations of readers in line with their valence, whereby positive reviews increase the evaluation and negative reviews dampen it" was rejected, this goes in line with the findings of the other two authors.

Regarding the importance of opinions almost all authors have discussed this subject, except for a few specific authors, namely Hung (2017), Peñalver-Martinez et al. (2014), Sharma and Dutta (2021) and Wu et al. (2021). Most authors reach conclusions that reviews impact the financial success of a movie. Regarding this column, several key observations emerge, Deng (2020) emphasizes that sole reliance on numeric ratings proves inadequate for predicting movie sales. Instead, it is the influence of critical reviews that significantly affects sales. Additionally, Vujić & Zhang (2018) and Wallentin (2016) point out that a higher volume of reviews correlates with a greater impact on box office performance and audience ratings, respectively.

The main topic/emotions found in previous studies were “acting”, “comedy”, “fun” and “story & plot” by Schneider et al. (2020), while Cheng and Huang (2020) used these categories “overall”, “screenplay”, “special effects”, “director” and “actors/actress”. Others (Beaudouin & Pasquier, 2017; Deng, 2020; Fu, 2022; Wu et al., 2021) identified the types of writing in reviews, things like if it is past-focused or if the review is centered in the movie or the feeling that caused in the viewer.

2.5 What were the conclusions of previous studies

By analyzing the final table 2.8 relating to this SRL, most authors discussed some type of limitation relative to the data, such as the lack of data, data limited to one platform or other type of data problem as previously mentioned.

Table 2.8 - Conclusions of previous studies

ID	Limitations	Contributions	Future Studies
1	N.S.	N.S.	N.S.
2	Does not consider complexity of Chinese semantic expression Limited dataset Classification results not ideal	N.S.	Aspect based sentiment analysis
3	Data limited to U.S. No causal evidence for the relationship between perceived novelty and product evaluation Evaluators assessments are influenced by product-unrelated information.	The relationship between product novelty and perceived novelty is nonsignificant. How perceived novelty influences customer engagement	Same study for different industries in other countries
4	LIWC should be used for entity extraction. LDA or LSA should be used for topic extraction. Data only from two countries	Empirical evidence about narrative content in terms of cultural differences extracted from online reviews	Examine same predictors in a hybrid culture. Same study but for different products
5	Data limited to 500 short reviews TF-IDF may be inaccurate in extracting keywords from short reviews Two words with different semantics may be identified as synonyms	A model based on Typical opinion mining applied to movie reviews	Deep learning sentiment-related concepts with LSTM

ID	Limitations	Contributions	Future Studies
6	N.S.	Lexicon creation by using distribution of keywords across star ratings Comparing the performance for 9 databases Proposing hybrid approach using SentiDraw along supervised methods	Remove named entities, like name of the movie, actor or director may be removed from the lexicon Need to normalize the score Word Sense Disambiguation helps identifying meaning of words
7	Only Reviews of USA. Only Reviews from on platform.	User and Critic reviews not consistent. More comprehensive view of the economic influence of WOM.	How users choose different channels of WOM.
8	Word count and readability are not sufficient to textual quality Dataset ML fail to achieve high accuracy with dataset that has more than 100 words per review.	How the textual quality affects the classification performance of the machine learners.	Do the same study in other data set and draw conclusion
9	Reviews may have noise. Target movies of the reviews are comprised of various types of quality; reviewers may have applied different criteria. Tv shows and documentaries are not covered by SMEC.	Online Reviews focus on 3 topics with only 2 that are common with SMEC. SMEC scales were developed to identify interindividual differences in what criteria viewers use and this study was aimed to what users write online.	Offers access to the scripts. Rule-based text extraction.
10	Only one source of data Could not investigate the causal relationship between reviews and movie sales	Analysis framework using opinion mining of consumer reviews Framework was used to analyze the relationship between user opinion and box-office The paper could provide movie companies with weekly reports to formulate appropriate marketing strategies	Investigate the causal relationship between reviews and movie sales with help of economic scholars Use of big data techniques
11	N.S.	The previous day's box office is significantly correlated with box office revenues in the current period	Micro data on Twitter Users Reviews from other sources Movie Ratings
12	Only studies cinema Only one platform	Examined the changes in cultural evaluation with the rise of aesthetic criteria.	Study other products and in various platforms
13	N.S.	Significance of WOM quality Lexicons built are adaptable to domains, fault tolerance and easy maintenance	LDA for selecting thematic words Relations between tokens and categories Collaborative filtering technique
14	Due to data limitations, it is not possible to conclude whether this is a case of predication or influence	Animated movies produced in Sweden or US make more money	Further investigation in stances where critics and consumers disagree

ID	Limitations	Contributions	Future Studies
15	N.S.	Considering the causality between WOM information and film performance Increase box office by stimulate positive critics in the short run (before week 7), and take care of negative critics in the long run (after 10 weeks)	N.S.
16	This model only tracks the complete withdrawal from theaters.	"… identification of a new mechanism through which reviews can impact the box office."	A model with both the survival time of a movie in a theater and weekly box office of the movie
17	Stimulus was different from live-action movies because actors were wearing mask	Creation of a psychological oriented model	Future research should therefore not only focus on a boarder audience in gathering samples, it should also control for their consumption patterns
18	Only used reviews from IMDb	Rise of peer-produced content challenges hierarchical model of cultural evaluation	Analyze other cultural forms other than films
19	Basic sentiment analysis based on SentiWordNet Validation is short Ontology of a predefine domain must be provided	Ontology-based feature Four different methods for polarity identification Vector analysis-based opinion mining approach	Large scale validation, with other domains

Many of the contributions are related to customer engagement or to economic scenarios (Deng, 2020; Cheng & Huang, 2020; Hsu & Jane, 2016; Vujić & Zhang, 2018; Wallentin, 2016), a few also consider their study to be a creation of a model to analyze data (Hung, 2017; Jacobs et al., 2015; Legoux et al., 2016; Sharma & Dutta, 2021; Wu, Hao, & Kim, 2021).

For future studies a few notes to take are to try previous techniques in new types of products or data sets, use data from new platforms and try new techniques. These authors (Vujić & Zhang, 2018; Wallentin, 2016) say that a possible route for a future study is to discover what are the stances of critics and consumers. Verboord (2014) recommends studying other cultural forms besides movies, and Beaudouin and Pasquier (2017) also recommend studying different products. Wallentin (2016) recommends studying where critics and audiences disagree, which is the basis of this research.

2.6 Articles evaluation

In table 2.9 it is possible to see the last step of the SRL. In this step the 19 articles are evaluated to how they responded to the criteria as it was explained previously, in this way the article is scored with one of the three possible scores 0 or 0,5 or 1. With a maximum score of 14, only one article was below the 50% mark, that was article 1. On average the articles had a score of 9,8 which is equal to say that on average each article had 70% mark in this evaluation and there are 10 articles with score higher than the average. Looking at the criteria 4.1, 4.2 and 4.4, these are the ones with the lowest score and they are also the ones where we try to find similarities with our study, this could mean that this theme is still not very well researched, the other criteria all had satisfactory scores with the lowest being 13 in the possible 19.

Table 2.9 - Article Evaluation

ID	1.1	1.2	2.1	2.2	3.1	3.2	3.3	4.1	4.2	4.3	4.4	5.1	5.2	5.3	Total
1	1	0	0	0	1	0,5	0,5	0	0	0	0	0	0	0	3
2	1	1	0	0	1	1	1	0	0	0	0	1	0	1	7
3	1	1	1	1	1	1	1	0	0	1	0	1	1	1	11
4	1	1	1	1	0	1	1	0,5	1	1	0	1	1	1	11,5
5	1	1	1	1	1	1	1	1	1	0	0	1	1	1	12
6	1	1	1	1	1	1	1	0	0	0	0	0	1	1	9
7	1	1	1	1	0,5	0	0,5	1	0	1	1	1	1	1	11
8	1	0,5	1	1	1	1	1	0	0	1	0	1	1	1	10,5
9	1	0,5	1	1	1	1	1	1	0	1	0	1	1	1	11,5
10	1	1	1	1	1	0,5	1	1	0	1	0	1	1	1	11,5
11	1	1	1	1	0,5	0,5	0,5	0	0	1	0	0	1	1	8,5
12	1	1	1	1	1	1	1	1	0	1	1	1	1	1	13
13	1	0,5	1	1	1	1	1	0	0	0	0	0	1	1	8,5
14	1	1	1	1	1	0,5	0,5	0	0	1	1	1	1	1	11
15	1	1	1	1	1	1	1	0	0	1	0	0	1	0	9
16	1	1	1	1	0	0,5	0	0	0	1	0	1	1	1	8,5
17	1	0,5	1	1	0	1	1	0	0	1	1	1	1	1	10,5
18	1	1	1	1	0	0	1	0	0	1	0	1	1	1	9
19	1	1	0,5	1	1	1	1	0	0	0	0	1	1	1	9,5
Total	19	16	16,5	17	14	14,5	16	5,5	2	13	4	14	17	17	

3. Chapter 3: Methodology

In the figure 3.1 it is possible to see the previous generalized process of the methodology (figure 2.2) adapted to this study, since the collection of the reviews until its analysis, the processing step is repeated for the number of project times two so that the TF-IDF only selects the important words for each project by audience and critics.

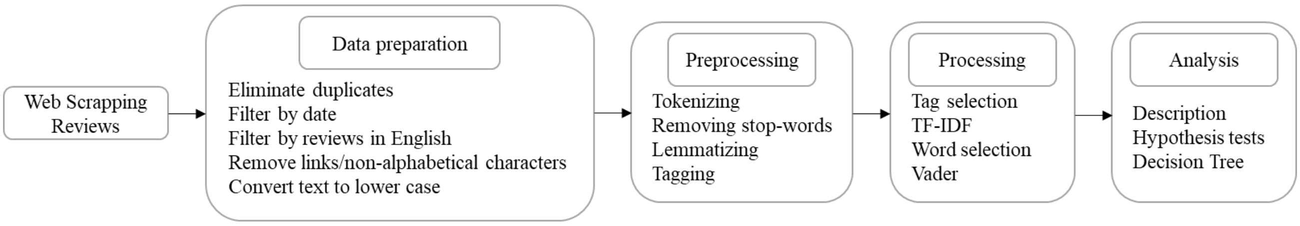

Figure 3.1 - Study Methodology

3.1 Understanding and Cleaning the data

The characteristics of the projects used in this study are presented in table 3.1, after the project name that is self-explanatory there is project type that indicates what type of project it is. Then there are three different date columns, starting with “Release Date” column this one refers to the date when the project was released, the “Start Date” column is the beginning of the period of collection for each project, it is necessary to have a start date before the release date because critics have access to early screenings of the projects, for this column it was considered the twenty days before the release date. The last date column “End Date” is the end of the period for this it was considered 10 weeks from the release date and then it was rounded to the next Sunday for movies, for tv shows it was given 15 days since the last episode aired and was also rounded to the next Sunday, for special presentations it was considered the next 30 days since the release it was not rounded to the next Sunday because coincided with one. In this table it is also possible to see the RT score for critics and audiences retrieved from their website. The reviews used in this thesis were extracted from various sources online and it was done in three different ways that are explained next.

Table 3.1 - Projects description

Project	Type	Start Date	Release Date	End Date	Critic's	Audience
WandaVision	TV Show	26/12/2020	15/01/2021	21/03/2021	91	87
The Falcon and The Winter Soldier	TV Show	27/02/2021	19/03/2021	09/05/2021	84	82
Loki	TV Show	20/05/2021	09/06/2021	01/08/2021	92	90
Black Widow	Movie	19/06/2021	09/07/2021	19/09/2021	79	91
What If...?	TV Show	22/07/2021	11/08/2021	24/10/2021	94	93
Shang-Chi and the Legend of the Ten Rings	Movie	14/08/2021	03/09/2021	14/11/2021	91	98

Project	Type	Start Date	Release Date	End Date	Critic's	Audience
Eternals	Movie	16/10/2021	05/11/2021	16/01/2022	47	77
Hawkeye	TV Show	04/11/2021	24/11/2021	09/01/2022	92	89
Spider-Man: No Way Home	Movie	27/11/2021	17/12/2021	27/02/2022	93	98
Moon Knight	TV Show	10/03/2022	30/03/2022	22/05/2022	86	89
Doctor Strange in the Multiverse of Madness	Movie	16/04/2022	06/05/2022	17/07/2022	74	85
Ms. Marvel	TV Show	17/05/2022	06/06/2022	31/07/2022	98	80
Thor: Love and Thunder	Movie	18/06/2022	08/07/2022	18/09/2022	63	77
She-Hulk: Attorney at Law	TV Show	29/07/2022	18/08/2022	23/10/2022	80	32
Werewolf By Night	SP	17/09/2022	07/10/2022	06/11/2022	90	89
Black Panther: Wakanda Forever	Movie	22/10/2022	11/11/2022	22/01/2023	84	94
The Guardians of the Galaxy: Holiday Special	SP	05/11/2022	25/11/2022	25/12/2022	93	80

Starting with the reviews from Critics, they were obtained by going to the RT page for each project, under the category "Tomatometer" and then by manually going to each critic review and copying it to a database. For each critic review it was extracted the name of the project, the critic's name, either the date when it was published or when it was last updated and a link for the source of the review. While trying to copy the review's, there were 5 occasions where it was not possible to retrieve the review "Copy Protection" (CP), "No Access" (NA), "Paywall", "Podcast" and "Video".

As it is demonstrated in table 3.2 it was possible to extract 3822 critic reviews across all seventeen projects, of all the reasons why it was not possible to extract reviews the more significant one was "Video" with 332 reviews.

Table 3.2 - Reviews from Critics

Project	CP	NA	Paywall	Podcast	Video	Review
WandaVision	0	10	7	2	31	306
The Falcon and The Winter Soldier	0	11	5	1	20	259
Loki	1	8	9	3	19	255
Black Widow	0	32	22	10	30	320
Marvel's What If...?	0	3	2	2	8	90
Shang-Chi and the Legend of the Ten Rings	1	19	15	8	25	241
Eternals	1	19	21	11	27	293
Hawkeye	0	4	8	3	9	119
Spider-Man: No Way Home	2	21	19	8	33	297
Moon Knight	2	3	9	2	9	179
Doctor Strange in the Multiverse of Madness	1	20	14	8	30	347
Ms. Marvel	2	2	7	0	6	138

Project	CP	NA	Paywall	Podcast	Video	Review
Thor: Love and Thunder	1	20	15	11	28	325
She-Hulk: Attorney at Law	0	2	5	1	14	206
Werewolf By Night	0	7	1	1	11	81
Black Panther: Wakanda Forever	1	20	10	7	27	319
The Guardians of the Galaxy: Holiday Special	1	1	1	1	5	47
Total	13	202	170	79	332	3822

Regarding audience's reviews initially it was supposed to have extracted audience's reviews from 4 platforms, but due to restrictions on the websites it was not possible to retrieve reviews from either RT or IMDb. The two other sources are the website Letterboxed (LB) this one being a social media with the intention for its users to share their opinions in movies and Tv Show's and the other source is a subreddit (forum dedicated to a community) r/marvelstudios where each project had threads at its release. With this it was possible to extract opinions from the website Letterboxed using a chrome extension (https://www.webscraper.io/) and from r/marvelstudios this extraction was done through python with access to Reddit api. The data collected from Letterboxed is almost equal in every project, 15360 reviews this is due to the method of extracting in this website it is only possible to go back 256 pages and each page has 12 reviews, facing this obstacle it was possible to extract data several times simply by changing the filters on the page. For each review it was also extracted the username, the date of the review and the project that it corresponds to. For the data collected from Reddit with its api it was possible to obtain the following fields for each review: review, date, and the corresponding project. The reviews were published in bookmarked threads for each project. In the table 3.3 is possible to see the distribution of reviews collect from Reddit and Letterboxed by project

Table 3.3 - Reviews from Audiences

Project	Letterboxed			Reddit		
	Original	English	Delta	Original	English	Delta
WandaVision	6802	5972	-830	26458	25953	-505
The Falcon and The Winter Soldier	8283	7117	-1166	15673	15403	-270
Loki	7611	6475	-1136	17662	17222	-440
Black Widow	6643	5806	-837	2574	2549	-25
What If...?	22	20	-2	12344	12095	-249
Shang-Chi and the Legend of the Ten Rings	6375	5429	-946	2697	2663	-34
Eternals	7776	6009	-1767	2661	2642	-19
Hawkeye	8501	7092	-1409	8576	8386	-190
Spider-Man: No Way Home	7001	5813	-1188	13456	13317	-139
Moon Knight	8254	6757	-1497	8282	8112	-170

Project	Letterboxed			Reddit		
	Original	English	Delta	Original	English	Delta
Doctor Strange in the Multiverse of Madness	9245	7545	-1700	9419	9294	-125
Ms. Marvel	8402	6914	-1488	4909	4810	-99
Thor: Love and Thunder	9248	7537	-1711	3655	3609	-46
She-Hulk: Attorney at Law	8950	7248	-1702	10092	9832	-260
Werewolf By Night	9967	8293	-1674	664	650	-14
Black Panther: Wakanda Forever	4356	3116	-1240	1158	1142	-16
The Guardians of the Galaxy: Holiday Special	9655	7943	-1712	736	719	-17
Total	127091	105086	-22005	141016	138398	-2618

To obtain the final data set, the first thing needed to do is to eliminate the duplicates reviews, and filter by the start and end date for each project in a way to only keep reviews inside of the desired time frame for each project. Afterward it was executed a python script using the “langid” library to identify the reviews of each project that are written in English this resulted in losing 22005 reviews from Letterboxed and 2618 reviews from Reddit as it can be seen in the previous table in the English columns. The next step was to remove links and non-alphabetical characters like emoji or punctuation present in the review, convert all the text in to lower case so that in a later stage we don’t have two similar tag’s that should be the same (Example: Action and action) and remove unnecessary extra spaces.

After getting the final data set it is still necessary to pre-process the reviews to perform the TF-IDF algorithm, using the python library “spacy” the reviews are tokenized, the stop words are removed, and the remaining tokens are lemmatized. Still using the same library, the next step is to tag words and select words with the following tags NOUN and PROPN.

3.2 TF-IDF and VADER

To execute the TF-IDF the first thing that was done was to create groups by audience or critic and project, for example “Spiderman: No Way Home and Critics”. After the review was grouped, the algorithm was executed by searching the words previously selected. This resulted in a file for each combination with the word and the corresponding score, in this study it was decided to only attribute a feeling to words with a score higher than 100 and if for a combination of a project and group of evaluators there were less than one hundred words with a score higher than 100, it would be selected the one hundred words with the highest score.

The next step was to create small text excerpts from the reviews. The way that the excerpts were created was that each review was matched with a file with words selected, this match was done by project and group of evaluators. If the reviews had words contained in the file of selected words, it was created an excerpt by adding the two words before and after for each matched word. This was done with the objective of reducing the quantity of text needed to analyze with VADER, the excerpt also kept the ID of the review where they were selected so that after the VADER algorithm was executed it would be possible to group by reviews.

By using VADER each excerpt was given as sentiment score between -1 and 1, to uniformize the scores it was considered that a score of 0 would mean that there was a neutral feeling associated to the word and would be represented with a 0. A score between 0 and 1 would mean that there was a positive feeling associated with that word and would be represented by 1. Finally, a score between -1 and 0 meant that there was a negative feeling associated with the word and would be represented by a -1.

After running VADER, it was necessary to group the excerpts by review in a way to have all the information on a single table. This table had the following fields, “ID”, “Review”, “Project”, “Review”, “Group of evaluators” and 770 of fields relating to the words analyzed. These fields were named using the word analyzed and it had 4 possible values, the three previously mentioned (-1, 0 and 1) and empty this would mean that that word was not represented in that review. By doing this process also meant that some reviews were excluded along the way because they didn’t include selected words in their text, so the final number of reviews analyzed was represented in the following table 3.4.

Table 3.4 - Selecting English words

Project	Letterboxed			Reddit			Critics		
	English	Processed	Delta	English	Processed	Delta	English	Processed	Delta
WandaVision	5972	5688	-284	25953	24798	-1155	306	305	-1
The Falcon and The Winter Soldier	7117	6574	-543	15403	14588	-815	259	258	-1
Loki	6475	6037	-438	17222	16120	-1102	255	255	0
Black Widow	5806	5389	-417	2549	2428	-121	320	319	-1
What If...?	20	18	-2	12095	11153	-942	90	90	0
Shang-Chi and the Legend of the Ten Rings	5429	5059	-370	2663	2509	-154	241	241	0
Eternals	6009	5474	-535	2642	2554	-88	293	293	0
Hawkeye	7092	6672	-420	8386	7842	-544	119	119	0
Spider-Man: No Way Home	5813	5213	-600	13317	12862	-455	297	297	0
Moon Knight	6757	6195	-562	8112	7486	-626	179	179	0
Doctor Strange in the Multiverse of Madness	7545	6805	-740	9294	8893	-401	347	347	0
Ms. Marvel	6914	6444	-470	4810	4393	-417	138	138	0
Thor: Love and Thunder	7537	6886	-651	3609	3440	-169	325	325	0

Project	Letterboxed			Reddit			Critics		
	English	Processed	Delta	English	Processed	Delta	English	Processed	Delta
She-Hulk: Attorney at Law	7248	6609	-639	9832	8878	-954	206	206	0
Werewolf By Night	8293	7606	-687	650	599	-51	81	80	-1
Black Panther: Wakanda Forever	3116	2860	-256	1142	1092	-50	319	319	0
The Guardians of the Galaxy: Holiday Special	7943	7251	-692	719	659	-60	47	47	0
Total	105086	96780	-8306	138398	130294	-8104	3822	3818	-4

In a way to reduce the number of fields they were aggregated in dimensions represented in the table below with their respective definition and percentage of the data where they are present, with the biggest dimensions being "Project", "Hero", "Feelings" and "Unidentified". In table 3.5 it is possible to see how each dimension is mentioned in the total of the reviews.

Table 3.5 - Dimensions definition

Dimension	Definition	Mentioned in Reviews
Unidentified	Difficult words to associate with the theme	76,8%
Feelings	Words used to express feelings	55,9%
Hero	Hero Names	50,5%
Project	Project Names	44,4%
Plot	Words related to projects plot	34,6%
Supporting	Words related to supporting characters	28,3%
Production	Words related to movie production	26,9%
Miscellaneous	Words related to theme but not easily identified	26,0%
Actor	Actor Names	17,6%
Villains	Villains Names	14,5%
Groups	Groups Names	9,1%
Genre	Genre Names	7,2%
Actress	Actress Names	6,8%
Location	Location Names	5,7%
Director	Director Names	5,5%
Audience	Mentioned word Audience	2,5%
Evaluations	Words that suggested an evaluation	1,3%
Critic	Mentioned word Critic	0,2%

3.3 Tests and Modeling

To test two hypotheses previously formulated, we ran two tests in spss statistics. One chi-square test for critics and audiences grouped by source (critics or audience), project and dimension that would test H1 and another chi-square test for audiences grouped by source (LB or Reddit), project and dimension to test H2. Even though the opinions of the "Unidentified" dimensions are statistically different in both tests, as this dimension was created when the word used could not be related to any theme this will not be further analyzed.

To be able to answer the third question it is necessary to find what each group thinks is more important, to do this it was decided to create a decision tree, where the rules are what each group considers important. To construct the decision tree first we had to balance the sample and split it into training and test samples. To balance the data set in a way that would be approximately 50% for each side, we multiplied the samples of critics by a factor of approximately fifty-nine. After the data set was split into 70% training and 30% testing, when this operation was executed, we made sure that the distribution between critics and audiences and distribution between audience's sources would stay the same in both partitions. We also created a few variables which were used in the decision tree, the variables are nothing more than a flag (variable that has two values 1 or 0) for each of the dimensions where every time a dimension is mentioned either be a positive, neutral or negative mention it has the value 1 in the flag.

The decision tree model used was C5.0 in spss modeler. The target of this decision tree is the variable "Source" which has two values "Audiences" or "Critics", the input values were the flags for each dimension and the variable "Type" that has the following values "Movie", "Tv Show" and "Special Presentation". In the parameters of the tree, it was only altered two from the default position, these were "Minimum records per child branch" this was set to 2000 in order to not have rules ending with small instances and the other change was in the "Misclassification costs" where when an audience is classified as Critic it was set have 8 of cost.

Table 3.6 - Example of confusion Matrix

	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

The decision tree will be evaluated according to a confusion matrix exemplified in table 3.6 and the following metrics, sensitivity (1) and precision (2) for audiences and specificity (3) and negative predictive (npv) (4) value, the formulas for each metric are described next.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \tag{1}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{2}$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \tag{3}$$

$$\text{NPV} = \frac{TP}{(TP + FN)} \tag{4}$$

4. Chapter 4: Analysis

4.1 Data Description

Before any analysis is performed it is important to understand the final data set, as it provides the essential context and insights necessary to unlock the potential of data. Understanding the complexities of data, such as its source and structure, helps individuals and organizations to make informed choices. In this subchapter we will start by describing the reviews and its sources, then we will cross this information with the projects. The other analysis is related to the mentions and the possible feelings encountered crossed with projects and types of people.

In table 4.1 is possible to visualize how the data is partitioned into different projects and from which source it comes from, with Letterboxed and Reddit representing the audiences. In total there are 230 892 reviews, the project most reviewed is the Tv Show "WandaVision" and the project with the least reviews is the Movie "Black Panther: Wakanda Forever".

Table 4.1 - Number of reviews by project

Project Name	Reviews	% of Total
WandaVision	30 791	13,3%
The Falcon and The Winter Soldier	21 420	9,3%
Loki	22 412	9,7%
Black Widow	8 136	3,5%
What If...?	11 261	4,9%
Shang-Chi and the Legend of the Ten Rings	7 809	3,4%
Eternals	8 321	3,6%
Hawkeye	14 633	6,3%
Spider-Man: No Way Home	18 372	8,0%
Moon Knight	13 860	6,0%
Doctor Strange in the Multiverse of Madness	16 045	6,9%
Ms. Marvel	10 975	4,8%
Thor: Love and Thunder	10 651	4,6%
She-Hulk: Attorney at Law	15 693	6,8%
Werewolf By Night	8 285	3,6%
Black Panther: Wakanda Forever	4 271	1,8%
The Guardians of the Galaxy: Holiday Special	7 957	3,4%
Total	230 892	100 %

When we analyze the reviews by source in the figure 4.1, the first thing that is noticeable is that number of reviews by critics are way smaller than the audiences but that is expected as there are a lot more audience members than critics.

Regarding the separation between the sources of audiences, Reddit has most of the reviews (approximately 56% of all reviews), but the other source Letterboxed (approximately 42% of all reviews) is not far behind.

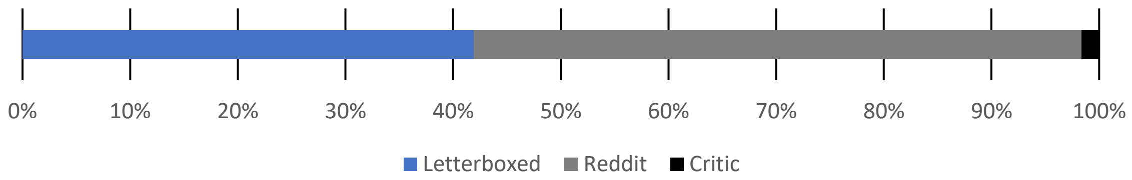

Figure 4.1 - Distribution of reviews by source

By separating the reviews by project, we obtain the following table 4.2 and image 4.2. When analyzing the differences between the sources that provided reviews for the audiences, there is a constant back and forth in the debate over which source boasts a higher quantity of reviews. However, the case of the "What If...?" project stands as an exception, where limitations in data extraction capabilities resulted in only a very limited number of reviews being extractable from Letterboxed. In the table previously mentioned it is also possible to explain why reddit is the bigger platform, as it has five projects with more than ten thousand reviews and one of them even has more than twenty thousand reviews.

Table 4.2 - Sources of reviews

Project Name	Letterboxed	Reddit	Critic
WandaVision	5 688	24 798	305
The Falcon and The Winter Soldier	6 574	14 588	258
Loki	6 037	16 120	255
Black Widow	5 389	2 428	319
What If...?	18	11 153	90
Shang-Chi and the Legend of the Ten Rings	5 059	2 509	241
Eternals	5 474	2 554	293
Hawkeye	6 672	7 842	119
Spider-Man: No Way Home	5 213	12 862	297
Moon Knight	6 195	7 486	179
Doctor Strange in the Multiverse of Madness	6 805	8 893	347
Ms. Marvel	6 444	4 393	138
Thor: Love and Thunder	6 886	3 440	325
She-Hulk: Attorney at Law	6 609	8 878	206
Werewolf By Night	7 606	599	80
Black Panther: Wakanda Forever	2 860	1 092	319
The Guardians of the Galaxy: Holiday Special	7 251	659	47
Total	96 780	130 294	3 818

Looking only at the image 4.2 it is evident that there are two projects were the number of reviews by Letterboxed is bigger than its audience counterpart, with these projects being "Werewolf By Night" and "The Guardians of the Galaxy: Holiday Special" curiously these two are only projects categorized as Special Presentations. Another thing that is observable is that the project where there is a bigger group of critics is in "Black Panther: Wakanda Forever", this is probably a result of an inferior number of reviews by the audiences.

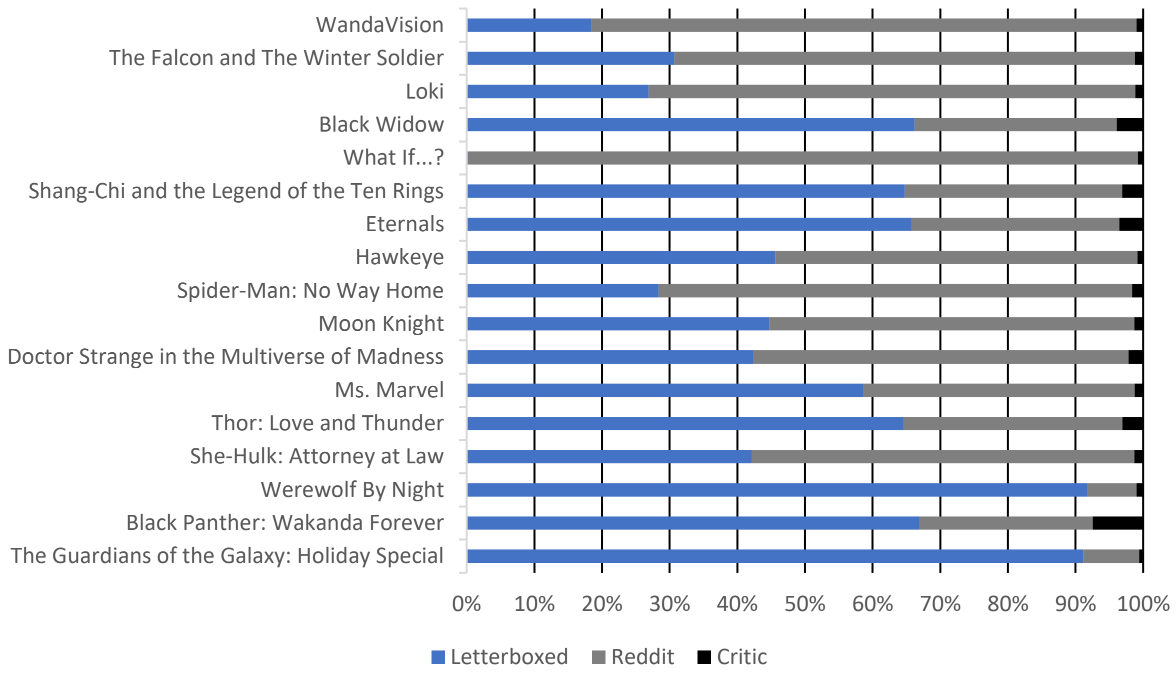

Figure 4.2 - Distribution of reviews by project and source

The next description of the data is related to the mentions of dimensions and the values corresponding to each feeling (positive, negative or neutral). When examining the following image 4.3, it becomes apparent that most of the mentions are positive (approximately 52% of all mentions), with the negative ones being in the minority (approximately 20% of all mentions) this means that for every negative mention there is approximately 2,5 positive mentions.

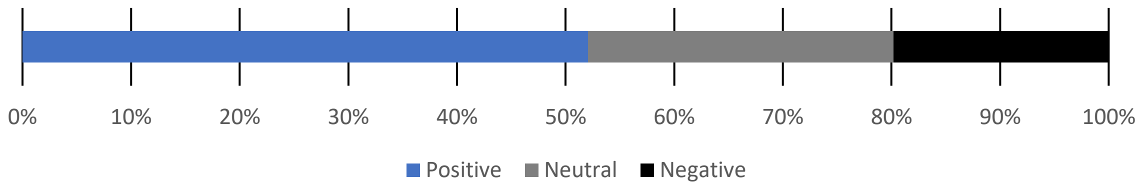

Figure 4.3 - Distribution of feelings

By observing the following table 4.3 it is apparent that on average the number of mentions by review is between 3 and 6, with the projects with the lowest values being "She-Hulk: Attorney at Law" and "What If…?" and the one with most mentions by review being

"Black Panther: Wakanda Forever". In total there were 955 313 mentions for 230 892 reviews. Most projects follow a rule where positive mentions are higher than neutral mentions and positive mentions are higher than negative mentions. However, there is an exception, "Doctor Strange in the Multiverse of Madness" has negative mentions higher than neutral mentions, this project is also the most negative project when talking about mentions in the same project. There are a few projects that stand out from the others, starting with the one that distinguishes themselves by the positive mentions there are 4 with more than 60% of positive mentions these ones being "Shang-Chi and the Legend of the Ten Rings", "Ms.Marvel", "Werewolf By Night" and "The Guardians of the Galaxy: Holiday Special", these projects are also a few of the ones that had higher ratings as it is possible to see in figure 1.2 . Lastly, there are two projects that have more than a third of their mentions as neutral feelings, these projects beings "Loki" and "What If…?".

Table 4.3 - Information relating to feelings

Project Name	Total Mentions	Mentions by Review	Positive	Neutral	Negative
WandaVision	122 150	4	46%	32%	22%
The Falcon and The Winter Soldier	83 668	4	47%	32%	22%
Loki	79 226	4	46%	34%	19%
Black Widow	44 631	5	54%	26%	19%
What If...?	37 313	3	43%	34%	24%
Shang-Chi and the Legend of the Ten Rings	42 098	5	60%	25%	15%
Eternals	43 045	5	56%	26%	18%
Hawkeye	55 141	4	54%	28%	18%
Spider-Man: No Way Home	81 779	4	49%	30%	21%
Moon Knight	52 290	4	53%	29%	18%
Doctor Strange in the Multiverse of Madness	72 866	5	43%	24%	32%
Ms. Marvel	44 051	4	64%	24%	12%
Thor: Love and Thunder	52 290	5	58%	23%	19%
She-Hulk: Attorney at Law	52 528	3	56%	25%	19%
Werewolf By Night	37 132	4	65%	21%	13%
Black Panther: Wakanda Forever	25 509	6	55%	27%	19%
The Guardians of the Galaxy: Holiday Special	29 596	4	68%	21%	11%

In table 4.4 it is possible to analyze how critics and audiences feel about each project, one of the first points to notice is that when there is a significant difference (bigger than 5%) between audiences and critics in positive mentions it normally tends relate to critics having more positive mentions as it happens in these projects "WandaVision" (57% to 46%), "Loki" (52% to 46%), "What If…?" (64% to 42%), "Eternals" (64% to 55%), "Hawkeye" (62% to 54%), "Ms. Marvel" (73% to 64%), "She-Hulk: Attorney at Law" (64% to 56%) and "The

Guardians of the Galaxy: Holiday Special" (78% to 68%). On the other hand, in the negative and neutral mentions this phenom is reversed, that means that when there is a significant difference between audiences and critics, here audiences are the ones that have the higher value. However, there is one exception in "Black Panther: Wakanda Forever" where critics are more negative than audiences (24 % to 18%).

Table 4.4 - Distribution of feelings by project and type of person

Project Name	Type	Positive	Neutral	Negative
WandaVision	Audience			
	Critic			
The Falcon and The Winter Soldier	Audience			
	Critic			
Loki	Audience			
	Critic			
Black Widow	Audience			
	Critic			
What If...?	Audience			
	Critic			
Shang-Chi and the Legend of the Ten Rings	Audience			
	Critic			
Eternals	Audience			
	Critic			
Hawkeye	Audience			
	Critic			
Spider-Man: No Way Home	Audience			
	Critic			
Moon Knight	Audience			
	Critic			
Doctor Strange in the Multiverse of Madness	Audience			
	Critic			
Ms. Marvel	Audience			
	Critic			
Thor: Love and Thunder	Audience			
	Critic			
She-Hulk: Attorney at Law	Audience			
	Critic			
Werewolf By Night	Audience			
	Critic			
Black Panther: Wakanda Forever	Audience			
	Critic			
The Guardians of the Galaxy: Holiday Special	Audience			
	Critic			

Lastly, by crossing the feelings mentioned and the source of the date we obtain the following table 4.5 and image 5.5, it is evident that Letterboxed mentions have more positive feelings (approximately 60%), while Reddit is more negative and neutral, with its neutral mentions being the largest of the three sources (approximately 32%). By further looking at the image it is noticeable that Critics and Letterboxed have a similar distribution of their mentions as they have on average 1,5 pp between the feelings of their mentions.

Table 4.5 - Mentions by source

Source	Positive	Neutral	Negative
Letterboxed	259 801	104 213	73 783
Reddit	208 714	151 615	106 222
Critic	29 086	12 730	9 149
Total	497 601	268 558	189 154

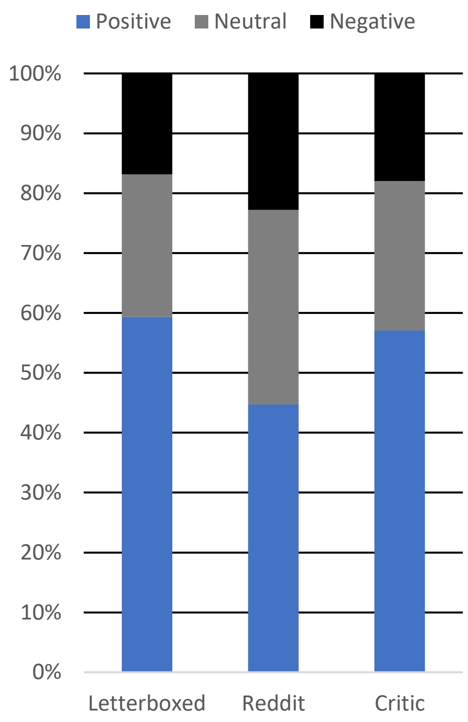

Figure 4.4 - Distribution of mentions by source.

4.2 Difference between critics and audiences

Our analysis began by testing the relationship between the opinions of critics and audiences by each dimension and project using the chi-square test to test H1. In the table 4.6 looking at the values near 0 (pvalue < 0,05) represented in green, it is possible to see where critics and audiences differ in opinions by rejecting the H1. The dimension "Critic" doesn't show up in the table because only mentioned by audiences. Looking at the total it is noticeable that only one dimension where audiences and critics opinions are not statistically different, with that dimension being Villains. Even though it is not considered statistically different if we go project by project, there are 6 projects where it is considered that the opinions are statistically different.

Table 4.6 - Chi-square test between Project, Dimensions and Person type

Project Name		Actor	Actress	Audience	Director	Evaluation	Feelings	Genre	Groups	Hero	Location	Miscellaneous	Plot	Production	Project	Supporting	Unidentified	Villains
WandaVision	N of Valid Cases	4056	2057	1023			13668	1987	2587	14349	4558	7984	10747	6230	11134	9352	26106	6312
	Pearson Chi-Square	0,00	0,00	0,00			0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,66
The Falcon and The Winter Soldier	N of Valid Cases	2338					10693	1775	1731	11909	1120	5497	5935	3856	8101	9158	16475	5080
	Pearson Chi-Square	0,30					0,00	0,00	0,34	0,00	0,88	0,00	0,25	0,15	0,01	0,00	0,00	0,03
Loki	N of Valid Cases	3702		124			10559		3595	11168	205	3975	8556	4566	8379	4013	17302	3082
	Pearson Chi-Square	0,00					0,00		0,00	0,00		0,00	0,00	0,04	0,00	0,00	0,00	0,00
Black Widow	N of Valid Cases	1918	2839	152	181	507	5284	1775	1623	4250	641	1978	3420	3537	4723	3640	6358	1805
	Pearson Chi-Square	0,55	0,06				0,01	0,00	0,17	0,00	0,00	0,00	0,69	0,15	0,73	0,00	0,00	0,02
What If...?	N of Valid Cases	692	45	44		34	5230	62	1321	4957	76	1983	4578	1360	4225	2413	8064	2229
	Pearson Chi-Square	0,02					0,00		0,91	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,74
Shang-Chi and the Legend of the Ten Rings	N of Valid Cases	2518	1209	96	214		5323	1728	912	4457	937	2115	4181	3993	4640	2619	6014	1142
	Pearson Chi-Square	0,00	0,01				0,01	0,00	0,71	0,00	0,89	0,01	0,01	0,00	0,10	0,00	0,00	0,00
Eternals	N of Valid Cases	1782	281	118	1558	604	5554	924	2747	3612	1398	2329	3136	3724	4776	2482	6618	1000
	Pearson Chi-Square	0,07			0,16		0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Hawkeye	N of Valid Cases	1678	1773	67			7733	911	821	7446	103	3067	3549	4109	6105	4765	10275	2739
	Pearson Chi-Square	0,26	0,10				0,04	0,00	0,31	0,00		0,00	0,01	0,00	0,05	0,04	0,00	0,51
Spider-Man: No Way Home	N of Valid Cases	7111	248	1712	250		10364	185	147	10582	217	6845	8031	5018	8751	3377	15285	3656
	Pearson Chi-Square	0,00		0,58			0,00			0,00		0,00	0,00	0,00	0,48	0,00	0,00	0,00
Moon Knight	N of Valid Cases	3521	104				7601	725	830	5918	144	5258	2897	3635	6454	4158	10177	868
	Pearson Chi-Square	0,00					0,01	0,00	0,67	0,00		0,00	0,02	0,26	0,11	0,00	0,00	0,00
Doctor Strange in the Multiverse of Madness	N of Valid Cases	1853	2116	1242	3282		9174	2006	1649	9757	193	3251	8124	4511	7991	4915	12051	751
	Pearson Chi-Square	0,00	0,04	0,29	0,05		0,00	0,00	0,17	0,00		0,01	0,00	0,00	0,00	0,00	0,00	
Ms. Marvel	N of Valid Cases	990	1371	552		465	7442		1199	5445	454	3228	2273	3390	5611	3046	8032	553
	Pearson Chi-Square		0,01	0,50			0,02		0,00	0,00	0,00	0,00	0,00	0,27	0,03	0,00	0,00	0,00
Thor: Love and Thunder	N of Valid Cases	3322	1306		2409	582	7237	1407	285	5079	586	2011	5026	3029	5645	4173	7612	2581
	Pearson Chi-Square	0,00	0,41		0,19		0,00	0,00		0,00	0,09	0,00	0,00	0,97	0,00	0,00	0,00	0,64
She-Hulk: Attorney at Law	N of Valid Cases	1429	1078	144			8930	706		7754	523	3622	2922	4036	7177	2516	11534	157
	Pearson Chi-Square	0,00	0,00				0,00	0,00		0,00	0,56	0,00	0,00	0,02	0,00	0,00	0,00	
Werewolf By Night	N of Valid Cases	1157	36	407	1487	492	5969	1859		4599	39	3059	2530	3216	3624	1897	6713	48
	Pearson Chi-Square	0,01		0,85	0,58		0,05	0,06		0,67		0,00	0,00	0,03	0,00	0,11	0,00	
Black Panther: Wakanda Forever	N of Valid Cases	1889	934		948	413	3036	617		2145	1391	1310	1878	2149	2550	1489	3249	1511
	Pearson Chi-Square	0,00	0,00		0,01		0,00	0,00		0,00	0,00	0,09	0,00	0,05	0,11	0,00	0,00	0,01
The Guardians of the Galaxy: Holiday Special	N of Valid Cases	717	312	24	2277		5221	19	1501	3090	474	2585	2097	1778	2656	1401	5444	
	Pearson Chi-Square	0,03	0,89		0,06		0,12		0,03	0,03	0,47	0,01	0,00	0,00	0,17	0,01	0,00	
Total	N of Valid Cases	40673	15709	5705	12606	3097	129018	16686	20948	116517	13059	60097	79880	62137	102542	65414	177309	33514
	Pearson Chi-Square	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,42

Not including “Unidentified”, the dimensions “Feelings”, “Hero”, “Miscellaneous” and “Supporting” are where critics and audience disagree the most as it can be seen in the previous table 4.6. For “Hero” and “Supporting” the only project where this dimension is not considered statistically different is “Werewolf by Night”, for “Miscellaneous” the project where H1 is not rejected is “Black Panther: Wakanda Forever” and lastly for “Supporting” the project that is not considered statistically different is “The Guardians of the Galaxy: Holiday Special”.

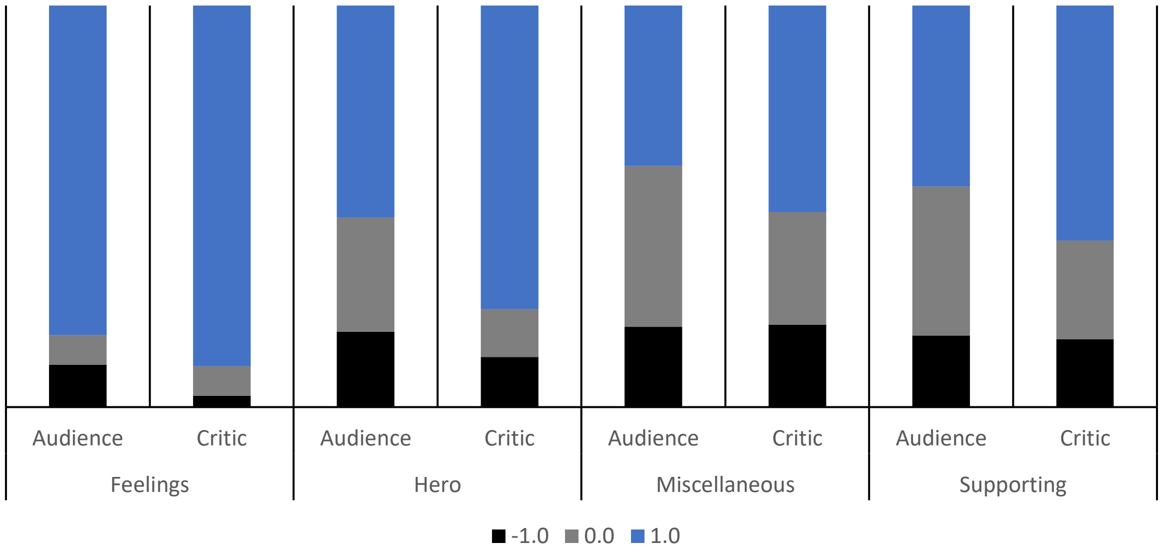

Figure 4.5 - Distribution of mentions between the most significant dimensions by person type

Considering the dimensions where both groups disagree the most, we went to see how the opinions differed as can be seen in figure 4.5. In “Feelings” critics are more positive whereas audience are more negative, and they seem to be on the same level regarding neutral feelings. In each of the following dimensions “Miscellaneous” and “Supporting”, critics tend to be more positive than the audiences. Lastly, critics are far more positive than the audience regarding “Hero”. In general critics tend to be more positive.

It is also noticeable that of the remaining dimensions there are five dimensions where critics and audiences disagreed in more than 50% of the projects, these dimensions being “Plot” (15 dimensions where H1 is rejected), “Actor” (12 dimensions where H0 is rejected) and the last three dimension “Genre”, “Production” and “Project” all have 11 dimensions where H1 is rejected.

These statistic tests cannot be looked without taking into consideration the number of observations, with this in mind it was decided to calculate a weighted average of the p-values by project where the weights are the number of valid cases for each corresponding p-values. This weighted average was crossed with the following metric (% Pos Critics - % Neg Critics) -

(% Pos Audiences - % Neg Audiences) by project and resulted in the graphic represented in figure 4.6. This means that when a project is more to the right it is considered that Critics are more positive towards that project, whereas when a project is more to the left it means that Audiences are more Positive.

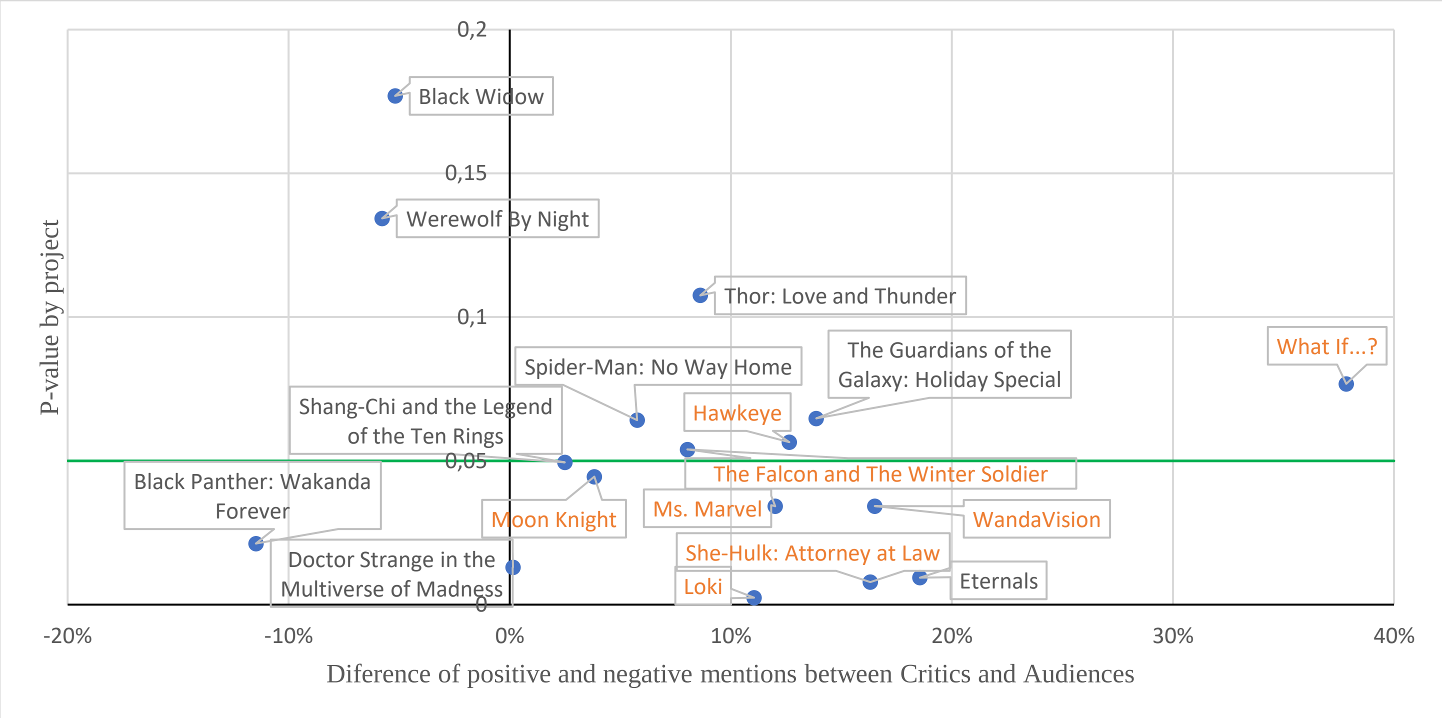

Figure 4.6 - P-values Vs Difference between Critics and Audiences

When analyzing this figure, it is important to notice the green line (0.05) as the opinions related projects below the line are considered statistically different, so we can reject H1 for the following projects “Black Panther: Wakanda Forever”, “Doctor Strange in the Multiverse of Madness”, “Shang-Chi and the Legend of the Ten Rings”, “Moon Knight”, “Loki”, “Ms. Marvel”, “She-Hulk: Attorney at Law”, “WandaVision” and “Eternals”. This means that the opinions of nine projects in the seventeen projects are considered statistically different, that is, in more than 50 % of the projects the H1 is rejected affirming that critics and audiences have different opinions. On average these projects had approximately 11 dimensions where H1 was rejected, with “WandaVision” being the one where Critics and audiences disagreed the most with 14 dimensions where H1 was rejected. Of the projects where we rejected H0 there is only one where Audiences are more positive than Critics, and that is “Black Panther: Wakanda Forever”. One interesting phenom that happens in this figure is that the projects that are Tv Shows (name is represented in orange) are concentrated in an area and in most of them the H1 is rejected, with one exception “What If…?” that is the project the furthest to the right, this might be related to this project being entirely in animation style.

4.3 Difference of opinions from different platforms

By doing the same type of analysis but by testing the relationship between the opinions of critic's or audience's sources and by dimension and project it is possible to test H2. In this research we have three different sources of reviews "Letterboxed" and "r/marvelstudios" for audiences and "RT" for critics. As we only have one source for critic's opinions, we will only test H2 for audiences.

Table 4.7 - Chi-square test between Project, Dimensions and Audience type

Project Name		Actor	Actress	Audience	Director	Evaluations	Feelings	Genre	Groups	Hero	Location	Miscelaneous	Plot	Production	Project	Supporting	Unidentified	Villains
WandaVision	N of Valid Cases	3784	1783	795			13395	1723	2354	14044	4280	7688	10442	5933	10829	9064	25801	6113
	Pearson Chi-Square	0,00	0,00	0,18			0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
The Falcon and The Winter Soldier	N of Valid Cases	2096					10447	1562	1486	11651	904	5241	5678	3604	7844	8900	16217	4918
	Pearson Chi-Square	0,00					0,00	0,00	0,52	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Loki	N of Valid Cases	3453					10312		3347	10913		3740	8302	4314	8124	3759	17047	2880
	Pearson Chi-Square	0,00					0,00		0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Black Widow	N of Valid Cases	1610	2520			507	4994	1494	1307	3931	337	1660	3101	3225	4406	3321	6039	1547
	Pearson Chi-Square	0,05	0,97			0,04	0,01	0,21	0,07	0,00	0,30	0,09	0,00	0,01	0,03	0,00	0,00	0,00
What If...?	N of Valid Cases	629					5145		1251	4868		1898	4489	1275	4135	2323	7974	2217
	Pearson Chi-Square	0,57					0,63		0,58	0,30		0,25	0,53	0,10	0,01	0,47	0,02	0,24
Shang-Chi and the Legend of the Ten Rings	N of Valid Cases	2277	973				5091	1529	731	4216	742	1877	3940	3755	4400	2378	5774	928
	Pearson Chi-Square	0,01	0,44				0,46	0,05	0,61	0,00	0,35	0,00	0,00	0,78	0,19	0,00	0,00	0,00
Eternals	N of Valid Cases	1568			1271	604	5272	723	2454	3320	1117	2040	2844	3435	4484	2189	6325	726
	Pearson Chi-Square	0,17			0,16	0,00	0,00	0,38	0,00	0,00	0,43	0,01	0,00	0,42	0,01	0,96	0,00	0,13
Hawkeye	N of Valid Cases	1564	1667				7617	822	705	7327		2952	3430	3993	5986	4648	10156	2709
	Pearson Chi-Square	0,00	0,13				0,00	0,00	0,08	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Spider-Man: No Way Home	N of Valid Cases	6822		1490			10070			10285		6548	7735	4731	8455	3091	14988	3378
	Pearson Chi-Square	0,00		0,59			0,00			0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Moon Knight	N of Valid Cases	3345					7431	602	677	5739		5079	2721	3466	6275	3982	9998	699
	Pearson Chi-Square	0,00					0,00	0,04	0,67	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Doctor Strange in the Multiverse of Madness	N of Valid Cases	1523	1780	1021	2941		8834	1698	1492	9410		2911	7777	4168	7644	4570	11704	751
	Pearson Chi-Square	0,00	0,00	0,98	0,00		0,03	0,05	0,00	0,00		0,01	0,00	0,01	0,00	0,00	0,00	0,00
Ms. Marvel	N of Valid Cases	990	1248	480		465	7309		1076	5307	323	3090	2136	3260	5473	2908	7894	523
	Pearson Chi-Square	0,00	0,03	0,98		0,06	0,00		0,23	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,38
Thor: Love and Thunder	N of Valid Cases	3001	991		2086	582	6913	1132		4754	417	1733	4701	2732	5324	3849	7288	2265
	Pearson Chi-Square	0,86	0,44		0,16	0,04	0,01	0,70		0,00	0,43	0,16	0,00	0,31	0,10	0,02	0,00	0,00
She-Hulk: Attorney at Law	N of Valid Cases	1256	918				8726	592		7548	405	3417	2717	3837	6971	2312	11328	
	Pearson Chi-Square	0,00	0,48				0,00	0,91		0,00	0,13	0,00	0,00	0,01	0,00	0,00	0,00	
Werewolf By Night	N of Valid Cases	1079		372	1415	492	5893	1784		4519		2979	2450	3136	3544	1820	6633	
	Pearson Chi-Square	0,63		0,91	0,08	0,50	0,02	0,18		0,00		0,11	0,62	0,02	0,31	0,15	0,00	
Black Panther: Wakanda Forever	N of Valid Cases	1571	620		647	413	2734	409		1826	1072	994	1560	1841	2233	1174	2930	1212
	Pearson Chi-Square	0,00	0,13		0,19	0,00	0,18	0,11		0,00	0,00	0,00	0,11	0,01	0,78	0,56	0,00	0,33
The Guardians of the Galaxy: Holiday Special	N of Valid Cases	671	267		2230		5175		1454	3043	431	2538	2050	1731	2610	1360	5397	
	Pearson Chi-Square	0,38	0,02		0,00		0,42		0,01	0,00	0,23	0,00	0,00	0,01	0,19	0,46	0,00	
Total	N of Valid Cases	37239	12767	4158	10590	3063	125358	14070	18334	112701	10028	56385	76073	58436	98737	61648	173493	30866
	Pearson Chi-Square	0,00	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

There is only one dimension where H2 is not rejected, this one being “Audience”, that means that in every other dimension the opinions expressed in them are statistically different between the two sources. Another situation that appears in this test is related to the project “What If…?”, where there are only two dimensions where H2 is rejected, this is probably a result of insufficient data from a source since from LB we were only able to extract eighteen reviews compared to the approximate eleven thousand.

As it was done in the previous analysis if we exclude the dimension “Unidentified”, “Hero” is the dimension where between audiences there is more conflict as such with a p-value <0,05 we can reject H2 for this dimension. As it is represented in figure 4.7, we can clearly see where audiences from different platforms differ from each other as LB has much more positive mentions than reddit, on the other hand reddit is far more negative and neutral.

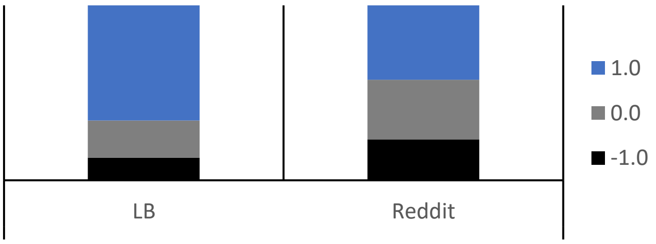

Figure 4.7 - Distribution of mentions in the Hero dimension by audience source

Excluding the two previously mentioned dimensions, there are 8 dimensions with statistically different opinions where audiences disagree in more than 50% of the projects as it is represented in the following figure 4.8, these dimensions are the following “Actor”, “Feelings”, “Miscellaneous”, “Plot”, “Production”, “Project”, “Supporting” and “Villains”.

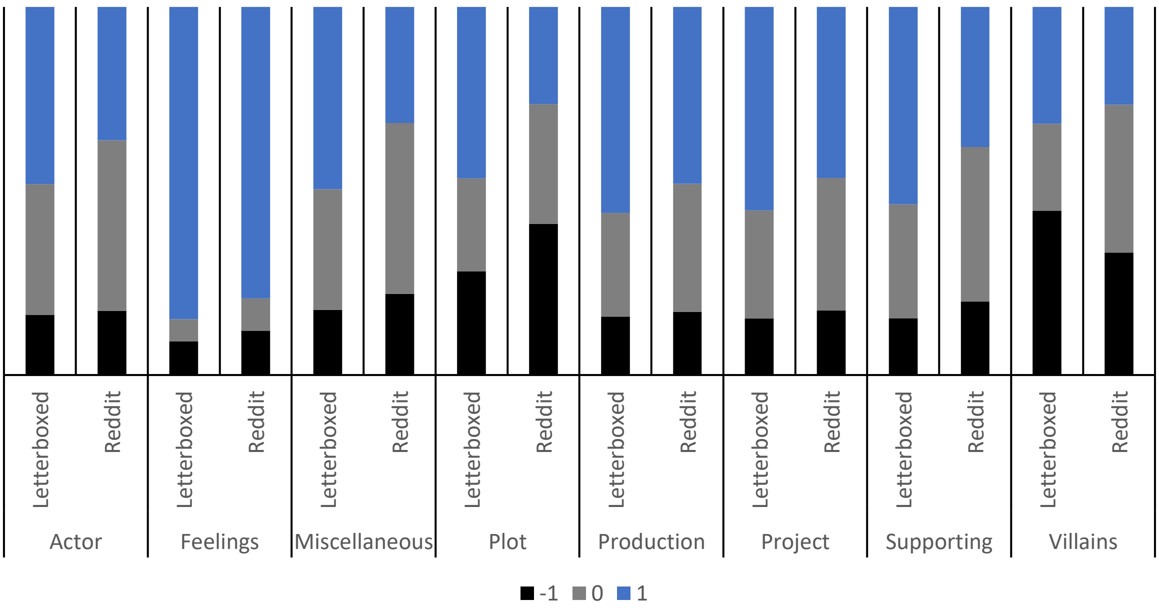

Figure 4.8 - Distribution of top 50% of significant dimension by audience members

By taking in consideration the last table we can make a few analyses, first the dimensions “Actor”, “Miscellaneous”, “Production”, “Project” and “Supporting” all seem similar with each having a few differences, but in these dimensions, Reddit always has more neutral mentions and less positive mentions, the negative mentions are balanced between the two sources. The dimensions “Feelings” also follows a similar trend to the previous dimensions but in this case, there is an overwhelming positive mention by both sources. In “Plot” Reddit has more negative mentions, while having fewer positive mentions and the neutral seem equal. Lastly, in “Villains” this is the only dimensions where Letterboxed has a higher level of negative mentions while having fewer neutral mentions, regarding the positive mentions they are almost the same as the ones from Reddit.

Using the same logic as in the previous test of H1, we also build a scatter plot (figure 4.9) with the weighted average and the same type of metric but now in between LB and Reddit instead of Critics and Audiences, (% Pos LB - % Neg LB) - (% Pos Reddit - % Neg Reddit) by project. When analyzing this figure, it is possible to observe that in nine projects we can reject H2 and assume that the opinions of audiences differ depending on its source. Of these nine projects seven are Tv Shows (“She-Hulk: Attorney at Law”, “Ms.Marvel”, “Moon Knight”, “The Falcon and the Winter Soldier ”, “WandaVision”, “Hawkeye” and “Loki”), that is almost all of the Tv Shows with the exception being “What if…?” that was explained previously. The two other projects where H2 is rejected are “Doctor Strange in the Multiverse of Madness” and “Spider-Man: No Way Home”. We can also examine that all the project’s reviews from LB tend to be more positive than the ones from Reddit.

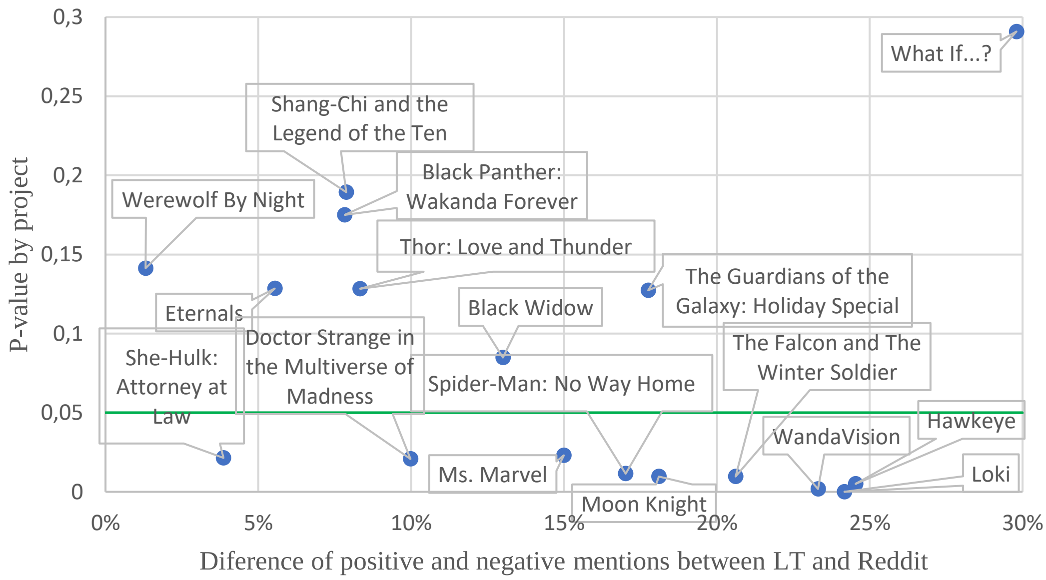

Figure 4.9 - P-values Vs Difference between different Audiences

4.4 What is important for both groups

As was previously mentioned, to study what is important to each group it was decided to do a decision tree, the elected model was the C5.0 and the parameters were previously disclosed. In the next figure 4.10 it is possible to examine what variables were considered and what it is important in the decision tree, the three more important variables are “Plot_Flag”, “Hero_Flag” and “Project_Flag” with importance scores near 0,2.

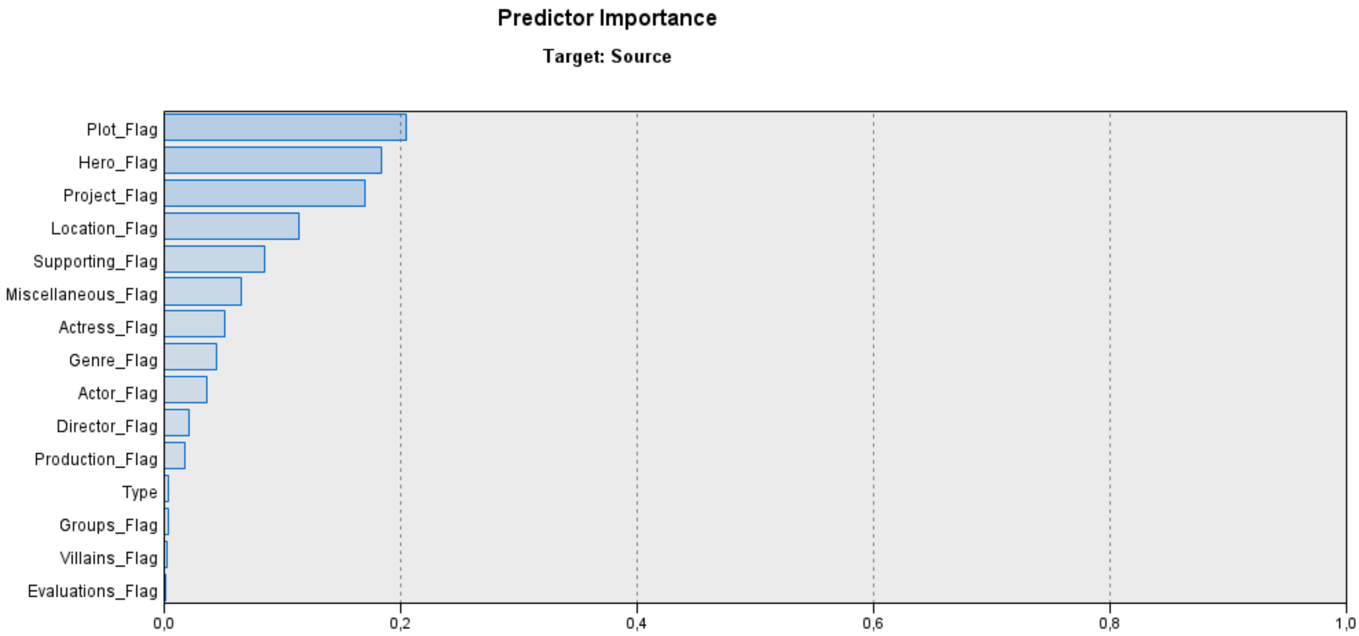

Figure 4.10 - Predictor importance

When analyzing the performance of this model we must look at two confusion matrix that are presented in table 4.8 and 4.9. At first sight most audiences and critics were correctly scored according to the model, but we must take a closer look at the values utilizing the metrics presented in the methodology.

Table 4.8 - Confusion matrix for training

Training	Audience	Critic
Audience	156976	1762
Critic	17329	141659

Table 4.9 - Confusion matrix for testing

Testing	Audience	Critic
Audience	67600	736
Critic	7477	60593

As we can see in the table 4.10 the values of the metrics don’t vary much between training and testing, the only difference being in precision and specificity with 0,1 pp difference.

Table 4.10 - Metrics of the model

	Sensitivity	Precision	Specificity	NPV
Training	98,9%	90,1%	89,1%	98,8%
Testing	98,9%	90,0%	89,0%	98,8%

To make sure this model is good for our study we must adapt the confusion matrix and the metrics to our real data set, has the values previously shown are related to the balanced data set, with this we get the following confusion matrix (table 4.11) and the respective metrics (table 4.12).

Table 4.11 - Confusion matrix adapted to data set

	Audience	Critic
Audience	224576	2498
Critic	417	3401

Table 4.12 - Metric adapted to data set

Sensitivity	Precision	Specificity	NPV
98,9%	99,8%	89,1%	57,7%

It is noticeable that the sensitivity and specificity didn't change much in the real data set, precision got higher, but this is simply an effect of the data not being balanced here. The biggest difference is in the NPV that went from a 98,8 % to 57,7 %, this means that of the reviews that were scored as critics, 2498 are from audiences. This might seem that we should not go forward with this model but because our specificity is high 89,1 % and a comment from a previous study where the author (Beaudouin & Pasquier, 2017) said that the more an author writes reviews the more they tend to appear to be written by a professional critic and when we look at the percentage of audience member classified as critics (1.1%) it is not impossible to think that they can write reviews similar to critics.

Now to see what critics and audiences feel is more important we will look at the rules generated by the model; we will only analyze rules with a confidence higher than 50% on the balanced data set. When selecting the rules with a confidence higher than 50% we get ten rules that lead to an audience review and eleven that lead to a critic review, because of this selection and that could not be defined by a rule, 8565 audience reviews and 471 critic reviews are left without a rule.

Looking at the rules that resulted in the next table 4.13 it is possible to see the rules that result in audience reviews, all of the rules have near 100% confidence, with the biggest difference being 0,5 pp. According to this table the easiest way to determine if a review is from the audience is if they talk about the locations in the movie and the plot but do not mention words contained in the project dimension, this dimension has words like “movie”, “premiere”, “episode”, “season” and others, as this rule leads to 86.6% of all audiences. All the other rules represent less than 5% of audiences, with none of these rules representing more than 3% of the audience, with this it was decided to not analyze this table any further.

Table 4.13 - Rules for audience

Rule nº	Rule	Support	Confidence
1	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 0 then Audience	568	100%
2	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 0 then Audience	196568	99,9%
3	if Location_Flag = 1 and Plot_Flag = 0 then Audience	1355	99,6%
4	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 0 and Miscellaneous_Flag = 1 and Project_Flag = 1 and Plot_Flag = 0 then Audience	4297	99,9%
5	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 0 and Miscellaneous_Flag = 1 and Project_Flag = 0 then Audience	578	99,7%
6	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 0 and Miscellaneous_Flag = 0 then Audience	532	100%
7	if Location_Flag = 0 and Actress_Flag = 0 and Genre_Flag = 1 and Actor_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 0 then Audience	6476	99,9%
8	if Location_Flag = 0 and Actress_Flag = 0 and Genre_Flag = 1 and Actor_Flag = 1 and Supporting_Flag = 0 then Audience	580	99,5%
9	if Location_Flag = 0 and Actress_Flag = 0 and Genre_Flag = 1 and Actor_Flag = 0 then Audience	836	99,9%
10	if Location_Flag = 0 and Actress_Flag = 0 and Genre_Flag = 0 then Audience	6719	99,5%

Next, we will examine table 4.14, which presents the rules defining a review by critics, comparing to the last table this one has several different results. This is due to the 2498 reviews from the audiences that were scored as critics reviews. One thing to remember is that we were not able to identify rules for 471 reviews from critics, this is equivalent to approximately 11% of all their reviews. Another thing that we can quickly notice is that the rules in this table are more complex than the ones represented in the previous table, this means that critics in general are more restrictive than the audiences in their reviews. As such there are only two reviews that we can comfortably say that mostly represent critics and those are rule one and five.

Starting by rule number one it has a support of 2859 with a confidence of 78,3%, this results in 2238 reviews by critics scored as critic reviews. Reviews that obey this rule talk about the location, the plot, the hero, the supporting characters, the production, the actress and mention words related to the project and miscellaneous dimensions, they also do not mention words in evaluations. This rule represents approximately 58% of critics.

Table 4.14 - Rules for critics

Rule nº	Rule	Support	Confidence
1	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 1 and Evaluations_Flag = 0 and Production_Flag = 1 and Actress_Flag = 1 then Critic	2859	78,3%
2	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 1 and Evaluations_Flag = 0 and Production_Flag = 1 and Actress_Flag = 0 and Actor_Flag = 1 and Director_Flag = 1 and Type in ["Movie"] then Critic	201	15,9%
3	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 1 and Evaluations_Flag = 0 and Production_Flag = 1 and Actress_Flag = 0 and Actor_Flag = 1 and Type in ["Special Presentation" "Tv Show"] then Critic	202	14,4%
4	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 1 and Evaluations_Flag = 0 and Production_Flag = 1 and Actress_Flag = 0 and Actor_Flag = 0 and Groups_Flag = 1 then Critic	146	18,5%
5	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 1 and Evaluations_Flag = 0 and Production_Flag = 0 and Actress_Flag = 1 then Critic	842	60,9%
6	if Location_Flag = 1 and Plot_Flag = 1 and Project_Flag = 1 and Hero_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 0 and Groups_Flag = 1 then Critic	232	15,1%
7	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 1 and Actor_Flag = 1 and Miscellaneous_Flag = 1 then Critic	102	48,0%
8	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 0 and Miscellaneous_Flag = 1 and Project_Flag = 1 and Plot_Flag = 1 and Type = Special Presentation then Critic	192	18,2%
9	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 0 and Miscellaneous_Flag = 1 and Project_Flag = 1 and Plot_Flag = 1 and Type = Tv Show and Actor_Flag = 1 and Villains_Flag = 1 then Critic	833	45,1%
10	if Location_Flag = 0 and Actress_Flag = 1 and Director_Flag = 0 and Miscellaneous_Flag = 1 and Project_Flag = 1 and Plot_Flag = 1 and Type = Tv Show and Actor_Flag = 1 and Villains_Flag = 0 and Genre_Flag = 1 then Critic	14	14,3%
11	if Location_Flag = 0 and Actress_Flag = 0 and Genre_Flag = 1 and Actor_Flag = 1 and Supporting_Flag = 1 and Miscellaneous_Flag = 1 and Groups_Flag = 1 and Villains_Flag = 1 then Critic	276	23,6%

Rule number has a support of 842 with a confidence of 60,9%, which results in 513 reviews wrote by critics being classified as a critic review. This rule is defined by reviews that talk about the locations, the plot, the hero, the supporting characters, and actress, they also mention words in the following dimensions project and miscellaneous. Finally, they do not talk about things related to the production of the project nor do they mention words related to evaluations. This rule is like the number one, with the exception that in this one they do not talk about the project production. This rule represents 13 % of all critic reviews.

Even though rule number 9 has a confidence of 45,1%, as it represents 10% of the critics reviews it was also considered important to analyze. This rule has a support of 833 and the confidence previously mentioned, this results in 376 reviews by critics scored as critics. The reviews selected by this rule talk about only Tv Shows and actress, actors, plot and villains, they also mention words in miscellaneous and projects dimensions, and they do not talk about the director and locations.

These three rules allow us to represent approximately 82% of the critics. The rule representing audiences has three restrictions, while the rules defining critics have an average of nine restrictions. This indicates that critics cover three times as many topics as audience reviews.

5. Chapter 5: Conclusions

5.1 Conclusions

In the movie and tv industry as it is demonstrated by previous studies critics and audiences are not in agreement, this dissertation proposed to study this phenomenon and for that created three investigation questions. In a way to study a specific population it was decided to study the "Marvel Cinematic Universe", more specifically the phase 4 of this cinematic universe, which was composed by 17 projects, of these 7 were movies, 8 were tv shows and 2 were special presentations. With this problem in mind the research tried to answer to three questions that can be summarized into the following: do audiences and critics have different opinions for different projects, do the opinion inside of a group (example: audiences) changes if we change platform and what is that each group considers important when sharing their opinion.

To understand the state of the art related to the theme of how the opinions of audiences and critics was presented online and how it was analyzed, for this the PRISMA methodology was utilized as a guiding hand on the execution of the SRL. We started with 119 scientific articles of the last decade from the platform web-of-science and through a series of inclusion and exclusion criteria we narrowed the articles to nineteen. Most of the articles were related to either opinion mining and model creation/testing or how the opinions affected sales, but we only found 3 articles where the difference between critics and audiences so we can assume that it is a topic not well explored. It was also through the SRL that we saw various similar methodologies and how it inspired the one used in this study.

For this research we collected data from 3 different sources, Letterboxed and Reddit for audiences and Rotten Tomatoes for critics, compared to previous studies this was the second largest data based used in a study about this subject. To answer the investigation question, first the data had to be cleaned and processed, afterward the algorithms TF-IDF and VADER were executed to identify what words were relevant in each review and if there was a sentiment associated to the relevant word for each review, before analyzing the data we still did one more step in which we reduce the number of fields (the relevant words) in to dimensions. Then the data was described and with hypothesis testing and clustering we answered the questions.

First, we confirmed that audiences and critics have different opinions in general answering the first question, especially when they talk about the heroes, feelings, supporting character and miscellaneous things in the projects, of all the dimensions there is one where they didn't disagree and one that couldn't be analyzed because only one side had data related to the dimension. There are 9 dimensions in where they disagree in more than 50% of the projects and the project where they did not agree the most was “WandaVision” with 14 dimensions where it was considered that critics and audiences have different opinions.

Curiously enough the audiences also disagree with themselves if they are on different platforms and with this answering the second question, the dimension where they agree the least is “Hero”. It is also noticeable that in this comparison the audiences do not agree with themselves in more than 50% of the projects in 9 dimensions. When comparing the differences where they disagreed more than 50% of the projects by dimensions between critics vs audiences and letterboxed vs reddit they have eight dimensions in common, with the odd ones being “Genre” for critics vs audiences and “Villains” for letterboxed vs reddit As it was between critics and audiences, when we face audiences against each other the project where they disagreed the most is also “WandaVision”.

Regarding the last investigation question, it was discovered one rule to identify audience’s reviews and 3 that can identify reviews from critics, even though that a few audiences reviews were considered critics reviews by the decision this it is probably related to the fact that as a person writes more they start to write more like a critic. Supporting this fact through the rules it is noticeable that audiences have three times less restrictions than critics, and it is that as a person starts to review more projects, they start to pay attention to other details as such their restrictions would grow. With the rules analyzed it is possible to identify 86,6% of audience’s reviews and 81,9% of critic’s reviews and with this we can assume that we can identify to where each review belongs.

5.2 Limitations and Future Studies

One of the biggest limitations of previous studies was the data, in this study this continues to be a limitation but not because of the lack of data in terms of quantity but as a lack of diversity in terms of sources, many possible sources are starting to lock the access to api and other methods of extracting the data behind paywalls, this is one of the reasons as why there are no reviews from IMDb and RT in this study. Even one data source used in this study, reddit, started to charge for access to their api this summer, but for now access to research for non-commercial and academic research is free.

To further support the findings in this thesis new studies with new platforms for audiences and critics would be interesting, it would also be interesting to gather data from previous phases and future phases of the MCU and see what analysis can be done, as the results found during this study are limited to these specific projects and platforms. Besides more projects and platforms, it would also be valuable to analyze how reviews evolve over time and whether they change their content in their content.

Another limitation found is the creation of a more complete dictionary about this matter, a lot of words are still categorized as unidentified or miscellaneous, one interesting path that could be followed is using artificial intelligence to categorize the words and maybe give the scripts of each project so that the artificial intelligence could have a deeper context of the meaning of each word used.

6. Bibliography

Basuroy, S., Abraham Ravid, S., Gretz, R. T., & Allen, B. J. (2020). Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues. *Journal of Cultural Economics, 44*(1), 57-96. doi:10.1007/s10824-019-09350-7

Beaudouin, V., & Pasquier, D. (2017). Forms of contribution and contributors' profiles: An automated textual analysis of amateur on line film critics. *New Media & Society, 19*(11), 1810-1828. doi:10.1177/1461444816643504

Box Office Mojo, M. B. (2023). *MCU Box Office Mojo*. Retrieved from Box Office Mojo by IMDbPro: https://www.boxofficemojo.com/franchise/fr541495045/?ref_=bo_frs_table_1

Caldas, R., Mundt, M., Potthast, W., Neto, F. B., & Markert, B. (2017). A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. *Gait & Posture, 57*, 204-210. doi:https://doi.org/10.1016/j.gaitpost.2017.06.019

Chen, C., Xu, B., Yang, J.-H., & Liu, M. (2022). Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning. *Computational Intelligence and Neuroscience, 2022*, 1-8. doi:10.1155/2022/8517205

Cheng, L.-C., & Huang, C.-L. (2020). Exploring contextual factors from consumer reviews affecting movie sales: an opinion mining approach. *Electronic Commerce Research, 20*(4), 807-832. doi:10.1007/s10660-019-09332-z

Deng, T. (2020). Investigating the effects of textual reviews from consumers and critics on movie sales. *Online Information Review, 44*(6), 1245-1265. doi:10.1108/OIR-10-2019-0323

Hsu, Y.-L., & Jane, W.-J. (2016). Bidirectional Causality for Word of Mouth and the Movie Box Office: An Empirical Investigation of Panel Data. *Journal of Media Economics, 29*(3), 139-152. doi:10.1080/08997764.2016.1208206

Hung, C. (2017). Word of mouth quality classification based on contextual sentiment lexicons. *Information Processing & Management, 53*(4), 751-763. doi:10.1016/j.ipm.2017.02.007

Jacobs, R. S., Heuvelman, A., Ben Allouch, S., & Peters, O. (2015). Everyone's a critic: The power of expert and consumer reviews to shape readers' post-viewing motion picture evaluations. *Poetics, 52*, 91-103. doi:10.1016/j.poetic.2015.07.002

Laureano, R. M., & Santos, M. R. (2021). COVID-19-Related Studies of Nonprofit Management: A Critical Review and Research Agenda. *VOLUNTAS: International*

Journal of Voluntary and Nonprofit Organizations, 33, 936-951. doi:https://doi.org/10.1007/s11266-021-00432-9

Legoux, R., Larocque, D., Laporte, S., Belmati, S., & Boquet, T. (2016). The effect of critical reviews on exhibitors' decisions: Do reviews affect the survival of a movie on screen? *International Journal of Research in Marketing, 33*(2), 357-374. doi:10.1016/j.ijresmar.2015.07.003

Li, L., Goh, T.-T., & Jin, D. (2020). How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Computing and Applications, 32*(9), 4378-4415. doi:10.1007/s00521-018-3865-7

Luan, Y., & Kim, Y. J. (2022). An integrative model of new product evaluation: A systematic investigation of perceived novelty and product evaluation in the movie industry. *PLOS ONE, 17*(3). doi:10.1371/journal.pone.0265193

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews, 10*, 2046-4053. doi:10.1186/s13643-021-01626-4

Peñalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M. Á., Moreno, V., Fraga, A., & Sánchez-Cervantes, J. L. (2014). Feature-based opinion mining through ontologies. *Expert Systems with Applications, 41*(13), 5995-6008. doi:10.1016/J.ESWA.2014.03.022

Sandwell, I., & Longridge, C. (2023, 5). *Marvel's 'Phases' explained: What is Phase 5? What happened in Phase 3?* Retrieved from Digital spy: https://www.digitalspy.com/movies/a871567/marvel-phase-4-3-2-1-mcu/

Schneider, F. M., Domahidi, E., & Dietrich, F. (2020). What Is Important When We Evaluate Movies? Insights from Computational Analysis of Online Reviews. *Media and Communication, 8*(3), 153-163. doi:10.17645/mac.v8i3.3134

Sharma, S. S., & Dutta, G. (2021). SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination. *Information Processing & Management, 58*(1). doi:10.1016/j.ipm.2020.102412

Verboord, M. (2014). The impact of peer-produced criticism on cultural evaluation: A multilevel analysis of discourse employment in online and offline film reviews. *New Media & Society, 16*(6), 921-940. doi:10.1177/1461444813495164

Vujić, S., & Zhang, X. (2018). Does Twitter chatter matter? Online reviews and box office revenues. *Applied Economics, 50*(34-35), 3702-3717. doi:10.1080/00036846.2018.1436148

Wallentin, E. (2016). Demand for cinema and diverging tastes of critics and audiences. *Journal of Retailing and Consumer Services, 33*, 72-81. doi:10.1016/j.jretconser.2016.08.002

Wu, T., Hao, F., & Kim, M. (2021). Kim, Mijin. *Entertainment Computing, 36*. doi:10.1016/j.entcom.2020.100391

Yu, X. (2022). The Identification and Dissemination of Creative Elements of New Media Original Film and Television Works Based on Review Text Mining and Machine Learning. *Mathematical Problems in Engineering, 2022*, 1-9. doi:10.1155/2022/5856069