# Black Panther vs The Help, the battle of Authenticity and Inclusiveness versus tokenism:

## An examination of representation of ethnic minorities in Hollywood films


**Name:** Bastiaan van Houten

**SNR:** 2082816

**Study program:** MSc Marketing Analytics

**Academic year:** 2023-2024

**Submission date:** January 5, 2024

**Name supervisor** Dr. Arjen van Lin

**Name co-reader** Prof. Dr. Marnik Dekimpe

Management summary

Racial Representation in Hollywood films is a subject sparking increasing debate, yet its true impact remains elusive. This paper delves into the complexities of Racial Representation in American films, emphasizing the necessity to move beyond tokenism towards Authentic and Inclusive Representation, capturing the nuances of effective portrayals.

The study challenges traditional metrics of film success, advocating for Long-term Audience Engagement as a more holistic measure compared to the conventional box office. This metric, rooted in lasting relationships between audiences and films, is argued to be more relevant in the social context of Racial Representation.

Focusing on three major ethnic minority groups – Asian, Black, and Hispanic – the study uncovers a nuanced relationship between Long-term Audience Engagement and Racial Authentic Inclusive Representation. The impact varies across ethnicities, emphasizing the importance of considering the specific cultural context and experiences of each group.

Furthermore, the dynamic nature of this relationship is observed in a second study in which time is incorporated as an external factor. The changing effects before and after #OscarsSoWhite suggest that the perception and impact of Racial Authentic Inclusive Representation differs over time, underscoring the need for continuous evaluation.

In essence, this thesis contends that a comprehensive understanding of Racial Representation requires a shift in perspective, considering nuanced metrics and accounting for the evolving nature of societal attitudes over time.

# Preface

I am delighted to present this master's thesis, which encapsulates my academic journey and reflections. I have always found a special interest in films and therefore I was delighted to hear I was able to perform research in this inspiring industry.

2023 was an extraordinary year for me. Besides working on my master's degree at Tilburg University, I also experienced the joy of becoming an uncle for the first time when my sister gave birth to a beautiful baby girl. It was a whirlwind of emotions, and just three months later, my brother welcomed his first child, a son.

Firstly, I extend my deepest gratitude to my thesis advisor, Arjen van Lin for his invaluable insights, constructive feedback, and encouragement throughout this research process. His expertise have been a guiding force, steering this project to a successful end. I like Arjen as a person as well and I am grateful he also brought me under the attention of an opportunity in academics at Tilburg University.

I would like to thank my volleyball coach Kasper Langendoek , he gave me the opportunity to play on a reasonably high level providing the sometimes much needed break while also prioritizing my studies throughout the year.

Lastly and most importantly, to my family, friends and my amazing girlfriend Jayliani, who have been a pillar of strength and a source of unwavering support, I extend my heartfelt thanks. Your encouragement and understanding have been a source of motivation during the challenging moments of this academic journey.

I hope this thesis is fun to read and inspiring. Thank you for your interest!

# Table of Contents

# Introduction

## Problem Indication

The film industry is a major part of our global economy. Films generate nearly $100 billion annually through theaters and home channels. Moreover, the economic impact is evident in successful products. For instance, Disney's "The Force Awakens" earned a net profit of $780 million (MPA, 2022), displaying that films are a serious, high-return industry despite their light-hearted nature. Furthermore, films do not only hold economic significance, they are related to cultural aspects of society.

Films are not only a form of entertainment, they reflect and shape the cultural context in which they are created. They mirror societal values, attitudes, and perspectives and can effectively convey ideas, emotions, and perspectives through their visual and narrative nature. As a result, they have the potential to shape how people perceive the world and the cultures around them (Belton, 1995; Kidd , 2015).

In today's increasingly diverse and socially conscious world, this responsibility currently translates that the Hollywood film industry is under increasing pressure to be socially conscious and to address issues of representation, particularly in relation to gender and people of color (Sperling, 2021). The common consensus is that representation in films is important because it counteracts narrow viewpoints regarding groups depicted in the media, which can reinforce negative stereotypes and biases that result in discrimination and marginalization (Neff et al., 2023; Castañeda, 2017; Ross, 2019; Kubrak, 2020; Buchanan and Hoffner, 2005; Kidd , 2015).

Moreover, representation in films is not only important for social factors, studies show films with diverse casts appeal to broader audiences and perform better at the box office (Neff et al., 2023). For example, a 2020 study by the University of Southern California found that films with diverse casts were 1.4 times more likely to be seen by broader audiences (Smith et al., 2020). Additionally, a 2021 McKinsey & Company study found that films with casts which consisted of at least 30% minority were 1.3 times more likely to be profitable (McKinsey, 2021). By prioritizing diversity and representation, filmmakers can create more inclusive and authentic stories, attract larger audiences, and boost their bottom line (Whitten, 2019). This suggests that racial diversity is both a moral and financial imperative for the film industry.

Efforts to increase diversity and representation in film have led to a significant increase in the proportion of films featuring minority actors. For example, UCLA data shows that between 2011 and 2021, the percentage of films with predominantly minority casts increased from 2% to 32.1% (Hunt and Ramón, 2021). Nevertheless, the public, media, and most studies adopt a narrow approach measuring diversity (e.g., Malik et al., 2021; Kuppuswamy and Younkin 2016; Kim et al., 2020).

Most studies do not make a distinction between different ethnic groups and measure diversity based on the share of ethnic minorities in the cast. Additionally, some studies focus solely on one minority group instead researching all the groups simultaneously. These narrow approaches are assumed to have caused two major issues.

Firstly, the approach of treating minority groups as a homogenous unit has resulted in uneven representation across the different minority groups. For instance, while African-Americans have been overrepresented in films for the past three consecutive years and the Asian community is 'rightfully' represented, the Latin community remains severely underrepresented, with 6.8 % of the actors and actresses being Hispanic compared to an actual population of 20% in the United States. (Hunt and Ramón, 2021; MPAA, 2021). As a result, this group still experiences the social consequences of being underrepresented.

Secondly, while measuring diversity based solely on minority share serves as a useful initial step, it falls short in capturing crucial information about the quality of representation. To comprehensively address the social dimensions of Racial Representation, Authenticity, and Inclusiveness become paramount(Malik et al., 2021; Lazar et al., 2020; Roughton, 2014). This underscores the necessity for films to transcend tokenism and instead craft meaningful storylines that authentically represent characters from ethnic minority backgrounds..

In addition to the limitations associated with the previous metrics used for Racial Representation, there are also constraints related to the metrics employed to measure a film's success. As previously discussed, Racial Representation in film has many social objectives. Despite this, most studies predominantly utilize box office revenue as the dependent variable to assess a film's success (e.g., Kim et al., 2020; Kuppuswamy and Younkin, 2016; Madongo and Zhongjun, 2023).

While box office revenue is a valuable indicator of commercial success, it may not adequately capture the film's quality or its impact on audiences. To comprehensively understand a film's influence, focusing on social aspects, Long-term Audience Engagement is chosen as the dependent variable. As elaborated later in this thesis, this concept encompasses both social and economic constructs, making it a more nuanced and fitting measure for a film's success within the framework of a social concept, such as Racial Representation.

Finally, through the analysis it was suggested that the main relationship is dynamic. For instance, the amount of films having Racial Authentic Inclusive Representation for ethnic minorities was significantly different prior and after 2015 (Lazar et al., 2020). This suggested that their might be external factors for which the main relationship might be sensitive to. Therefore, in this thesis there are two studies. The relationship between Authentic Inclusive Representation for ethnic minorities and Long-term Audience Engagement will be analyzed in Study 1. Study 2 will provide more dimensionality to the relationship, conducted to capture the possible dynamic nature of the main relationship researched.

Building on this perspective, the current literature has yet to undertake a fully comprehensive analysis of the effects of Racial Representation. While emphasizing diversity in the cast is undoubtedly a positive step toward showcasing Racial Representation, studies must avoid relying on overly simplistic methods to measure its impact. The critical factors of authenticity and inclusiveness demand careful consideration in the analysis of Racial Representation. It becomes evident that a more socially nuanced construct is warranted to measure a film's success. Furthermore, incorporating possible external factors is essential in this analysis to capture the broader context and influences at play.

**Problem Statement**

Following the problem background this studies problem statement is formulated as follows: "*What is the relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement in Hollywood films?"*

**Academic Contribution**

Previous studies on racial diversity in films have taken a narrow approach, either by focusing on only one or two ethnic groups (e.g., Patel, 2015; Hall, 2020; Dixon and Linz, 2000; Kuppuswamy and Younkin, 2016) or by grouping all minority groups together (e.g., Aumer, 2017, Lazar et al., 2021). This study addresses these limitations by focusing on three ethnic minority groups (Asian, Black and Hispanic) distinguishing between them and analyzing them simultaneously. This is important because different minority groups have different experiences and perspectives, and their representation in films should be considered separately. Moreover, analyzing the minority groups with this approach improves the generalization of the results.

Second, as noted by Malik et al. (2021, p. 1), "there are no clearly defined, standardized, and scalable metrics for taking stock of racial minorities cinematographic representation". Previous studies such as Neff et al. (2023) and Lazar et al. (2020) have used manual annotation, which produces high-quality insights, but is time-consuming and expensive. This study builds on prior research on gender biases in film (Agarwal et al., 2015), and seeks to standardize the concept of Racial Authentic Inclusive Representation using the Bechdel-Wallace test (1985).

Third, by using Long-term Audience Engagement, this study uses a more comprehensive measure of success, as it takes into account factors such as word-of-mouth, social media engagement, and re-watching. This contributes to existing literature because it provides a more informed understanding of how audiences are engaging with films with diverse casts. Moreover, this study in contrast to other studies identifies films that are having a lasting impact on audiences, even if they may not have been box office blockbusters.

By combining the work of previous researchers, this study aims to make the metric for Racial Representation and a film's success as inclusive as possible while keeping the automated nature in its approach, making it easily applicable to a large number of films. Using this approach, this study provides a more comprehensive and in-depth analysis of the relationship between Racial Representation and a film success.

### Managerial Contribution

This study provides valuable insights for filmmakers, studios, and stakeholders, offering them informed guidance in making decisions that can profoundly influence a film's success within societal contexts. By investigating the correlation between authentically representing characters from various backgrounds, this research aims to demonstrate how these efforts can enhance audience engagement. While the primary focus of this study lies in understanding the societal impact of Racial Representation, it is essential to underscore that Long-term Audience Engagement not only aligns with the broader societal goals but also carries significant financial implications, thereby making it a compelling incentive for the industry.

Moreover, as streaming services surge, the traditional reliance on box office revenue diminishes (Orankiewicz and Bartosiewicz, 2023). This study demonstrates Long-term Audience Engagement potential as a viable measure. Thus, it provides film studios and creators with an additional metric to evaluate the success of their films, acknowledging the industry's shift towards a more comprehensive assessment of sustained impact.

### Structure of the Thesis

To explore the research question posed earlier, the two studies of this thesis adopt a quantitative research approach. The initial study centers on the primary relationship, while the second study delves into the potential influence of an additional factor on this relationship.

The thesis unfolds in a structured manner, commencing with an introduction of a theoretical framework to establish a contextual foundation for the research. This is succeeded by an exploration of the research methodology employed in Study 1. Following these steps, the findings are presented and discussed. After the first study the focus shifts to the second one, presented in a similar order as the first. Both studies are then subjected to two robustness checks, specifically designed to display the differences between these studies and prior research. In the final section, a comprehensive synthesis and discussion are provided, serving as a conclusion to the thesis.

## Literature Review and Hypotheses

This literature review embarks on an examination of Long-term Audience Engagement and Racial Representation, followed by a discussion of the significance of Authenticity and Inclusiveness in ensuring effective Racial Representation.

## Long-term Audience Engagement

Audience engagement is a complex and multifaceted concept that encompasses the involvement of a film's audience. In this study, the focus is on Long-term Audience Engagement which goes beyond merely watching a film and encompasses factors such as word-of-mouth, cultural impact, and sustained interest over time (Broersma, 2019; Kumar and Rubin, 2022). With respect to financial implications, Long-term Audience Engagement is positively correlated with ancillary revenues, including engrossed viewing, prolonged distribution longevity, and the acquisition of film-related merchandise (Kumar and Rubin, 2022). From a social perspective, Long-term Audience Engagement is a crucial element in a film's capacity to influence attitudes, beliefs, and behaviors. It fosters a deeper and more enduring connection between the audience and the film, potentially having a long-lasting impact on their lives (Tan, 2018).

## Racial Representation in Films

The social consequences and opportunities of Racial Representation in films[1] emphasize the influence that filmmakers wield through their creative choices on audiences. The following sections explore existing literature regarding the anticipated impact of Racial Representation on a film's success. It kicks off with a general discussion of Racial Representation in the film industry followed by an explanation of why Authenticity and Inclusiveness deem pivotal attributes.

[1] Representation of minorities in films pertains to the presence and portrayal of characters from minority groups, including people of color, people with disabilities, LGBTQ+ individuals, and other marginalized groups (Buckingham, 2008). Representation is essential because it allows people from different groups to see themselves on screen, which can lead to a greater sense of inclusion, empowerment, and validation (Neff et al., 2023; Dixon, 2000). In contrast, the absence of representation has negative impacts on self-worth, as individuals from underrepresented groups do not see themselves or their experiences reflected in mainstream media (Castañeda, 2015; Ross, 2019; Kubrak, 2020; Buchanan and Hoffner, 2005). Moreover, negative portrayals, such as associations with criminality, can exacerbate negative stereotypes and widen divisions among ethnic groups (Abraham and Appiah, 2006; Hurley et al. 2015). In multicultural societies, building positive relationships among diverse groups presents a significant challenge. In densely populated areas characterized by ethnic segregation, individuals often encounter other cultures and ethnicities solely through media portrayals (Kidd, 2015). Mastro et al. in 2007 found that white people exposed to negative racial stereotypes in the media were more likely to hold those stereotypes themselves. This was especially true for the ones who do not have much real-life contact with people of color, When we're constantly exposed with negative portrayals of certain groups of people, it is hard not to start believing them. In contrast, positive depictions of communities of color can diminish feelings of threat and social distance among white audiences (Dalisay and Tan, 2009; Ortiz and Harwood, 2007) The spread of positive and accurate portrayals in the media is therefore essential for people to explore their identities with regards to race.

# Study 1 : The Relationship between racial Authentic Inclusive Representation and Long-term Audience Engagement

**Conceptual framework and hypothesis.**

This section delves into the existing literature concerning the central relationship under examination. As the following sections explore existing literature regarding the anticipated impact of Racial Representation on a film's success, it becomes evident that portraying minority characters with Authenticity and Inclusiveness is crucial. This approach helps prevent the perpetuation of harmful stereotypes and may potentially mitigate the negative effects associated with Racial Representation

***The effects of Racial Representation on a film's success***

The power of storytelling lies in its ability to create connections between viewers and characters who possess relatable qualities and admirable traits (Abraham and Appiah, 2006; Hall, 2020). This connection is strengthened when there are similarities in demographic factors such as ethnicity, age, and gender, creating a sense of affinity between viewers and the on-screen portrayals (Hall, 2020). As a result, when individuals see themselves or their own experiences represented in a story, they are more likely to form a strong emotional bond and become engaged with the film and its characters (Abraham and Appiah, 2006; Hall, 2020). This suggests that when a film embraces racial diversity including more ethnic groups, it has the potential to attract a broader audience and foster greater Long-term Audience Engagement.

Moreover, it is argued that diverse casts can better reflect the diversity of the real world, which can help viewers to connect with the characters and the story (Neff and Smith, 2023). Furthermore, ethnic minority groups tend to be more prolific creators of online content (Correa, 2011). When a film successfully resonates with these audiences through Racial Representation, it is more likely to stimulate online discussions, which leads to higher Long-term Audience Engagement (Kumar and Rubin, 2022). This connection is likely to become even more impactful in the future as minority groups continue to grow as a percentage of the total U.S. population (U.S. Census Bureau, 2015).

However, the call for heightened racial diversity in film encounters resistance. Patel (2015) contends that initiatives to enhance racial representation have faced criticism, particularly from individuals apprehensive about change and the heightened visibility of people of color. This resistance is rooted in a persistent culture of colonialism or systematic racism. The preference for the status quo has given rise to the industry practice of whitewashing, a phenomenon illustrated in studios' decisions to allocate smaller production budgets (Weaver, 2011; Smith and Choueit, 2020), a factor shown to be a significant predictor of box-office sales (Eliashberg et al., 2014; Clement et al., 2014).

Nevertheless, recent studies have raised doubts about the idea that white actors are necessary for the financial success of films (Aumer et al., 2017). Causing that the prevalence of whitewashing may be more a product of industry habit than an accurate reflection of audience preferences. Nonetheless, due to the underfunding of diverse films, accurately assessing their potential appeal to audiences becomes a challenging task.

To conclude, it seems there is not a general consensus of the influence racial diversity on a film's success. Numerous studies mention that including a diverse number of ethnicities is expected to broaden audiences, increasing success. However, including diverse characters in films is not without its challenges. The most important concern, tokenism, is discussed thoroughly in the subsequent section.

*Authenticity and Inclusiveness as essential parts of Racial Representation*

The representation of minority groups in films takes on various forms. Some films and studies focus on the experiences of minority characters, while others simply include them as part of a larger cast (Malik et al., 2021). Nevertheless, underrepresented ethnic characters who are introduced without fully developed storylines can have negative influences on a film’s success. King (2020) identifies a segment of the audience that does not oppose Racial Representation per se, but rather prioritizes the quality of the narrative. They argue that tokenistic inclusion of minority characters detracts from the overall story, creating the perception of a forced "racial agenda", that often alienates viewers.

Moreover, overlooking Authenticity and Inclusiveness in Racial Representation can result in films being mistakenly categorized as racially diverse while still perpetuating stereotypes and contributing to marginalization. (Umaña-Taylor et al., 2014; Abraham and Appiah, 2006; Hurley et al., 2015). This can lead to backlash and damage Long-term Audience Engagement if viewers perceive the diversity as being insincere, which can ultimately affect the perceived quality of the film (Smith et al., 2016). Additionally, these harmful portrayals can lead to minority groups disengaging from the media altogether. El Hazzouri's (2019) found that ethnic minorities who viewed public health ads featuring members of their own group were less likely to follow the advice provided compared to those who saw ads with white individuals. The researchers attributed this to the feeling of being negatively stereotyped by the advertisers, having ultimately discouraging ethnic minorities from following the message.

Beyond the general audience, these stereotypes are widely condemned by critics who advocate for avoiding films that perpetuate harmful portrayals, regardless of the representation. Given the influence of critical reception on Long-term Audience Engagement, misrepresentation can significantly impact audience enthusiasm and viewership (Hofmann et al., 2016; Ghiassi et al., 2017; Kuppuswamy and Younkin, 2016; Kumar and Rubin, 2022).

In contrast, to the potential pitfalls of misrepresentation, Authentic and Inclusive portrayals of diverse groups can yield positive consequences. Studies have shown that high-budget films featuring inclusive and accurate representations of underrepresented communities tend to perform better commercially (Lazar et al., 2020). This research highlights the significance of Authenticity and Inclusiveness in Racial Representation, as it fosters empathy, understanding, and a deeper connection between viewers and the characters they encounter on screen.

The pursuit of authentic representation of minorities in films is a longstanding endeavor, driven by a growing body of research suggesting that it has a profound and transformative impact on audiences (Roughton, 2014; Bamford, 2018). Advocates for diversity argue that its multifaceted nature provides a wealth of perspectives and experiences, enriching the Authenticity of characters and challenging stereotypes and assumptions associated with different cultures. This leads to more nuanced and complex representations of people from diverse backgrounds (Smith and Choueit, 2020), which is expected to foster greater Long-term Audience Engagement (Hall, 2020).

Moreover, when diverse individuals are portrayed without the shackles of stereotypes, the media's educational and socially engaging potential is amplified. Genuine understanding and empathy, untainted by prejudice, are key to fostering positive change. This holds true across various media platforms, as demonstrated by Roberts' (2021) study that found news stories depicting diverse cultures and identities authentically were more likely to cultivate empathy and understanding among readers. This likely stems from the ability of authentic and inclusive representation to foster a deeper emotional connection with the characters and the story, rendering the media more compelling and impactful.

In conclusion, while a common consensus on the impact of Racial Representation on a film's success remains elusive, the exploration of diverse perspectives in this study suggests that Authentic and Inclusive representation may effectively counter potential negative effects associated with Racial Representation. Therefore, the hypothesis is:

**H1:** Racial Authentic Inclusive Representation is positively correlated with Long-term Audience Engagement for films.

***Conceptual framework study 1***

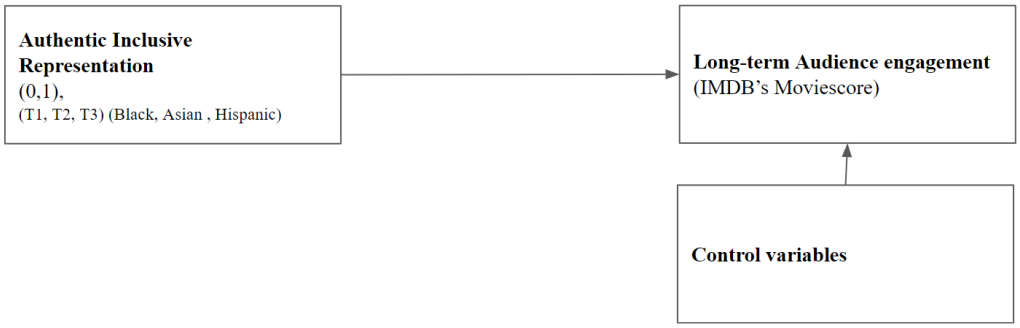



*Figure 1 Conceptual model study 1*

Figure 1 indicates the conceptual model consists of one relationship spanning three distinct levels (see the independent variable operationalization) within three ethnic minority groups: Hispanic, Asian, and Black[2]. These relationships will undergo examination through regression models, as elaborated upon in the subsequent section.

[2] While it is recognized that this study focuses on three ethnicities (Black, Hispanic, and Asian) due to the limitations of the software programs used (Rethnicity and Kairos), it is important to acknowledge that there are many nuances in ethnicity and racial identity.

## Research Methodology study 1

This study aims to investigate the relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement in Hollywood films. The following sections will provide a detailed overview of the data collection and sampling procedures, as well as a clear operationalization of the variables. Given the subjective nature of Racial Authentic Inclusive Representation, the Bechdel Test was employed as a quantitative measure for this concept. Notably, it will be discussed across three distinct levels, as defined by the Bechdel test, elaborated in the operationalization section.

### *Main data sources*

The main data sources for this study are the Internet Movie Database (IMDb), The Numbers, and The Movie Database (TMDB). IMDB is the most important source, as it will be used to measure Long-term Audience Engagement. While IMDb users and the general filmgoing population differ somewhat in demographics, previous research has used IMDB data numerous times demonstrating the capability of using it for the broader filmgoing audience (e.g. Ghiassi et al., 2017; Partha et al., 2019; Apala et al., 2013).

### *Sample*

The films for the sample were chosen carefully and based on specific criteria. IMDB started in 1996 but was not widely used until 2003 so the sample was limited to films released after the year 2000 (for a further explanation, see the dependent variable operationalization).

Films produced outside the United States and animated films were also filtered out, because Hollywood is the focus of this study and voice actors are not represented on screen, undermining representation. Following Joshi and Mao (2012), the analysis only included films that received a wide release, requiring a minimum of 500 screens at their launch.

To measure Racial Authentic Inclusive Representation, data on interactions between actors is required (for a further explanation, see Racial Authentic Inclusive Representation variable operationalization). Observations with missing data which appeared to be random were eliminated. After applying all criteria, the sample consisted of 1,178 films.

### *Variable operationalization*

*Long-term Audience Engagement (Dependent Variable)*

Long-term Audience Engagement was operationalized using IMDb's MovieMeter, a metric derived from popularity rankings. A score of one means that the film was the most popular with regards to clicks, page views and reviews on IMDB in that week compared to other films. Therefore, this film score includes direct indicators of Long-term Audience Engagement, such as online discussions, reviews, and word-of-mouth conversations. To measure sustained engagement, I measured Long-term Audience Engagement as the average ranking of a film over a one-year period, starting in the third year after its release. I chose this metric because it shows how engaged audiences are with a film after a few years, and the average reduces the influence of any spikes or certain drops in the engagement of audiences with a film.

*Ethnicity*

To determine actors' ethnicity, I used the Kairos API, a deep learning algorithm that can detect ethnicity through facial recognition. I collected the profile pictures of the actors and actresses with a scraper from IMDb and processed these with the API. The Kairos API is chosen because of its efficiency, accuracy (99.63%), and ability to handle a large dataset (Kairos, 2023). In the films included in the sample of this study a total of 62,834 characters were present. This count includes instances where an actor or actress appeared in multiple films. It also included uncredited people (7,251). The Kairos API analyzed 26,891 images found on IMDb of the total of 32,513 unique actors and actresses.

For the remaining imageless people the ethnicities were based on their first and last name using the R package Rethnicity, which has an accuracy of around 80%. People not having an image on IMDB is correlated with not having played a significant role in a film [3]. Therefore, they are assumed to be less important for the analysis.

[3]The filtering process showed that imageless characters were more likely to play non named characters. After the filter for generic names, 83% of characters without images were removed, compared to 66% of characters with images. This suggests that important characters were mostly played by people with images. As a result, imageless people are also more likely to play characters that were not Authentic and Inclusive with regards to the criteria used in this study. This suggests that the decision to use Rethenicity for the imageless people is less troublesome due to these actors probably not being important for the analysis regardless.

*Racial Authentic Inclusive Representation (Independent Variables)*

Quantifying the concept of Racial Authentic Inclusive Representation in films has resulted in researchers using the Bechdel-Wallace test (1985), originally designed to measure representation of women in a film (Agarwal, 2015). The Bechdel test is chosen as a tool for measuring Racial Authentic Inclusive Representation in films because of its adaptability to this study's specific focus (Lazar et al., 2020), its potential for automation, and its quantifiability (Argarwal, 2015). The original Bechdel test is structured around three levels, (T1) The movie has to have at least two named women in it (T2) who talk to each other, (T3) about something other than a man (Bechdel-Wallace, 1985)

Building on this foundation, the Bechdel test was adapted, drawing inspiration from Lazar et al. (2020). The measurement of Racial Authentic Inclusive Representation for the three distinct minority ethnicities in this study allows for an examination of the concept across three distinct levels (T1, T2, T3):

(T1) Two named {ethnicity} characters appear in film X.

(T2) Two named {ethnicity} characters appear in a scene together.

(T3) Two named {ethnicity} characters appear in a scene together without a white character.

An example for such a conversation would be a conversation between the main character and his wife collected from the film script '12 years a slave' :

**ANNE**
Solomon...
**SOLOMON**
Come, Anne. Jump.
**ANNE**
I will not ruin my dress. Catch me!
**SOLOMON**
I will catch you, Anne. I will.
**ANNE**
You will.

Even though the conversation is not very thorough or meaningful, it is between two named ethnic minority characters, and no white people are present in the scene. Therefore, the film will have the condition for all levels for the Black ethnicity.

As can be seen the modified Bechdel test similar to the original requires the actors to be named. Therefore, the first step of conducting the modified Bechdel test was to create a process which removed ‘generic’ characters from the dataset. To filter generic characters from the dataset, I identified frequently occurring tokens, where a token is a segment of a name divided by spaces. I then removed characters which only included tokens within a stop word list. The stop word list eventually contained 915 words, such as "doctor," "agent," and "the." The entire list can be seen in Appendix A.

I made an exception to the character filter: I did not remove a character if the stop word was the first token in their name for example "Colonel Rich Bron" and “Doctor Johnson” were characters kept. However, the addition to this exception was when the first token was the only token. For instance, a character such as "Colonel" would be removed.

To delve deeper into Authentic and Inclusive characters I implemented steps 2 and 3 of the modified Bechdel test. Previous research such as Argawal (2015) used screenplays to conduct the Bechdel test. However, film scripts can be changed during production, so they may not match the final film. Therefore, I decided to take a different approach and use subtitles for the hearing impaired. Subtitles represent the final version of the film's dialogue, capturing it exactly as it appears in the film.

By using the subtitles.org API I was able to find subtitles for the hearing impaired in a format that could be standardized for testing. Each subtitle file was in an a strict formatted .srt file, which has a unique identifier, precise start and end times, and one or two lines of text. As illustration, the opening of the film "300: Rise of an Empire":

1
00:00:38,363 --> 00:00:40,698
(HORSE NICKERS)
2
00:01:02,654 --> 00:01:07,024
QUEEN GORGO: The oracle's
words stand as a warning.
3
00:01:07,026 --> 00:01:08,225
A prophecy.
4
00:01:08,227 --> 00:01:11,796
"Sparta will fall.

In these subtitle files, when it is unclear who is speaking, the person is identified and labeled ("QUEEN GORGO"). To ascertain the individuals in the scenes, I employed the en_core_web_md model from the spaCy Natural Language package, which can recognize entities in the text. To identify scenes, I wrote code which went through the subtitle file and looked for pauses in dialogue. If a pause exceeded five seconds, the code identified it as a new scene. I chose five seconds because it allows the audience to mentally adjust to the new scene without disrupting the narrative flow. Because this is a threshold, even if a scene switch took longer, it would still count it as one scene switch. Finally, I conducted a fuzzy merge, aligning the people mentioned list with their respective ethnicities within the dataset.

After processing the subtitles, the dataset comprised scene indices, speakers, individuals mentioned, and the ethnicities of the characters. The subsequent step involved applying the revised Bechdel Test to this dataset. Films were classified based on whether they satisfied the conditions (T1), (T2) and (T3) for different ethnicities. Films which hold the condition T3 directly hold the other conditions as well. These classification served as the independent variables utilized in the regression models.

***Covariates***

It is important to account for additional factors that have been identified as influencing a film's success, in doing so this study draws upon previous research. By controlling for the impact of these variables, more accurate estimations can be derived for the variables under investigation. Moreover, by including these covariates the chances for an omitted variable bias is reduced.

The variables considered alongside Racial Authentic Inclusive Representation in this study include Sequel, the star power of both Actors and Directors, MPAA rating, Number of Opening Screens, Critical Acclaim, Awards, Production Budget, Genre, Source and Seasonality. Table 1 provides detailed information including which sources identified which variables, the following paragraphs briefly discuss how the variables are measured in this study.

The measure for $STARPOWER_i$ is based on the measurement from Nelson and Glotfelty (2012), it is the four highest-grossing actors' ranking on the website The Numbers for one year before the film. For $DIRECTORPOWER_i$ the directors' ranking on The Numbers is also used, also the year before the film.

With regards to critical acclaim, the $CRITICS_i$ value is the average rating on metacritic.com. Moreover, this research will use the actual number of awards $NOMINATIONS_i$ as a proxy for award nomination. $WINS_i$ will also be added to the model to represent awards wins. Because there is an abundance of available film awards this study uses the awards mentioned on the website of IMDB which makes it very easy and accessible to account for a vast amount of awards internationally and nationally.

$MPAA_i$ rating is given by the Motion Picture Association of America and is used to rate a film's suitability for certain audiences based on its content. These ratings are encoded as an interval variable as [0 = unrated; 1 = G; 2 = PG; 3 = PG-13; 4 = R;5 = C-17]. $SCREENS_i$ is the amount of opening theater screens the film had according to the Numbers. $BUDGET_i$ was the production budget available at one of the three data sources used (IMDB, The Numbers and TMDB), if multiple production budgets were available across the sources the average was taken. In this study, $SEQUEL_i$ indicates where the film is a sequel. It is a dummy variable, which takes the value one if the film is a sequel.

Moreover, this study introduced 19 genres through dummy variables: Action, Adventure, Comedy, Crime, Drama, Family, Fantasy, Horror, Romance, Musical, Sci-Fi, Mystery, Thriller, Western, Biography, Documentary, History, Music, Sport and War. Because a film could have multiple genres, these dummy variables are not mutually exclusive.

Within this study the four seasons ($SPRING_i$, $SUMMER_i$, $FALL_i$, $WINTER_i$) are encoded as the following. Spring[March, April, May] Summer [June, July, August] Fall [September, October, November], Winter [December, January, February]. $RUNTIME_i$ is included as the actual numerical value in minutes, following.

Furthermore, following Hofmann, Clement, Völckner, and Hennig-Thurau (2016), multiple dummy variables were added to control for whether the film was $\{BASED\ ON\}_i$ a book, comic, novel, short story, or TV series. Moreover, whether the film is a $REMAKE_i$ or $SPINOFF_i$ . Similar, to variables representing the genres of the film, these based on dummy variables are not mutually exclusive because a film could be based on multiple sources. In order to account for potential variations over time, dummy variables $YEAR_i$ were introduced for each year within the sample, enhancing the model's ability to control for temporal effects.

Table 1: Measures of Variables

| Variable Name | Description | Measure | Data Source | Academic research |
|---|---|---|---|---|
| **Independent variable** | | | | |
| Model 1: $ASIANT1_i$ $BLACKT1_i$ $HISPANICT1_i$ etc. | Racial Authentic Inclusive Representation of the three ethnicities analyzed in this study | Dichotomous variable with three separate levels for the three different ethnicities. | Subtitles.org | L, R |
| **Dependent variable** | | | | |
| $Log(LTAE_i)$ | IMDB ranking where a higher score means being less popular. | Log (Average ranking over a one year period of time in the third year after release). | IMDB | |
| **Covariates** | | | | |
| $Log(STARPOWER_i)$ | The top box office stars are based on the cumulative worldwide box office of all the films a star has had a leading role in over their lifetime. | Log (Sum of the top 4 actors, one year before the release * 100 + 1) | The Numbers | KRM, H, KY, G, CM |
| $Log(DIRECTORPOWER_i)$ | Derived from the list of the highest grossing Directors based on the worldwide box office of the films they worked on. | Log ( The director ranking * 100 + 1) one year before the film release . | The Numbers | KRM, H, G, KY, CM |
| $Log(CRITIC_i)$ | Average rating the film received from professional film reviewers on Metacritic.com. | Log (Average rating the film received from professional film reviewers + 1 ) | Metacritic | H, KRM, KY, G |
| $Log(NOMINATIONS_i)$ | Number of award nominations the film received. | Log(Number of award nominations the film received + 1) . | IMDb | KY,KRM, HTS |
| $Log(WINS_i)$ | Number of award wins the film received. | Log( Number of award nominations the film received + 1). | IMDB | KY,KRM, HTS |
| $Log(MPAA_i)$ | MPAA rating, the film received 1 = Not Rated, 2 = G, 3 = PG, 4 = PG-13, 5 = R, 6 = NC-17 | Log transferred of the interval ranking | IMDb | H, KY, L, BCR, KRM , SD, CM |
| $Log(SCREENS_i)$ | Number of screens at release | Log(Number of screens) | Numbers | H , KY, KRM , HTS, CM |
| $Log(BUDGET_i)$ | Production budget of film | Log(The average production budget of the data sources) | IMDB, Numbers, OMDB | KY,KRM, HTS, CM |
| $Log(RUNTIME_i)$ | Duration of a film in minutes | Log( Duration of a film in minutes) | IMDb | KY |
| $SEQUEL_i$ | Indiciation if the films is a sequel or not | Sequel = 1, Not a sequel = 0 | IMDB | KRM, H , KY, CM |
| $ACTION_i$ | Genre Action | Genre Action = 1, Other = 0 | IMDb | KRM, H , KY, G,CM |
| $ADVENTURE_i$ | Genre Adventure | Genre Adventure = 1, Other = 0 | IMDb | - |
| $ANIMATION_i$ | Genre Animation | Genre Animation = 1, Other = 0 | IMDb | - |
| $COMEDY_i$ | Genre Comedy | Genre Comedy = 1, Other = 0 | IMDb | - |

| | | | | |
|---|---|---|---|---|
| $CRIME_i$ | Genre Crime | Genre Crime = 1, Other = 0 | IMDb | - |
| $DRAMA_i$ | Genre Drama | Genre Drama = 1, Other = 0 | IMDb | - |
| $FAMILY_i$ | Genre Family | Genre Family = 1, Other = 0 | IMDb | - |
| $FANTASY_i$ | Genre Fantasy | Genre Fantasy = 1, Other = 0 | IMDb | - |
| $HORROR_i$ | Genre Horror | Genre Horror = 1, Other = 0 | IMDb | - |
| $MUSICAL_i$ | Genre Musical | Genre Musical = 1, Other = 0 | IMDb | - |
| $MYSTERY_i$ | Genre Mystery | Genre Mystery = 1, Other = 0 | IMDb | - |
| $ROMANCE_i$ | Genre Romance | Genre Romance = 1, Other = 0 | IMDb | - |
| $SCI\text{-}FI_i$ | Genre Sci-Fi | Genre Sci-Fi = 1, Other = 0 | IMDb | - |
| $THRILLER_i$ | Genre Thriller | Genre Thriller = 1, Other = 0 | IMDb | - |
| $WESTERN_i$ | Genre Western | Genre Western = 1, Other = 0 | IMDb | - |
| $BIOGRAPHY_i$ | Genre Biography | Genre Biography = 1, Other = 0 | IMDb | - |
| $DOCUMENTARY_i$ | Genre Documentary | Genre Documentary = 1, Other = 0 | IMDb | - |
| $MUSIC_i$ | Genre Music | Genre Music = 1, Other = 0 | IMDb | - |
| $HISTORY_i$ | Genre History | Genre History = 1, Other = 0 | IMDb | - |
| $SPORT_i$ | Genre Sport | Genre Sport = 1, Other = 0 | IMDb | - |
| $SPRING_i$ | Film released in the spring | [Mar, Apr, May]= 1, Other = 0 | IMDb | V |
| $SUMMER_i$ | Film released in the summer | [Jun, Jul, Aug]= 1, Other = 0 | IMDb | - |
| $AUTUM_i$ | Film released in the autumn | [Sep, Oct, Nov] = 1, Other = 0 | IMDb | - |
| $WINTER_i$ | Film released in the winter | [Dec, Jan, Feb]= 1, Other = 0 | IMDb | - |
| $BOOK_i$ | Film is based on a book | Film is based on a book = 1, other = 0 | IMDb | H |
| $COMIC_i$ | Film is based on a comic | Film is based on a comic = 1, other = 0 | IMDb | - |
| $NOVEL_i$ | Film is based on a novel | Film is based on a novel = 1, other = 0 | IMDb | - |
| $SHORTSTOR_i$ | Film is based on a short story | Film is based on a short story = 1, other = 0 | IMDb | - |
| $TVSERIES_i$ | Film is based on TV series | Film is based on a TV series = 1, other = 0 | IMDb | - |
| $REMAKE_i$ | Film is a remake | Film is a remake = 1, other = 0 | IMDb | - |
| $SPINOFF_i$ | Film is a spinoff | Film is a spinoff = 1, other = 0 | IMDb | - |
| $YEAR2000_i$ | Year2000 | 2000 = 1, other = 0 | IMDb | |
| $YEAR2001_i$ | Year2001 | 2001 = 1, other = 0 | IMDb | |
| $\{.....\}_i$ | Years 2002 – 2017 | | IMDb | |
| $YEAR2018_i$ | Year2018 | 2018 = 1, other = 0 | IMDb | |
| $YEAR2019_i$ | Year2019 | 2019 = 1, other = 0 | IMDb | |

*Notes: CM = Clement et al. (2014) , KRM = Kumar et al. (2022) ; H = Hofmann et al.(2016) ; L = Lazar et al. (2020) ; KY = Kuppuswamy and Younkin (2016) ; HTS = Hunter et al. (2016) ; SD = Ramesh & Delen (2006); BCR = Basuroy et al. (2003), L = Litman , (1983); G = Ghiassi et al. (2017), HL = Hall, (2022) , SM = Smith and Choueit. (2020)*

***Models study 1***

Following the approach outlined by Clement, Wu, and Fischer (2014), this research uses a log-log regression model. As such, all variables that are not dummy variables were log-transformed. However, it is worth noting that numerous continuous variables had zero values, and taking the logarithm of '0' would result in an error. To address this, a small constant value of '1' was added to all variables which were continuous before taking the logarithm, ensuring meaningful results. For the variables director power and star power which were numbers between 0 and 1.5 the value was first multiplied by one hundred before the number was added, assuring the small integer added did not have too much effect on the value. Because Racial Authentic Inclusive Representation provides different levels, models were made for all three levels with the ethnicity conditions as independent variables.

Model T1 :

$$\log(LTAE_i) = \beta0 + \beta1 \times ASIANT1_i + \beta2 \times BLACKT1_i +$$
$$\beta3 \times HISPANICT1_i + \beta4 \times \log(DIRECTORPOWER_i) +$$
$$\beta5 \times \log(CRITIC_i) + \beta6 \times \log(NOMINATIONS_i) + \beta7 \times \log(WINS_i) +$$
$$\beta8 \times \log(MPAA_i) + \beta9 \times \log(SCREENS_i) + \beta10 \times \log(BUDGET_i) +$$
$$\beta11 \times ACTION_i + \beta12 \times ADVENTURE_i + \beta13 \times ANIMATION_i +$$
$$\beta14 \times COMEDY_i + \beta15 \times CRIME_i + \beta16 \times DRAMA_i + \beta17 \times FAMILY_i +$$
$$\beta18 \times FANTASY_i + \beta19 \times HORROR_i + \beta20 \times MUSICAL_i +$$
$$\beta21 \times MYSTERY_i + \beta22 \times ROMANCE_i + \beta23 \times SCI\text{-}FI_i +$$
$$\beta24 \times THRILLER_i + \beta25 \times WESTERN_i + \beta26 \times BIOGRAPHY_i +$$
$$\beta27 \times DOCUMENTARY_i + \beta28 \times MUSIC_i + \beta29 \times HISTORY_i +$$
$$\beta30 \times SPORT_i + \beta31 \times SPRING_i + \beta32 \times SUMMER_i +$$
$$\beta33 \times AUTUMN_i + \beta34 \times WINTER_i + \beta35 \times \log(RUNTIME_i) +$$
$$\beta36 \times BOOK_i + \beta37 \times COMIC_i + \beta38 \times NOVEL_i + \beta39 \times SHORTSTOR_i +$$
$$\beta40 \times TVSERIES_i + \beta41 \times REMAKE_i + \beta42 \times SPINOFF_i + \beta43 \times SERIES_i +$$
$$\beta44 \times \log(STARPOWER_i) + \beta45 \times YEAR2000_i + \beta46 \times YEAR2001_i + \beta47 \times YEAR2002_i +$$
$$\beta48 \times YEAR2003_i + \beta49 \times YEAR2004_i + \beta50 \times YEAR2005_i + \beta51 \times YEAR2006_i +$$
$$\beta52 \times YEAR2007_i + \beta53 \times YEAR2008_i + \beta54 \times YEAR2009_i + \beta55 \times YEAR2010_i +$$
$$\beta56 \times YEAR2011_i + \beta57 \times YEAR2012_i + \beta58 \times YEAR2013_i + \beta59 \times YEAR2014_i +$$
$$\beta60 \times YEAR2015_i + \beta61 \times YEAR2016_i + \beta62 \times YEAR2017_i + \beta63 \times YEAR2018_i + \beta64 \times YEAR2019_i + \varepsilon i$$

Model T2:

$$\log(LTAE_i) = \beta0 + \beta1 \times ASIANT2_i + \beta2 \times BLACKT2_i + \beta3 \times HISPANICT2_i + \beta4 \times \log(DIRECTORPOWER_i) + \ldots \beta60 \times YEAR2015_i + \beta61 \times YEAR2016_i + \beta62 \times YEAR2017_i + \beta63 \times YEAR2018_i + \beta64 \times YEAR2019_i + \varepsilon i$$

Model T3:

$$\log(LTAE_i) = \beta0 + \beta1 \times ASIANT3_i + \beta2 \times BLACKT3_i + \beta3 \times HISPANICT3_i + \beta4 \times \log(DIRECTORPOWER_i) + \ldots \beta60 \times YEAR2015_i + \beta61 \times YEAR2016_i + \beta62 \times YEAR2017_i + \beta63 \times YEAR2018_i + \beta64 \times YEAR2019_i + \varepsilon i$$

In these models, LTAEi represents the Long-term Audience Engagement for film i. ASIANT , BLACKT and HISPANICT represent the Authenticity and Inclusiveness, identification, of Hispanic, Black, and Asian representation in film i. The other variables are the control variables.

**Results study 1**

In this section, the results of the study are discussed. First, to establish a comprehensive overview of the dataset, descriptive and frequency statistics for the variables are presented. It is crucial to note that, as mentioned previously, Long-term Audience Engagement represents the average rank of a film in the third year post-release. Therefore, a higher value of the dependent variable signifies lower Long-term Audience Engagement. Consequently, negative estimates in the results denote positive effects, while positive estimates indicate negative effects on Long-term Audience Engagement.

***Descriptive Statistics***

Table 2 presents descriptive statistics for the dependent variable and the key control variables used in the models (before being logged). This table displays the number of observations, mean, standard deviation, minimum and maximum values.

Table 2: Descriptive statistics

| Variable Name | N | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|
| *Dependent variable* | | | | | |
| Avg Rank Third Year | 1178 | 3410 | 4387 | 75.92 | 103633 |
| *Control Variables* | | | | | |
| Opening screens | 1178 | 2650.06 | 826.48 | 502 | 4662 |
| Runtime | 1178 | 108.69 | 17.19 | 75 | 219 |
| Budget | 1178 | 54,673,873.16 | 53,393,953.25 | 250,000 | 356,000,000 |
| Nominee | 1178 | 16.26 | 36.30 | 0 | 462 |
| Winner | 1178 | 5.99 | 23.02 | 0 | 490 |
| Director power | 1178 | 0.03 | 0.09 | 0 | 1.0 |
| Star Power | 1178 | 0.06 | 0.15 | 0 | 1.2 |
| Metascore | 1178 | 47.66 | 16.46 | 0 | 96 |
| MPAA | 1178 | 3.24 | 0.69 | 1 | 4 |

The table presented below offers a comprehensive overview of the independent variable at levels T1, T2, and T3. This table displays the amount of films passing the condition. In Appendix B tables representing the descriptives for the dummy control variables are presented.

Table 3: Number of films passing T1, T2 and T3 condition,

| Ethnicity : Condition | T1 | T2 | T3 |
|---|---|---|---|
| Black | 503 | 335 | 206 |
| Asian | 334 | 149 | 77 |
| Hispanic | 326 | 199 | 97 |

***Log-log regression model estimates***

With information available for different levels of Racial Authentic Inclusive Representation (T1, T2, and T3), regression analysis was conducted for all three levels. Before running the models, the variance inflation factor (VIF) of all variables in all models were checked to get an understanding of potential multicollinearity. None of the models showed any VIF values exceeding the threshold of 10. To deal with potential heteroscedasticity issues robust standard errors were implemented.

Table 4 presents the results for each model. The last rows show the focal variables and the model fits. It should be noted (again) that the dependent variable, average ranking, is inversely related to film Long-term Audience Engagement. This means that a higher ranking corresponds to lower engagement. As a result, a negative coefficient indicates a positive effect of the variable on film Long-term Audience Engagement and vice versa.

Table 4: Results regression analysis T3, T2, T1

| Variable | T1 | T2 | T3 |
|---|---|---|---|
| Action | -0.064 | -0.066 | -0.07 |
| Adventure | -0.054 | -0.041 | -0.054 |
| Biography | 0.293 *** | 0.289 *** | 0.291 *** |
| Comedy | 0.144 ** | 0.148 *** | 0.152 *** |
| Crime | 0.052 | 0.029 | 0.037 |
| Documentary | 0.936 ** | 0.986 *** | 0.989 *** |
| Drama | 0.137 *** | 0.138 *** | 0.137 *** |
| Family | 0.241 ** | 0.239 ** | 0.228 ** |
| Fantasy | -0.043 | -0.033 | -0.035 |
| History | 0.135 | 0.15 | 0.142 |
| Horror | -0.013 | -0.002 | -0.003 |
| Music | 0.263 ** | 0.248 ** | 0.233 ** |
| Musical | 0.102 | 0.091 | 0.073 |
| Mystery | 0.008 | 0.027 | 0.038 |
| Romance | -0.15 *** | -0.143 ** | -0.15 *** |
| Sci.Fi | 0.009 | 0.002 | 0.005 |
| Sport | 0.037 | 0.002 | -0.006 |

| | | | |
|---|---|---|---|
| Thriller | 0.03 | 0.031 | 0.023 |
| War | 0.081 | 0.068 | 0.063 |
| Western | 0.503 ** | 0.482 ** | 0.487 ** |
| Fall | -0.01 | -0.011 | -0.02 |
| Spring | -0.016 | -0.012 | -0.022 |
| Summer | -0.045 | -0.044 | -0.052 |
| Based on book | -0.101 | -0.108 | -0.102 |
| Based on comic book | -0.215 ** | -0.199 ** | -0.227 *** |
| Based on novel | -0.152 *** | -0.128 ** | -0.131 ** |
| Based on play | 0.311 | 0.343 | 0.352 |
| Based on short story | -0.002 | 0.042 | 0.028 |
| Sequel | 0.102 * | 0.097 * | 0.096 * |
| Spinoff | -0.036 | -0.033 | -0.031 |
| Remake | 0.007 | 0.006 | 0.007 |
| Log(Budget) | -0.031 | -0.032 | -0.025 |
| Log(Screens) | -0.476 *** | -0.464 *** | -0.457 *** |
| Log(Runtime) | -0.737 *** | -0.733 *** | -0.733 *** |
| Log(MPAA) | -0.504 *** | -0.509 *** | -0.533 *** |
| Log(Director Power) | -0.089 *** | -0.102 *** | -0.108 *** |
| Log(Metascore) | -0.327 *** | -0.34 *** | -0.331 *** |
| Log(Nominee) | -0.154 *** | -0.152 *** | -0.155 *** |
| Log(Star power) | -0.034 † | -0.035* | -0.031 † |
| Log(Winner) | -0.147 *** | -0.142 *** | -0.142 *** |
| | | | |
| Asian | -0.002 | 0.013 | 0.036 |
| Black | 0.034 | 0.183 *** | 0.247 *** |
| Hispanic | -0.041 | 0.052 | 0.084 |
| $R^2$ | 0.7211 | 0.7294 | 0.7329 |
| Adjusted $R^2$ | 0.7054 | 0.7142 | 0.7082 |
| F Statistic | 44.51*** | 46.664*** | 47.354*** |

*Significance levels: † p < 0.10 * p < 0.05, ** p < 0.01, *** p < 0.001*

Starting with an assessment of model fit, the R-squared values are notably high across all models. However, comparing the model's adjusted R-squared values with the adjusted R-squared of a model without the T variables (0.70), it could be seen that the variables of focus did not substantially enhance explanatory power.

The control variables were based on studies which mostly used box office as their dependent variable. Most control variables show significant effects in the models from this study, emphasizing their importance not just for short-term success but also for understanding why a film keeps audiences engaged in the long term.

Focusing on the dummy control variables, out of 19 genres, 8 exhibited significant effects, demonstrating diverse positive and negative impacts. Moreover, films based on existing sources such as novels and comic books consistently exhibited negative coefficients across most models, with coefficient estimates ranging from -0.131 to -0.227. Notably, this trend was not observed for films based on short stories, books or plays. Finally, Sequel, a dummy variable which also indicates an existing audience had a positive estimate in all three models.

Zooming in on the continuous variables, consistent significant negative coefficients were observed for Nominee, Winner, Metascore, Director Power, Screens, MPAA, Star Power, and Runtime across all models. Notably, the estimate of Star Power exhibited only marginal significance in two models ($0.05 < \text{p-value} < 0.10$), implying a potential influence on a film's success. These findings align with previous research, except for Budget, indicating that, apart from the box office, all these variables positively correlate with a film's Long-term Audience Engagement.

A noteworthy observation is the insignificance of the seasonality variables. However, this can be rationalized by considering the nature of the Long-term Audience Engagement metric which spans over a period of a year controlling for seasonality. Nevertheless, this does suggest that the initial success of a film due to seasonality, as observed in previous research, may diminish over time and is likely more closely tied to success at the time of release.

Zooming in on the key variables, an examination of the ethnicity variables reveals several interesting findings. The variables indicating Asian and Hispanic representation showed no significant effect in any of the models. Moreover, the variables representing the Black ethnicity exhibited a positive coefficient with average ranking in two out of the three models, indicating a negative effect. In the context of a logged dependent variable, dummy variables should be interpret as a exp(x) – 1 * 100 = percentage change. The modules suggest when two named black characters engage in a conversation the films experiences an overall average ranking increase of 20.6% (T2 module), when this conversations is not accompanied by a white character this is 30% (T3 module) remaining all other variables constant.

For instance, consider the case of the mean film ranking, 3410. Under the specific condition T2, holding all other variables constant, a film is projected to have an average ranking of (3410 * 1.206 =) 4,112. This implies that films meeting the criteria for both T2 and T3 concerning Black ethnicity appear to exhibit reduced Long-term Audience Engagement.

These results, particularly the negative correlation for the Black ethnicity, challenge the hypothesis, which anticipated a positive relationship between Racial Authentic Inclusive Representation in a film and Long-term Audience Engagement. Therefore, while this study has shed light on various factors influencing film success the unexpected outcomes might indicate considering an additional factor. Subsequently, an additional study was conducted, introducing an extra variable to assess the impact of an external factor. This subsequent study, delving into the nuances of Racial Authentic Inclusive Representation's influence on film success, is presented in the following section of this thesis.

Table 5: Hypothesis study 1 table

| **Hypotheses** | **Ethnicity** | **β (T1, T2 , T3)** | **P-Value (T1, T2 , T3)** | **Decision** |
|---|---|---|---|---|
| H1a: Racial Authentic Inclusive Representation is positively correlated with Long-term Audience Engagement for films. | Asian<br>Black<br>Hispanic | (-0.002 ,0.013 , 0.036)<br>(0.034 , 0.183 , 0.247)<br>(-0.041, 0.052 , 0.084) | (P = 0.966, 0.784, 0.556)<br>(P = 0.303, < 0.001, < 0.001)<br>(P = 0.252, 0.202, 0.130) | Rejected |

# Study 2 : Introducing an external factor to the relationship

While the marketing field recognizes the growing influence of external factors on consumer behavior (Cruz-Cardenas et al. 2021), many film studies, particularly those investigating racial diversity, have confined their focus to inherent film characteristics, neglecting a broader context. While some studies attempt to address temporal variations by incorporating dummy variables, these methods primarily establish that film success fluctuates across time periods. The limited focus on inherent film characteristics, as demonstrated in Study 1 , Malik et al. (2021) and Kuppuswamy and Younkin (2016), may overlook significant external factors influencing the relationship between Racial (Authentic Inclusive) Representation and Long-term Audience Engagement.

Study 1 revealed instances of a negative relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement. To delve deeper into the dynamics of this relationship, I explored whether an external factor could moderate the impact of Racial Authentic Inclusive Representation on Long-term Audience Engagement. The #OscarsSoWhite movement, which emerged in January 2015, marked a significant shift in public discourse on Racial Representation in the film industry as can be seen in Appendix C. The event emerged as a social media hashtag to highlight the lack of racial diversity and representation in the nominations and awards of the Academy Awards, particularly in regard to people of color. Therefore, the following hypothesis is proposed:

H2: The relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement differs prior and after #OscarsSoWhite.

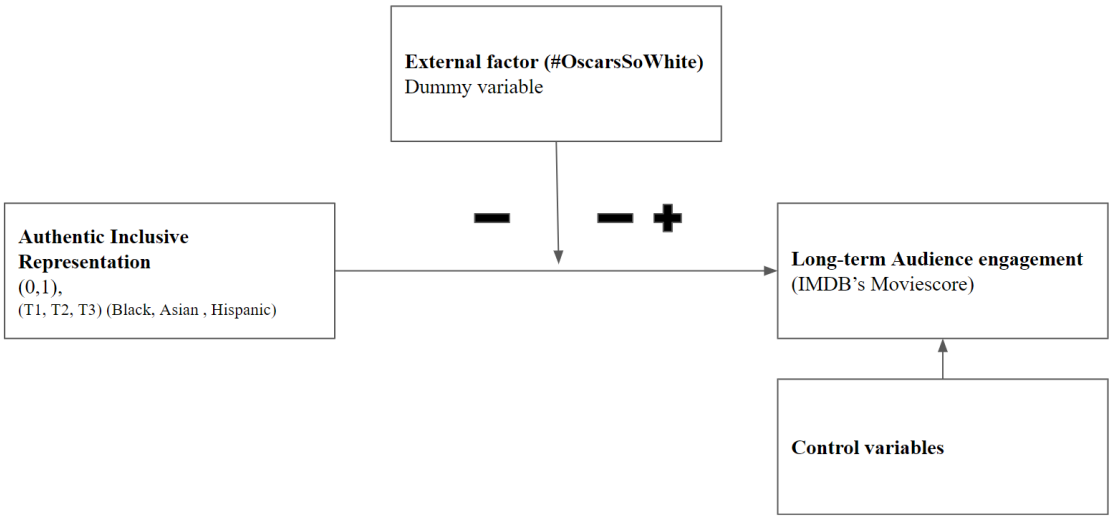



*Figure 2 conceptual model study 2*

The conceptual model shown in figure 2 replicates the original negative relationship observed in Study 1. This study investigates whether appreciation of Racial Representation is time-sensitive, testing how the effects change after the #OscarsSoWhite movement started.

**Research Methodology 2**

The research methodology of Study 2 closely mirrors that of Study 1. As previously mentioned, the #OscarsSoWhite movement was selected as a reference point for this study. To capture the temporal dynamics, a step dummy variable, After_Jan_2015, was generated. Films released in or after 2015 were assigned a value of one, indicating the period when Racial Representation gained heightened cultural significance. The assumption, supported by continued emphasis on belonging and community in 2020 (Neufeld, 2020), and findings by Lazar et al. (2021), is that this period of cultural significance extended at least from 2015 to 2019. This assumption obviated the necessity to explicitly model a moment when Racial Representation ceased to be a prominent cultural topic.

***Models***

To investigate the potential moderating effect of an external factor on the relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement the interaction term after_jan_2015 was incorporated into the three T1, T2, and T3 models. For example, the T1 model:

$$\log(LTAE_i) = \beta0 + \beta1 \times ASIANT1_i + \beta2 \times BLACKT1_i + \beta3 \times HISPANICT1_i + \underline{\beta4 \times FILMSAFTERJAN2015_i * ASIANT1_i} + \underline{\beta5 \times FILMSAFTERJAN2015_i * BLACKT1_i} + \underline{\beta6 \times FILMSAFTERJAN2015_i * HISPANICT1_i} + \{Control\ Variables\} + \varepsilon_i$$

In this model, LTAE similar to the previous study represents the Long-term Audience Engagement for film i, FILMSAFTERJAN2015 is a dummy variable that takes a value of 1 if the film was released in or after February 2015. The interaction terms $FILMSAFTERJAN2015_i \times ASIANT1i$, $FILMSAFTERJAN2015_i * BLACKT1_i$ and $FILMSAFTERJAN2015_i * HISPANICT1_i$ allow to capture whether the effect of Racial Authentic Inclusive Representation on Long-term Audience Engagement differs before and after January 2015.

What needs to be noted is that while the interaction terms are included in the model, the main effect of the dummy variable of after January 2015 is not included. This is because it would have multicollinearity issues with the year dummies incorporated as control variables.

**Results study 2**

The results of the models incorporating the interaction term are presented in the following table, showcasing only the variables under investigation to avoid redundant display of the control variables for which significance and influence was thoroughly discussed in study 1.

Table 6: Results step dummy regression analysis

| Variable | Model step T1 | Model step T2 | Model step T3 |
|---|---|---|---|
| Asian | 0.073 † | 0.099 † | 0.038 |
| Black | 0.041 | 0.182 *** | 0.262 *** |
| Hispanic | - 0.067 | 0.041 | 0.031 |
| | | | |
| Interaction term Asian | -0.256 *** | - 0.228* | -0.014 |
| Interaction term Black | -0.032 | 0.003 | -0.049 |
| Interaction term Hispanic | 0.075 | 0.046 | 0.153 |
| | | | |
| Effect of Asian After Jan 2015 | -0.184* | -0.129 † | 0.023 |
| Effect of Black After Jan 2015 | 0.009 | 0.184* | 0.212* |
| Effect of Hispanic After Jan 2015 | 0.008 | 0.087 | 0.184* |
| $R^2$ | 0.7240 | 0.7308 | 0.7330 |
| Adjusted $R^2$ | 0.7074 | 0.7147 | 0.7171 |
| Degrees of Freedom | 1116 | 1116 | 1116 |
| F Statistic | 44.83*** | 45.73** | 46.18*** |

*Significance levels: † p < 0.10 * p < 0.05, ** p < 0.01, *** p < 0.001*

The three models, representing different time periods, demonstrated higher adjusted R-squared values compared to the corresponding models in Study 1 (T1: 0.70, T2: 0.71, T3: 0.71). This indicates that incorporating the interaction term allowed the model to capture more variance in the dependent variable. However, the marginal increase in values suggests that introducing the dummy variable did not substantially enhance explanatory power.

Analyzing parameter estimates for Racial Authentic Inclusive Representation (t variables) before 2015 , reveals positive coefficients for all ethnicities in at least one of the model's. This finding diverges from the insignificant relationships observed in Study 1. To provide clarity, the estimates of interaction terms are incorporated. A comprehensive assessment involves constructing the actual effect of the interaction terms by scrutinizing the linear combination. The Delta Rule is then applied to gauge the significance of the observed effects.

A noteworthy finding is the effect of the interaction terms for the Asian ethnicity. Model T1 exhibited a reversed effect applying the Delta rule, indicating a reversal of the initial positive coefficient. In T2, the effect for Asian ethnicity was also reversed, with significance at a p-value of 0.10 for both the prior and after January 2015 estimates. A possible explanation for this change in effect for Asian ethnicity is an increase in films with Racial Authentic Inclusive Representation in 2015-2019, as noted by Lazar et al. (2020). This rise may have resulted in better authentic portrayals of Asian characters, improving the resonating with audiences.

The interaction term estimates related to Black T2 and T3 model remain significant. Suggesting for both T2 and T3 there is still a negative relationship after 2015 between Black representation and Long-term Audience Engagement. Moreover, the Hispanic T3 interaction term becomes significant, with an estimate of 0.184. This suggests that films with the T3 condition of Hispanic representation in 2015-2019 were more likely to experience less Long-term Audience Engagement. This finding is surprising, given the previously insignificant relationship for Hispanic ethnicity and the heightened focus on Racial Representation after January 2015.

In conclusion, the results suggest a more challenging path for Black and Hispanic actors, in contrast to Asian actors, in gaining acceptance and resonance with audiences. Nevertheless, this study doesn't want to make claims for causality per se. This study primarily focusses on highlighting the intricate relationship between an external factor, Racial Representation, and film Long-term Audience Engagement, suggesting that factors beyond film characteristics influence attitudes towards Racial Authentic Inclusive Representation.

Table 7: Hypothesis study 2

| **Hypotheses** | **Ethnicity** | **β (T1, T2 , T3)** | **P-Value (T1, T2 , T3)** | **Decision** |
|---|---|---|---|---|
| H2a: The relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement differs prior and after #OscarsSoWhite. | Asian<br>Black<br>Hispanic | (-0.184, -0.129, 0.023)<br>(0.009,0.184,0.212)<br>(0.008, -0.087, 0.184) | (P = 0.021, 0.008, 0.990)<br>(P = 0.695, 0.002, 0.004)<br>(P = 0.278, 0.552, 0.032) | Partially Accepted |

# Robustness checks study 1 and 2

To ensure the reliability and validity of the findings, two robustness checks were implemented. These checks primarily target the two significant divergences between our study and prior research. First, one check involves altering the dependent variable to box office performance. The second check explores whether employing simplistic methods would yield comparable results to both our study and previous research. To not derive too much from the main focus of this study, the complete analysis is provided in Appendix D. The most important findings from this robustness checks were;

The t variables showed similar patterns of effects to Box Office as what they did to Long-term Audience Engagement.

Furthermore, the divergence in findings between this research and other studies can be attributed to the varied metrics employed for assessing Racial Representation, rather than indicating a fundamental difference in the effects of Racial Representation itself. When utilizing the simplistic measurements from previous studies, the dataset in this study yielded comparable results. Intriguingly, this implied that, upon comparison, there was a positive correlation between Long-term Audience Engagement and these straightforward measurements. While this underscores the imperative to scrutinize Authenticity and Inclusiveness within the context of Racial Representation, it also challenges the assumption that these attributes are universally pivotal for positive effects. The analysis, however, revealed the presence of potential omitted biases, such as total cast size, which could significantly impact the outcomes.

# Discussion

## General conclusion and discussion

This study delves into the impact of Racial Representation on Long-term Audience Engagement in films, introducing the dimensions of Authenticity and Inclusiveness to enhance the Racial Representation metric. Additionally, IMDB ranking was employed as a quantifiable measure for Long-term Audience Engagement. Two research designs were employed: the first examined the overall relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement, while the second investigated the impact of an external factor on this relationship.

The most notable finding of this study is that Racial Authentic Inclusive Representation significantly affects a film's Long-term Audience Engagement, albeit differently across ethnicities. Black representation, for instance, demonstrated a negative association with Long-term Audience Engagement, while Asian Authentic Inclusive Representation exhibits a positive association in specific contexts.

Interestingly, the negative and insignificant relationships contradicted the initial study hypothesis. The initial study suggested that increased audience engagement over time results from a greater number of individuals being able to relate to characters based on shared demographic characteristics. However, insights from Lazar et al. (2020) suggest a potential underlying reason for the observed negative associations. Lazar et al. found a positive correlation between box office performance and Racial Authentic Inclusive Representation, contingent upon the film having a substantial budget. This highlights a potential drawback in the inclusion of Authentic Inclusive Representation for a specific ethnicity without broader representation of various ethnicities, potentially leading to the underrepresentation of other ethnic groups, such as white actors. Notably, this drawback may be less pronounced in large-budget films due to their expansive casts, allowing for a more comprehensive representation of multiple ethnicities.

Furthermore, a plausible explanation for the overall negative association with the Black ethnicity may arise from the notable overrepresentation of Black actors, the only minority group discussed in the introduction to exhibit such a phenomenon within the industry. The abundance of Black characters introduced in films could potentially be perceived as tokenism. In this context, Authentic Inclusive representation positively influences the film, but it might be achieved through the inclusion of a large number of Black actors, some of whom may be included tokenistically. Such a perception may suggest the presence of a perceived 'racial agenda,' potentially exerting a detrimental influence on a film's potential success.

The second study in this research explores the possibility that the relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement was influenced by an external factor, the social movement #OscarsSoWhite. This specific period, marked by heightened awareness of Racial Representation, exhibited intriguing dynamics. While negative correlations between Racial Authentic Inclusive Representation and Long-term Audience Engagement were prevalent before #OscarsSoWhite, these effects appeared to lessen or even reverse afterward. While the study does not directly attribute this change to #OscarsSoWhite, it highlights the evolving nature of perceptions of Racial Representation across time periods, challenging the conventional approach of solely analyzing film characteristics and emphasizing the potential impact of external factors.

What is particularly promising is that the findings of the second study align with Malik et al. (2021), who used screen time as a proxy for Racial Representation. When using International Rent (IntRent) as their dependent variable, they found that films featuring more East and South Asian characters were significantly more likely to generate higher IntRent compared to films with white characters. Conversely, films with Black and African American and Latin characters were significantly less likely to achieve higher IntRent compared to films with white characters. These findings mirror the relationships observed in the present research.

The second study also aligned with the findings of Lazar et al. (2020), who identified 2015-2019 as a heightened period of focus on Racial Authentic Inclusive Representation in films. This could have influenced the contexts in which these actors were portrayed or may have contributed to these films garnering more critical acclaim after #OscarsSoWhite, attracting a broader audience seeking diverse cinematic experiences.

However, the shifting appreciation for Racial Representation over time raises questions about the underlying motivations driving audience engagement. If external factors can influence our preferences, it challenges the idea that our desire to see ourselves represented on screen is solely based on internal factors.

**Theoretical implications.**

The study's critical evaluation of previous research prompts a re-examination of the suitability of simplistic metrics like actor count in capturing the intricacies of Racial Representation. When utilizing news metric it was seen that the results differ from previous research. Nevertheless, the transition from mere actor counts to the more refined tool did not align with the theoretical underpinnings of Authenticity and Inclusiveness in Racial Representation.

This study builds upon existing research by demonstrating how Racial Authentic Inclusive Representation can be quantified. While Authentic Inclusive Representation was previously applied to studies of gender representation, this study demonstrates its applicability to Racial Representation as well. The process of introducing this metric suggested a trade-off in the realm of Racial Representation metrics: the balance between scalability and comprehension. This trade-off is evident when comparing this study's metric to those employed by Lazar et al. (2020) and Malik et al. (2021). While their metrics arguably provide a more comprehensive assessment of Racial Representation, they have only been applied to a small number of films (under 100). It is assumed this is caused by limitations with regards to scalability.

Finally, this study offers fresh theoretical insights into measuring a film’s success. It demonstrates that Long-term Audience Engagement, as reflected by ranking on IMDB, can be used as a viable indicator of a film's success. While the study doesn't specifically compare its effectiveness to other measurements, it successfully captures a film's enduring Long-term Audience Engagement over time. Moreover, this study shows that the metric used for Long-term Audience Engagement is correlated with the established factors of a film's success.

**Managerial implications.**

The findings of this study hold significant implications for decision-makers in the film industry, emphasizing the importance of a nuanced and forward-thinking approach to Racial Representation.

Industry professionals must recognize that different metrics can lead to varying outcomes. Therefore, careful consideration should be given to the metrics employed to assess Racial Representation. Recognizing the limitations of exclusively relying on the total number of actors from a specific ethnicity is essential. A more expansive view of Racial Representation should be embraced. Nevertheless, as the studies of this thesis show the variables Racial Representation seems to only have marginal impacts, which argues against managerial relevance.

Notably, this study introduces a novel metric for measuring film success, offering an alternative to the limitations of box office revenue. While many films are primarily driven by financial objectives, which can be better evaluated through box office revenue, Long-term Audience Engagement measured through IMDB rankings serves as a valuable metric for films that aspire to leave a lasting impression on audiences and foster meaningful connections. Moreover, the Racial Authentic Inclusive Representation while not as exhaustive as some studies provides concrete measurements based on quantitative data rather than relying on subjective evaluations. Its simplicity and adaptability make it a highly practical tool for industry professionals, enabling them to assess the Racial Representation on a large body of films

Finally, the temporal dimension of the study's findings underscores the fluidity of audience perceptions regarding Racial Representation. The study's longitudinal analysis revealed a change in the association between Racial Representation and Long-term Audience Engagement following the #OscarsSoWhite movement. This evolving dynamic signifies the importance of filmmakers adopting a long-term perspective, continuously monitoring trends, and adapting casting strategies accordingly. This dynamic nature also cautions against sweeping singular solutions regarding racial representation. To suggest, for instance, that reducing the presence of Black actors would somehow optimize Long-term Audience Engagement would be demonstrably reductive and inaccurate. The path forward lies in embracing the complexity of the relationship and tailoring casting decisions to specific contexts and narratives, while remaining attentive to the evolving landscape of audience preferences.

**Limitations and suggestions for further research**

Despite its significant contributions, this study has certain limitations. Notably, the study did not account for the impact of a film's initial success on its Long-term Audience Engagement. Early success displayed through Box Office, often driven by factors like marketing strategies, can have a lingering influence on audience engagement. Sadly, I had no data on this. Incorporating a model that considers the cumulative effect of past marketing efforts could have improved the study's ability to isolate the effects of Racial Authentic Inclusive Representation on Long-term Audience Engagement.

Another possibility not researched in this study was an interaction effect between a film characteristic and Racial Authentic Inclusive Representation. Lazar et al. (2020) found that Racial Authentic Inclusive Representation positively influences the box-office of large-budget films. This suggests that other film characteristics might also impact the relationship between Racial Authentic Inclusive Representation and Long-term Audience Engagement, a possibility that wasn't explored in this study.

Finally, to further enhance our understanding of Racial Representation in films, I propose exploring the application of social network analysis as a sophisticated method to measure the centrality of characters. This approach leverages neural networks to capture the dynamic and intricate relationships between characters, providing a more nuanced and comprehensive assessment of Racial Representation. While the Racial Authentic Inclusive Representation metric has proven valuable, social network analysis offers a distinct perspective, enabling us to delve deeper into the impact of each character on the narrative and overall engagement.

# Bibliography


Abraham, L., & Appiah, O. (2006). Framing news stories: The role of visual imagery in priming racial stereotypes. *Howard Journal of Communications, 17*(3), 183–203.

Agarwal, A., Zheng, J., Shruti, B., & Sriramkumar, S. (2015). Key Female Characters in Film Have More to Talk About Besides Men: Automating the Bechdel Test. *Conference: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Aumer, K., Blas, D., Huston, K., Mabuti, C., & Hsu, N. (2017). Assessing Racial Preferences in Movies: The Impact of Mere-Exposure and Social Identity Theory. *Psychology, 8*(9), 1314-1332.

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing, 67*(4), 103–117.

Belton, J. (1995). *Movies and Mass Culture.* The Athlone Press. From https://www.goodreads.com/book/show/1639322.Movies_and_Mass_Culture

Broersma, M. (2019). Audience Engagement. *The International Encyclopedia of Journalism*, 1-6.

Buchanan, M., & Hoffner, C. (2005). Young Adults' Wishful Identification With Television Characters: The Role of Perceived Similarity and Character Attributes. *Media Psychology, 7*, 325-351.

Buckingham, D. (2008). *Youth , identity and digital media.* MIT Press, Cambridge, Mass.

Castañeda, M. (2017). *Challenging Inequalities: Readings in Race, Ethnicity, and Immigration.*

Chakraborty, P., Rahman, S., & Rahman Zahid, Z. (2019). Movie Success Prediction using Historical and Current Data Mining. *International Journal of Computer Applications, 178*(47), 1-5.

Clement, M., Wu, S., & Fischer, M. (2014). Empirical generalizations of demand and supply dynamics for movies. *International Journal of Research in Marketing, 31*(2), 207-223.

Correa, T., & Jeong, S. (2011). RACE AND ONLINE CONTENT CREATION. *Information Communication and Society*.

Dalisay, F., & Tan, A. (2009). Assimilation and Contrast Effects in the Priming of Asian American and African American Stereotypes through TV Exposure. *Journalism & Mass Communication Quarterly, 86*(1), 7–22.

Dixon, T. L., & Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication,, 50*, 131-154.

Ghiassi, M. S. (2017). Predicting Movie Success Based on IMDB Data. *Journal of Big Data*, 1-14.

Hall, A. E. (2022). Audience Responses to Diverse Superheroes: The Roles of Gender and Race in Forging Connections With Media Characters in Superhero Franchise Films. *Psychology of Aesthetics, Creativity, and the Arts, 16*, 414–425.

Hofmann, J., Clement, M., Völckner, F., & Hennig-Thurau, T. (2016). Empirical Generalizations on the Impact of Stars on the Economic Success of Movies. *Journal of Research in Marketing, 34*(2), 442-461.

Hunt, D., & Ramón, A. (2021). *Hollywood Diversity Report 2021*. UCLA.

Hunter, S., Smith, S., & Singh, S. (2016). Predicting Box Office from the Screenplay: A Text Analytical Approach. *Journal of Screenwriting, 7*(2), 135-154.

Hurley, R. J., Jensen, J. J., Weaver, A., & Dixon, T. (2015). Viewer Ethnicity Matters: Black Crime in TV News and Its Impact on Decisions Regarding Public Policy. *Journal of Social Issues, 71*(1), 155–170.

Kidd, M. A. (2015). Archetypes, stereotypes and media representation in a multi-cultural society. *Procedia, 236*, 25-28.

Krushikanth, R., Jose, M., Motnam, S., Chien-Chung, C., & Liszka, k. (2013). Prediction of movies box office performance using social media. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.*

Kubrak, T. (2020, May 2). Impact of Films: Changes in Young People's Attitudes after Watching a Movie. *Behav Sci (Basel)*.

Kuppuswamy, V., & Younkin, P. (2016). Testing the theory of consumer discrimination as an explanation for the lack of minority hiring in Hollywood films. *Management Science, 66*(3).

Litman, B. (1983). Predicting Success of Theatrical Movies: An Empirical Study. *The Journal of Popular Culture, 16*, 159-175.

Madongo, C. Z. (2023). A movie box office revenue prediction model based on deep multimodal features. Multimed Tools Appl. *82*, 31981–32009.

Mastro, D. B.-M. (2007). The cultivation of social perceptions of Latinos: A mental models approach. *Media Psychology*, *9*(2), 347-365.

McKinsey. (2021). The bottom line: movies with more diverse casts perform better at the box office.

MPA. (2022). *Theatrical and Home Entertainment Market Environment Report.*

Neff, K., & Smith, K. (2023). *Inequality Across 1,600 Popular Films: Examining Gender, Race/Ethnicity & Age of Leads/Co Leads From 2007 to 2022.* University of Southern California, USC Anneberg Inclusion Initiative.

Nelson, R., & Glotfelty, R. (2012). Movie stars and box office revenues: An empirical analysis. *Journal of Cultural Economics, 36*(2), 141-166.

Orankiewicz, A., & Bartosiewicz, A. (2023). The COVID-19 Pandemic and the Adoption Factors of Film Distribution Business Models in the Context of Sustainability. *Sustainability, 15*, 11-18.

Patel, L. (2015, November). Desiring Diversity and Backlash: White Property Rights in Higher Education. *Regional & Urban Planning*, *47*(4).

Ross, T. (2019). *The Palgrave Handbook of Ethnicity.*

Rubin, D., Mohr, I., & Kumar, V. (2022). Beyond the box office: A conceptual framework for the drivers of audience engagement,. *Journal of Business Research, 151*(3), 473-488.

Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications, 30*(2), 243-254.

Smith, S. L., & Choueit, L. (2020). *Inequality in 1,300 popular films: Examining portrayals of gender, race/ethnicity, LGBTQIA, and disability from 2007–2019.* University of Southern California , USC Annenberg Inclusion Initiative.

Sperling, N. (2021, March). Hollywood Loses $10 Billion a Year Due to Lack of Diversity, Study Finds.

Tan, E. S. (2018). A psychology of the film. *Humanities and Social Sciences Communications, 4*(1), 82-82.

Umaña-Taylor, A., Quintana, S., Lee, R., Syed, M., & Seaton, E. (2014, February). Ethnic and Racial Identity in the 21st Century Study Group. Ethnic and racial identity during adolescence and into young adulthood: an integrated conceptualization. *Child Development, 85*, 21-39.

Weaver, A. J. (2011). The Role of Actors' Race in White Audiences' Selective Exposure to Movies. *Journal of Communication, 61*, 369-385.

Whitten, S. (2019). More diversity, more money: Theater owners want more inclusive film. From https://www.cnbc.com/2019/04/04/more-diversity-more-money-theater-owners-want-more-inclusive-films.html

Appendix A: Stop word list

"the", "and", "a", "an", "man", "woman", "girl", "boy", "person",

"doctor", "nurse", "teacher", "student", "police", "officer",

"captain", "detective", "driver", "waiter", "waitress",

"bartender", "chef", "pilot", "scientist", "engineer",

"professor", "lawyer", "firefighter", "writer", "artist",

"singer", "dancer", "actor", "actress", "model",

"president", "king", "queen", "prince", "princess",

"doctor", "sir", "madam", "lord", "lady",

"mr", "mrs", "ms", "miss", "(uncredited)", "self", "barman", "mom", "dad",

"teammate", "cop", "director", "character", "role", "extra", "child",

"player", "customer", "passenger", "soldier", "villager", "audience",

"guest", "customer", "victim", "stranger", "citizen", "neighbor",

"client", "assistant", "host", "companion", "artist", "musician", "performer",

"guest", "teacher's pet", "bodyguard", "consultant", "helper", "coworker",

"bystander", "partygoer", "attendee", "athlete", "official", "associate",

"patient", "driver", "veterinarian", "coach", "driver", "employee", "scientist",

"assistant", "stranger", "examiner", "salesperson", "waitstaff", "witness",

"shooter", "victim", "spy", "veteran", "trader", "guard", "hiker", "shopper",

"actor", "writer", "adventurer", "explorer", "admiral",

"professor", "conductor", "fireman", "sheriff", "reporter", "editor",

"thief", "student", "resident", "colleague", "reporter", "buyer", "driver",

"rider", "leader", "teacher", "hunter", "butler", "stewardess", "crew",

"manager", "waitstaff", "server", "waiter", "waitress", "administrator",

"scientist", "technician", "fighter", "runner", "visitor", "pedestrian",
"stranger", "witness", "operator", "instructor", "inspector", "spectator",
"visitor", "spectator", "celebrity", "policeman", "housekeeper", "housewife",
"househusband", "deliveryman", "scientist", "guide", "designer", "commander",
"commando", "cowboy", "scientist", "scholar", "gentleman", "nanny", "professor",
"missionary", "nun", "missionary", "chaplain", "coach", "airman", "admiral",
"navigator", "soldier", "fighter", "nobleman", "servant", "neighbor", "sailor",
"secretary", "assistant", "professor", "clerk", "nobleman", "parolee", "neighbor",
"kidnapper", "mentor", "nurse", "reporter", "maid", "student", "scientist",
"scholar", "trainee", "supervisor", "official", "deliverywoman", "reporter",
"neighbor", "lawyer", "assistant", "reader", "guest", "witness", "villager",
"visitor", "crew", "engineer", "guest", "helper", "professor", "villager",
"guest", "team", "scientist", "student", "assistant", "chef", "waiter",
"waitress", "restaurantgoer", "attendant", "interviewer", "officer", "professor",
"assistant", "administrator", "assistant", "mentor", "assistant", "escort",
"client", "professor", "manager", "celebrity", "photographer", "tourist",
"presenter", "listener", "expert", "scientist", "doctor", "professor",
"philosopher", "veterinarian", "wife", "homemaker", "househusband",
"housekeeper", "neighbor", "guest", "stranger", "assistant", "teacher",
"student", "servant", "companion", "assistant", "professor", "shopper",
"customer", "bystander", "visitor", "neighbor", "colleague", "assistant",
"seller", "customer", "pedestrian", "passerby", "tourist", "clerk", "singer",
"scientist", "actor", "detective", "professor", "chef", "photographer",
"model", "driver", "manager", "assistant", "attendant", "professor",

"guest", "client", "student", "visitor", "stranger", "neighbor", "witness",

"driver", "athlete", "reporter", "official", "scientist", "shopper",

"customer", "clerk", "manager", "employee", "guest", "assistant", "teacher",

"student", "scientist", "expert", "host", "participant", "judge", "lawyer",

"scientist", "professor", "scientist", "student", "assistant", "singer",

"dancer", "model", "reporter", "scientist", "professor", "attendant",

"professor", "guest", "audience", "technician", "athlete", "clerk", "professor",

"consultant", "buyer", "tourist", "shopper", "client", "customer", "expert",

"scientist", "performer", "actor", "actress", "student", "director", "waitstaff",

"bartender", "actor", "actress", "model", "singer", "dancer", "musician",

"athlete", "attendant", "manager", "agent", "waitstaff", "assistant", "clerk",

"seller", "guest", "customer", "professor", "client", "designer", "expert",

"expert", "participant", "scientist", "speaker", "professor", "manager",

"operator", "technician", "bartender", "bystander", "assistant", "observer",

"producer", "villian", "hero", "sidekick", "mentor", "architect", "on",

"#1", "#2","#3", '#4', '#5' , "commentator", "comedian", "in", "surgeon",

"host", "contestant", "reporter","executive", "(voice)", "guy", "receptionist", "tech", "anchor", "janitor", "ranger", "1", "2", "3", "4", "worker", "member", "banker", "stockbroker", "paperboy", "footman", "daughter", "principal", "school", "friend", "addict", "mechanic", "traveler", "officer", "chief", "owner", "uncle",

"deputy","aunt", "security" , "woman", "brother", "female", "priest", "walker", "thug", "master", "king", "girl", "vendor", "russian", "minister", "baker", "baby", "boss", "reporter", "husband", "mayor", "dr", "father", "mother",

"announcer", "sister", "girlfriend", "senator", "band", "party", "club", "teen", "teenage", "teenager" "gardner","archer", "warrior", "fbi", "teller", "councilman", "prostitute", "fan", "salesman", "attorney", "grandma", "grandpa", "footage)", "dealer", "cashier", "dj", "farmer", "german", "newscaster", "chinese", "dog", "cat", "cheerleader"

"trainer", "co-worker", "partner", "stewart", "restaurant", "translator", "queen", "with", "quarterback", "paramedic"

"quarterback", "paramedic", "veteran", "villager", "plumber", "electrician",

"mechanic", "butcher", "librarian", "historian", "accountant", "banker", "clerk",

"cashier", "waiter", "waitress", "hostess", "host", "presenter", "journalist",

"anchor", "receptionist", "janitor", "worker", "employee", "executive", "owner",

"manager", "partner", "assistant", "supervisor", "consultant", "scientist",

"researcher", "technician", "designer", "architect", "engineer", "doctor", "nurse",

"therapist", "patient", "pharmacist", "student", "teacher", "professor", "scholar",

"lawyer", "attorney", "judge", "jury", "witness", "reporter", "detective",

"officer", "sergeant", "captain", "lieutenant", "commander", "director", "producer",

"actor", "actress", "musician", "singer", "dancer", "artist", "model", "athlete",

"coach", "player", "character", "hero", "villain", "sidekick", "comedian", "commentator",

"host", "contestant", "participant", "audience", "listener", "viewer", "reader",

"spectator", "guest", "fan", "customer", "client", "shopper", "buyer", "seller",

"vendor", "consumer", "passerby", "pedestrian", "tourist", "traveler", "driver",

"rider", "pilot", "navigator", "adventurer", "explorer", "admiral", "sheriff",

"thief", "fireman", "paramedic", "surgeon", "doctor", "nun", "missionary",

"chaplain", "missionary", "gentleman", "lady", "sir", "madam", "lord", "woman",

"man", "girl", "boy", "child", "baby", "infant", "toddler", "teen", "teenager",

"elder", "senior", "junior", "boss", "manager", "owner", "chief", "deputy",

"sarge", "capt", "sergeant", "colleague", "associate", "partner", "expert",

"specialist", "professional", "novice", "amateur", "enthusiast", "expertise",

"authority", "expert", "guru", "entertainer", "performer", "singer", "dancer",

"musician", "artist", "comedian", "actor", "actress", "model", "character", "role", "persona", "identity", "individual", "personality", "citizen", "resident", "inhabitant", "neighbor", "neighborhood", "community", "colony", "society", "culture", "civilization", "world", "planet", "earth", "universe", "cosmos", "galaxy", "star", "celestial", "space", "astronomy", "galactic", "extraterrestrial", "alien", "creature", "being", "entity", "thing", "object", "item", "element", "substance", "material", "matter", "stuff", "commodity", "product", "good", "merchandise", "service", "assistance", "help", "aid", "support", "benefit", "advantage", "profit", "gain", "value", "worth", "quality", "feature", "attribute", "trait", "characteristic", "property", "possession", "ownership", "possession", "belonging", "object", "thing", "item", "artifact", "creation", "invention", "innovation", "discovery", "finding", "uncovering", "revelation", "disclosure", "knowledge", "wisdom", "insight", "information", "data", "fact", "reality", "truth", "certainty", "certitude", "surety", "confidence", "assurance", "guarantee", "promise", "commitment", "responsibility", "obligation", "duty", "requirement", "necessity", "essential", "must", "need", "want", "desire", "wish", "hope", "dream", "ambition", "goal", "objective", "purpose", "intention", "plan", "strategy", "tactic", "method", "approach", "technique", "way", "means", "mode", "manner", "fashion", "style", "form", "expression", "communication", "message", "idea", "thought", "notion", "concept", "perception", "understanding", "cognition", "knowledge", "wisdom", "insight", "intelligence", "genius", "brilliance", "creativity", "imagination", "innovation", "invention", "discovery", "inspiration", "influence", "impact", "effect", "outcome", "result", "consequence", "significance", "importance", "relevance", "meaning", "value", "worth", "quality", "feature", "characteristic",

"trait", "aspect", "dimension", "factor", "element", "component", "ingredient",
"constituent", "part", "portion", "segment", "section", "piece", "fragment",
"bit", "particle", "molecule", "atom", "microscopic", "subatomic", "infinitesimal",
"imperceptible", "unobservable", "indistinguishable", "invisible", "hidden",
"secret", "mysterious", "unknown", "obscure", "ambiguous", "unclear", "uncertain",
"(segment", "dentist", "pharmacist", "veterinarian", "plumber", "electrician",
"carpenter", "mechanic", "pilot", "engineer", "architect", "designer",
"artist", "musician", "athlete", "coach", "sir", "madam", "lord", "lady",
"king", "queen", "prince", "princess", "duke", "duchess", "mother", "father",
"daughter", "son", "brother", "sister", "aunt", "uncle", "cousin", "niece",
"nephew", "grandmother", "grandfather", "robot", "alien", "monster", "creature",
"spirit", "ghost", "vampire", "witch", "wizard", "sorcerer", "fairy", "elf", "double",
"dentist", "pharmacist", "veterinarian", "plumber", "electrician",
"carpenter", "mechanic", "pilot", "engineer", "architect", "designer",
"artist", "musician", "athlete", "coach", "sir", "madam", "lord", "lady",
"king", "queen", "prince", "princess", "duke", "duchess", "mother", "father",
"daughter", "son", "brother", "sister", "aunt", "uncle", "cousin", "niece",
"nephew", "grandmother", "grandfather", "robot", "alien", "monster", "creature",
"spirit", "ghost", "vampire", "witch", "wizard", "sorcerer", "fairy", "elf",
"god", "goddess", "deity", "myth", "legend", "hero", "heroine", "warrior",
"saint", "martyr", "historical", "mythical", "legendary", "ancient", "wise",
"sage", "prophet", "messenger", "angel", "demon", "creature", "beast", "minion",
"avatar", "monarch", "sovereign", "emperor", "empress", "dictator", "president",
"chancellor", "prime", "minister", "mayor", "governor", "senator", "congressman",

"congresswoman", "representative", "ambassador", "diplomat", "envoy", "delegate", "official", "officer", "commander", "admiral", "captain", "lieutenant", "sergeant", "private", "colonel", "general", "marshal", "sheriff", "deputy", "officer", "agent", "inspector", "detective", "investigator", "trooper", "sergeant", "captain", "officer", "agent", "inspector", "detective", "investigator", "trooper", "sergeant", "captain", "soldier", "fighter", "warrior", "combatant", "brawler", "champion", "winner", "loser", "villain", "sidekick", "henchman", "assistant", "partner", "companion", "friend", "foe", "enemy", "rival", "opponent", "stranger", "neighbor", "citizen", "resident", "vagrant", "wanderer", "traveler", "tourist", "explorer", "adventurer", "pioneer", "settler", "migrant", "nomad", "drifter", "visitor", "spectator", "audience", "viewer", "listener", "watcher", "reader", "auditor", "reviewer", "critic", "commentator", "comedian", "host", "presenter", "interviewer", "moderator", "panelist", "expert", "analyst", "commentator", "reporter", "journalist", "correspondent", "photographer", "cameraman", "director", "producer", "editor", "executive", "manager", "administrator", "coordinator", "supervisor", "foreman", "boss", "leader", "chief", "owner", "proprietor", "landlord", "landlady", "tenant", "resident", "guest", "visitor", "client", "customer", "shopper", "buyer", "seller", "merchant", "salesman", "clerk", "cashier", "teller", "baker", "butcher", "grocer", "chef", "waiter", "waitress", "bartender", "barista", "hostess", "steward", "stewardess", "attendant", "server", "employee", "worker", "laborer", "technician", "engineer", "scientist", "researcher", "scholar", "professor", "teacher", "instructor", "lecturer", "educator", "tutor", "trainer", "coach", "mentor", "advisor", "counselor", "therapist", "psychologist", "psychiatrist", "doctor", "physician", "nurse", "pharmacist", "therapist", "patient", "client",

"resident", "inmate", "prisoner", "detainee", "hostage", "victim", "survivor", "sufferer", "patient", "witness", "bystander", "spectator", "onlooker", "observer", "participant", "contestant", "competitor", "candidate", "nominee", "winner", "loser", "champion", "challenger", "challenger", "contender", "opponent", "rival", "opponent", "adversary", "foe", "enemy", "opponent", "combatant", "athlete", "runner", "jogger", "swimmer", "cyclist", "athlete", "participant", "competitor", "winner", "loser", "champion", "medalist", "record", "holder", "challenger", "contender", "favorite", "underdog", "outsider", "front-runner", "candidate", "nominee", "competitor", "contender", "champion", "medalist", "record", "holder", "challenger", "contender", "favorite", "underdog", "outsider", "front-runner", "candidate", "nominee", "player", "contestant", "competitor", "challenger", "champion", "finalist", "sem", "dude"

Appendix B : Dummy variables descriptives

Tables 8, 9, and 10 present the frequency statistics for the control dummy variables for year, genre, seasonality, and the 'based-on' variables. These tables provide the number of observations per variable (N) and the percentage this dummy variable represents compared to the total number of 1,187 observations.

Table 8: Year dummies

| **Year** | **N** | **Percentage** | | **Year** | **N** | **Percentage** |
|---|---|---|---|---|---|---|
| 2000 | 117 | 4.82 | | 2010 | 119 | 4.91 |
| 2001 | 122 | 5.03 | | 2011 | 121 | 4.99 |
| 2002 | 120 | 4.95 | | 2012 | 106 | 4.37 |
| 2003 | 116 | 4.78 | | 2013 | 109 | 4.49 |
| 2004 | 127 | 5.24 | | 2014 | 118 | 4.87 |
| 2005 | 134 | 5.53 | | 2015 | 110 | 4.54 |
| 2006 | 143 | 5.90 | | 2016 | 134 | 5.53 |
| 2007 | 133 | 5.48 | | 2017 | 104 | 4.29 |
| 2008 | 130 | 5.36 | | 2018 | 132 | 5.44 |
| 2009 | 125 | 5.15 | | 2019 | 105 | 4.33 |

Table 9: Genre dummies

| **Genre** | **N** | **Percentage** | | **Genre** | **N** | **Percentage** |
|---|---|---|---|---|---|---|
| Drama | 489 | 42.30 | | Fantasy | 160 | 13.84 |
| Thriller | 430 | 37.20 | | Family | 115 | 9.95 |
| Comedy | 428 | 37.02 | | Biography | 74 | 6.40 |
| Action | 408 | 35.29 | | Sport | 50 | 4.33 |
| Adventure | 253 | 21.89 | | History | 44 | 3.81 |
| Romance | 236 | 20.42 | | Music | 43 | 3.72 |
| Crime | 220 | 19.03 | | War | 42 | 3.63 |
| Sci.Fi | 202 | 17.47 | | Musical | 18 | 1.56 |
| Mystery | 181 | 15.66 | | Western | 10 | 0.87 |
| Horror | 162 | 14.01 | | Documentary | 5 | 0.43 |

Table 10 : Remaining dummies

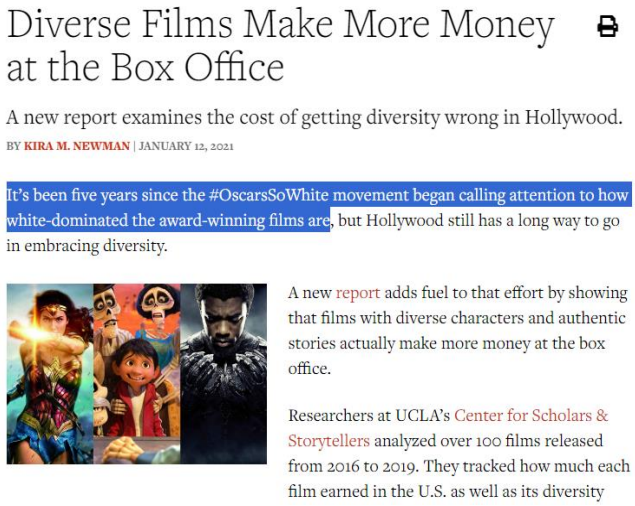
| Season | N | Percentage | | Variable | N | Percentage |
|---|---|---|---|---|---|---|
| Fall | 311 | 26.90 | | Based on novel | 190 | 16.44 |
| Summer | 304 | 26.30 | | Based on comic book | 75 | 6.49 |
| Spring | 301 | 26.03 | | Based on book | 56 | 4.84 |
| Winter | 240 | 20.76 | | Spinoff | 41 | 3.55 |
| | | | | Based on book series | 23 | 1.99 |
| | | | | Based on short story | 8 | 0.69 |
| | | | | Based on play | 7 | 0.61 |

Appendix C: Protests Oscars so white


New York times article:


Can be found at https://www.nytimes.com/2020/02/06/movies/oscarssowhite-history.html

Article from the university of Berkeley



Diverse Films Make More Money at the Box Office

A new report examines the cost of getting diversity wrong in Hollywood.

BY KIRA M. NEWMAN | JANUARY 12, 2021

It's been five years since the #OscarsSoWhite movement began calling attention to how white-dominated the award-winning films are, but Hollywood still has a long way to go in embracing diversity.

A new report adds fuel to that effort by showing that films with diverse characters and authentic stories actually make more money at the box office.

Researchers at UCLA's Center for Scholars & Storytellers analyzed over 100 films released from 2016 to 2019. They tracked how much each film earned in the U.S. as well as its diversity

Can be found at https://greatergood.berkeley.edu/article/item/diverse_films_make_more_money_at_the_box_office

# How #OscarsSoWhite changed the Academy Awards

9 maart 2023 om 07:13 EST

Share

**By Sophie Long**
BBC News, Los Angeles


Can be found at

https://www.bbc.com/news/world-us-canada-64883399

Appendix D: Robustness checks

*Short term success (box office as DV):* Short-term success, measured by box office performance, has traditionally been the focal point of film industry studies. This research, however, extends its focus to the broader social implications of Racial Representation, necessitating a metric that goes beyond financial performance. To ensure the robustness of our findings, we initially examined box office success as the dependent variable, diverging from the original emphasis on Long-term Audience Engagement.

As expected, the film's ranking on IMDB during its release week displayed a significant correlation with box office success. The release week ranking accounted for approximately 38% of the variance in box office performance ($R^2 = 0.3824$).

When applying the modules to box office success, the observed patterns for box office mirrored those from Long-term Audience Engagement. Notably, Black ethnicity exhibits a consistent negative relationship with box office performance across all three modules (T1, T2, and T3), while Hispanic and Asian ethnicities display similar, insignificant relationships. A noteworthy discovery from the robustness check is the insignificance of interaction terms applying the Delta rule, except for the Hispanic T1 interaction term.

It's crucial to note that metrics for T2 and T3 are based on the conversations that happen in the film. Given that box office success is likely influenced by advertising, marketing, and external factors, it is expected that the actual film content may be correlated but may not be the cause.

*Simplistic measure for Racial Representation:* The initial study's findings challenge previous research by Kuppuswamy and Younkin (2016) and Kim et al. (2020), who observed a positive correlation between the number of Black actors in a film and box office success. To address the contradictory findings between these studies and that of Kuppuswamy and Younkin (2016) and Kim et al. (2020) the focal variables in the model were replaced with variables representing the total number of Black, Asian, and Hispanic actors.

This analysis yielded similar positive effects the Asian ($\beta = -0.016$, $p = 0.004$), Black ($\beta = -0.005$, $p = 0.012$) and Hispanic ($\beta = -0.018$, $p = 0.006$) ethnicities all showed positive correlations with regard to Long-term Audience Engagement. This suggests that a higher number of actors from these ethnicities had a positive relationship with Long-term Audience Engagement, aligning with Kuppuswamy and Younkin's (2016) and Kim and other's (2020) findings.

However, examining only the total number of actors fails to consider the overall cast size, introducing a bias caused by the omission of this crucial factor. While Kim et al. (2020) partially control this issue, Kuppuswamy and Younkin (2016) neglects the total cast size altogether. To overcome this limitation, I utilized the Simpson diversity index to quantify racial diversity throughout the entire cast. The complete method is outlined on the next page. Incorporating this metric uncovered a negative correlation between racial diversity and Long-term Audience Engagement ($\beta = 0.268$, $p < 0.001$).

Therefore, upon this examination, it became apparent that the discrepancy in the findings stemmed from the different metrics used for Racial Representation rather than a fundamental difference in the effects of Racial Representation itself.

{Racial diversity through Simpson diversity index}

Inverse Simpson diversity index: After determining the ethnicities of individuals, following previous research I assessed the racial diversity of the entire cast. To measure racial diversity, I used the inverse Simpson index (1949). The Simpson diversity index is a measure of diversity that takes into account the number of ethnicities present, as well as the relative abundance of each ethnicity. It is calculated using the following formula:

$$D = 1 - / (\Sigma\ ni(ni-1)/N(N-1))$$

Where: n is the number of actresses and actors that belong to an ethnicity i and N is the total number of actresses and actors. These were the results of the regression for clarity I only added the significant coefficients. No violations of the assumptions of a regression analysis were detected.

| *Variable* | *Coefficient* | *Sign* | Variable | *Coefficient* | *Sign* |
|---|---|---|---|---|---|
| *Inverse Simpson index* | 0.268 | < 0.001 *** | | | |
| *Log(Screens)* | -0.461 | < 0.001 *** | *Biography* | 0.163 | 0.006 ** |
| *Log(Runtime)* | -1.351 | < 0.001 *** | *Sport* | 0.136 | 0.025 * |
| *Log(MPAA)* | -0.711 | < 0.001 *** | *War* | 0.143 | 0.043 * |
| *Log(Average budget)* | -0.047 | 0.008 ** | *Family* | 0.208 | < 0.001 *** |
| *Log(Sequel)* | 0.006 | < 0.001 *** | *Musical* | -0.292 | 0.007 ** |
| *Based on book* | -0.132 | 0.034 * | *History* | 0.159 | 0.028 * |
| *Based on play* | 0.375 | 0.012 * | *Documentary* | 0.749 | < 0.001 *** |

| | | | | | |
|---|---|---|---|---|---|
| *Based on comic book* | -0.314 | < 0.001 *** | *Western* | 0.492 | < 0.001 *** |
| *Based on novel* | -0.108 | 0.002 ** | *2009* | 0.172 | 0.023 * |
| *Log(Nominee)* | -0.009 | < 0.001 *** | *2010* | 0.201 | 0.009 ** |
| *Log(Winner)* | -0.014 | < 0.001 *** | *2011* | 0.351 | < 0.001 *** |
| *Log(Director Power)* | -0.010 | < 0.001 *** | *2012* | 0.502 | < 0.001 *** |
| *Log(Metascore)* | -0.036 | < 0.001 *** | *2013* | 0.617 | < 0.001 *** |
| *Log(Star power)* | -0.004 | 0.032 ** | *2014* | 0.791 | < 0.001 *** |
| *Adventure* | -0.135 | < 0.001 *** | *2015* | 0.791 | < 0.001 *** |
| *Comedy* | 0.130 | < 0.001 *** | *2016* | 0.879 | < 0.001 *** |
| *Drama* | 0.149 | < 0.001 *** | *2017* | 0.823 | < 0.001 *** |
| *Romance* | -0.085 | 0.013 * | *2018* | 1.021 | < 0.001 *** |
| *Sci.Fi* | -0.109 | 0.006 ** | *2019* | 0.968 | < 0.001 *** |
| *R2* | 0.626 | | *Adj R2* | 0.617 | |
| *F Statistic* | 65.980*** | (df = 60; 2364) | | | |