Black Panther vs The Help, the battle of inclusion versus tokenism:

A Comprehensive Examination of Underrepresentation of Minorities in Hollywood Films

MASTER THESIS MARKETING ANALYTICS

FALL 2023

A logo with blue and yellow squares

Description automatically generated

**Name:** Bastiaan van Houten

**SNR:** 2082816

**Study program:** MSc Marketing Management/Analytics

**Academic year:** 2023-2024

**Submission date:** January 5, 2024

**Name supervisor** Dr. Arjen van Lin

**Name co-reader** Dr. Kunthi Kusumawardani

Management summary

In recent years, the issue of minority representation in Hollywood films has become increasingly prominent. Countless studies have been published on this topic, and there is a growing demand for clarity about the relationship between representation and its effects.

This paper addresses the pressing issue of diversity and representation within the film industry, specifically focusing on Hollywood films. Films are powerful cultural influencers that shape societal beliefs. The lack of representation in films can perpetuate negative stereotypes and marginalization, which can have a harmful impact on individuals and communities. The study highlights that films with diverse casts tend to perform better at the box office due to their wider appeal. This suggests that diversity is not only a moral imperative, but also a financial necessity. The study highlights the economic significance of the film industry, which generates billions of dollars in annual revenue and produces some of the most profitable products in the world.

Despite recent improvements in representation, the study underscores that existing measures often overlook the nuances of representation and the need for authentic inclusive representation (AIR). AIR is a comprehensive approach to representation that accounts for effective representation rather than mere tokenism. The study argues that AIR is essential for creating films that resonate with diverse audiences and positively impact society.

Preface

Table of contents

Chapter 1 . Introduction

* 1. 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. Besides this economic significance, films are related to cultural aspects of society.

Films are not only a form of entertainment, they also 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)

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 (The Annenberg Foundation, 2018; Castañeda, 2015; Ross, 2019; Kubrak, 2020; Buchanan, 2005).

Moreover, representation in films is not only important for social factors, studies show that films with diverse casts appeal to broader audiences and perform better at the box office. (Annenberg Foundation, 2018) For example, a 2018 University of Southern California study found that films with diverse casts were 1.4 times more likely to be seen by broader audiences, and a 2021 McKinsey & Company study found that films with casts that had at least 30% minority actors were 1.3 times more likely to be profitable. By prioritizing diversity and representation, filmmakers can create more inclusive and authentic stories, attract larger audiences, and boost their bottom line. (Whitten, 2019; Bunche, 2018; Reporter, 2021).This suggests that 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, a study by UCLA found that in the years from 2011 to 2021, films went from 51.2% having less than 11% of colored actors to 28.8% of films featuring a cast that has a majority of colored actors.

Nevertheless, the public, media, and most studies adopt a narrow approach measuring diversity. (Malik, 2022) Most studies do not make a distinction between different minority groups and measure diversity based on the share of minorities in the cast. This approach has 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 a gap of 13.2% (compared to an actual population of 20%) respectively. (UCLA, 2021; MPAA, 2021). As a result, this group still experiences the social consequences of being underrepresented.

Secondly, measuring diversity based solely on minority share overlooks crucial information about the nature of the representation. (Malik, 2022; Lazar, 2020) To effectively address the social dimension associated with diversity, authentic inclusive representation (AIR) is necessary. (Lazar, 2020; Roughton , 2014). AIR is in this study referred to the practice of accurately and respectfully depicting characters from diverse backgrounds, this means going past tokenism and trying to create meaningful storylines for the characters. (Lazar, 2020). With AIR being arguably crucial for effective representation this study argues that by neglecting to consider AIR alongside diversity, previous studies have fallen short to conduct a comprehensive analysis of the effects of racial diversity. This causes that the true effects on both the social dimensions and financial implications are not being properly studied.

Lastly, as discussed previously racial representation in film is a social construct. Nevertheless, most studies use box office revenue as the dependent variable to analyze film success (Malik, 2022; Kuppuswamy, 2016 ; Madongo, 2023; Eliashberg, 2014; Gemser, 2007). While box office revenue is a useful measure of commercial success, it does not necessarily reflect the quality of a film or its impact on audiences, as it is highly influenced by marketing , advertising and hype. (Eliashberg, 2014; Michel Clement, 2014) Therefore, this study will choose long-term audience engagement as dependent variable. As discussed later this concept entails both social and economical constructs which makes it a better measurement for a film’s success in the context of a social concept such as racial representation.

1.2. Problem Statement

Following the problem background this studies problem statement is formulated as follows: “*What is the relationship between racial diversity in Hollywood films and LTAE, and to what extent does the AIR mediate this relationship?”*

1.3 Academic contribution

Previous studies on racial diversity in films have taken a narrow approach, either by focusing on only one or two ethnic groups (Patel, 2015; Hall, 2020; Dixon, 2000 ; Kuppuswamy , 2016) or by grouping all minority groups together (Aumer, 2017). This study addresses these limitations by including multiple minority groups and distinguishing between them. This is important because different minority groups have different experiences and perspectives, and their representation in films should be considered separately. Moreover, analyzing all minority groups separately but at the same time displays real world scenario’s, improving the generalization of the results.

Second, as noted by Malik (2021, p. 1), 'there are no clearly defined, standardized, and scalable metrics for taking stock of racial minorities’ cinematographic representation'. Methodologies used in previous work such as Weitzman et. Al (1972) and Smith et al. (2013) has been done by manual annotation. Which does produce high quality insights, but requires intensive time and are often expensive. This study uses an automated process to measure the concept of AIR and racial diversity. By doing so, this study not only aims to be as inclusive as possible, going beyond mere inclusion of actors but also makes it applicable for automation. Therefore, it can be easily applied to a large number of films.

Building upon prior research on gender biases in film (Agarwal et al., 2015; Kagan et al., 2020), this study seeks to standardize the concept of AIR through the Bechdel-Wallace test (1985). This study is not the first time the Bechdel test regards to racial representation is applied. The REM test was developed by UCLA CSS in 2020. Nevertheless, CSS has never tried to automate this test. More importantly the CSS doesn’t make a distinction between the different minorities groups in their test. By combining the works of previous researchers this study tries to automate the Bechdel test for racial representation in films.

Third, as mentioned before by using Long-term audience engagement (LTAE), 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 nuanced understanding of how audiences are engaging with films with diverse casts. Moreover, this causes that 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.

Using this approach, this study builds upon previous research and provides a more comprehensive and in-depth analysis of the relationship between representation and long term audience engagement in films.

1.4 Managerial contribution

This study offers valuable insights for filmmakers, studios, and stakeholders, helping them make informed decisions that can significantly impact a film's success in both financial and cultural terms. By intentionally casting diverse roles and developing characters that genuinely represent a variety of backgrounds, ethnicities, and cultures, filmmakers can attract a wider audience, which can increase a film's potential for financial success and profitability—a key goal in this commercial industry.

Furthermore, by using a measurement which entails long-term success it allows studios to assess whether embracing diversity in their casts and themes can expand their audience reach over time. This would signal a necessary shift in focus going from short-term box office results to long-term audience engagement. Moreover, if the research findings confirm that minority groups respond more favorably to films where they are authentically represented, marketers and distribution teams can leverage these insights to develop targeted marketing strategies. This could lead to more effective marketing activities.

Lastly, the study underscores the significance of Authentic Inclusive Representation (AIR) in films. By emphasizing this concept, filmmakers can understand that simply including minority characters is not enough; authenticity is essential. Focusing on crafting narratives that genuinely reflect the diversity of human experiences enables filmmakers to create stories that resonate emotionally with audiences. Ultimately, this understanding can lead to better practices of racial representation.

1.5 Structure of the Thesis

To address the research questions stated earlier, a quantitative research approach was employed in this study. This paper will first present a theoretical framework to contextualize the research, followed by a detailed description of the research methodology. Afterwards it will be discussed how the data collection and processing went. Finally, the results will be discussed.

Chapter 2. Literature review and hypotheses

This literature review delves into the current existing literature on the relationship between racial diversity in Hollywood films and LTAE. To provide clarity before delving into the relationships we first define LTAE and racial representation, and discuss the importance of AIR in the latter.

2.1 LTAE (Long-term audience engagement )

Audience engagement is a complex and multifaceted concept that encompasses the active involvement of an audience. In this study, the focus is on long-term audience engagement (LTAE), which goes beyond passive consumption and encompasses factors such as word-of-mouth, cultural impact, and sustained interest over time. (Broersma, 2019; Kumar, 2022)

Before delving into the social attribute of LTAE it’s important to clarify that LTAE has financial purposes as well. LTAE encompasses dimensions of ancillary revenues such as engrossed viewing, longevity in distribution channels, and the purchase of film-related merchandise (Kumar, 2022). These factors signify a lasting connection with the film beyond its initial screening and address the mechanism that LTAE is able to provide financial components over a longer period of time. These financial constructs exist through the social components of LTAE.

LTAE is a key factor in the media's capacity to shape attitudes, beliefs, and behaviors. It allows for a deeper and more sustained connection between the audience and the film, which can have a lasting impact on their lives. (Tan, 2018) The effect of LTAE and changing behavior is demonstrated by Bard research in 2006, which discovered that individuals highly interested in violent video games for a longer period of time were more prone to displaying heightened levels of aggressive behavior, aggressive cognitions, feelings of anger, and physiological arousal.

Furthermore, the effects of long term engagement with media and changing our beliefs was displayed by Mastro et al. in 2007 which found that white people who were constantly exposed to negative racial stereotypes in the media were more likely to hold those stereotypes themselves. This was especially true for white people who don't have much real-life contact with people of color. When we're constantly exposed with negative portrayals of certain groups of people, it's hard not to start believing them. Therefore, the importance of racial representation in media will be discussed in the following paragraph.

2.2 Racial representation in films

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. (Annenberg Foundation., 2018; 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, 2005) The spread of positive and accurate portrayals in the media is therefore essential for people to explore their identities also with regard to race.

Racial identity is a complex concept that is constantly evolving. This study uses the defintion of Umaña-Taylor AJ, (2014 p. 3) which defines racial identity as “a multidimensional psychological construct that reflects the beliefs and attitudes that individuals have about their ethnic group memberships”. The four ethnicities groups discussed in this study are Black, Hispanic, White and Asian.

Our racial identity is not solely a product of our own perspective but is also shaped by how others perceive us. 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).Therefore, the powerful impact of media representations on shaping perceptions cannot be underestimated.

Positive depictions of communities of color can diminish feelings of threat and social distance among white audiences (Dalisay & Tan, 2009; Ortiz & Harwood, 2007), whereas negative portrayals, such as associations with criminality, can exacerbate negative stereotypes and widen divisions among ethnic groups (Abraham & Appiah, 2006; Hurley et al., 2015).

Further in this study the ongoing debate regarding the extent to which films should mirror the diversity of the real world will be discussed. Ultimately, filmmakers must understand the potential impact their creative choices can exert on viewers. Moreover, as highlighted in the next few paragraphs for filmmakers to understand how to represent ethnic minorities in a successful manner is as equally important.

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, 2021). Nevertheless, overlooking authenticity in representation can result in films being mistakenly categorized as racially diverse while still perpetuating stereotypes and contributing to marginalization. As discussed previously these stereotypes in the film can perpetuate biases and misconceptions about minority groups. (Umaña-Taylor AJ, 2014; Abraham & Appiah, 2006; Hurley et al., 2015) Therefore, the concept of Authentic Inclusive Representation (AIR) emerges as a crucial element in portraying marginalized groups.

The pursuit of authenticity in the representation of minorities in films is not a recent development. In 2014, Ralph Roughton stressed that genuine understanding and empathy, free from stereotypes, are key to changing attitudes. Effective representation requires viewers to truly understand and empathize with characters. (Roughton, 2014) AIR, as a concept, describes how accurately and respectfully a film portrays underrepresented groups in a nuanced manner. To make this concept more tangible, researchers often use the Bechdel-Wallace test (1985), which was originally designed to measure the authentic representation of women in a film.

The Bechdel-Wallace test can be adapted to measure the representation of ethnic underrepresented groups which will result in the following question: Do two {ethnicity} named characters in {movie} have a conversation that does not center around a white character. An example for such a conversation would be a conversation between the main character and his wife in the movie ‘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 it is not a very thorough and meaningful conversation the characters both are African Americans they are named characters and not a white person is present or mentioned in the conversation. So this movie would have AIR for black people. With an understanding of the concepts, the following paragraphs will explain how they relate to each other.

2.2.1 Racial diversity and LTAE:

The power of storytelling lies in its ability to create connections between viewers and characters who possess relatable qualities and admirable traits (Murray, 1999; Appiah, 2001; Hall, 2020). This connection is strengthened when there are similarities in demographic factors such as race, 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 deeply engaged with the film and its characters. (Murray, 1999; Appiah, 2001; Hall, 2020) This suggests that when a film embraces racial diversity, it has the potential to attract a broader audience and foster greater overall LTAE. (USC Annenberg, 2018)

The demand for greater racial diversity in film is not universally embraced. A study by King (2020) identifies a group while not opposing racial representation in film, prioritizes the quality of the storyline. Arguing that when minorities are introduced without developed characters it influences the quality of the storyline. Patel (2015) argues that the efforts to increase racial diversity in films have faced criticism, particularly from those who fear change and the increased visibility of people of color due to a persistent culture of colonialism or systematic racism

Moreover, some studies have argued that the industry's practice of whitewashing is based on an assumption that white majority audiences prefer racially homogeneous casts and that diversity would not positively impact LTAE (Weaver, 2011). This assumption is believed to have led to reduced budgets allocated by studios for projects with higher racial diversity, as studios aim to mitigate financial risks.

Supporting this trend, research indicates that films featuring racial minorities in lead roles often receive significantly less production budget (Smith et al., 2020). It's important to note that lower production budgets are a significant predictor of lower box-office sales (Eliashberg, 2014; Michel Clement, 2014), making these films less likely to achieve success, regardless of their diversity.

Nevertheless, it's essential to emphasize that the preference for whitewashing doesn't hold true for all scenarios (Katherine-Aumer, 2017). Recent studies have also raised doubts about the idea that white actors are financially indispensable for the success of films (Chow, 2016). 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.

Furthermore, Teresa Correa's study in 2011 delved into the connection between racial diversity and social media engagement, revealing that minority groups tend to be more prolific creators of online content. This heightened interaction between minority communities and media is further evidenced by the active involvement of African-American and Latin communities in both traditional and digital entertainment, surpassing their proportional representation in the U.S. population (Gonzalez, 2014; MPAA, 2014).

Furthermore, this trend is poised to become even more impactful in the future as minority groups continue to grow as a percentage of the total U.S. population (Desilver, 2015). This underscores the potential benefit of appealing to these minority communities. When a film successfully resonates with these audiences through representation, it is more likely to stimulate online discussions, thereby fostering a deeper sense of LTAE (Kumar, 2022).

Finally, 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. (USC Annenberg, 2018) Despite the challenges of tokenism which would be discussed in the next paragraph, the prominent consensus is that increased racial diversity in films has a positive influence on LTAE. Based on the considerations, the hypothesis for the relationship is as follows:

**H1:** Increased racial diversity has a positive influence on LTAE.

2.4.1 Diversity , LTAE and AIR:

There is an intuitive belief that racial diversity leads to authentic inclusive representation (AIR). The rationale behind this notion is straightforward. The more people from different races are included, the more likely it is that one of them will provide an authentic representation for that ethnicity. Moreover , others argue that diversity encompasses a variety of perspectives and experiences, which can enhance authenticity of characters. This is because diverse casts can help to challenge stereotypes and assumptions about different cultures. As a result, they can help to create more nuanced and complex representations of people from different backgrounds. (Smith, 2020).

However, including diverse characters in films is not without its challenges. One concern is tokenism, which occurs when underrepresented characters are introduced without fully developed storylines. This can lead to backlash and damage LTAE if viewers perceive the diversity as being insincere , which can ultimately affect the perceived quality of the film (Smith, 2016) . For example, when minorities are introduced through tokenism, it can create the perception of a "racial agenda," which is often disliked by viewers (King, 2020).

Specifically, negative stereotypes are seen as a gross misrepresentation by balanced critics who advocate for avoiding films with stereotyping regardless of representation. Since critic reception plays a crucial role in shaping LTAE, misrepresentation is likely to lead to a decline in LTAE (Hofmann, 2016; Ghiassi, 2017; Kuppuswamy, 2016; Kumar et al., 2022).

Moreover, these negative depictions can cause minority groups to disengage completely from the media. For example, a study by El Hazzouri (2019) found that ethnic minorities who saw public health ads featuring people from their own ethnic group were less likely to follow the advice in the ads than those who saw ads featuring white people. The authors explained this by saying that minorities felt like they were being negatively stereotyped by the advertisers.

In contrast AIR can have a positive effect if done correctly. According to research, large-budget films in 2021 performed better when they had more racial authentic representation. (Lazar, 2020) The study emphasizes that racial representation must be authentic in order to foster empathy, understanding, and connection among viewers.

Moreover, there is a growing body of research that suggests that authentic representation of diverse individuals on screen, free from stereotypes, has a profound educational and socially engaging impact on audiences. (Bamford, 2018) This transfers to all types of media, as shown by a study by Roberts, (2021) which found that news stories that portrayed diverse cultures and identities in an authentic way were more likely to foster empathy and understanding among readers. This is likely because seeing oneself and one's own experiences reflected in the media can lead to a deeper emotional connection with the characters and the story, making the media more engaging Strong predilection for the idea that accurate representation is paramount and that misrepresentations are unethical. Therefore, the hypothesis will be the following:

**H2:** The relationship between racial diversity and LTAE is mediated through AIR.

2.7 Conclusion

This literature review provides a comprehensive overview of the research on the relationship between racial diversity in Hollywood films and long-term audience engagement (LTAE). The review defines and elaborates on key concepts, such as LTAE, racial representation in films, authentic inclusive representation (AIR) and cultural resonance. The review also included factors that have previously shown to have an effect on a films succes. Even though these were initially focused on a movie’s finanical succes these most likely provide important input in this study. The review concludes that there is a complex and nuanced relationship between racial diversity and LTAE. And most importantly it identified the gaps in current literature. The findings of this study could have important implications for the film industry and for society as a whole.

2.8 Theory

A diagram of a person's relationship

Description automatically generated

Figure 1 Conceptual model paper

The conceptual model consists of three multiple regressions. Thorough model specification can be found in the research methodology. For the relationship between Diversity and AIR, it will be a logistic regression. The models will e a log-log regression model.

3. Research Methodology

The main objective of this study is to examine the relationship between diversity and LTAE , with an additional focus on authentic and inclusive representation (AIR). The following paragraphs describe the data collection process and the decisions made in the sampling process.

3.1. Data sources

*Film characteristics*: We collected data on film characteristics from IMDb, complemented by the Numbers and The Movie Database (TMDB) as alternative sources. These are all trusted platforms with millions of registered users, ensuring credibility and diverse perspectives (Ghiassi, 2017).

*Race determination:* Data on race was collected using the Kairos API, a deep learning algorithm that can detect the ethnicity of individuals in images. There were in total 121,991 actors and actresses in all films. By using the API for around 100,000 of them the ethnicity could be determined for which the images could be found on TMDB or IMDb. This list includes duplicates if actors were listed in more than one film.

It is important to note that the absence of an actor or actress's image on these platforms generally means that they did not have a significant role in any film. This means it can be almost be determined from this that these actors and actresses played characters that are not considered authentic and inclusive (AIR). As a result, they are less influential or of no purpose in the analysis.

In the few cases where all actors were used for an analysis, such as for robustness checks, the ethnicities of the imageless people were based on their first and last name using the R package Rethnicity, which has an accuracy of around 80%. After removing the uncredited (13.491) and the actors/actresses which could not be assigned (77) the dataset consisted of 108423 people.

*AIR determination*: There are two automated ways to conduct the Bechdel test to determine the level of authentic and inclusive representation (AIR) in the films studied. The first method, consistent with prior studies (Argawal, 2015), involved the analysis of film scripts, sourced from websites used in previous studies (Hunter, 2016; Eliashberg, 2014). The second approach uses subtitles for the hearing impaired, readily accessible from opensubtitles.org, which offered a user-friendly API that allowed for the rapid collection of a large number of subtitles.

The primary distinction between these two approaches lies in the nature of the data. Subtitles for the hearing impaired represent a finalized version of the film and capture dialogues precisely as they appear in the final version of the film. Film scripts, on the other hand, can change during filming, so the script dialogue may not match the final film exactly. Screenplay however offer a better indication of switch in scene’s allowing for a more accurate way of scene identification.

Of the around twenty-two hundred films in our dataset, we were able to find subtitles for around eleven hundreds of them, which were in a format which could be standardized to do the tests. Details on subtitle processing are discussed later in this section.

3.2 Sample

The films for the sample were chosen carefully and based on specific criteria. The sample was limited to a particular time frame, starting from 1998, due to the dependent variable (DV) became publicly available. The films were filtered on being produced by companies within the United States. The films with animation as a genre were excluded because voice actors are not represented on screen, which undermines representation.

Following the approach of Joshi and Mao (2012), the analysis only included films that received a wide release, requiring a minimum of 500 screens at their launch, thereby filtering out independent films. During dataset analysis, it was observed that IMDb popularity rankings, although present between 1998-2000, was not widely used, as shown in Figure 2. Additionally, ratings from that era appeared to lack reliability. Therefore, the films from these years were filtered out. To further refine the dataset observations with missing data were eliminated. Following these criteria, the final sample consisted of 2,201 unique films.

3.3: Variable operationalization

*LTAE* : We operationalized (LTAE) using IMDb's movie score, a metric derived from popularity rankings. This movie score includes direct indicators of LTAE, such as online discussions, reviews, and word-of-mouth conversations. By examining movie scores over time, our study aims to measure sustained engagement, where lower scores indicate higher popularity. Specifically, the metric used is the average popularity rankings over a one-year period, starting in the third year after a film's release.

IMDb Pro was used to collect the popularity scores of the films. The scraper used for this process is detailed in the Appendix. However, it is important to note that this scraper is specific to the computer used for scraping and is not transferable to other systems. This limitation arises because IMDb Pro requires user login credentials. Figure 2 displays the average popularity rankings for the films per year, highlighting that rankings before 2000 were not particularly representative, as previously discussed.

A graph with a line going up

Description automatically generated

Figure 2 Average popularity per year

To assess the generalizability of our findings, considering the DV is rooted in the behavior of IMDb users, demographic data on IMDb users was collected and compared to moviegoers.

IMDb has around 503 million monthly visits, 62% of those users are men.(Similarweb, 2023) This is a gender imbalance, given that movie audiences are generally split evenly between men and women with 53% of moviegoers being men. (MPA, 2022) However, the gender distribution of ticket sales is more comparable to that of IMDB users, with 59% of tickets sold to men. This suggests that male viewers mostly pay for their female companions. More importantly with the large user base and the rather small gap the differences seem neglectable.

In terms of age distribution, as shown in figure 3 and 4 IMDb's demographics align with those of moviegoers. Comparison data from Similarweb and MPAA, although employing different age groups, consistently highlight a spike in the 25-39 age group. The only notable difference between IMDB users and moviegoers is the underrepresentation of older age groups on IMDb.

It's essential to acknowledge that IMDb's user base is predominantly American (Similarweb, 2023). Nevertheless, considering the study's focus on multicultural societies, the Hollywood film industry, and the cultural aspects of racial representation, it is logical to conduct the analysis with a United States-centric approach, with the possibility of transferring results to countries with a similar cultural landscape.

Overall, IMDb is believed to be a reliable source of data for this study, given its large user base and diverse perspectives. To enhance the validity of our claims, we will address the nationality imbalance by primarily concentrating on the United States. Additionally, it is important to acknowledge the limitations of the data, considering the disparities in gender and age representation.

A screenshot of a graph

Description automatically generated

Figure 4 Age distribution Moviegoears MPAA (2022)

Figure 3 Age distribution IMDB users

A graph with numbers and a number of people

Description automatically generated with medium confidence

*Racial diversity*: With regards to measuring diversity , the inverse Simpson index (1949) will be used. The Simpson diversity index is a measure of diversity that takes into account the number of species present, as well as the relative abundance of each species. It is calculated using the formula:

D = 1 - / (Σ ni(ni-1)/N(N-1))

Where:

-n is the number of organisms that belong to species i   
- N is the total number of organisms.

With using the Kairos API the dataset with regards to ethnicity finally consisted of the actors, their characters and four ethnicities all with their respective probabilities. On average, these probabilities were as follows: Asian (0.086), Black (0.1220), Hispanic (0.074), and White (0.7148), indicating a notable skew towards white actors. Figure 5 below show the distribution of probabilities for ethnicity, displaying an unequal distribution among different ethnicities. This is caused due to majority of actors being white in the industry. Because mostly when a person is assigned the high probability of white it assigns lower probabilities to the other ethnicities. (COMPARE TO anneberg) 1600 films

A graph of different sizes and shapes

Description automatically generated with medium confidence

Given these probabilities, an initial filtering process was essential to establish appropriate thresholds for ethnicity assignments. With regards to the uneven distributions among the ethnicities two thresholds were introduced:

* A probability exceeding 0.6 led to an ethnicity assignment value of 1.
* Values between 0.25 and 0.6 received an assignment value of 0.5.

Utilizing the 0.6 threshold, 98,905 individuals were successfully identified an ethnicity. The additional 0.25 threshold was used to categorize the remaining observations. This approach aimed to retain meaningful data while eliminating less informative values. [[1]](#footnote-1)

After determining the ethnicities of individuals, various methodologies for calculating the reverse Simpson index were explored. This index could be used to assess the diversity of the entire cast, regardless of whether the cast included characters with generic names (e.g., "repair man"). The second method was to only focus on the named characters.

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 these tokens using a stop word list. The stop word list contained 915 words, such as "doctor," "agent," and "the." However, I made an exceptions to the stop word filter, I did not remove the character if the stop word was the first token in a name. For example, "Colonel Rich Bron" is kept. The exception to this rule was if the character existed of a single word for example "Colonel" would be removed.

The initial approach yielded an average Simpson diversity index of 0.4001. When considering only the diversity of named characters, the average of the index was 0.3953. This suggests that when focusing solely on named characters, ethnic diversity in films tends to decrease, suggesting a tendency for films to feature important characters of the same ethnicity.

To delve deeper into the relationship between more inclusive characters and racial diversity, we introduce the concept of Racial Diversity in Authentic Inclusive Representation (AIR).

Previous claim made that actors and actresses which did not have an image mostly do not have a significant role. This became apparent in the filtering process, where 77% of non-named characters with an image were removed, while only 66% of non-named characters without an image were removed. As a result, the percentage of characters which where assigned an ethnicity through name instead of facial recognition decreased from 15% to 11% after filtering.

*AIR*: In addition to the primary objective of this study, we aim to explore the viability of conducting a partial mediation analysis to examine the role of Authentic Inclusive Representation (AIR) within the primary relationship. Building upon previous research, notably Agarwal's work in 2015, which focused on automating the Bechdel test, our study takes a step further. Our approach extends the work of the CSS UCLA, which has also ventured into automating the Bechdel test. However, it's worth noting that their approach lacks differentiation among various demographic groups and does not incorporate automation.

In our study, AIR is treated as a dichotomous variable, with a value of 1 signifying the presence of AIR and a value of 0 denoting its absence. Nonetheless, to ensure a more comprehensive analysis, we will apply different criteria for assessing the Bechdel test. These criteria include, two named characters of {ethnicity} ethnicity appear in the film, (T2) who speak to one another, (T3) about something besides a white character. This approach allows us to examine and quantify the presence of AIR across various measurements, facilitating a more nuanced evaluation of representation in films.

*Subtitle file parching:* Subtitles for the hearing impaired where used. These subtitle files can be extracted from a movie in the form of .srt files. These are text files with very strict formatting. Each subtitle within these files features a unique identification, precise start and end times, and one or two lines of text, indicating when the subtitles should appear on screen. For instance, consider the following excerpt from the opening of the movie "300: Rise of an Empire" in 2014:

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.

These subtitle files where cleaned and a logic for a scene identification was applied. As can be seen in file snippet when it’s not clear when a person is talking that person is mentioned. This was referred to in our datasets as the ‘speaker’. Moreover, to identify more people related in the to the scene entities were identified with model en\_core\_web\_md of the Natural Language package spacy. This meant that our dataset consisted of speakers, people mentioned and scene index.

The scene identification was done by going through the file and looking for pauses in dialogue. If a pause in dialogue exceeded five seconds, it was identified as a new scene. Five seconds was chosen because it allows the audience to mentally adjust to the new scene without disrupting the narrative flow. However, because this is a threshold, even if a scene switch took longer, it was still counted as one scene switch.

This approach has two ways of causing errors:  
1. It does not identify scene changes with had dialogue pause shorter than five seconds.  
2. It does not identify dialogues that did not change scene even after a five-second break in dialogue.

To see if these errors occurred a lot a robustness check was done manually.

After identifying the scene’s the ‘speakers’ in the scene and the people mentioned in the scene the three separate AIR tests could be done. To provide some context about the AIR tests table 2 is provided. Here can be seen that if the character has to be ‘named’ the films which receive a pass on test already drop dramatically in step one. And by comparing this first step already indicates numerous films have acted in tokenism because this already indicates a type of tokenism. All the films which have surpassed step 1 with regards to all characters but not for named characters indicates that these films which have more than one character from an ethnicity did this by introducing at least one generic character.

3.4 Covariates:

Although demand factors are not the substantive focus of the paper, controlling for their effect allows for more precise testing of the effects of the market factors. In order to account for additional factors that have been identified as influencing a film’s success, this study draws upon previous research. Combinations of the following variables explained at 60% of the variation in box office revenue for certain studies. The variables considered in this study include Series, the star power of Actors and Directors, MPAA rating, Number of screens, critical acclaim as measured by ratings from Critics, Awards, Budget, Genre, and Release date. Table 2 provides detailed information on how these variables are measured.

*Series:* Films belonging to a series, such as sequels or prequels, tend to enjoy greater success. Fans of the series are already familiar with the characters and story, making them more inclined to watch the new installment (Hofmann, 2016; Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022). In this study, (SERIESi) represents the number of sequels associated with a film.

*Star power of actors and directors:* Films featuring well-known actors and directors are more likely to succeed as audiences are drawn to familiar and esteemed talent (Hofmann, 2016; Michel Clement, 2014; Ghiassi, 2017; Kuppuswamy, 2016; Kumar et al., 2022).

The star power score (STARPOWERi) is derived from Nelson and Glotfelty's (2012) measure, based on the highest-grossing actors' ranking on The Numbers. (DIRECTORPOWERi) is similarly measured using the highest-grossing directors' ranking on The Numbers.

*Critical acclaim*(CRITICSi): Films that receive positive reviews from critics are more likely to be successful than films that receive negative reviews. This is because positive reviews can help generate interest in the film and encourage people to see it (Hofmann, 2016; Ghiassi, 2017; Kuppuswamy, 2016; Kumar et al., 2022).

*Awards:* Awards (NOMINATIONSi, WINSi): Films that win prestigious awards, such as Oscars and Golden Globes, tend to be more successful, as these accolades lend credibility and appeal to audiences. (Kuppuswamy, 2016; Kumar et al., 2022; Hunter et al., 2016).

This research will use the actual number of award (NOMINATIONSi) as a proxy for award nomination. (WINSi) will also be added to the model to represent awards. Because there is an abundance of available movie awards of which some are more known and valuable than others, there is focused on a select group of awards following the research of Gemser, Leenders and Wijnberg (2008). [[2]](#footnote-2)

*MPAA rating:* Films with a PG-13 or R rating are more likely to be successful than films with a G or PG rating. This is because audiences are more likely to see films that are considered to be more mature and exciting (Hofmann, 2016; Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022). (MPAAi) rating is given by the Motion Picture Association of America and is used in the United States and its territories to rate a film's suitability for certain audiences based on its content. These ratings are encoded [0 = unrated; 1 = G; 2 = PG; 3 = PG-13; 4 = R;5 = C-17].

*Number of Screens* (#SCREENSi): Movies shown on more screens are more likely to succeed because they reach a wider audience (Hofmann, 2016; Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022; Hunter et al., 2016).

*Budget(BUDGETi):* Films with a higher budget are more likely to be successful than films with a lower budget. This is because higher budgets can be used to produce higher-quality films with better special effects and production values (Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022; Hunter et al., 2016). The production budget available at one of the three data sources was used if multipleproduction multiple budgets where available across the sources the average was taken.

*Genre:* Some genres, such as action and comedy, are more popular than others, such as drama and documentary (Hofmann, 2016; Michel Clement, 2014; Ghiassi, 2017; Kuppuswamy, 2016; Kumar et al., 2022). This study introduces 20 genres , Action, Adventure, Comedy, Crime, Drama, Family, Fantasy, Horror, Romance, Musical, Sci-Fi, Mystery, Thriller, Western, Biography, Documentary, History, Music, Sport and War. Because a movie could have multiple genres, these dummy variables are not mutually exclusive.

*Seasonality*(SPRINGi, SUMMERi, FALLi, WINTERi): Films that are released during the holiday season or during summer are more likely to be successful than films that are released at other times of the year (Hofmann, 2016; Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022). Within this study the four seasons are encoded as the following. Spring[March, April, May] Summer [June, July, August] Fall [September, October, November], Winter [December, January, February].

In addition to the control variables listed above, this study also includes dummy variables for whether the movie is a (REMAKEi) or (SPINOFFi) (Hennig-Thurau and Houston, 2019; Hofmann, Clement, Völckner, and Hennig-Thurau, 2017). These variables are important to control for because remakes and spinoffs tend to be more successful at the box office than original films due to their established familiarity with audiences and existing fan bases.

*Runtime*(RUNTIMEi): is included as the actual numerical value in minutes, following (Holbrook, 1999).

Furthemore, Following Hofmann, Clement, Völckner, and Hennig-Thurau (2017), multiple dummy variables were added to control for whether the movie was (BASED ONi) a book, comic, novel, short story, or TV series. Similar , to the genre variables these dummy variables are not mutually exclusive because a movie could be based on multiple sources.

To control for possible differences over time, (YEARi) dummies for every year in the sample were added to the model. Because log-log linear regressions are used in this research, all control variables that are not dummy variables were log-transformed.

It is crucial for this study to be cautious in every step of the process to possibly achieve causality. Therefore, every factor previously mentioned will also be analyzed on its influence on the racial diversity of the movie before it is included as a control variable. Two significant factors that have demonstrated a correlation with racial diversity in films are the film genre and budget. In 2019, action films exhibited a notably higher likelihood of including characters from underrepresented racial and ethnic groups. Additionally, films with larger budgets tend to have more diverse casts, as they possess greater resources to hire a wide range of actors (Smith, 2020).

Table 1: Measures of Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Measure | Data Source |
| Simpson diversity indexi | Measure of diversity which takes into account the number of species present, as well as the relative abundance of each species. | Continuous value between 0-1 | IMDB, TMDB , Kairos |
| AIRi | Authentically Inclusive Representation. | Dichotomous variable with three separate level | Subtitles hearing impaired |
| Log(LTAEi) | Long-Term Audience engagement. | IMDB popularity score where a higher score means being less popular. Average ranking over a one year period of time in the third year after release logged. | IMDB |
| Covariates | - | **-** | **-** |
| SERIESi | Number of films associated with the film |  | IMDB |
| Log(STARPOWERi) | Log-transformed score of the top three actors in the movie + 0.001. | Log-transformed sum of 100 divided through the star ranking + 1 (e.g., 100/ rank 2 = 50 + 1) for the top 3 actors one year before the movie release. | The Numbers |
| Log(DIRECTORPOWERi) | Log-transformed score of the number one director in the movie + 0.001. | Log-transformed sum of 100 divided through the director ranking + 1 (e.g., log(100/rank 2 +1)) for the director one year before the movie release. | The Numbers |
| Log(CRITICi) | Log-transformed average rating the movie received from professional movie reviewers + 0.001. | Log-transformed average rating the movie received from professional movie reviewers | Metacritic |
| Log(NOMINATIONSi) | Log-transformed number of award nominations the movie received + 0.001. | Log transformed number of award nominations the movie received + 1 (e.g., log(50+1)). | IMDb |
| Log(WINSi) | Log-transformed number of award nominations the movie received + 0.001. | Log transformed number of award nominations the movie received + 1 (e.g., log(50+1)). | IMDB |
| Log(MPAAi) | Log-transformed MPAA rating the movie received 1 = Not Rated, 2 = G, 3 = PG, 4 = PG-13, 5 = R, 6 = NC-17 | - | IMDb |
| Log(SCREENSi) | Number of screens at release | Total amount of screens logged | Numbers |
| Log(BUDGETi) | Production budget of film | The average production budget of the data sources | IMDB, Numbers, OMDB |
| ACTIONi | Genre Action | Genre Action = 1, Other = 0 | IMDb |
| ADVENTUREi | Genre Adventure | Genre Adventure = 1, Other = 0 | IMDb |
| ANIMATIONi | Genre Animation | Genre Animation = 1, Other = 0 | IMDb |
| COMEDYi | Genre Comedy | Genre Comedy = 1, Other = 0 | IMDb |
| CRIMEi | Genre Crime | Genre Crime = 1, Other = 0 | IMDb |
| DRAMAi | Genre Drama | Genre Drama = 1, Other = 0 | IMDb |
| FAMILYi | Genre Family | Genre Family = 1, Other = 0 | IMDb |
| FANTASYi | Genre Fantasy | Genre Fantasy = 1, Other = 0 | IMDb |
| HORRORi | Genre Horror | Genre Horror = 1, Other = 0 | IMDb |
| MUSICALi | Genre Musical | Genre Musical = 1, Other = 0 | IMDb |
| MYSTERYi | Genre Mystery | Genre Mystery = 1, Other = 0 | IMDb |
| ROMANCEi | Genre Romance | Genre Romance = 1, Other = 0 | IMDb |
| SCI-FIi | Genre Sci-Fi | Genre Sci-Fi = 1, Other = 0 | IMDb |
| THRILLERi | Genre Thriller | Genre Thriller = 1, Other = 0 | IMDb |
| WESTERNi | Genre Western | Genre Western = 1, Other = 0 | IMDb |
| BIOGRAPHYi | Genre Biography | Genre Biography = 1, Other = 0 | IMDb |
| DOCUMENTARYi | Genre Documentary | Genre Documentary = 1, Other = 0 | IMDb |
| MUSICi | Genre Music | Genre Music = 1, Other = 0 | IMDb |
| HISTORYi | Genre History | Genre History = 1, Other = 0 | IMDb |
| SPORTi | Genre Sport | Genre Sport = 1, Other = 0 | IMDb |
| SPRINGi | Movie is released in the spring 1 = Movie released in the spring, 0 = other season | Spring release = 1, Other = 0 | IMDb |
| SUMMERi | Movie is released in the summer 1 = Movie released in the summer, 0 = other season | Summer release = 1, Other = 0 | IMDb |
| AUTUMi | Movie is released in the autumn 1 = Movie released in the autumn, 0 = other season | Autum release = 1, Other = 0 | IMDb |
| WINTERi | Movie is released in the winter 1 = Movie released in the winter, 0 = other season | Winter release = 1, Other = 0 | IMDb |
| Log(RUNTIMEi) | Log-transformed duration of a movie in minutes | Log-transformed duration of a movie in minutes | IMDb |
| BOOKi | Movie is based on a book 1 = Based on a book, 0 = not based on a book | - | IMDb |
| COMICi | Movie is based on a comic 1 = Based on a comic, 0 = not based on a comic | - | IMDb |
| NOVELi | Movie is based on a novel 1 = Based on a novel, | - | IMDb |
| YEAR2000i | Year2000 | 2000 = 1, other = 0 | IMDb |
| YEAR2001i | Year2001 | 2001 = 1, other = 0 | IMDb |
| YEAR2002i | Year2002 | 2002 = 1, other = 0 | IMDb |
| YEAR2003i | Year2003 | 2003 = 1, other = 0 | IMDb |
| YEAR2004i | Year2004 | 2004 = 1, other = 0 | IMDb |
| YEAR2005i | Year2005 | 2005 = 1, other = 0 | IMDb |
| YEAR2006i | Year2006 | 2006 = 1, other = 0 | IMDb |
| YEAR2007i | Year2007 | 2007 = 1, other = 0 | IMDb |
| YEAR2008i | Year2008 | 2008 = 1, other = 0 | IMDb |
| YEAR2009i | Year2009 | 2009 = 1, other = 0 | IMDb |
| YEAR2010i | Year2010 | 2010 = 1, other = 0 | IMDb |
| YEAR2011i | Year2011 | 2011 = 1, other = 0 | IMDb |
| YEAR2012i | Year2012 | 2012 = 1, other = 0 | IMDb |
| YEAR2013i | Year2013 | 2013 = 1, other = 0 | IMDb |
| YEAR2014i | Year2014 | 2014 = 1, other = 0 | IMDb |
| YEAR2015i | Year2015 | 2015 = 1, other = 0 | IMDb |
| YEAR2016i | Year2016 | 2016 = 1, other = 0 | IMDb |
| YEAR2017i | Year2017 | 2017 = 1, other = 0 | IMDb |
| YEAR2018i | Year2018 | 2018 = 1, other = 0 | IMDb |
| YEAR2019i | Year2019 | 2019 = 1, other = 0 | IMDb |

3.5 Models:

Following Clement, Wu and Fischer (2014), this research uses log-log linear regressions.

However, because table 2 shows that many films did not receive any award nomination and taking the log of ‘0’ would lead to an error, all films receive a ‘+0.001’ on award nomination before taking the log.

Due to hetregneoty this study uses a log-log model. Due to with a log log model the vairablea are not able to be negative or equal to zero or control variables therefore to every control variables small constant of 0.001 was added.

log(LTAE)i = β0 + β1 × Simpson diversity indexi + β2 × AIRi +

β3 × SERIESi + β4 × log(STARPOWERi) + β5 × log(DIRECTORPOWERi) +

β6 × log(CRITICi) + β7 × log(NOMINATIONSi) + β8 × log(WINSi) +

β9 × log(MPAAi) + β10 × log(SCREENSi) + β11 × log(BUDGETi) +

β12 × ACTIONi + β13 × ADVENTUREi + β14 × ANIMATIONi +

β15 × COMEDYi + β16 × CRIMEi + β17 × DRAMAi + β18 × FAMILYi +

β19 × FANTASYi + β20 × HORRORi + β21 × MUSICALi +

β22 × MYSTERYi + β23 × ROMANCEi + β24 × SCI-Fii +

β25 × THRILLERi + β26 × WESTERNi + β27 × BIOGRAPHYi +

β28 × DOCUMENTARYi + β29 × MUSICi + β30 × HISTORYi +

β31 × SPORTi + β32 × SPRINGi + β33 × SUMMERi +

β34 × AUTUMNi + β35 × WINTERi + β36 × log(RUNTIMEi) +

β37 × BOOKi + β38 × COMICi + β39 × NOVELi + β40 × YEAR2000i +

β41 × YEAR2001i + β42 × YEAR2002i + β43 × YEAR2003i +

β44 × YEAR2004i + β45 × YEAR2005i + β46 × YEAR2006i +

β47 × YEAR2007i + β48 × YEAR2008i + β49 × YEAR2009i +

β50 × YEAR2010i + β51 × YEAR2011i + β52 × YEAR2012i +

β53 × YEAR2013i + β54 × YEAR2014i + β55 × YEAR2015i +

β56 × YEAR2016i + β57 × YEAR2017i + β58 × YEAR2018i +

β59 × YEAR2019i + εi

Main relationship:

Mediation:

Relationship diversity and AIR (A path mediation):

p(AIR) = e ^ (β0+ β1*seriesi*+ β2*actori*+ β3*directori* + β4*ratingi* + β5*MPAAratingi* + β6*numberofscreensi* + β7*criticali* + β8*awardsi* + β9*budge*t*i* + β10- β22*genrei* + β22*releasei* + β23*diversityi)/ (1 +* e ^( β0+ β1*seriesi +* β2*actori …)*

Relationship AIR and LTAE (B path mediation):

LTAE = β0 + β1*seriesli +* β2*actori +* β3*directori +* β4*ratingi* + β5*MPAAratingi +* β6*numberofscreensi +* β7*criticali +* β8*awardsi +* β9*budgeti +* β10- β22*genrei* + β22*releasei +* β24*AIRi* + ε

What do you want to show:  
1. Differences in diversity in main leads and non main leads. The already finished dataset of IMDB has only fur characters those are the four most important players.

2. Point in time, difference in diversity in the films will probably not change. Because 2016 but how it was perceived is that changed. So rating before, just right after and a longer period of time after. Plus the diversity in the films . before just right after and after a longer period of time.

3.

3.1.5 Final sample/ Data description

Data summary

After collection and cleaning we had four major datasets.

1. One with the films and the characteristics of the films based on the control variables.
2. One with the films and their popularity scores
3. One with the actors of the films and their ethnicity distribution
4. One with the movie’s subtitles for the hearing impaired.

Relationship cultural resonance and diversity

- Prove no multicollinearity

Year = significant , weird because it is 3 years after release so it should not be affected by . IS this due to the growth of IMDB? And that more people look up more films pushing less relevant films back?

Na 2016 niet meer sig voor 2016 wel een negatieve relatie. Komt dit door AIR?

Regressions models

Main relationship: Is there a relationship between racial diversity and LTAE

Affect cultural resonance: Does cultural resonance cause more diversity in films.

* Is the affect of diversity on LTAE more when cultural resonance appreciates representation more.

Mediator: AIR, does part of the relationship run through AIR?  
**Stat video: How to test for moderation if you have one categorical variable and one in scale?**

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated  
Check for multicollinearity

Check observation independent

Check if the variances equal across treatment groups

Check if the dependent variable is normally distributed.

Basically is racial diversity higher after oscarssowhite.

Distinguish three groups?

Simple effects is possible to do a diff in diff with simple effect,   
A picture containing text, font, screenshot, algebra

Description automatically generated

Stereotyps

**African americans**

(NMAAHC) National museum of african american history and culture. “laziness, as well as characteristics of submissiveness, backwardness, lewdness, treachery, and dishonesty, historically became stereotypes assigned to African Americans”. Mammy’s is also seen as stertypes related to African American woman these are overweight, self-sacrificing domestic worker.

The stereotype of Uncle Tom is innately submissive, obedient and in constant desire of white approval. depicted African American women as aggressive, loud, and angry - in direct violation of social norms.

Watermelon , Mandingo as animalistic and brutish

**Hispanics** , lazy, less financial independence   
  
positive characteristics; strong family, tradition-loving, religious, and old-fashioned. uneducated, poor, rebellious, physically violent, dirty/smelly, noisy, and criminally inclined. Additional characteristics that distinguished Hispanics were more ambiguous with respect to their valence. Hispanics were viewed as less conservative, materialistic, competitive, self-centered, progressive, and individualistic, and as more oldfashioned.

less value on physical fitness, mature love, recognition by the community, a good life for others, national greatness, delaying pleasure to achieve success, the pursuit of knowledge, good health, economic prosperity, and financial independence.

(Jackson, 1995)

Asian:  
The so-called “model minority” stereotype, one of the most pervasive and harmful assumptions about Asian Americans, holds that Asian Americans are a uniformly high-achieving racial minority that has assimilated well into American society through hard work, obedience to social mores and academic achievement. Asian American men are often seen as effeminate or asexual, while Asian American women are viewed as sexually desirable, exotic and passive.

The difficult part about this factor is that the movie will most of the time only receive a sequel if the initial movie was a success. This makes it difficult to analyse it’s impact on LTAE because is a movie popular because it’s part of a collection or the movie was successful and therefore it’s part of a collection.

4. Results

Additional analysis

The last remark this study would like to make is that as explained previously, films and culture are deeply intertwined. In today’s society the following values are seen as highly important belonging, community, and personal growth, which are all closely related to representation. (Neufeld, 2020) It is proven that films that properly represent the cultural zeitgeist in which the film is made are more successful. (Ettema, 2005)

Therefore, it is reasonable to assume that films that reflect the values related to representation will be more successful in today's cultural landscape. However, as societal norms and values evolve, this appreciation might wonder off. This study aims to explore how much of a movie's success is due to the effects caused by representation, and how much is due to the degree the films is displaying the current cultural zeitgeist which is focused on representation.

Finally, prior studies (Malik, 2022; Kuppuswamy, 2016) have overlooked cofounding variables and macro environmental factors, and have largely focused solely on the relationship between racial diversity and movie success. This study introduces cultural resonance to the analysis, arguing that it is a crucial factor in the success of films with diverse casts, as it may influence how one's cultural values shape their perception of these films.

2.5 Cultural resonance

Cultural resonance, as discussed in this study, refers to how well a film connects with the cultural values, experiences, and zeitgeist of its audience. It can be achieved by either reflecting or challenging the audience's own culture, values, and experiences in a thought-provoking way. (Ettema, 2005) Representation is key, as films that represent different cultures and experiences can broaden audiences and create a sense of connection. (Bamford, 2018)

For the sake of clarity and illustration, the film Black Panther was praised for its positive portrayal of African culture and its representation of African people. It was released at a time when representation was a major topic of discussion, and its popularity amplified the conversation about race and representation. Additionally, its positive portrayal of African culture helped to promote African fashion. Thus, Black Panther is both a product of culture, parts of its success is caused by culture and it is a force for change in culture.

2.5.1 Culture as cofound/moderator

Culture has a big impact on how long audiences stay engaged with a film. Racially diverse films often explore important social and cultural issues, which can lead to lasting discussions about race, identity, and social justice. As Garrett (2020) points out, "Films about race offer us the chance to grapple with past and present constructs of racism, power, and oppression." This shows that cultural shifts are driving conversations about racial diversity, which are essential for keeping audiences engaged in the long term. This underscores the idea that when films activate these conversations they will remain more under the attention for longer periods of time.

*Brokeback Mountain, https://www.tandfonline.com/doi/full/10.1080/19359705.2013.792128*

Moreover, contemporary film critics place a high value on representation and inclusion in media. (Akser, 2021) As a result, films featuring racially diverse casts tend to receive more favorable critical reception, which in itself contributes to LTAE. Finally, cinema's transformative capacity on culture extends to its influence on the hiring practices within the industry, resulting in increased racial diversity in films. (Khrebtan-Hörhager, 2011) With regards to the above considerations, the following hypothesis is proposed:

**H4**: The relationship between racial diversity and LTAE is cofounded by culture resonance.

.2.3 Study 3: Cultural resonance as cofounding variable

Oscars so white   
https://www.tandfonline.com/doi/full/10.1080/15205436.2017.1409356?src=recsys

*Variable operationalization* : for the relationship studied a diff in diff analysis was seen as the most fitting. Quantifying whenever racial representation started to become a major topic in today’s culture. Eventually the event #oscarssowhite was chosen as natural schock. n the context of a difference-in-differences (DiD) analysis, a "natural shock" refers to an external event or occurrence that affects one group or entity being studied but not another group, and it is not caused by any deliberate intervention or treatment by the researchers.

*Model* :

# Bibliography

Agarwal, A. &. (2015). Key Female Characters in Film Have More to Talk About Besides Men: Automating the Bechdel Test.

Akser, M. (2021). Diversity and Inclusion in Film, Television and Media Sector: Policy Alternatives for an Inclusive Film Industry and Training. *CINEJ Cinema Journal, 9*(1), 1-13. doi:10.5195/cinej.2021.414

Alexander Dhoest, N. S. (2015). Still ‘Watching’ TV? The Consumption of TV Fiction by. *Media and Communication*, 176-184. doi:10.17645/mac.v4i3.427

Association., M. P. (2021). Theatrical Market Statistics. Retrieved from https://www.motionpictures.org/wp-content/uploads/2022/03/MPA-2021-THEME-Report-FINAL.pdf

Bamford, N. (2018). The Strangers’ Case: harnessing the power of screen entertainment to. *In: Cross-Cultural Communication Conference*. Retrieved from https://eprints.bournemouth.ac.uk/30284/

Basuroy, S. C. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing*, 103–117. Retrieved from http://www.jstor.org/stable/30040552

Belton, J. (1995). *Films and Mass Culture.* The Athlone Press. Retrieved from https://www.goodreads.com/book/show/1639322.Films\_and\_Mass\_Culture

Benshoff, H. M. (2016). America on Film: Representing Race, Class, Gender, and Sexuality at the Films.

Boboltz, S. &. (2017). Why On-Screen Representation Actually Matters.

Brad J. Bushman, P., & L. Rowell Huesmann, P. (2006). Short-term and Long-term Effects of Violent Media. *Personality and Social Psychology*.

Broersma, M. (2019). *Audience Engagement. .* doi:10.1002/9781118841570.iejs0060.

Buckingham, D. (2008). *Youth , identity and digital media.* MIT Press, Cambridge, Mass. Retrieved from https://library.oapen.org/bitstream/handle/20.500.12657/26085/1/1004001.pdf

Bunche, R. J. (2018). "The Business Case for Diversity in the Entertainment Industry". Retrieved from https://bunchecenterold.pre.ss.ucla.edu/2018/02/28/new-hollywood-diversity-report-2018/

Castañeda, M. (2015). The power of (mis)representation: Why racial and ethnic stereotypes in the media matter.

Dan Rubin, I. M. (2022). *Beyond the box office: A conceptual framework for the drivers of audience engagement,* (Vol. 151). Journal of Business Research.

Dittmar, L. &. (2016). Racial and gender biases in magazine advertising persisted, and in some cases increased, from 1983 to 2014. *Marketing Science*(35), 911-926.

Dixon, T. L. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication,, 50*, 131-154. doi:https://doi.org/10.1111/j.1460-2466.2000.tb02866.x

Einav, L. (2007). Seasonality in the U.S. Motion Picture Industry. *The RAND Journal of Economics, 38*, 127–145. doi:http://www.jstor.org/stable/25046296

Elberse, A. &. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science., 22*, 329-354. doi:10.1287/mksc.22.3.329.17740.

Eliashberg, J. H. (2014). Assessing Box Office Performance Using Movie.

Ettema, J. S. (2005). Crafting cultural resonance: Imaginative power in everyday journalism. *Journalism , 6*(2). doi:https://doi.org/10.1177/1464884905051005

Foundation, T. A. (2018). Inclusion in the Director’s Chair? Gender, Race & Age of Directors across 1,000 films from 2007-2017. Retrieved from https://www.bing.com/ck/a?!&&p=6955b1041ed8beadJmltdHM9MTY4MzY3NjgwMCZpZ3VpZD0wZTJmYjE3ZS04MWM3LTZiOWMtMWU2OC1hMzE5ODA0NDZhOGMmaW5zaWQ9NTE5Ng&ptn=3&hsh=3&fclid=0e2fb17e-81c7-6b9c-1e68-a31980446a8c&psq=Inclusion+in+the+Director%e2%80%99s+Chair%3f+Gender%2c

Garrett, G. (2020). *A Long, Long Way.* Oxford University Press. Retrieved from https://global.oup.com/academic/product/a-long-long-way-9780190906252?cc=nl&lang=en&

Gemser, G. V. (2007). The impact of film reviews on the box office performance of art house versus mainstream motion pictures. *Journal of Cultural Economics*, 43-63.

Ghiassi, M. S. (2017). Predicting Movie Success Based on IMDB Data. *Journal of Big Data*, 1-14. doi:https://doi.org/10.1186/s40537-017-0072-4

Govindaraju, W. W. (2010). Explaining Global Box-Office Tastes in Hollywood Films: Homogenization of National Audiences’ Movie Selections. *Communication Research, 37*, 215 - 238.

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. Retrieved from Psychology of Aesthetics, Creativity, and the Arts

Hunter, S. D. (2016), June 1). Predicting Box Office from the Screenplay: A Text Analytical Approach .

Initiative., A. I. (2018). Inclusion in Netflix scripted series and films. Retrieved from https://annenberg.usc.edu/sites/default/files/2019/02/11/MDSCI\_Inclusion\_in\_Netflix\_2.11.19.pdf

Jackson, L. A. (1995). Stereotypes, Emotions, Behavior, and.

Katherine Aumer, D. B. (2017). Assessing Racial Preferences in Films: The Impact of Mere-Exposure and Social Identity Theory. Retrieved from https://www.scirp.org/pdf/PSYCH\_2017071814195150.pdf

Khrebtan-Hörhager, J. (2011). The Art of Creating a Cross-Cultural Dialogue on Screen: "The Way Home "& "The Colour of Fear". *The International Journal of the Arts in the Society*. Retrieved from https://www.academia.edu/29706441/The\_Art\_of\_Creating\_a\_Cross\_cultural\_Dialogue\_on\_Screen\_The\_Way\_Home\_and\_The\_Colour\_of\_Fear\_

Kidd, M. A. (2015). Archetypes, stereotypes and media representation in a multi-cultural.

King, J. &. (2020). Representing race: the race spectrum subjectivity of diversity in film. *Ethnic and Racial Studies, 44*, 1-18. doi:10.1080/01419870.2020.1740290.

Kohl, B. R. (1989). Predicting financial success of motion pictures: The '80s experience. *Journal of Media Economics}, 2*(2), 35-50. doi:10.1080/08997768909358184

Kubrak, T. (2020, May 2). Impact of Films: Changes in Young People's Attitudes after Watching a Movie. *Behav Sci (Basel)*. doi:10.3390/bs10050086.

Kumar, D. I. (2022). Beyond the box office: A conceptual framework for the drivers of audience engagement. *Journal of Business Research*.

Kuppuswamy, V. &. (2016). Testing the theory of consumer discrimination as an explanation for the lack of minority hiring in Hollywood films. Retrieved from https://ssrn.com/abstract=2738728

Kwon, O. L. (2020). Movie Success Prediction using Historical and Current Data Mining. *Advances in Intelligent Systems and Computing*, 491-498. doi:https://doi.org/10.1007/978-981-15-5566-4\_59

Lazar, L. H. (2020). Driving box office performance through authentically inclusive storytelling. Retrieved from https://www.fullstoryinitiative.com/Full\_Story\_Research.pdf

Lemmens, J. V. (2011). Psychosocial causes and consequences of pathological gaming. Computers in Human Behavior. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0747563210002116

Leonard, D. J. (2006). *Screens Fade to Black:.* Praeger Publishers. Retrieved from https://cultureinjection.files.wordpress.com/2020/07/david-j.-leonard-screens-fade-to-black-contemporary-african-american-cinema.pdf

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

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

Mestyán, M. Y. (2013). Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data. doi:https://doi.org/10.1371/journal.pone.0071226

Michel Clement, S. W. (2014). Empirical generalizations of demand and supply dynamics for films. *International Journal of Research in Marketing*, 207-223.

MPA. (2022). *Theatrical and Home Entertainment Market Environment Report.* Retrieved from https://www.motionpictures.org/wp-content/uploads/2022/03/MPA-2021-THEME-Report-FINAL.pdf

Musa Malik, F. R. (2022, February). Representations of Racial Minorities in Popular Films, A Content-Analytic Synergy of Computer Vision and Network Science. *Computational Communication Research, 4*(1). doi:https://doi.org/10.5117/CCR2022.1.006.MALI

Neufeld, D. (2020). LEADERSHIPThe World’s Most Influential Values, In One Graphic. *Visual Capital*. Retrieved from https://www.visualcapitalist.com/most-influential-values/

NMAAHC. (n.d.). Popular and Pervasive Stereotypes of African Americans. Retrieved from https://nmaahc.si.edu/explore/stories/popular-and-pervasive-stereotypes-african-americans

Patel, L. (2015, November). Desiring Diversity and Backlash: White Property Rights in Higher Education. *Regional & Urban Planning, 47*. doi:https://doi.org/10.1007/s11256-015-0328-7

Prag, J. C. (1994). An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. . *Journal of Cultural Economics*, 217–235. doi:https://doi.org/10.1007/BF01080227

Ramesh Sharda, D. D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications, 30*(2), 243-254. doi:https://doi.org/10.1016/j.eswa.2005.07.018.

Reporter, T. H. (2021). Audiences Prefer Hollywood Films With Diverse Casts. Retrieved from https://newsroom.ucla.edu/releases/2021-hollywood-diversity-report

Roberts, J. (2021). Empathy Cultivation through (Pro)Social Media: A Counter to Compassion Fatigue. *Journal. Media*, 819-829. doi:https://doi.org/10.3390/journalmedia2040047

Smith, S. L. (2020). Inequality in 1,300 popular films: Examining portrayals of gender, race/ethnicity, LGBTQIA, and disability from 2007–2019.

Sperling, N. (2021, March). Hollywood Loses $10 Billion a Year Due to Lack of Diversity, Study Finds. *New York times*. Retrieved from https://www.nytimes.com/2021/03/11/films/hollywood-black-representation.html#:~:text='Ahsoka'-,Hollywood%20Loses%20%2410%20Billion%20a%20Year%20Due%20to%20Lack%20of,representation%20can%20lead%20to%20change.

Tan, E. S. (2018). A psychology of the film. . *Humanities and Social Sciences Communications*. Retrieved from https://doi.org/10.1057/s41599-018-0111-y

Teresa Correa, S. H. (2011). RACE AND ONLINE CONTENT CREATION. *Information Communication and Society*. Retrieved from https://www.researchgate.net/publication/232924586\_RACE\_AND\_ONLINE\_CONTENT\_CREATION

UCLA. (2021). Hollywood Diversity Report 2021. Retrieved from https://socialsciences.ucla.edu/hollywood-diversity-report-2021/

Umaña-Taylor AJ, Q. S.-D. (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. 21-39. doi:10.1111/cdev.12196.

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

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

1. I acknowledge that race ethnicity is not binary and reducing it to only four separate ethnicities is simplistic, however the software programs used don’t allow for a more complex understanding of race. [↑](#footnote-ref-1)
2. The awards used in this research are the following: Academy Awards, Critics’ Choice Awards, Directors Guild Awards, Golden Globe Awards, Golden Laurel Awards, Independent Spirit Awards, Los Angeles Film Association Awards, MTV Awards, National Board of Review Awards, National Society of Film Critics Awards, New York Film Critics Circle Awards, People’s Choice Awards and the Screen Actors Guild Awards. [↑](#footnote-ref-2)