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 yellow and blue text

Description automatically generated

Marketing Department

Tilburg School of Economics and Management

­­

Name : Bastiaan van Houten

SNR: 2082816

Supervisor: Arjen van Lin

Co-Reader: Kunthi Kusumawardani

Abstract

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 movies. Movies 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 movies 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. Movies generate nearly $100 billion annually through theaters and home channels. Moreover, a film’s 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, movies are related to cultural aspects of society as well.

Movies are not only a form of entertainment, they also reflect and shape the cultural context in which they are created. They can mirror societal values, attitudes, and perspectives. Moreover, they 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 diversity and representation, particularly in relation to 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. Additionally, when movies feature characters from various backgrounds, ethnicities, genders, and cultures, they more accurately reflect the diversity of the real world. (The Annenberg Foundation, 2018)

Representation in films is not only important for social reasons, but also for financial reasons. Studies have shown that films with more diverse casts tend to perform better at the box office. For example, a 2021 study by McKinsey & Company found that films with casts that were at least 30% minority actors were 1.3 times more likely to be profitable. This is supported by a study by the University of Southern California, which found that films with diverse casts were 1.4 times more likely to be seen by diverse audiences.

This suggests that diversity is both a moral and financial imperative for the film industry. A diverse cast can appeal to a broader demographic of viewers, making the film more relatable and interesting to a wider range of people. By prioritizing diversity and representation, filmmakers can not only create more inclusive and authentic stories, but they can also attract larger audiences and boost their bottom line. (Whitten, 2019; Bunche, 2018; Reporter, 2021).

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 ten 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. They focus on the inclusion of actors 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 and Native American communities remain severely underrepresented, with a gap of 13.2% (compared to an actual population of 20%) and 0.7% (compared to a population share of 2%), respectively. (UCLA, 2021; MPAA, 2021). As a result, these groups still experience the social consequences of being underrepresented.

Secondly, measuring diversity based solely on minority share overlooks crucial information about the nature of the representation (Lazar, 2020). It is argued that to effectively address the social dimension associated with diversity, authentic inclusive representation (AIR) is necessary. 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.

As explained previously, movies and culture are deeply intertwined. Our society currently values belonging, community, and personal growth, which are all closely related to representation (Neufeld, 2020). Therefore, it is reasonable to assume that movies that reflect these values will be more successful in today's cultural landscape.

However, as societal norms and values evolve, the types of representation that are valued may also change. This study aims to explore how much of a movie's success is due to intrinsic factors, such as seeing characters that we can relate to, and how much is due to the degree to which it aligns with current cultural beliefs.

1.2. Problem Statement and Research Questions

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 and cultural resonance affect this relationship?”*

The problem statement translates in the following research questions:

1. How does racial diversity in Hollywood films affect LTAE?
2. How does racial diversity affect the AIR in Hollywood films?
3. How does AIR in Hollywood films affect LTAE?
4. How does the relationship between representation LTAE and racial diversity get affected by culture resonance.

1.3 Academic relevance

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) 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.

Second, as noted by Malik , 2021 “there are no clearly defined, standardized, and scalable metrics for taking stock of racial minorities’ cinematographic representation”. This study uses an automated process to measure the concept of authentic inclusive representation (AIR) and the inverse Simpson diversity index (1949) . By doing this the study tries to include an approach that is as inclusive as possible but also standardized and can be automated and therefore easily applied to large sums of films.

Third, most studies use box office revenue as the dependent variable to analyze the success of a movie (Madongo, C. Z., 2023; Eliashberg, J. H., 2014; Gemser, G. V., 2007). Box office revenue is a useful measure of commercial success, but it does not necessarily reflect the quality of a film or its impact on audiences. Bu 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 streaming views.

Finally, prior studies (Malik, 2022; Kuppuswamy, 2016) have overlooked cofounding variables and macro environmental variables, 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.

By taking these innovative approaches, this study builds upon previous research and provides a more comprehensive and in-depth analysis of the relationship between racial diversity and representation in films.

1.4 Managerial relevance

The managerial relevance of this study lies in its potential to inform the decisions of filmmakers, studios, and other stakeholders about how to represent diversity in films and the effects of this representation on a film's success. By exploring the interplay between racial diversity, authentic inclusive representation (AIR), cultural resonance and sustained audience engagement over time (LTAE), the study reveals practical implications for crafting narratives that genuinely resonate with diverse viewer demographics. This information can be used to make decisions about casting, writing, marketing, and distribution.

For example, this study could help filmmakers identify the types of characters and stories that are most likely to resonate with diverse audiences. It could also help studios make more informed decisions about which films to greenlight and how to market them. By prioritizing LTAE as a key metric, the study underscores its importance in evaluating a film's lasting impact. This, in turn, contributes to the broader industry effort to create narratives that genuinely connect with diverse audience segments and foster long-term success.

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. Before delving into the relationships for clarity purposes an explanation of the main components LTAE, racial diversity , AIR and cultural is provided.

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, we focus 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) As discussed in the following two paragraphs it consists of social and financial concepts.

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 Brad's research in 2006, which discovered that individuals highly interested in violent video games were more prone to displaying heightened levels of aggressive behavior, aggressive cognitions, feelings of anger, and physiological arousal.

LTAE also encompasses various dimensions that contribute to ancillary revenues, such as engrossed viewing, longevity in distribution channels, and the purchase of film-related merchandise (Kumar, 2022). All factors signify the presence of a lasting connection with the film beyond its initial screening. More importantly it addresses the financial dimension LTAE encompasses of. In addition to LTAE, this study will also examine short-term audience engagement and how it compares to long-term engagement. We will explore how short-term engagement is influenced by factors such as the cultural climate at the time of release.

2.2 Racial representation in movies

Representation of minorities in movies refers to the presence and portrayal of characters from minority groups in film. This includes characters who are people of color, people with disabilities, LGBT+ people, and other marginalized groups. Portrayal of characters, themes, and ideas in films is fundamentally connected to psychological mechanisms and considerations making it first and foremost a psychological concept . (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) The absence of representation in turn has negative impacts on the self-esteem and self-worth. As individuals belonging to that group 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 racial identities.

Moreover, one of the major challenges for a multicultural society is promoting positive relationships among different people. In large populations **people may never interact with members of other groups due to economic isolation or ethnic segregation. In these situations, the only "interaction" people may have with other cultures is through media depictions. Therefore, the perception of some ethnic groups by some people is solely depicted on what they have seen in the media. Displaying the responsibility of media creators.** (Kidd, 2015)

Racial identity is a complex concept that is constantly evolving, and it allows people from different backgrounds to connect with others who share their experiences. Umaña-Taylor AJ, (2014) define racial identity as “a multidimensional psychological construct that reflects the beliefs and attitudes that individuals have about their ethnic–racial group memberships, as well as the processes by which these beliefs and attitudes develop over time”.

The demand for greater racial diversity in film is not universally embraced. A study by King (2020) identifies a group termed who, while not opposing racial representation in film, prioritize the quality of the storyline. Moreover, recent efforts to increase racial diversity in movies have faced criticism, particularly from those who fear change and the increased visibility of people of color. Some scholars, such as Patel (2015), argue that backlash for initiatives regarding racial equality may stem from a persistent culture of colonialism or systematic racism.

This suggests that while there is a growing appetite for diversity in film casts, not everyone agrees on its importance. The debate over the extent to which films should reflect the diversity of the real world continues to evolve, with differing perspectives on its necessity. Ultimately filmmakers should be aware of the potential impact that their choices can have on viewers.

2.3 Simpson diversity index

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. D = 1 - / (Σ ni(ni-1)/N(N-1)) where n is the number of organisms that belong to species i and N is the total number of organisms. Nevertheless, representation is more than just mere including actors which is discussed in the following paragraph.

2.4 AIR (Authentic Inclusive Representation)

There are many different ways to represent minorities in movies. Musa Malik (2022) states “as manual annotation of racial and gender biases in films is a costly and labor intensive task, it is no surprise that large-scale attempts at understanding the structural and visual biases, latent in Hollywood movies, have not been undertaken”.

Some films and studies focus on the experiences of minority characters, while others simply include them as part of the larger cast. Overlooking authenticity in representation can result in movies being mistakenly categorized as racially diverse while still perpetuating stereotypes and contributing to marginalization. (Dixon, 2000) Stereotypes in film can perpetuate biases and misconceptions about minority groups. (Umaña-Taylor AJ, 2014)

*(AIR)* is a measure of how accurately and respectfully a film portrays underrepresented groups in a nuanced way. The REM test is a tool for assessing AIR in films. It was developed by the UCLA CSS and is based on the Bechdel-Wallace test (1985), which was originally designed to measure the representation of women in film. The REM test asks a series of questions the many questions in the test can be simplified to one question: Do two named non-white characters in {movie} have a conversation that does not reinforce positive or negative stereotypes, and that is not centered around a white character? [[1]](#footnote-1)

2.5 Cultural resonance

Cultural resonance is meant in this study to the extent to which a film connects with the cultural values, experiences, and zeitgeist of its audience. It can be achieved by 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.

For example, 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. In this way, Black Panther is both a product of culture and a force for change in culture.

2.6 Factors films success

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 least 60% of the variation in box office revenue for 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 1 provides detailed information on how these variables are measured.

*Series:* A film that is part of a series, such as a sequel or prequel, is more likely to be successful than a stand-alone film. This is because fans of the series are already familiar with the characters and the story, and they are more likely to see the new film. (Hofmann, 2016; Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022).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.

*Star power of actors and directors:* Films with well-known actors and directors are more likely to be successful than films with lesser-known talent. This is because audiences are more likely to see films with actors and directors that they know and like (Hofmann, 2016; Michel Clement, 2014; Ghiassi, 2017; Kuppuswamy, 2016; Kumar et al., 2022).

*Critical acclaim:* 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:* Films that win awards, such as Oscars or Golden Globes, are more likely to be successful than films that do not win awards. This is because awards can help legitimize the film and make it more appealing to audiences (Kuppuswamy, 2016; Kumar et al., 2022; Hunter et al., 2016).

*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). *Number of screens:* Films that are shown on more screens are more likely to be successful than films that are shown on fewer screens. This is because more screens mean that more people will have the opportunity to see the film (Hofmann, 2016; Michel Clement, 2014; Kuppuswamy, 2016; Kumar et al., 2022; Hunter et al., 2016).

*Budget:* 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).

*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). *Release date:* 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).

2.7 Hypotheses/ Theory

2.7.1 Racial diversity and LTAE:

It is widely recognized that viewers tend to connect with characters who possess relatable qualities or admirable traits. 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 movie embraces racial diversity, it has the potential to attract a broader audience and foster greater overall LTAE. (USC Annenberg, 2018)

A study by Teresa Correa (2011) explored the relationship between racial diversity and social media buzz. The study found that minority groups – African-Americans, Latinos, and Asians – tend to create online content more frequently than white students. This suggests that when a film attracts these audiences, it is more likely to generate online discussion, which is one of the factors of LTAE. (Kumar, 2022)

Some studies have argued that whitewashing is an audience preference, meaning that diversity will not have a positive effect on LTAE. However, a study by Katherine Aumer (2017) found that this preference is only found for Western stories and by monoracial People of Color Minorities when they are exposed to Asian actors. These findings suggest that other factors, specifically cultural values, may be important in understanding preference for White actors in the media. (Chow, 2016).

However, while the positive effects of racial diversity on LTAE are evident, some scholars have also noted potential challenges and critiques. For example, questions have arisen regarding tokenism, where characters from underrepresented racial backgrounds may be included merely for the sake of diversity, without fully developed storylines. Such instances can lead to backlash and diminish LTAE if viewers perceive the diversity as insincere. (Smith, 2016).

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, the prominent consensus is that increased racial diversity in films has a positive influence on LTAE. Based on these considerations, the hypothesis for the main relationship is as follows:

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

2.7.2 Diversity and AIR:

There is an intuitive belief that racial diversity leads to authentic inclusive representation (AIR). This is because the more people from different races are included, the more likely it is that one of them will provide an authentic representation of the culture. However, some argue that culturally-specific narratives, such as those seen in films such as "Coco" can offer a deeper exploration of cultural complexities through a racially homogenous cast, thereby enhancing AIR. This means that a movie does not necessarily need to be racially diverse to be AIR, as long as it provides an authentic representation of a particular culture.

Others argue that diversity encompasses a variety of perspectives and experiences, which can enhance AIR. 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). As a result, the hypothesis is as follows:

H2: The likelihood of a film featuring AIR increases with a rise in racial diversity.

2.7.3 AIR and LTAE:

According to the UCLA Center for Scholars and Storytellers, Creative Artists Agency, and Full Story Initiative report "AIR 2.0: Driving Box Office Performance Through Authentically Inclusive Storytelling," large-budget films in 2021 performed better when they had more authentic representation. The report also emphasizes that representation must be authentic in order to foster empathy, understanding, and connection among viewers (Lazar, 2020).

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. For example, a study by Roberts, (2021) 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 this helps viewers to see themselves and their own experiences reflected in the media. This can lead to a deeper emotional connection with the characters and the story, which can make the film more engaging. Therefore, the hypothesis will be the following:

H3: The presence of AIR in a movie positively influences its ability to capture LTAE.

2.7.4 Culture as cofound/moderator

Culture is an important factor to consider when examining the relationship between racial diversity and LTAE. Films that reflect the contemporary social and cultural zeitgeist can often connect more deeply with viewers because they address current issues, concerns, and trends. This is especially true for films that tackle important social and cultural issues through diverse characters and narratives.

In today's cultural zeitgeist, where representation is a major topic, racially diverse films often touch upon important social and cultural issues. This can have a lasting impact on society by promoting discussions on topics such as race, identity, and social justice.

Research by Martinez (2018) demonstrates that films that tackle these issues through diverse characters and narratives are more likely to become cultural touchstones, leading to sustained audience engagement. This is because they are more likely to connect with audiences on an emotional level and to leave a lasting impression.

Additionally, film critics currently highly value and advocate for representation and inclusion in the media. As a result, movies featuring a racially diverse cast tend to receive more positive critical reception, which is positively associated with LTAE.

Finally, it is believed that current cultural forces have caused changes in the hiring of minorities in the industry, leading to increased racial diversity in films. This suggests that culture resonance may influence the relationship between racial diversity and LTAE. With regards to the above considerations, the following hypothesis is proposed:

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

2.7.5 Additional factors

As previously mentioned there are numerus factors that have been proven to affect a movie’s success. The main factor that is expected to have an influence on the main relationship is culture. Nevertheless, 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. Factors that have previously shown proven relationships are the genre of the film, action films were significantly more likely to feature characters from underrepresented racial/ethnic groups in 2019 (Smith, 2020), the budget of the film: films with larger budgets are more likely to have more diverse casts, as they have more resources to hire a variety of actors.

2.8 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 movies, authentic inclusive representation (AIR), cultural resonance. It also provided an overview on the previoused researched factors influencing film success. Eventhough 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 litrature. The findings of this study could have important implications for the film industry and for society as a whole. By understanding the factors that influence the relationship between racial diversity and LTAE, we can work to create films that are more inclusive and that have a positive impact on audiences.

The following part of this are the hypotheses and the framework of this study, which explains the relationships which will be analyzed in this study trying to fill the gap in the existing literature.

2.9 Theory

A diagram of a model

Description automatically generated

Figure 1 Conceptual model paper

The conceptual model consists of three multiple regressions and one diff in diff which also will be used in a regression all can be found on the research methodology. For the relationship between Diversity and AIR, it will be a logistic regression.

4. Research Methodology

The main objective of this study is to examine the relationship between cast diversity and LTAE.

4.1 Data

4.1.1 Variables

LTAE will be measured using IMDb's movie score, which is derived from popularity rankings. The movie score includes direct indicators of LTAE, capturing factors such as online conversations, reviews, and word-of-mouth discussions. By analyzing movie scores over time, the study aims to track enduring engagement, where lower scores indicate higher popularity. This approach goes beyond a purely financial perspective, and instead utilizes a more comprehensive predictor. Additionally, comparing movie to movie directly helps eliminate variations caused by factors such as seasonality, providing a more accurate assessment.

Movie ranks from release date, also controlled by ranks one week before release.

Demographics imdb users,   
https://nycdatascience.com/blog/author/stephensh/

<https://nycdatascience.com/blog/student-works/web-scraping/whos-behind-imdbs-ratings/>

In addition to the main objective, this study seeks to investigate the feasibility of conducting a partial mediation analysis to examine the role of AIR in the main relationship. AIR is treated as a dichotomous variable, with a value of 1 indicating the presence of AIR and a value of 0 indicating its absence. To determine the presence of AIR, the study will utilize the R.E.M. (Race in Entertainment Media) test developed by Lazar et al. (2020) inspired by the Bechdel Test (1985). This test consists of four questions that can be condensed into a single question, which increases accessibility and objectivity in measuring and analyzing (refer to Appendix B)

The results for the R.E.M test can be collected using two possible approaches. Firstly, the test can be conducted through text analysis of the scripts obtained from websites that were utilized in previous studies (Hunter, 2016; Eliashberg, 2014). Alternatively, the single summarized question can be posed to the language model ChatGPT, which has shown an accuracy rate of around 95 percent during a small preliminary test involving a subset of 99 films (refer to Appendix C).

4.1.2 Data sources

IMDb is selected as the primary data source for this study, alternative data source is Numbers and The Movie Database (TMDB). All due to their reputation as trusted platforms with millions of registered users. This ensures credibility and allows for the inclusion of diverse perspectives (Ghiassi, 2017). To check the generalizability the demographics of IMDB users were analyzed.

The IMDB dataset also includes information on actor ethnicity, which is used to compute the Simpson Diversity Index.

Finally, in addition to collecting most variables from the IMDB database, MovieScore data will also be scraped using a methodology similar to that of collecting answers for the R.E.M. test. The sample will be carefully chosen, likely limited to a specific timeframe of a few years, to minimize heterogeneity. Factors such as the impact of Covid-19, the emergence of streaming platforms, and other relevant considerations will be taken into account during the sample selection process.

Test for AIR:   
Test 1: are there at least two named people from the ethnicity {} in the movie?

Resources for determining race

Kairos uses deep learning algorithms to detect ethnicity from pictures. This API was used to detect around 100000 from the 120000 actors this included duplicates if actors were listed in more movies than one. The pictures used were collected through TMDB and IMDB. If an actor or actress does not have a picture on of these platforms it mostly means that this person has not played in a significant role in any movie yet. Therefore, they are not seen as too important. Nevertheless, the remaining acotrs and actresses were assigned ehtnicities based on their first and last name through the R package Rethnicity. Which has an accuracy around eighty percent.

The dataset finally consisted of four ethnicities all with their respective probabilities. These probabilities where on average Asian(0.248), black (0.25) , Hispanic (0.249) and White (0.25) meaning that looking at overall probabilities that people belong to an ethnicity was somewhat similar. These probabilities meant that there was a initial filtering process needed to have a threshold to assign people different ethnicities. Sadly Kairo’s is way better in predicting ethnicity when someone is white. The distribution of the different ethnicities can be seen in the table below . This means when we do different thresholds and filter people out this will not be equally distributed among the different ethnicities. Therefore, we assigned two thresholds. One of 0.6 if the probability is higher than this the person is assigned a value one for the ethnicity. If the value is between 0.25 and 0.6 the person is assigned a value of 0.5. This could mean that some people have four ethnicities assigned. This threshold is aimed to eliminate all values with no meaning while still trying to remain most of the information.

A graph of different colored lines

Description automatically generated with medium confidence

Test 2 : Do two characters from the ethnicity {} talk to each other

So far we have the ethnicity for the different characters and their actors or actresses. The following step was to automate the process of analyzing screenplay or subtitles to identify if these characters in the movies had an interaction with one another. The choice between the different files either subtitle or screen play was a difficult one. While screen plays where mostly not available and possibly not displaying the actual dialogue in the movie. The subtitle file even though they would be for the hearing impaired where not always going to be naming the person talking.

4.1.3 Data collection

The results for the R.E.M test can be collected using two possible approaches. Firstly, the test can be conducted through text analysis of the scripts obtained from websites that were utilized in previous studies (Hunter, 2016; Eliashberg, 2014). Alternatively, the single summarized question can be posed to the language model ChatGPT, which has shown an accuracy rate of around 95 percent during a small preliminary test involving a subset of 99 films (refer to Appendix C).

We started by identifying what movies we would be interested in. The dataset we used as a base to build upon was ‘The Movies Dataset’[[2]](#footnote-2) found on the website Kaggle. Which was filtered as mentioned in the next section of this paper. The dataset consisted of forty five thousand movies which after cleaning was brought down to eight thousand six hundred and sixty. The dataset consisted among other off genre, production company etc. The dataset was constructed with the API of The Movie Database further referred to as TMDB.

We used IMDB pro to collect the popularity scores of the movies. The scraper for this process can be seen in the Appendix. Nevertheless, this scraper was specific for the computer used for scraping and therefore would not work for anyone else. This is due to IMDB pro acquiring a login.

The names of the cast where collected by the API of TMDB as well. After collecting the names we used the R package called “rethnicity” . Which divides the full named of a person and provides the probability of belonging to one race or the other. What is important to mention is that due to the ethnicity being based on first and last name people from southern Europe are mostly always classified as Hispanic. Hispanic having the definition “relating to Spain or to Spanish-speaking countries, especially those of Central and South America”. Also the perplexion of the skin of southern European are mostly also more relating southern American than Northern Europe.

4.1.4 Cleaning data

We chose to only have movies with the country of the production company to be from the United States to indicate that it was a least influenced by Hollywood.

We chose to remove the genre animation due to the fact that due to voice actors not being displayed on the screen.

We had a total of around 20000 movies. First there were some key filtering. Such as the movie having a value for all of the control variables in this study. Lastly we controlled for all movies having at least 500 opening theaters. After all cleaning the dataset concluded of 2418 movies. Budget was the variable where we used multiple sources. Due to differences in the budget value from these different sources the average was taken.

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.

4.1.5 Final sample/ Data description

Data summary

After collection and cleaning we had four major datasets.

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

4.2 Variable operationalization

4.3 Models

Main relationship:

LTAE = β0 + β1*series+* β2*actor+* β3*director +* β4*rating* + β5*MPAArating +* β6*numberofscreens +* β7*critical+* β8*awards+* β9*budget+* β10- β22*genre* + β22*release +* β23*diversity +* ε

Relationship diversity and AIR (A path mediation):

p(AIR) = e ^ (β0+ β1*series*+ β2*actor*+ β3*director* + β4*rating* + β5*MPAArating* + β6*numberofscreens* + β7*critical* + β8*awards*+ β9*budge*t+ β10- β22*genre* + β22*release* + β23*diversity)/ (1 +* e ^( β0+ β1*series +* β2*actor …)*

Relationship AIR and LTAE (B path mediation):

LTAE = β0 + β1*seriesl+* β2*actor+* β3*director +* β4*rating* + β5*MPAArating +* β6*numberofscreens +* β7*critical+* β8*awards+* β9*budget+* β10- β22*genre* + β22*release +* β24*AIR* + ε

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 movies pushing less relevant movies 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 movies.

* 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.

1. This study is not the first in trying to automate the Bechdel test see (Agarwal, 2015) for reference. [↑](#footnote-ref-1)
2. https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset [↑](#footnote-ref-2)