The impact of competition between popular videos on chart success in the Streaming Video on Demand industry

A study of the Netflix top 10

Name: Rob van der Wielen

SNR: 2084445

Study program: MSc Marketing Analytics

Academic year: 2023

Submission date: June 2, 2023

Name supervisor: Dr. George Knox

Name co-reader: Nazli Alagoz

MANAGEMENT SUMMARY

The streaming industry has experienced massive growth in recent times. However, due to the lack of transparency in viewing data, understanding this industry remains a challenge. The Netflix top 10 feature, which lists the most popular movies and TV shows, provides valuable insights into this relatively unknown industry. In this thesis, I aim to investigate how the presence of similar (i.e., same genre or MPAA rating as a Netflix video) popular videos on other SVOD platforms affect the popularity of videos on Netflix, which is measured by the cumulative weeks on the Netflix top ten, and whether this varies with seasonality. While the role of competition among popular videos in the motion picture industry has been extensively studied, the streaming industry has received little attention in this regard.

I utilised data from a third-party provider, FlixPatrol, to collect information of Netflix and its competitors. Additionally, I used the IMDB API to enhance the dataset with video characteristics. This resulted in 408 unique Netflix titles featuring on the Netflix top ten in the US during the year 2022. A multiple regression was conducted on the data.

The results show that competition between similar popular videos in the SVOD industry can affect the cumulative rank of these videos. More specifically, releasing a popular video with a similar genre on a competitor SVOD platform can lead to a decrease in cumulative rank of a popular Netflix video, while competitor videos with the same MPAA rating can have a small positive effect on the cumulative rank. Furthermore, I found a marginally significant effect that the negative effect of the number of competitor videos with a similar genre on the cumulative rank is strongest during holiday seasons and becomes weaker the further we move away from a holiday. For videos with a similar MPAA rating the positive effect also becomes weaker as we move further away from a holiday.

These findings add to the existing literature by showing the effects of competition between popular videos in the SVOD industry, partially confirming what was already found in the movie industry. That the genre does have a negative impact while MPAA rating does not. With these findings, movie studios and streaming platforms can carefully consider the timing of their video releases. Specifically, they may choose to delay or push forward the release of their videos to avoid simultaneous releases of competing and possibly popular videos, which may negatively affect the popularity of their content. Finally, content creators and providers can make informed decisions about their release schedule and avoid releasing similar videos during highly competitive seasons like holidays, where the negative effect of competition is stronger. However, the latter needs to be taken with caution.

PREFACE

I am pleased to present my thesis, which marks the end of my master's program in Marketing Analytics at Tilburg University.

I would like to express my gratitude to everyone who helped me along the way. First and foremost, I would like to thank my thesis supervisor, Professor George Knox, who provided me with valuable guidance, feedback, and expertise throughout the thesis process.

I would also like to thank my fellow students, parents and friends. Your belief and support have been invaluable throughout my academic journey.

TABLE OF CONTENTS

INTRODUCTION

1.1 Business problem background

Ranked lists (e.g., top ten charts) are important to many consumers. It provides the most significant source of information for a given product domain, information that can have a major impact on a brand's future prospects (Bradlow & Fader, 2001). This is also seen by Netflix, the world's leading Streaming Video On Demand (SVOD) service, with over 223 million subscribers worldwide (Stoll, 2022). Operating in a market that has witnessed a booming growth and great success in the recent decade, followed by a new generation of users that are willing to pay more to watch videos online. But also faces challenges of customers churning, due to the overabundance of content available on such SVOD platforms (Bai et al., 2020). Due to the big increase in the number of providers such as Amazon Prime Video, Hulu, Disney+ and HBO Max, Netflix' share of subscribers in the US dropped by a staggering 31% in 2022 (Lucas, 2021). This shows that, more than ever, consumers are in command and providers need to fight harder to retain them.

This reflects Netflix' strategy with the return of an opaque ranking system, which is perhaps perceived as more transparent than the already existing recommendation algorithms (Scarlata, 2022). Studies have uncovered widespread ambivalence about the credibility of these recommendation algorithms (Johnson et al., 2020). Therefore, Netflix introduced the "Netflix top 10", a feature that is updated daily, that ranks movies and TV-shows (i.e., videos) based on the popularity within a user's country. Ending up in a chart creates a hype around the titles that it lists, which creates social pressure among people to keep up with what is popular and thereby can increase popularity of the title dramatically (Scarlata, 2022). Extensive research in the motion picture industry into movie popularity has been done. Yet, all these studies are based on box office sales, and these are not disclosed by SVOD platforms (Ainslie, et al., 2005; Elberse & Eliashberg, 2003; Fetscherin, 2010). The only existing metric that is currently available to measure the popularity of a video in the streaming industry are chart lists like the Netflix top ten. Therefore, it is crucial to study factors that influence the Netflix top ten for content producers, competitors, and other stakeholders in the SVOD market.

When examining the Netflix top ten there is a close correlation between the top ten titles and content that was recently added to Netflix. About 88% of the titles that were added to the platform and that were featured in the top ten made their debut on the rank within just four days after being added (Scarlata, 2022). These findings are similar to what is found in the movie industry where box office revenues, which often are used to reflect popularity of a video, are mostly earned in the first few weeks after releasing (Ainslie et al., 2005; Krider & Weinberg, 1998; Mukherjee & Kadiyali, 2011). Hence, charts like the Netflix top ten mostly reflect the popular content that was recently released on these platforms.

Moreover, extant research in the movie industry has found that movie popularity is also strongly correlated with the amount of video releases of similar movies. Competition plays a significant role in the understanding of box-office dynamics and is crucial for videos (Fetscherin, 2010). When two movies, that are perceived as good, are released at the same time, their box office revenues tend to be negatively correlated (Yeung et al., 2011). Releasing a movie with the same genre and MPAA (Motion Picture Association of America) rating as competitors will hurt the box office sales all around (Ainslie, et al., 2005; Elberse & Eliashberg, 2003). The MPAA is a rating system that categorises videos based on their appropriate age group, including levels such as General audience (G), Parental guidance Suggested (PG), Parents Strongly Cautioned (PG-13), Restricted (R), and Adults only (NC – 17) (Motionpictures, n.d.; Shafaei et al., 2020). Although competition and video popularity among comparable videos has been extensively studied in the motion picture industry, the video streaming industry has received no attention on this matter.

People frequently maintain subscriptions with multiple streaming providers simultaneously. The average US household is subscribed to 5.2 different SVOD streaming platforms (Sangari, 2022) (e.g., about 90% of the HBO Max subscribers are also subscribed to Netflix (Stoll, 2021)). Furthermore, besides monthly subscription fee, there are no extra marginal costs when choosing a video in the SVOD industry. This makes it easier than in the motion picture industry to switch between a vast amount of videos among all the subscribed platforms (Sangari, 2022 ; Stoll, 2021). These differences with the motion picture industry could either benefit the popularity of a video when released simultaneously with another video of similar characteristics (i.e., genre or MPAA rating). Because consumers do not have to choose between movies as often is the case in the motion picture industry, but instead can watch both. On the other hand, the ease of switching between a vast amount of videos makes it effortless to switch to a similar competitor video when released. However, due to the lack of empirical analysis in the streaming industry, as also pointed out by many studies in this field (Cody, 2021; Jiang et al., 2019; Jeon et al., 2021; Ainslie et al., 2005), it is unknown if competition plays a role in the popularity of a video in the SVOD industry.

The effect of similar competing videos on video popularity may depend on seasonality. In the case of theatres, programming similarities are correlated with the market size (i.e., during the holidays there is an increased overlap in content). Mainly because during these periods there are more content releases (Chisholm et al., 2010). In the motion picture industry, the biggest movies are released at times when the demand is at its highest which is mostly during the holidays (Einav, 2007). There are intensely competitive high-revenue seasons (e.g., during the Christmas and summer season) and there are low revenue, less-competitive seasons (Krider & Weinberg, 1998). People tend to connect their viewing to the season (Johnston, 2018).

While current literature has mainly focussed on video popularity between similar and often competing videos in the movie industry, very little is known about video popularity between competing videos that are seen as good in the Streaming Video on Demand industry. Due to the gap in existing literature, it is interesting to quantify how competition, or rather the amount of similar popular competitor videos, influence the video popularity of videos on the Netflix top ten, and whether this depends on seasonality.

1.2 Problem statement

Against this background, the problem statement of this study is: "To what extent do the number of videos with certain video characteristics (genre & MPAA rating) of popular chart titles on competitor SVOD platforms influence the video popularity of a video with the same video characteristics (genre & MPAA rating) on the Netflix top ten in the United States, and to what extent does this depend on seasonality"?

To answer this problem statement, the following research questions are formulated:

Theoretical research questions:

- What is the effect of a bestseller list?
- What does the current streaming environment look like?
- How are video content releases related to video popularity?
- How does the availability of videos with certain video characteristics (genre & MPAA rating) affect the popularity of videos with similar characteristics?
- How is the relationship between the availability of videos with certain video characteristics (genre & MPAA rating) on the popularity of videos with similar characteristics affected when taking into account seasonality?

Empirical research questions:

- To what extent do the number of videos with certain video characteristics (genre & MPAA rating) of popular chart videos on competitor SVOD platforms influence the video popularity of videos on Netflix?

- To what extent is the relationship between the number of videos with certain video characteristics (genre & MPAA rating) of popular videos on competitor SVOD platforms and the video popularity of videos on Netflix affected taking into account seasonality?
- How can managers implement the findings of this study?

1.3 Contribution

1.3.1 Theoretical contribution

From an academic point of view, this research adds to the literature by looking at the impact of popular content on different SVOD platforms on video popularity. Existing literature in the streaming industry contains many theoretical analyses of content strategies, but relatively few empirical studies (Kennedy, 2002; Ainslie et al., 2005). This is mainly due to the fact that streaming platforms and direct-to-consumer models change rapidly. What makes analysis more difficult and constantly changing, and as a result, any analysis is only a snapshot of the current state of the streaming landscape (Cody, 2021). Consequently, there is a lack of direct real-world comparison between the SVOD platforms in the academic field, as also pointed out by many studies (Cody, 2021; Jiang et al., 2019; Jeon et al., 2021).

What is known from the literature is that there are strong correlations between content that is released and content that is popular, also in the streaming industry (Ainslie et al., 2005; Krider & Weinberg, 1998; Mukherjee & Kadiyali, 2011; Scarlata, 2022). Furthermore, from the movie industry it is known that competition, and especially competitor releases, play a significant role in the popularity of a video (Ainslie, et al., 2005; Elberse & Eliashberg, 2003; Yeung et al., 2011). However, it is not known in the literature if this also holds in a streaming or subscription setting. The SVOD industry distinguishes itself from the motion picture industry primarily through its subscription-based pricing model, in contrast to the per-ticket pricing employed by theatres (Kim & Kim, 2017). Additionally, while theatres offer a limited selection of options at a certain moment in time, SVOD platforms provide a wide range of content that can be accessed at any time, without any additional costs (Dogruel, 2018). Given these differences, it is important to examine the impact of competition within the SVOD industry.

This research will study the influence of competition on the video popularity of videos on the Netflix top ten to get a better understanding of the drivers of this functionality. Adding seasonality as a moderator, this study will also find out if seasonal differences strengthen or weaken this relationship.

1.3.2 Managerial contribution

Accurate prediction of the popularity of videos can benefit SVOD platforms a lot, e.g., in online advertisement, copyright procurement, recommendation strategies and so on (Bai et al., 2019). Furthermore, prior research to the bestseller lists of music platforms found that bestseller lists can increase the music revenues as well as the product consumptions. Thus, finding the drivers affecting these charts is valuable to Netflix, its competitors, and the content producers (Sim et al., 2022).

However, accurate prediction of movie popularity is very difficult, because it is influenced by a variety of dynamic factors. If one changes, the current movement trend will be completely altered (Divakaran et al., 2017). In the case of streaming service, revenues are based on subscribers, making existing models that predict movie popularity and that primary focus on box office sales, useless.

Understanding the impact of videos on competitor SVOD platforms on the success of similar videos on Netflix, as well as identifying any seasonal patterns that may be around, affect the choice of the timing of a video release. The time of release of videos on these platforms may be pushed back or brought forward to avoid coming out simultaneously with competing videos that may be strong players (Ainslie et al., 2005). This research will provide managers with insights on the possible influence of competition on video popularity on Netflix and how it differs looking at seasonality. Which is especially useful in the current market where there are many SVOD providers fighting for viewer attention.

1.4 Outline of the thesis

The remainder of this thesis is organised as follows. Chapter 2 presents the existing literature that is relevant for this thesis, and it explains the relationships between the variables followed by the research hypotheses. In chapter 3 I will describe the research methodology and the data that will be used for the analysis. This is followed by the analysis and results in chapter 4. Finally in chapter 5 the results will be discussed along with the theoretical and practical implications of this study. The limitations and suggestions in chapter 5 for future research will finalize this thesis.

THEORETICAL FRAMEWORK

2.1 Background literature

This chapter provides an overview of the existing relevant literature. A brief overview of the most relevant literature of this chapter is given in table 1. In this chapter, information about hit lists (Sim et al. 2022), the SVOD industry, video releases (Scarlata, 2022; Kennedy 2022), video competition in the movie and streaming industry (Krider et al., 1998; Cha et al., 2009; Ainslie et al., 2005; Elberse et al., 2003; Chiou, 2007; Yeung et al., 2011; Jiang et al., 2019), and seasonality (Einav, 2007; Chisholm et al., 2010) is given. Based on the prior research findings, hypotheses are formulated, and these will be explained using the available literature. The goal of this chapter is to use prior research to develop hypotheses that I will empirically test in this thesis.

Table 1: Comparison of relevant literature

	Effect of	**Effect on**
Sim et al. (2022)	Hourly-updated bestseller lists	Music discovery
Kennedy (2002)	Type of program introduction	Demand
Scarlata (2022)	Netflix top 10	Popularity
Krider et al. (1998)	Timing	Product introduction
Ainslie et al. (2005)	Movie release	Box-office sales
Cha et al. (2009)	User generated content systems	Commercial and technical implications
Elberse et al. (2003)	New product launch	Box office performance in domestic and foreign markets
Fetscherin (2010)	Indian movie sales overseas	Total box office sales
Chiou (2007)	Underlying seasonality demand or quality of movies released	Theatrical revenues
Einav (2007)	Seasonality of box-office revenues	Weekly demand for movies
Yeung et al. (2011)	Movie competition	Recommender system
Jiang et al. (2019)	Product differentiation	Profits
Rios et al. (2018)	Introduction Netflix	Local SVOD creation
C. Chisholm et al. (2010)	Product similarity	Demand
This thesis	**Video competition**	**Video popularity on Netflix**

2.1.1 Hit lists

For long, people have been under the spell of ranked lists of celebrities, places, movies and countless other entities. The psychological and economic effects of being “on the list” can be enormous, with far-reaching influences on perceptions, profits, and awareness (Bradlow & Fader, 2001).

Many studies in a variety of settings have pointed out the benefits of being placed on a list. Most studies have found that being on a list can improve the sales substantially (Aguiar & Waldfogel, 2018). Sorensen (2007) found that a book’s listing on the New York Times Bestseller Lists increases its sales by 4.3% on average. The effect is stronger for new authors. When being added to a big music Top Hits list, a list with 18.5 million followers, music streams rose by almost twenty million (Aguiar & Wildfogel, 2018). The place in a hit list does also sometimes matter. In a study where they looked at the willingness to pay for apps, the value attributable to rank one is about twice as large as the corresponding value for rank two (Carare, 2012). In the industry of streaming services, possible evidence was found that a song ranked on the top 10^{th} or higher is perceived slightly differently from other songs. They showed an increase of 2.5% in total streaming counts, but no significant rise in new streamers. It seems that in a purchase setting the position on the list does matter more than in a subscription setting (Sim et al., 2022).

Observational learning, when people draw quality inferences from direct observations of peer choices of others (Zhang, 2010), has been identified as the main reason people choose to follow these lists instead of using their own knowledge or information (Carare, 2012; Sorensen, 2007; Cai et al., 2007; Zhang & Tucker, 2011). However, recent research indicated that in the case of streaming services that have a contractual business model, the effect of bestseller lists is different. Sim et al. (2022) found that bestseller lists influence the consumer choice significantly, but the salience effect contributes to the effects of the ranking charts more significantly than observational learning. This means that people give more weight to information of stimuli that are more noticeable, prominent or attention-grabbing. It is suggested that this is due to zero marginal costs and large production catalogue for streaming services. Streamers might be insensitive to quality signals. Sim et al. (2022) also concluded that the position of a ranked item does not matter, only being on the list matters. Furthermore Godinho de Matos et al., (2016) found that artificial swaps of rank positions have a significant impact on the short-term demand of an item, but consumers rapidly adjust their choices toward the true quality of a video based on outside information.

2.1.2 SVOD industry

SVOD services have attracted exceptional attention in recent years for their novelty and growth. The emergence of global players such as Netflix, Disney, Amazon, Hulu, HBO Max and many other national and regional services (e.g., Videoland), have transformed the entertainment landscape (Lobato & Lotz, 2021). Recent business press called it a "streaming war", where platforms compete intensively for subscribers (Nash et al., 2022). The key distinction among these video services is the variation in content they offer, including their specialty genres, original productions, and the depth and breadth of their catalogues (Lobato & Lotz, 2021). The analysis conducted by Jeon et al. (2021), and Chen et al. (2022) highlights the importance of age groups (MPAA ratings) and genres. With new services entering the industry all the time, the content strategy of SVOD platforms plays a vital role in their success. The main reason for subscribing to multiple SVOD platforms in the US is to expand the content that people have available (Stoll, 2022).

2.1.3 Video popularity on Netflix

Many studies have focused on what thrives video popularity, but these are all based on box-office revenues (Ainslie, et al., 2005; Elberse & Eliashberg, 2003; Yeung et al., 2011). In the streaming industry the box-office sales of videos are not disclosed. As streaming platforms are reluctant to disclose streaming data, the lack of transparency into this data is possibly a reason why so little academic research has focused on video popularity in the SVOD industry (L Wayne, 2021). The only metric that is currently publicly available to measure the popularity of a video, are charts like the Netflix top ten. These reflect the popularity of a video at any given time on any SVOD platform. Furthermore, as previously described, charts in the SVOD industry in some way quantify the popularity of newly released videos on these platforms.

2.1.4 Video releases

An analysis of the Netflix top ten has shown that about 88% of the unique titles that were featured in the top ten made their debut within less than a week of being added to the Australian Netflix catalogue (Scarlata, 2022). Moreover, in the motion picture industry when substantial investments were made in a video production, the opening weekend is extremely critical. This is because most of the box office revenues are often earned in the first few weeks after releasing (Ainslie et al., 2005; Krider & Weinberg, 1998; Mukherjee & Kadiyali, 2011). Cha et al. (2009) found that movie popularity is mostly determined at the early stage of the video age. While the goals of the movie industry are different from those of the streaming industry, these findings indicate that for both industries video popularity is strongly correlated with the moment of release of a video.

Movie industry releases

Furthermore, research has found that movie popularity is correlated with releases of similar movies. A significant concern in the film industry is to prevent simultaneous releases of similar movies. As even though films often attract different audiences, there are always close substitutes available (Krider & Weinenberg, 1998).

According to Yeung et al, (2011) the likelihood of a movie becoming a hit is greater when its competitors (i.e., movies with similar characteristics) at the time of its release are bad movies. Good and bad movies were classified by their movie score. Additionally, Krider & Weinenberg (1998) found that strong movies (i.e., high in marketability and playability) should compete head-to-head during peak weeks, while weak movies should delay their releases when facing intense competition.

Videos experience particularly strong competition from releases of videos with similar genres and MPAA ratings (Elberse & Eliashberg, 2003). Releasing a movie against a movie with the same genre as competitors will hurt the box office performance all around. When a movie is released alongside other movies with the same MPAA rating, the box office sales will be hurt at the beginning, but the loss of sales is less severe in the long run (Ainslie et al. 2005). Genre informs consumers about the basic dramaturgic and aesthetic patterns they can expect from a certain TV-show or movie (Dogruel, 2018). Viewers often can guess the story, setting, and mood of a video by its genre(s) (Hwang et al., 2016). The MPAA is a system that establishes the age that is appropriate for video viewers (Shafaei et al., 2020).

The genre and MPAA rating of a video were also found to be strong predictors for revenues in the opening weeks in many domestic markets (Elberse & Eliashberg, 2003; Fetscherin, 2010). According to Chiou (2007), while it can hurt the sales, the introduction of a video of a given genre does not necessarily lead to a decline in the share of videos of the same genre.

Streaming video on demand releases

In the streaming industry, there is often little distinction between larger and smaller productions as videos are simply placed on the platform without much differentiation. SVOD platforms primarily operate on flat-rate subscription models, focusing on the overall value provided by the entire content library rather than individual content pieces (Kübler et al., 2020). Consequently, the emphasis is placed on the collective library rather than the release of specific videos.

While in the past, all-at-once TV show releases were very popular, today firms are adopting a more drip distribution for their content, which means releasing content one period at a time (e.g., once every week). This approach encourages users to explore other content available on the platform (Godinho de Matos et al., 2023).

In 2022, the average US household was subscribed to 5.2 different streaming platforms, which suggests a considerable overlap in the consumer base among SVOD platforms (Sangari, 2022). For instance, data shows that about 90% of the HBO Max subscribers are also subscribed to Netflix (Stoll, 2021). Besides monthly subscription fee, there are no extra marginal costs when choosing a video in the SVOD industry, which makes it very easy to switch between a vast amount of videos among all the subscribed platforms (Dogruel, 2018). Furthermore, literature states that under multi product purchase, when consumers subscribe to multiple providers that provide the same service, the profits are higher when the content is sufficiently differentiated (Jiang et al., 2019). Moreover, an empirical analysis on programming release decisions by networks found that networks with an imitative program of competitors' introduction underperforms differentiated introductions (Kennedy, 2002). This study was done by dividing the programs of broadcasting networks up in 15 “genres” (e.g., comedy, sports, news etc.). In 2019, the majority of the content offered by streaming services such as Netflix and Hulu are acquired from third-party creators. Yet, as more SVOD platforms begin to produce original content (exclusives), it appears that these platforms are attempting to differentiate themselves from one another (Jiang et al., 2019).

2.1.5 Seasonality

Releasing dates are critical in a marketing mix decision. As described before, one of the most important things in a movie’s run is “staying away” from similar strong videos. This is particularly dominant at the early part of a video run. However, videos should also try to capture as much of the peak primary demand as possible (Krider & Weinenberg, 1998). In the motion picture industry, there are intensely competitive high-revenue seasons (e.g., during the Christmas and summer season) and there are low revenue, less-competitive seasons (Krider & Weinberg, 1998; Elberse & Eliashberg, 2003). Movie distributors release their biggest hits in the summer, especially during the 4th of July, and during the Christmas holiday (Einav, 2007). During the summer most videos are released during July the 4th as opposed to Labour day, Einav (2007) found that the demand remains stable throughout the entire summer. Litman & Kohl (1989) found that the peak season is in favour of the summer.

During these highly competitive seasons (i.e., when there is an increase in new releases) the programming similarity increases, as seen in the case of theatres. During the holidays there is an increase in the overlap of content (Chisholm et al., 2010). Providers should carefully think about the moment of release of a video because of the seasonal patterns in demand (Chiou, 2008). As people tend to connect their viewing habits to a specific season. Also, producers see this when broadcasting new programmes with particular genres (Johnston, 2018). A study into the consumption of video games found that not only the valence of content increases during certain times, but also the genres that were preferred. For instance, battle-genre games are more popular during the spring to summer seasons. The abundance of vacation time during summer allows consumers to spend more leisure time together, making fighting games an appealing choice due to their intuitive nature and minimal investment in learning controls (Palomba, 2019).

2.2 Conceptual Framework and hypotheses

2.2.1 Competitor videos

According to Jiang et al. (2019) and Kennedy et al. (2002) SVOD platforms should strive for differentiation in their content releases. This could potentially lead to avoiding simultaneous releases of similar popular content as is seen as beneficial in the motion picture industry. On the other hand, because of the large amount of SVOD subscriptions in the US and the low additional costs of switching between movies, it takes very less effort to switch between competitor videos when a similar popular video appears on one of the subscribed platforms (Sangari, 2022 ; Stoll, 2021). But the low additional cost could also lead to watching both videos, instead of choosing between similar videos. However, none of the existing studies on content releases in the streaming industry do specifically account for competition among SVOD platforms when examining release strategies. The lack of real-world empirical analysis in the SVOD industry leaves this as an open question. That is why I will look at the findings in the motion picture industry.

When looking at the movie industry, it is found that competition plays a significant role in understanding the box-office dynamics. Movie popularity is correlated with the amount of video releases of competing similar movies that are perceived good (Krider & Weinenberg, 1998; Yeung et al., 2010). Specifically, when more similar videos that are perceived as good are released, their box-office performances will be negatively affected. When looking at video characteristics of similar competitor videos, then releasing a video with the same genre or MPAA rating as its competitors at the moment of release will hurt the box office revenues (Ainslie et al. 2005; Chiou, 2007; Elberse & Eliashberg, 2003).

Because of these findings it is expected that also in the streaming industry the genre and MPAA rating of competitor videos have a negative impact on the video popularity of a video with a similar genre or MPAA rating. While in the streaming industry, no box-offices are displayed, I will use the Netflix top ten to quantify the effect of similar popular competing videos on video popularity. As a result, the following hypothesis is drawn up:

H1a: *The number of popular chart videos on competitor platforms with the same genre(s) as a video on the Netflix top 10 negatively affect the popularity of the Netflix video (i.e., the more similar competitor videos the fewer weeks on the Netflix top 10).*

H1b: *The number of popular chart videos on competitor platforms with the same MPAA rating as a video on the Netflix top 10 negatively affects the popularity of the Netflix video (i.e., the more similar competitor videos the fewer weeks on the Netflix top 10).*

2.2.2 Seasonality

In the motion picture industry, movie distributors release their biggest hits, often videos that appear on chart lists, during highly competitive and revenue-generating seasons such as summer and the Christmas holiday (Einav, 2007; Krider & Weinberg, 1998). As a result, the programming similarities tend to increase during these periods (Chisholm et al., 2010).

On the other hand, during the holidays, people have more leisure time to watch movies, resulting in an increased movie consumption which can lead to a decrease in movie competition. However, seasonality also has been found to influence viewers' preferences and viewing behaviour, as individuals tend to connect their viewing habits to specific seasons (Johnston, 2018). Which leads to more of the same content that will be viewed.

Given these observations, it is expected that the popularity of videos on Netflix may be influenced by the season. During specific seasons when certain genres are popular and studios tend to release their biggest hits (i.e., videos that are likely to appear on a chart list), there is likely to be a higher concentration of popular videos with similar genres or MPAA rating on competitor chart lists. Consequently, the negative effect of competitor genre or MPAA rating videos on the popularity of similar videos on Netflix is expected to be strongest during these highly competitive seasons.

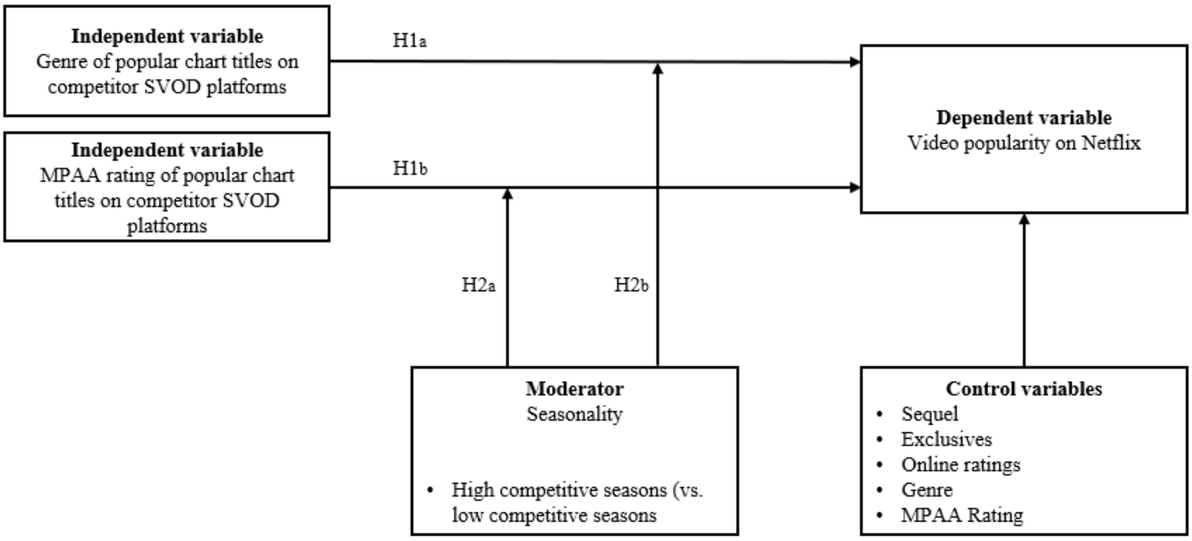
H2a: *The negative effect of the number of popular chart videos on competitor platforms with the same genre as a video on the Netflix top ten on the video popularity on Netflix is strengthened during highly competitive seasons (vs. low competitive seasons).*

H2b: *The negative effect of the number of popular chart videos on competitor platforms with the same MPAA rating as a video on the Netflix top ten on the video popularity on Netflix is strengthened during highly competitive seasons (vs. low competitive seasons).*

2.3 Conceptual framework

A schematic representation of the variables that I intend to research is given in figure 1. This study will look into the effect that competition has on video popularity on Netflix. In addition, the effect of competition on video popularity on Netflix is expected to depend on seasonality. Other variables that also possibly have an impact on the video popularity on Netflix are given as control variables, these will be further explained in chapter three.

Figure 1: Conceptual Framework



METHODOLOGY

After the problem statement of this study was given and the hypotheses were formulated with the existing literature, chapter three will include the methodology of my thesis. First, this chapter includes a small introduction of the research context, followed by the data sources and the sample. Then I will operationalize the variables and give the data transformation and cleaning. Lastly, a description of the model that will be used for analysis is given in this chapter.

3.1 Research context

First, when looking into the movie or streaming industry, I have to account for differences between countries. In a study done by Dogruel (2018) it was found that not only cross-cultural differences affect the preferences for cultural products as such, but it also affects the way that individuals conduct their movie selection. To control for the differences in video selection between countries I choose to only focus on the SVOD market in the United States, the biggest SVOD market in the world (Statista, 2022). Moreover, the motion picture market in the US has received by far the most attention by academics (Elberse & Eliashberg, 2003). In my analysis I will include the five biggest SVOD platforms in the US by market share in 2022. Netflix (21%), and for the competitor videos I will use Amazon Prime Video (19%), HBO Max (15%), Disney+ (15%) and Hulu (10%) (Stoll, 2022).

3.2 Data sources and sample

3.2.1 Top ten charts

The data collection process in general can be divided into two parts: first, to get the most popular videos for all the platforms, data on the Weekly top ten most popular movies and TV-shows of each SVOD platform was obtained via web scraping the website Flixpatrol. FlixPatrol gathers all possible streaming data about movies and TV shows (e.g., the official Netflix top 10 charts, Disney's trending titles and Amazon's Bestseller charts) (FlixPatrol, n.d.). They gather this information on a daily and weekly basis for all countries and every SVOD platform around the world. Hence, for my scraping purposes this is particularly useful. The sample covers a period between January 1st, 2022 and December 31, 2022. The weekly US top ten charts contain the following elements: SVOD platform name, rank, title, type (TV-show/movie), exclusive (yes/no) (for the Netflix data), chart week number, URL, and cumulative weeks in top ten (for the Netflix data). After scraping, the total dataset contained the top ten movies and TV shows for every week of 2022 for all the five SVOD platforms. This total dataset contains 5,062 rows, where each row means the weekly ranking of one video on one of the five platforms. For Disney+, FlixPatrol did not always show a top ten, sometimes it showed a chart list with less videos. However, this did not have an impact on the analysis. Furthermore, FlixPatrol only gathered the data for Amazon Prime from the 20th week in 2022 onwards. For the weeks prior to this week, the charts of Amazon Video were used. The Python script that was used for scraping the FlixPatrol website can be found in Appendix A1 as well as the resulting raw datafile.

3.2.2 Video characteristics

Second, because Flixpatrol offered limited video information, the data on the video characteristics were retrieved from IMDB via its API by using the Python library IMDBpy. IMDB is an online database of information related to films, television programs, celebrities, and streaming content (IMDB, n.d.). One way of getting the data from the IMDB database is by using its API (Geeksforgeeks, 2022). The name of each title that was scraped on FlixPatrol was used to collect the data via the IMDBpy API. For every title in the dataset, the genre, MPAA rating and the online rating was collected. The Python coding that was used for retrieving the above mentioned data via the IMDBpy API can be found in Appendix A1 as well as the resulting datafile.

3.3 Variable operationalization

3.3.1 Video popularity on Netflix

Video popularity on Netflix is the dependent variable in this research. To the knowledge of the researcher, this variable has not yet been used in academic research. While there are no box-office sales or any other data disclosed to measure video popularity of a video in a streaming or subscription setting, this research will focus on the “Weekly Netflix top 10” as a dependent variable. The Netflix top ten reflect the videos that are perceived as good and are popular on Netflix for a given week. If content reached the US weekly top ten then this title was among the most popular content for at least one week (Scarlata, 2022). To measure this dependent variable, the count variable “Cumulative weeks in the top 10” of Flixpatrol.com is used. This variable counts the instances that a video appeared on the Netflix top ten in the United States for every week between January 1st, 2022 and December 31, 2022. Video popularity on Netflix is a numerical variable that is given in the number of weeks a title appeared in the Netflix top ten.

3.3.2 Genre and MPAA rating of competitor videos

When comparing the titles among the SVOD platforms, the platforms often do not offer the exact same titles, and therefore I cannot compare titles directly to each other. To demonstrate, among the top 60 TV shows on Netflix, 78% of them are not available on Hulu, the other way around this is 87% (Jiang et al., 2019). Thus, in this thesis I will look at the characteristics that were found in the movie industry to affect the sales of similar competing titles, namely genre and the MPAA ratings (Ainslie et al. 2005; Chiou, 2007; Elberse & Eliashberg, 2003).

To directly incorporate the effect of competition from closer substitutes, I will construct two variables, NGenre and NMPAA. These variables reflect the number of videos with the same genre or MPAA rating that are popular at the same time as the video of consideration from the Netflix top ten. To get the videos that are popular and thus are perceived as good on these competitor platforms I will use top ten charts of Netflix' competitors (e.g., the Disney's trending titles and Amazon's Bestseller charts) which can be compared to the Netflix top ten functionality. Also in other studies, to measure the strength of the competitor environment, the number of videos with similar characteristics from the top 25 or top 10 were counted (Ainslie et al. 2005; Elberse & Eliashberg, 2003). Because a video can have many different genres, to measure the presence of similar videos, Elberse & Eliashberg (2003) in their study grouped the videos in five different genres, namely action, comedy, drama, romance and/or thriller. In this study, to accurately measure the presence of close substitutes, the genres will also be grouped. This will be further described in chapter 3.4, in the data cleaning and transformation. When a video has a negative coefficient on the variables NGenre and NMPAA, this would allow for the impact of close substitutes on video popularity, expressed in cumulative weeks on the Netflix top ten, thus reducing the attractiveness of videos of the same genre or MPAA rating that run simultaneously (Ainslie et al., 2005).

Moreover, a movie can be popular for multiple weeks, so I will take the average NGenre and NMPAA of all competing videos across all the SVOD platforms that were popular during the cumulative weeks that the video of consideration was popular, as also was done in the study of Chintagunta et al. (2010).

3.3.3 Seasonality

Seasonality on movie performance has been extensively researched. Most videos are released during the intense competitive high revenue seasons, mostly around the holiday seasons. Also during these periods, the program similarities between video providers increase (Einav, 2007; Krider & Weinberg, 1998; Chisholm et al., 2010). To quantify the effect of these seasons I will use a proximity to major holidays. I will construct the variable *Holiday_distance*, which is defined as the number of weeks t to the nearest holiday. If the current week is a holiday then this variable equals zero. The variable will increase till it reaches the midpoint between two holidays, from then it will decline until it reaches zero again at the next holiday. I will use the same holidays that were used by Einav (2007) and Chisholm et al. (2010), these are Memorial Day (week 22), Fourth of July (week 27), Thanksgiving (week 47) and Christmas (week 52).

3.3.4 Control variables

To minimise the endogeneity in my analysis I will add some control variables in my model that were found to impact video popularity. In the literature it was found that exclusives are perceived differently than content produced by traditional broadcasters, this is because platform originals are often "binged" oriented. (Laban et al., 2020). Exclusives are only available on a particular platform and have not yet been viewed by anyone else prior to their release, therefore they may affect the video popularity. Because of these findings, exclusives will be added as a control variable and will be a binary variable, with either a 1 (i.e., exclusive) or a 0 (i.e., no exclusive).

Furthermore, it is known from the literature that sequels do better than non-sequels in generating more attendance and popularity in the first week and in total (Tirtha et al., 2011; Basuroy & Chatterjee, 2007; Dhar et al., 2012). Sequels are often seen as less risky and often reduce the uncertainty as well (Eliashberg et al. 2006). Also, in the SVOD market, results from empirical studies show that sequels are more likely to be downloaded than non-sequels (Jang et al., 2021). Because of these findings, sequel will be included as a control variable. Sequel will also be a binary variable, with either a 1 (i.e., sequel) or a 0 (i.e., no sequel).

Often after a consumer has inspected the product information, whether it meets his or her personal taste, the second step is looking at the ratings, which also significantly impacts the video popularity (Dogruel, 2018; Legoux et al., 2016). For SVOD platforms, movie ratings are an effective measure for member's satisfaction and it reflects the positivity of the word-of-mouth surrounding the movie. (Legoux et al., 2016). It is mostly the valence of online reviews (user ratings) that has a significant positive effect on the box office revenues, and not the volume of reviews (Chintagunta et al., 2010). Dellarocas et al. (2007) found that the valence, and volume of the user ratings have a positive and significant effect in the opening weekend on the box office performance. Because online ratings have an impact on video popularity, and especially in the opening weekend, this variable will also be included as a control variable. Online ratings will be measured on a ratio scale.

Lastly, the genre and MPAA rating of the video of consideration will be added as control variables. This is done because it could be that videos with certain genres or MPAA ratings are simply more popular regardless of the number of competing videos with similar characteristics. This could influence both the number of competitor videos and the cumulative rank of a video on Netflix. When controlling for these genres and MPAA rating I can isolate the effect of the number of competitor genre or MPAA rating videos on the cumulative rank. Table 2 provides a summary of the variables that will be used in the analysis and how they are operationalized.

Table 2: Operationalization of the variables

Variable name	Operationalization	Data source
Cumulative_rank	Video popularity by the total cumulative weeks on the "Netflix weekly top ten" during sample period	FlixPatrol
Title	The title of the video	FlixPatrol
Platform_name	The name of the competitor SVOD platform	FlixPatrol
Type	The type of content: movie or TV show	FlixPatrol
Chart_week	The week number of the chart	FlixPatrol
Netflix_exclusive	1 = Netflix exclusive, 0 = no Netflix exclusive	FlixPatrol
AVG_Ngenre_comp	Average number of videos with the same genre as focal Netflix video on competitor platforms	IMDB API
AVG_NMPAA_comp	Average number of videos with the same MPAA rating as focal Netflix video on competitor platforms	IMDB API
Holiday_distance	The number of weeks to the nearest holiday (0 = holiday)	
Online_ratings	The average IMDb rating of a video on the Netflix top ten (1-10)	IMDB API
Sequel	1 = sequel, 0 = no sequel	IMDB API

3.4 Data transformation and cleaning

After all the data was gathered into two different datafiles, I cleaned and transformed the data. This was done in the program R. Before merging the datasets together, I did some cleaning on the data as is described below.

3.4.1 IMDBpy API data

I gathered the data via the IMDB API for every row in the FlixPatrol dataset, and therefore I had many duplicate rows. In the FlixPatrol data, a title could be on a chart list for several weeks. After removing all duplicates, I was left with 1,408 unique titles for all the platforms. There were no NAs for the rating variable. Furthermore, three titles had no genre. However, because these titles were not Netflix' chart titles I could ignore these empty cells. These videos have no other videos with a similar genre and thus were not counted. In total there were 23 genres in this dataset.

A video can have many genres, in my dataset up to 12. Simply counting the number of videos for each level of genre of a Netflix top ten title and summing all these videos could result in many similar videos, while most videos probably are not actual substitutes. Therefore, I will transform the genre variable into groups and delete the excessive subgenres, as is also done by almost every paper that uses genre as a variable. Every study uses a slightly different classification. Elberse & Eliashberg (2003) used 5 main genres to measure the presence of similar videos, namely action, comedy, drama and children. I will add a few more classifications. I will keep the original levels of the Comedy, Drama, Romance and Sci-Fi genres (Mukherjee & Kadiyali, 2018; Dhar et al., 2011; Dellarocas et al., 2007; Chiu et al., 2018; Hinz, 2010). Then, I will merge the action and adventure genres together into the action genre (Mukherjee & Kadiyali, 2018). Then the family genre will include all videos that were listed as family or animated under their IMDB classification (Chiou, 2007). Furthermore, I will combine the thriller, crime and mystery genres under the label of suspense, which is a classification that is used regularly (Chiou, 2007; Mukherjee & Kadiyali, 2018). Lastly, the other genres that remained and that are mostly sub-genres, were removed from the dataset. The titles that had no genre after removing the remaining genres got "Other" as a genre, which accounted for about 10% of the Netflix titles (Fetscherin, 2010).

The MPAA rating had 23 different levels. The levels that were not part of the official ratings systems, these levels only had a maximum of four instances assigned to them and were used with another level of the MPAA rating, were removed. The remaining levels on this variable were now divided in the programming rating system (TV-...), which are for TV-shows, and the MPAA (Motion Picture Association of America) rating system, which are for movies. The MPAA rating system is used by film studios and streaming services. Because there is great overlap between the programming rating system and the MPAA rating system, these levels were joined together into groups that more or less are the same. The following classification was used: G - General Audiences (G, TV-Y, TV-G), PG - Parental Guidance Suggested (PG, TV-Y7, TV-Y7-FV, TV-PG), PG-13 - Parents Strongly Cautioned (PG - 13, TV-14, TV-13), R - Restricted (R), NC - 17 - Adults only (NC - 17, TV-MA) and Not rated (Not Rated, Unrated, NA) (motionpictures, n.d.; tvguidelines, n.d.). The videos that had NA on this variable fall under "Not Rated" videos.

3.4.2 FlixPatrol Data

After this transformation, the IMDB data was joined by the title with the FlixPatrol dataset. Now, every chart title had the genres, MPAA ratings and online ratings attached to it. This dataset was split into a dataset that contained all Netflix' competitor's data and a dataset that contained the data of Netflix. Then, I computed the total count of videos for each week of 2022, categorised by each level genre and MPAA rating, from the competitor data.

For the Netflix data, I first transformed the exclusive variable as previously described. Furthermore, for the sequel variable I went through the information on IMDb.com. A movie got a 1 on this variable if the title is a sequel and a 0 otherwise. A TV-show got a 1 if the season that appeared on the Netflix top ten was season two or higher, and a 0 otherwise. Next, to obtain the dependent variable "*Cumulative_rank*," I filtered the dataset to retain only the rows with the highest cumulative rank. Moreover, for every title I kept all the week numbers of the weeks that the title appeared on the Netflix top ten. By doing this, I now could link the exact genre combination and MPAA rating for a Netflix title with the corresponding total number of competitors titles with the same genres or MPAA rating during each week that a title appeared on the Netflix top ten. I joined the number of competitor videos with a certain genre or MPAA rating on a given week with the Netflix data and for every title I summed all videos and divided by all weeks a title was on the Netflix top ten. A title is only a close substitute if it matches on the exact same genre combination, e.g., when a video has the genres Action & Suspense, only videos with that exact genre combination are substitutes. This resulted in the variables AVG_Nmpaa_comp and AVG_Ngene_comp. This left me with 408 titles in the Netflix dataset. Lasty, I computed the Holiday_distance variable as specified in the variable operationalization and added it to the dataset. For titles that appeared on the Netflix' top ten list for multiple weeks, I selected the first week they appeared on the list as the reference point to compute the Holiday_distance variable. I choose the first week because this is the earliest point at which the competition could potentially affect a title's cumulative rank. The resulting cleaned datafile can be found in Appendix A1.

3.5 Model

The objective of this thesis is to quantify the effect of competitor videos on video popularity in a streaming setting. In particular, I want to analyse if the popularity of a video on the Netflix top ten is affected when a similar video in terms of video characteristics on a competitor platform is also popular. Past research in the movie industry has shown that competitor videos do affect the box office revenues of similar videos. The dependent variable (DV) of this study is video popularity on Netflix which is measured in the cumulative weeks a video is on the Netflix top ten. After all data was collected, I conducted some descriptive analysis to gather some first insights on the obtained data. Afterwards, I used a statistical analysis to study the hypotheses that were formulated. The statistical analysis that I used was a multivariate regression.

A multivariate regression is a tool with which I can measure the effect that the independent variables and the moderator have on the dependent variable, or outcome variable. In the motion picture industry (multiple) regression has been widely used to predict video popularity (Dhar et al., 2011; Fetscherin, 2010; Einav, 2007). I will apply three multivariate regression models in this research. The first model contains the dependent variable '*Cumulative_rank,*' which is measured by the cumulative weeks on the Netflix top ten. Furthermore, it will contain the independent variables and their interactions: *'AVG_Ngenre_comp', 'AVG_NMPAA_comp', 'AVG_Ngenre_comp* Holdiay_distance'* and *'AVG_NMPAA_comp* Holdiay_distance'* . Then for the second model the control variables will be added, namely *'Type'*, *'Netflix_exclusive'*, *'Sequel'* and *'Online_ratings*. Lastly, in the third model I will also add the genre and MPAA rating of the focal video as additional control variables. The statistical model of this thesis can be defined as follows:

$$\begin{aligned} Cumulative\_rank_{it} = \ & \beta_0 + \beta_1 AVG\_Ngenre\_comp_{it} + \beta_2 AVG\_NMPAA\_comp_{it} \\ & + \beta_3 (AVG\_Ngenre\_comp * Holiday\_distance)_{it} \\ & + \beta_4 (AVG\_NMPAA\_comp * Holiday\_distance)_{it} \\ & + \beta_5 Netflix\_exclusive_{it} + \beta_6 Sequel_{it} + \beta_7 Online\_ratings_{it} \\ & + \beta_8 Type_{it} + \theta_{it} + \gamma_{it} + \varepsilon_{it} \end{aligned}$$

Where:

- t denotes the week number
- i denotes the movies (i = 1, 2, …., 408)
- β_0 is the intercept of *Cumulative_rank.* This is the value of *Cumulative_rank* when all coefficients are zero
- $\beta_1, \ \beta_2 \ldots., \beta_8$ are all the estimated regression coefficients of the independent variables. It measures the effect size. This is the change of the mean of the dependent variable when the independent variable increases with one.
- θ_{it} *and* γ_{it} are the levels of the control variables for the genre and MPAA rating of the focal video.
- ε_{it} is the model's error term, which is known as the residual. It represents the unexplained variation in the dependent variables, namely *Cumulative_rank.*

RESULTS

After all the cleaning and data transformation has been done as described in chapter three, this chapter will contain the results. First the descriptive statistics are given, followed by the evaluation of the regression assumptions. Then the results of the multiple regression model are given. At the end of this chapter, I conduct a robustness check on the model.

4.1 Descriptive statistics

First, to get a view on the data that has been collected, Table 3 lists some frequency statistics for all the platforms that were included in the analysis. HBO has the most titles in the dataset, which means that titles that are listed on the HBO charts probably stay on this chart the shortest. Disney+ has the least titles in the dataset, which indicates that titles that are listed on the Disney+ charts probably stay on there the longest. Furthermore, the table includes the genres that were used in the analysis and all the levels of the MPAA rating. When looking at the genre distribution, Amazon, Hulu, Netflix, and HBO broadly speaking, have quite a similar distribution. Disney+ has a slightly different distribution, with more focus on the family genre. The same holds for the MPAA rating, Disney+ has more videos for all age groups, while the other platforms have lesser videos for all age groups.

Table 3: Frequency statistics

	Disney+	Amazon	Hulu	Netflix	HBO
Total videos	***N = 141***	***N = 210***	***N = 362***	***N = 408***	***N = 435***
Action	98	80	129	140	152
Comedy	87	54	141	125	156
Drama	39	108	164	209	170
Family	88	15	33	45	62
Other	5	24	19	41	67
Romance	12	29	50	69	57
Sci-Fi	45	35	60	43	77
Suspense	26	81	156	184	158
G - General audiences	40	10	8	18	20
PG – Parental guidance suggested	80	35	53	68	73
PG13 - Parents strongly cautioned	42	100	139	140	171
R - Restricted	1	41	115	85	100
NC17 - Adults only	4	41	90	142	119
Not Rated	9	20	34	17	56

Then table 4 lists the descriptive statistics of the final dataset. The total dataset includes 408 unique titles. This table includes the number of observations (N), the mean, the standard deviation (SD), the minimum (Min), and maximum (Max) of all continuous variables.

The dependent variable *Cumulative_rank*, has a mean of 2.39. This means that on average a video title stays on the Netflix top ten for 2.39 weeks, with a standard deviation of 2.09 weeks. No title has stayed on the Netflix top ten for longer than 20 weeks.

During the weeks that a video appeared on the Netflix top ten, on average a total of 3.06 videos with similar genres appeared on a chart list of all four competitor streaming platforms. The standard deviation for this variable is 3.02 videos. Every video has a minimum of 0 videos with a similar genre on chart lists of competitors and a maximum of 20. Furthermore, during all weeks that a video appeared on the Netflix top ten, on average a total of 25.21 videos with a similar MPAA rating appeared on a chart list of all four competitor platforms. For this variable, the standard deviation is 13.34 videos.

Holiday_distance means the number of weeks that a title for the first time appeared on the Netflix top ten from the nearest holiday in the United States. The average distance to a holiday for all titles in the dataset is 4.52 weeks, with a standard deviation of 3.16.

The *type* variable is a dummy coded variable with the levels 1, referring to movies and 0, referring to TV-shows. The majority of videos in this dataset are movies, about 63% (i.e., N = 257). The other 37% are TV-shows (i.e., N = 151). In Appendix B, figure B1, the cumulative rank is broken down into the two video types. Because of a few extreme values on cumulative rank, it is difficult to observe, however TV-shows have a slightly higher mean cumulative rank (2.96) compared to movies (2.05).

Sequel and exclusives are both dummy coded variables. About 23% of the titles in the dataset are sequels. TV-shows account for the majority of sequels, about 63% (i.e., N = 58). Furthermore, the majority of videos are Netflix exclusives, about 55% (i.e., N = 224). Lastly, the variable rating gives the average IMDB rating for each title. In general, the average rating of a title on the Netflix top ten is 6.53, with a standard deviation of 1.1.

Table 4: Descriptive statistics

Variable	N	Mean	SD	Min	Max
Cumulative_rank	408	2.39	2.09	1	20
AVG_Ngenre_comp	408	3.06	3.02	0	12

AVG_Nmpaa_comp	408	25.21	13.34	3	114
Holiday_distance	408	4.52	3.16	0	11
Type	408	0.63	0.48	0	1
Sequel	408	0.23	0.42	0	1
Exclusive	408	0.55	0.5	0	1
Rating	408	6.53	1.1	1.9	8.9

4.2 Assumptions

Statistical tests often rely on assumptions about the variables used in the analysis. When these assumptions are not met then the results may not be fully trustworthy, meaning an over- or under-estimation of the significance or effect size(s). Below four assumptions of the linear regression model will be tested (Osborne & Waters, 2002).

4.2.1 Homoscedasticity

The first assumption that has to be met is the homoscedasticity assumption. Meaning that the residuals should have a constant error across all the levels of the independent variable. The best way to detect potential heteroscedasticity are residual plots (Osborne & Waters, 2022). A residual plot is added in Appendix B, Figure B2. The plot shows that the homoscedasticity assumption is not met. The error term has an unequal variance, because the data points are distributed like a cone shape. This indicates that heteroscedasticity is probably present. To solve the problem of heteroscedasticity, the dependent variable can be log transformed. The residual plot with the log transformed dependent variable can be found in Appendix B, Figure B3. Now, it looks like the homoscedasticity assumption is met. Furthermore, the assumption of linearity is met, because the variance of the error term is now more evenly spread out and the red line is almost horizontal at zero. These assumptions were met for all models.

4.2.3 Normality

The error terms should be normally distributed. The best way to check this assumption is by looking at the normal Q-Q plot, which quantiles the residual distribution against the, in theory, ideal quantiles. In the plot (Appendix B, Figure B4) it looks like that the errors are not entirely normal distributed (Das & Imon, 2016). Also, a skewness test was conducted and from this test it could be concluded that there is no substantial non-normality, because the value (0.51) did not exceed the cut-off of 1 or -1. The data has a moderate negatively skewed distribution, which means that in the normal distribution it has a longer left tail.

4.2.3 Multicollinearity

The third assumption is to check if multicollinearity is present. Multicollinearity means that two or more independent variables are highly intercorrelated. Multicollinearity makes the coefficients unreliable, which decreases the precision (Alin, 2010). The measure that will be used to check this assumption is the Variance Inflation Factor or VIF. VIF measures the increase in variance of a coefficient relative to the variance if there was no multicollinearity. VIF scores that exceed the threshold of 10 are considered as high (Alin, 2010). In Appendix C, table C1, two VIF values exceed this threshold, which means that there is probably multicollinearity present in the data. To find out which variable is causing the multicollinearity, I also created a correlation table (Appendix C, table C2). Often a cut-off of 0.8 marks high correlation between variables. No value in the table exceeds this threshold. However, there is a level of the control variable MPAA rating that correlates relatively high with the *"AVG_Nmpaa_comp"* variable. Because this is just one level of a control variable, this level will be omitted from the analysis. After removing this level, all the VIF-values did not exceed the threshold. Thus, there is no multicollinearity present in the data anymore.

4.2.4 Influential observations

Lastly, I want to check for influential observations. These are observations that contaminate the analysis. Either by exerting too much influence on the fitted regression model or by violating the assumptions. Often this is an observation with extreme outcome values, relative to other data. However, these are not necessarily always influential, and omitting this variable often may hardly change the fitted equation at all (Draper & John, 1981; Chatterjee & Hadi; 1986). A data point is influential when it also has high leverage, meaning that it is also extreme on the Y-axis (Chatterjee & Hadi, 1986). To obtain the high influential points, I looked at the outliers, the high leverage points and I used the Cook's D. The Cook's D measure works with the logic by comparing models estimated from the whole sample to models that are estimated from samples excluding individual observations (Draper & John, 1981). From the plot in Appendix B, Figure B6 it can be concluded that there are two points that are seen as both outliers and high leverage points. As can be seen in the Cook's D plot in Appendix B, Figure B5, there are two highly influential observations, one of these points did not occur in the plot in Figure B6. After some analysis, which will be further explained later in the results, it was concluded to remove three observations from the dataset. When these observations were removed, the R-squared from the cleaned model was larger as were all of the slopes, which means that these observations exerted great impact on the fitted equation. When removing these observations, I was left with 405 rows in the dataset.

4.3 Results

In Table 5 the results for the multiple regression for models 1, 2 and 3 are given. Model 1 contains only the independent variables to test if these significantly determine a proportion of explained variance in the dependent variable. Model 2 also includes the control variables that were described in chapter 3. Model 3 includes some additional control variables, namely the levels of the genre and MPAA rating of the video of consideration on the Netflix top 10. This research will use a significance level of 0.05 to accept or reject my hypotheses, which indicates a 5% risk of concluding that a difference exists when actually there is no difference.

Table 5: Regression outcome

Dependent variable: Log(Cumulative rank)	Model 1	Model 2	Model 3
Intercept	0.577***	-0.073	-0.238
Amount competitor MPAA rating videos	0.009*	0.009*	0.008*
Amount competitor genre videos	-0.058***	-0.060***	-0.040*
Amount competitor MPAA videos * Holiday Distance	-0.002*	-0.001*	-0.002*
Amount competitor genre videos * Holiday Distance	0.006 ·	0.006 ·	0.005 ·
Rating		0.100***	0.106***
Netflix Exclusive		0.209***	0.250***
Movie		-0.181*	-0.254***
Sequel		0.123 ·	0.092
Drama			0.040
Comedy			0.072
Suspense			0.081
Family			0.128
Action			-0.020
Sci.Fi			0.179 ·
Romance			0.159 ·
Other			-0.085
R - Restricted for children			0.096
G - General Audiences			0.142
PG - Parents guidance suggested			0.013
NC17 - Adults only			-0.044

No MPAA rating			-0.233
N	405	405	405
R^2	0.035	0.172	0.215
Adjusted R^2	0.023	0.153	0.170
F statistic	2.937*	9.113***	4.765***

Note: · p <0.1, *p<0.05, **p<0.01, ***p<0.001. The dependent variable in all the models is the cumulative weeks spent in the Netflix top ten. The dependent variable is log-transformed.

When looking at the F statistic of the three models, it can be concluded that all the models are statistically significant. Which means that all models explain a significant proportion of the variance in the dependent variable. The R squared, that determines the proportion of variance explained by the independent variables, for model 1 is 0.035. Thus, 3.5% of all the variation in the cumulative rank is explained by only the independent variables of this study. Furthermore, the R squared increased for both model 2 and 3 when adding the additional control variables. Almost all of the coefficients have the same level of significance for the three models. While the R squared increases going from model 2 to model 3, the increase in R-squared is not statistically significant ($F(13, 382) = 1.6243$, $p = 0.07602$). Thus, the added levels of the genre and MPAA rating of the video did not significantly improve the model's ability to explain the variance in the dependent variable. Because of the insignificant increase in R squared and almost no change in significance for any of the coefficients going from model 2 to model 3, model 2 will be used for the interpretation of the coefficients.

Figures 2 and 3 show the interactions between the independent variables with holiday distance on the cumulative rank. Holiday distance is set to three different levels. Mainly one SD below the mean, 1 SD within the mean, and one SD above the mean. In model 2, the interaction between the amount of competitor MPAA videos and holiday distance is statistically significant at a 0.05 significance level. Furthermore, the direct effect of the amount of competitor MPAA videos is significant, with a positive effect size of 0.009. Hence, there is moderation present in the model. More specifically, the amount of competitor MPAA videos on the cumulative rank depends on holiday distance. The interaction term estimate has a negative coefficient of -0.001. Which means that for every unit increase in the holiday distance, the positive effect of the amount of competitor MPAA rating videos on the cumulative rank decreases with 0.1%, this is a relatively small effect. This can also be seen in figure 2. When we go 1 standard deviation below the mean, which means that we move closer to a holiday, the cumulative rank of a video goes up faster relative to the other levels. The coefficient of the direct effect of the competitor MPAA rating videos on the cumulative rank is positive, what is the opposite to what was found in the literature (Ainslie et al. 2005; Chiou, 2007; Elberse & Eliashberg, 2003; Fetscherin, 2010). As a consequence, hypothesis 1b, indicating that the amount of competitor MPAA rating videos negatively affects the cumulative rank of a video on Netflix, is not supported by the data. Furthermore, hypothesis 2b, namely that the negative effect of competitor MPAA rating videos on the cumulative rank would be at its strongest during the holidays, is not supported by the data. The effect is strongest during the holidays, but on the positive side.

The estimate of the interaction between the amount of competitor genre videos and holiday distance is only significant at a 0.1 significance level with a negative effect size of 0.006. Which means that for every week increase in the holiday distance, the negative effect of the amount of competitor genre videos on the cumulative rank of a video on the Netflix top 10 increases by 0.6%. The direct effect for the genre variable is negative. So, with a positive interaction the direct effect becomes less negative.

This effect can also be seen in figure 3. For 1 standard deviation below the mean, meaning that when we move closer to a holiday, the cumulative rank of a video goes down faster relative to the other levels. This is supported by what was found in the literature, that the effect of competition is strengthened during intense competitive seasons like holidays (Einav, 2007; Krider & Weinberg, 1998; Chisholm et al., 2010). However, because the moderation

Figure 2: Interaction effect MPAA rating

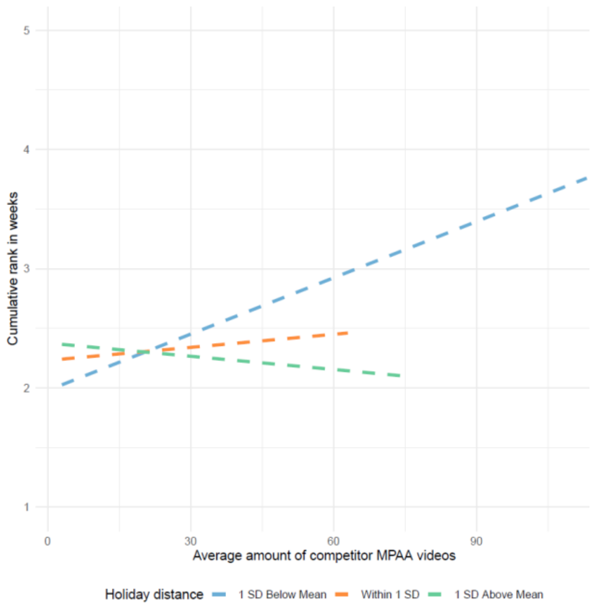
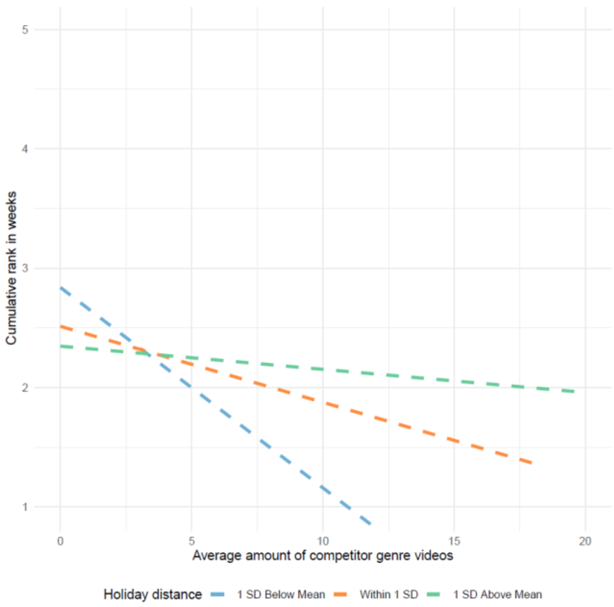


Figure 3: Interaction effect genre



effect is only marginally significant, hypothesis 2a, indicating that the negative effect of competitor genre videos on the cumulative rank would be at its strongest during the holidays, is not supported by the data. Therefore, I continue with interpreting the direct effect.

The direct effect of the amount of competitor genre videos on the cumulative rank is significant, with a negative effect size of -0.060. This means that the cumulative weeks of a video on the Netflix top ten decrease by 5.82% when one video with similar genres appears on a chart list of a competitor, keeping all other variables constant. This is quite a substantial effect. This finding is in line with what was found in chapter 2. Namely, when good movies are released simultaneously that have the same genre, the popularity will be harmed (Ainslie et al., 2005; Chiou, 2007; Elberse & Eliashberg, 2003). This means that hypothesis 1a, indicating that the amount of competitor genre videos would negatively affect the cumulative rank of a video on Netflix, is supported by the data.

Control variables

As for the control variables, only the coefficient for the variable sequel is not significant. This means that I found no effect for the sequel on the cumulative rank of a video on the Netflix top ten. This is contrary to the existing literature that states that sequels would do better than non sequels (Tirtha et al., 2011; Basuroy & Chatterjee, 2007; Dhar et al., 2012). The effect of online ratings on the cumulative rank is statistically significant with an effect size of 0.100. This means that a one unit increase in the IMDB rating leads to an increase in cumulative rank of a video of 9,52%, keeping all other variables constant. This result is in line with what was found by Dogruel (2018) and Legoux et al. (2016), that online rating positively affects the popularity of a video. Furthermore, the variable Exclusive is also significant with an effect size of 0.209. Which means that the cumulative rank of a video on the Netflix top ten on average increases with 23.44% when a title is a Netflix exclusive, keeping all other variables constant. This is also in line with what was found in the study done by Laban et al., 2020. Lastly, the variable movie, which describes whether a video is a movie or a TV-show, is also statistically significant with an effect size of -0.181. Which means that the cumulative rank of movies is 16.56% lower than for TV-shows.

4.4 Robustness check

Robustness check: influential observations

The analysis that was conducted as previously described was done with 405 observations, out of the 408 total observations that were in the dataset. The three observations that were deleted from the analysis were marked as highly influential points. In the Cook's Distance test, when using the cut-off 4/sample_size, I found 25 high influential observations. This robustness check will compare the analysis with three different models. The entire model with all 408 observations, the model used for analysis with 405 observations and the model that excludes the 25 highly influential observations.

To find the highly influential observations, I first plotted an outlier plot using the studentized residuals (Appendix B, Figure B8). Only one value exceeded the threshold of 3 with great distance (observation 269). Furthermore, I created a plot that gives the observations that are seen as both outliers and high leverage points (Appendix B, Figure B6). An observation is seen as influential if it is both an outlier and it has high leverage. In the plot, there are two observations that have both high leverage and are seen as an outlier (observations 72 & 285).

Then the Cook's Distance was calculated, it found 25 observations that were possibly influential (Appendix B, Figure B7). However, I decided to only remove three observations from the dataset. The points that had, with great distance, the highest Cook's D (observations 72 & 269; Appendix B, Figure B5) and the additional point that was also seen as an outlier and a highly influential point (observations 285)

For model 1 in Appendix C, Table C3, a table is given which includes three models. The first model is the model that excludes the 25 observations that were seen as influential by the Cook's Distance. The second model is the model that I used in my analysis that excludes three influential observations. The third model includes all the data points in the data set. The R squared of the first and the second model are both higher than the model that includes all the data points, as well as almost all of the coefficients and the F statistics. When removing only the three most influential observations, the R squared of the model increases with 0.019. When removing the 22 other influential observations, the R squared only increases further with 0.017, which is a relatively small increase compared to the second model. The values of the coefficients of the first model with the least observations are in general slightly higher. Furthermore, the interaction term between the amount of genre videos and the holiday distance is significant in the first model. However, when I would remove all the 25 highly influential observations I would omit more than 6% of the total data points. Because of this fact and the relatively small increase of the R squared I decided to use the model that only omitted the 3 most influential observations from the dataset.

Table 6 summarises all hypotheses with their result and conclusion.

Table 6: Results hypotheses

Hypothesis	Result	Conclusion
H1a: : The number of popular chart videos on competitor platforms with the same genre(s) as a video on the Netflix top 10 negatively affect the popularity of the Netflix video	Negative significant effect	Supported
H1b: The number of popular chart videos on competitor platforms with the same MPAA rating as a video on the Netflix top 10 negatively affects the popularity of the Netflix video	Positive significant effect	Not supported
H2a: The negative effect of the number of popular chart videos on competitor platforms with the same genre as a video on the Netflix top ten on the video popularity on Netflix is strengthened during highly competitive seasons (vs. low competitive seasons).	Negative marginally significant effect	Not supported
H2b: The negative effect of the number of popular chart videos on competitor platforms with the same MPAA rating as a video on the Netflix top ten on the video popularity on Netflix is strengthened during highly competitive seasons (vs. low competitive seasons).	Positive significant effect	Not supported

DISCUSSION

In this final chapter of my thesis, I will first discuss the results from the analysis that was done in chapter 4. Then I will give the theoretical and managerial implications of these results. Finally, the limitations and suggestions for further research will conclude this chapter.

5.1 Discussion of the results

In this study I investigated the impact of competition between videos with similar characteristics (i.e., genre and MPAA rating) on the video popularity on Netflix. The popularity was measured by the cumulative weeks that a video ranked in the Netflix top 10. Specifically, I first examined the effect of competitor videos with similar genres on the cumulative rank. Existing literature in the movie industry suggested that releasing a video against a video with similar genre(s) would lead to a decrease in popularity all around (Ainslie et al. 2005; Chiou, 2007; Elberse & Eliashberg, 2003). The results of my analysis revealed that the same effect exists in the SVOD industry, indicating that releasing a video that is perceived as good or is popular on one platform can negatively affect the cumulative rank of a video with a similar genre on another platform. An explanation for this effect can be that, when more videos with a similar genre appear on competitor chart lists, people that are interested in that genre divide their attention among more options on the SVOD platforms that they are subscribed to, which can decrease the popularity of a video with that genre on Netflix. Moreover, because currently in the United States, the average US household is subscribed to 5.2 different streaming platforms, switching between platforms to see other (competitor) videos takes little effort and adds no additional costs (Sangari, 2022).

Next, it was expected that the same would hold for the MPAA rating. Although prior research did suggest that the effect for the MPAA rating would be less severe in the long run (Ainslie et al., 2005). My analysis showed that competitor videos with the same MPAA rating have a relatively small positive effect on the cumulative rank of a similar popular video. This finding is unexpected. The observed difference in the effects of competitor genre videos and competitor MPAA rating videos on cumulative rank is difficult to explain and suggests that there are other factors at play that could be further investigated.

In addition to examining the impact of competitor videos on video popularity, I also accounted for the effect of seasonality on this relationship. More specifically, I looked at whether the effect of competition on video popularity would be stronger during highly competitive seasons, such as the holidays. The biggest hits are typically released during these periods, which increases the competition between strong videos. For the MPAA rating, the results indicated that the effect was strongest during the holiday season and weakened as we moved further away from the holiday period. However, the direct effect of the MPAA rating was positive, which is contrary to what was initially hypothesised. On the other hand, for the genre variable, the effect of seasonality was negative. This suggests that the negative impact of the number of videos with a similar genre on the cumulative rank of a video with the same genre is stronger during holiday seasons than during non-holiday seasons. Which means that the further we move away from a holiday the weaker the negative effect becomes. This finding is in line with previous literature (Einav, 2007; Krider & Weinberg, 1998; Chisholm et al., 2010). However, because this effect is only marginally significant.

5.2 Theoretical contribution

While the effect of competition on video popularity has been heavily researched in the movie industry, there is a lack of research in the SVOD industry, which is notably different from the traditional movie industry (Ainslie et al., 2005; Krider & Weinberg, 1998; Mukherjee & Kadiyali, 2011). In my study, I found statistical evidence that competitor videos can negatively influence the video popularity of a similar video in terms of genre. This is in line with the existing literature from the movie industry. So, this finding adds to the literature that the effect of competition between videos that was described before also holds in the streaming industry and it adds to understanding the drivers of the Netflix top ten. Moreover, I found that seasonality plays a role in this relationship in the SVOD industry. However, this effect was only marginally significant, which implies that further investigation is necessary.

Furthermore, I found contradictory evidence that competitor videos with the same MPAA rating positively influence the popularity of a similar video. Due to the limited research on the MPAA rating in the SVOD industry, it is difficult to explain the discrepancy between the literature and my findings. This discrepancy could be an interesting topic for future research.

5.3 Managerial relevance

The findings of this research have important managerial implications for content creators and providers (e.g., SVOD platforms) in the SVOD industry. As streaming platforms are reluctant to disclose streaming data, the insights gained from this study can be particularly valuable (L Wayne, 2021). The negative effect of releasing a video together with a similar video in terms of genre on the video popularity suggests that movie studios and streaming platforms should carefully consider the timing of their video releases. Specifically, they may choose to delay or push forward the release of their videos to avoid simultaneous releases of similar competing popular videos, which may negatively affect the popularity of their content. Furthermore, the findings suggest that releasing videos with similar MPAA ratings does not negatively impact the cumulative rank of a popular video and may even have a slightly positive effect. Finally, it is better to not release similar videos in terms of genre(s) during highly competitive seasons (e.g., holidays), because around this time the effect is at its strongest. This can guide content creators and providers to make informed decisions about their release schedule and avoid releasing similar videos during highly competitive seasons, where the negative effect of competition is stronger.

5.4 Limitations and suggestions for future research

This study has several limitations that should be taken into account when interpreting its results. First, the data used in this study only covered the United States, and it is uncertain whether these findings are generalizable to other countries and cultures. To address this limitation, future research could extend this study to other countries and cultures to see if similar patterns hold.

Another limitation of this thesis is related to the lack of research in the SVOD industry. In my study, I used the five biggest competitor platforms of Netflix and all top 10 chart movies and TV shows to gather competitor videos for each Netflix top 10 title. However, due to the lack of research into the SVOD industry it is very difficult to determine what constitutes a direct video competitor in this industry. Elberse & Eliashberg 2003, to measure the presence of similar popular movies, took movies that appeared on the U.S. box office top 25 and were released for every week in 1999. To see how movies in theatres compete with each other, Chisholm et al. (2010) used the average number of theatres that showed the movie that had the same genre as the focal movie for 13 theatres around Boston in the period prior to the movie release. Unfortunately, there has been no academic research yet in the SVOD industry into competition between similar videos. To address this limitation, future research could investigate how videos compete with each other in the SVOD industry and use these findings to validate the results of this study.

Another limitation is that this study used the cumulative rank on the Netflix top 10 as the dependent variable, and it is unclear if these findings would hold for the cumulative rank of a video on another platform. While other platforms do not disclose the cumulative rank of a video, this information can be gathered from the data itself. Therefore, future research could expand this study by examining whether these results also hold for the cumulative rank on other platforms.

REFERENCES

Aguiar, L., Waldfogel, J. (2018). Platforms, promotion, and product discovery : evidence from spotify playlists. *The Journal of Industrial Economics*. 69(3). 653-691. https://doi.org/10.1111/joie.12263

Ainslie, A., Drèze Xavier, & Zufryden, F. (2005). Modeling movie life cycles and market share. *Marketing Science*, 24(3), 508–517.

Alin, A. (2010). Multicollinearity. Wiley Interdisciplinary Reviews: *Computational Statistics,* 2(3), 370–374. https://doi.org/10.1002/wics.84

Bai, W., Zhang, Y., Huang, W., Zhou, Y., Wu, D., Liu, G., & Xiao, L. (2020). Deepfusion: predicting movie popularity via cross-platform feature fusion. *Multimedia Tools and Applications : An International Journal*, 79(27-28), 19289–19306. https://doi.org/10.1007/s11042-020-08730-y

Basuroy, S., & Chatterjee, S. (2008). Fast and frequent: investigating box office revenues of motion picture sequels. *Journal of Business Research*, 61(7), 798–803. https://doi.org/10.1016/j.jbusres.2007.07.030

Bradlow, E. T., & Fader, P. S. (2001). A bayesian lifetime model for the "hot 100" billboard songs. *Journal of the American Statistical Association, 96*(454), 368–368.

Business Strategy Hub (2020, October 10). *Netflix Business Model (2023) | How does Netflix make money.* Retrieved from https://bstrategyhub.com/netflix-business-model-how-does-netflix-make-money/

Cai, H., Chen, Y., & Fang, H. (2009). Observational learning: evidence from a randomized natural field experiment. *The American Economic Review,* 99(3), 864–882.

Carare, O. (2012). The impact of bestseller rank on demand: evidence from the app market. *International Economic Review,* 53(3), 717–742.

Chatterjee, S., & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science,* 1(3), 379–393.

Chen, Q., Kotadia, M., Huang, Y., Kim, H. Saleem, A. (2022, June 13) . *Streaming Platforms and Movies Analysis*. Retrieved from Medium; https://medium.com/@AshBlogs/streaming-platforms-and-movies-analysis-a02f7de506b1

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5), 944–957. https://doi.org/10.1287/mksc.1100.0572

Chiou, L. (2008). The timing of movie releases: evidence from the home video industry. *International Journal of Industrial Organization,* 26(5), 1059–1073. https://doi.org/10.1016/j.ijindorg.2007.11.005

Chisholm, D. C., McMillan, M. S., & Norman, G. (2010). Product differentiation and film-programming choice: do first-run movie theatres show the same films? *Journal of Cultural Economics,* 34(2), 131–145.

Cody, H. (April, 2021). Disney, Netflix, and Amazon Oh My! An Analysis of Streaming Brand Competition and the Impact on the Future of Consumer Entertainment. - *Findings in Sport, Hospitality, Entertainment, and Event Management*. Retrieved from ;https://www.researchgate.net/publication/350922785_Disney_Netflix_and_Amazon_Oh_My_An_Analysis_of_Streaming_Brand_Competition_and_the_Impact_on_the_Future_of_Consumer_Entertainment_-_Findings_in_Sport_Hospitality_Entertainment_and_Event_Management

Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: the case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23–45. https://doi.org/10.1002/dir.20087

Dhar, T., Sun, G., & Weinberg, C. B. (2012). The long-term box office performance of sequel movies. Marketing Letters : *A Journal of Research in Marketing*, 23(1), 13–29. https://doi.org/10.1007/s11002-011-9146-1

Divakaran, P. K. P., Palmer, A., Søndergaard, H. A., & Matkovskyy, R. (2017). Pre-launch prediction of market performance for short lifecycle products using online community data. *Journal of Interactive Marketing,* 38(1), 12–28. https://doi.org/10.1016/j.intmar.2016.10.004

Dogruel, L. (2018). Cross-cultural differences in movie selection. decision-making of german, u.s., and singaporean media users for video-on-demand movies. *Journal of International Consumer Marketing*, 30(2), 115–127. https://doi.org/10.1080/08961530.2017.1384710

Draper, N. R., & John, J. A. (1981). Influential observations and outliers in regression. *Technometrics,* 23(1), 21–26.

Einav, L. (2007). Seasonality in the u.s. motion picture industry. *The Rand Journal of Economics,* 38(1), 127–145.

Elberse, A., & Eliashberg, J. (2003*).* Demand and supply dynamics for sequentially released products in international markets: the case of motion pictures. *Marketing Science*, 22(3), 329–354. https://doi.org/10.1287/mksc.22.3.329.17740

Eliashberg, J., Elberse, A., & Leenders, M. A. A. M. (2006). The motion picture industry: critical issues in practice, current research, and new research directions. *Marketing Science,* 25(6), 638–661. https://doi.org/10.1287/mksc.1050.0177

Fetscherin, M. (2010). The main determinants of bollywood movie box office sales. *Journal of Global Marketing, 23*(5), 461–476. https://doi.org/10.1080/08911762.2010.521117

Geeksforgeeks (2022, August 24). *Python IMDbPY- Searching a movie.* Retrieved from Geeksforgeeks; https://www.geeksforgeeks.org/python-imdbpy-searching-a-movie/

Godinho de Matos, M., Ferreira, P., Smith, M. D., & Telang, R. (2016). Culling the herd: using real-world randomized experiments to measure social bias with known costly goods. *Management Science,* 62(9), 2563–2580. https://doi.org/10.1287/mnsc.2015.2258

Godinho de Matos, M., Mamadehussene, S. & Ferreira, P. (2023), When Less is More: Content Strategies for Subscription Video on Demand. http://dx.doi.org/10.2139/ssrn.4352446

Hwang, T.-G., Park, C.-S., Hong, J.-H., & Kim, S. K. (2016). An algorithm for movie classification and recommendation using genre correlation. Multimedia Tools and Applications : *An International Journal*, 75(20), 12843–12858. https://doi.org/10.1007/s11042-016-3526-8

IMDb (n.d.). IMDb Datasets. Retrieved from https://www.imdb.com/interfaces/.

Jang, M., Baek, H., & Kim, S. (2021). Movie characteristics as determinants of download-to-own performance in the korean video-on-demand market. Telecommunications Policy, 45(7). https://doi.org/10.1016/j.telpol.2021.102140

Jeon, C., Soo Ha, J., Ju Rick Kim, Y. (2021, May 20). *Data Analysis on OTT Platforms: Which Service Should I Choose.* Retrieved from Towardsdatascience; https://towardsdatascience.com/data-analysis-on-ott-platforms-which-service-should-i-choose-8eed953ff7d2

Jiang, B., Tian, L., & Zhou, B. (2019). Competition of content acquisition and distribution under consumer multipurchase. *Journal of Marketing Research,* 56(6), 1066–1084. https://doi.org/10.1177/0022243719870803

Johnson, C., Dempsey, L., Hills, M. (2020, January 28). *Routes to content: how people decide what TV to watch.* Retrieved from University of Huddersfield; https://research.hud.ac.uk/media/assets/document/schools/artdesignandarchitecture/research/RoutestoContentReportFinal28Jan20.pdf

Johnston, D. (2018). Exploring Television Seasonality. *Critical Studies in Television Conference State of Play: Television Scholarship in 'TVIV.'*

Kennedy, R. E. (2002). Strategy fads and competitive convergence: an empirical test for herd behavior in prime-time television programming. *The Journal of Industrial Economics,* 50(1), 57–84.

Kim, E., & Kim, S. (2017). Online movie success in sequential markets: determinants of video-on-demand film success in korea. *Telematics and Informatics,* 34(7), 987–995. https://doi.org/10.1016/j.tele.2017.04.009

Krider, R. E., & Weinberg, C. B. (1998). Competitive dynamics and the introduction of new products: the motion picture timing game. *Journal of Marketing Research,* 35(1), 1–15.

Kübler, R., Seifert, R., & Kandziora, M. (2020). *Content valuation strategies for digital subscription platforms. Journal of Cultural Economics* : Published in Cooperation with the Association for Cultural Economics International, 45(2), 295–326. https://doi.org/10.1007/s10824-020-09391-3

Laban, G., Zeidler, C., & Brussee, E. (2020). Binge-watching (Netflix) product placement: A content analysis on different product placements in Netflix originals vs. non-Netflix originals, and drama vs. comedy shows. https://doi.org/10.33767/osf.io/hxjgf

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

Litman, R. & Kohl, S. (1989) Predicting financial success of motion pictures: The '80s experience, Journal of Media Economics, 2:2, 35-50, DOI: 10.1080/08997768909358184

Lucas, S. (2021, April 19. *Netflix Is Losing Market Share. But Is It Losing Customers?*. Retrieved from Bloomberg; https://www.bloomberg.com/news/newsletters/2021-04-18/netflix-is-losing-market-share-but-is-it-losing-customers

Moon, S., Bergey, P. K., & Iacobucci, D. (2010). Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of Marketing*, *74*(1), 108–121. https://doi.org/10.1509/jmkg.74.1.108

Motionpictures, n.d. Film Ratings. Retrieved from; https://www.motionpictures.org/film-ratings/

Mukherjee, A., & Kadiyali, V. (2011). Modeling multichannel home video demand in the u.s. motion picture industry. *Journal of Marketing Research*, 48(6), 985–995. https://doi.org/10.1509/jmr.07.0359

Mukherjee, A., & Kadiyali, V. (2018). The competitive dynamics of new dvd releases. *Management Science*, 64(8), 3536–3553. https://doi.org/10.1287/mnsc.2017.2795

Nash, K., Trott, V., Allen, W., & Pajkovic, N. (2022). Algorithms and taste-making: exposing the netflix recommender system's operational logics. *Convergence*, 28(1), 214–235. https://doi.org/10.1177/13548565211014464

Osborne, J & Waters, E. (2002) "Four assumptions of multiple regression that researchers should always test," *Practical Assessment, Research, and Evaluation:* 8(2). DOI: https://doi.org/10.7275/r222-hv23

Palomba, A. (2019). Digital seasons: how time of the year may shift video game play habits. Entertainment Computing, *Entertainment Computing*, 30. https://doi.org/10.1016/j.entcom.2019.100296

Porter, R. (2019, July 8). *'Stranger Things 3' Racks Up Record Viewership, According to Netflix.* Retrieved from Hollywoodreporter; https://www.hollywoodreporter.com/tv/tv-news/stranger-things-racks-up-record-viewership-netflix-1223028/

R., Lobato & A., Lotz (2021). Beyond Streaming Wars : Rethinking Competition in Video Services. *Media Industries*, 8(1), pp. 89-108. https://doi.org/10.3998/mij.1338

Rios, S., & Scarlata, A. (2018). Locating svod in australia and mexico: stan and blim contend with netflix. *Critical Studies in Television*, 13(4), 475–490. https://doi.org/10.1177/1749602018798158

Sangari, N. (2022, October 20). *Percentage of US streaming households drops while cable TV declines slow*. Retrieved from Kantar; https://www.kantar.com/north-america/inspiration/technology/percentage-of-us-streaming-households-drops-in-q3-while-cable-tv-declines-slowed

Scarlata, A. (2022). 'what are people watching in your area?': interrogating the role and reliability of the netflix top 10 feature. *Critical Studies in Television, (20220915)*. https://doi.org/10.1177/17496020221127183

Shafaei, M., Samghabadi, N., Kar, S. & Solorio, T. (2020). Age Suitability Rating: Predicting the MPAA Rating Based on Movie Dialogues. *Proceedings of the 12th Conference on Language Resources and Evaluation* (LREC 2020), 1327–1335

Sim, J., Park, J. G., Cho, D., Smith, M. D., & Jung, J. (2022). Bestseller lists and product discovery in the subscription-based market: evidence from music streaming. *Journal of Economic Behavior and Organization*, 194, 550–567. https://doi.org/10.1016/j.jebo.2021.12.030

Sorensen, A. T. (2007). Bestseller lists and product variety. *The Journal of Industrial Economics*, 55(4), 715–738.

Stoll, J. (2022). *Reasons for subscribing to multiple paid video streaming services among adults in the United States as of February 2020.* Retrieved from Statista; https://www.statista.com/statistics/1131877/why-to-subscribe-paid-video-streaming-services-us/

Stoll, J. (2022, October 19). *Number of Netflix paid subscribers worldwide from 1st quarter 2013 to 3rd quarter 2022.* Retrieved from Statista; https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/

Stoll, J. (2022, September 2). *Ranking of countries with the highest penetration rate of subscription video-on-demand (SVOD) services worldwide in 2022.* Retrieved from Statista; https://www.statista.com/statistics/813698/svod-reach-by-country/#:~:text=In%202022%2C%20the%20penetration%20rate,Canada%2C%20at%20around%2039%20percent.

Tvguidelines, n.d., Understanding the TV ratings and Parental controls. Retrieved from; http://www.tvguidelines.org/resources/TV_Parental_guidelines_Brochure.pdf

Wayne, M. L. (2018). Netflix, amazon, and branded television content in subscription video on-demand portals. *Media, Culture & Society*, 40(5), 725–741. https://doi.org/10.1177/0163443717736118

Yeung, C., Cimini, G., Jin, C. (2011). Dynamics of movie competition and popularity spreading in recommender systems. *Psysical review*. http://dx.doi.org/10.1103/PhysRevE.83.016105

Zhang, J. & Tucker, C. (2011) How Does Popularity Information Affect Choices? A Field Experiment. *Management Science.* 57(5):828-842. https://doi.org/10.1287/mnsc.1110.1312

Zhang, J. (2010). The sound of silence: observational learning in the u.s. kidney market. *Marketing Science*, 29(2), 315–335. https://doi.org/10.1287/mksc.1090.0500

APPENDIX

Appendix A

Appendix A1: GitHub link

A GitHub link to the repository that includes the Python web scraper, the Python code for retrieval of the data from the IMDbPy API, all raw and cleaned datasets and all R code which includes cleaning, and all analysis.

https://github.com/Robvdwielen/ThesisRob

Appendix B

Figure B1: Cumulative rank by video type

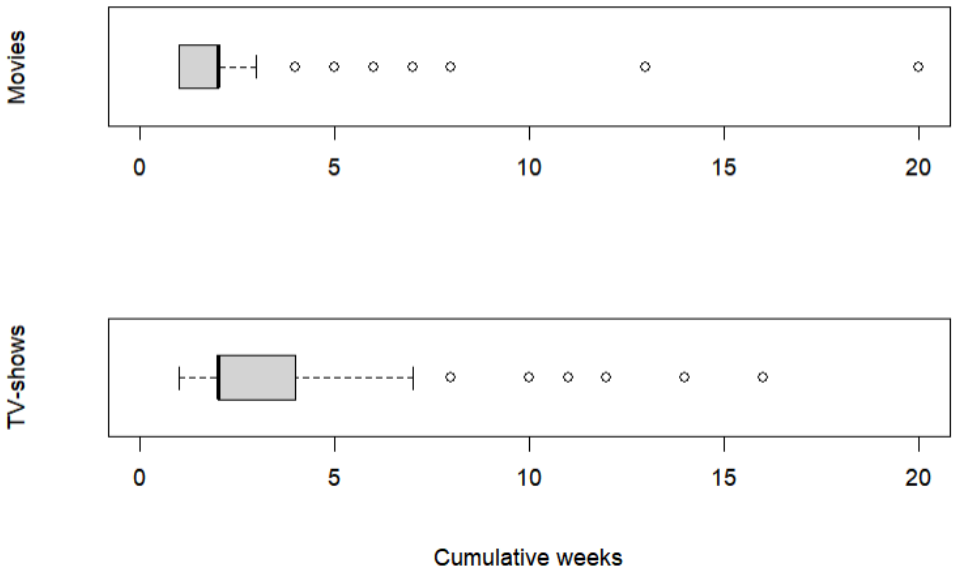


Figure B2: Homoscedasticity without log() transformation

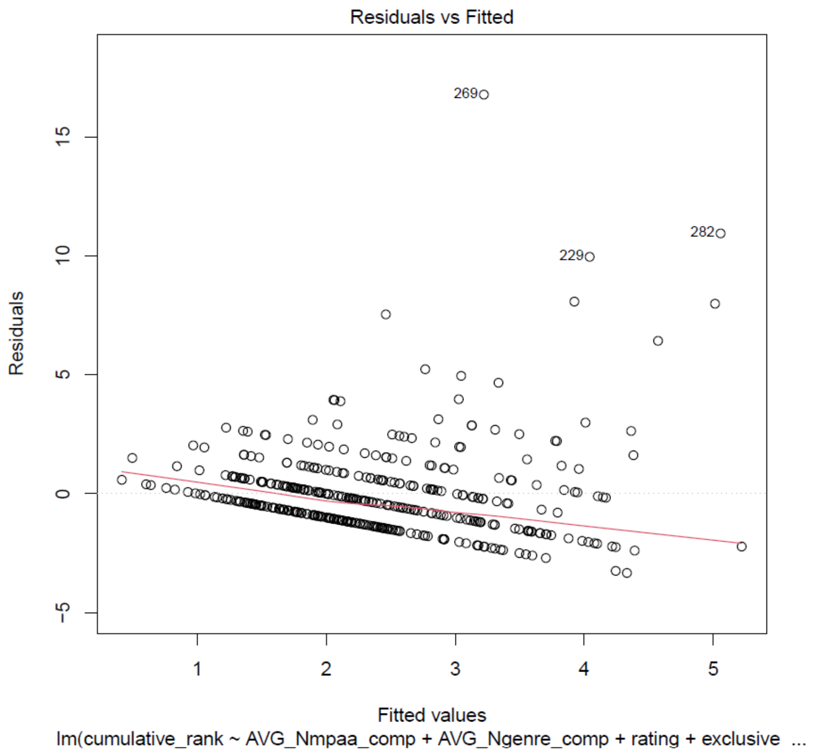


Figure B3: Homoscedasticity with log() transformation

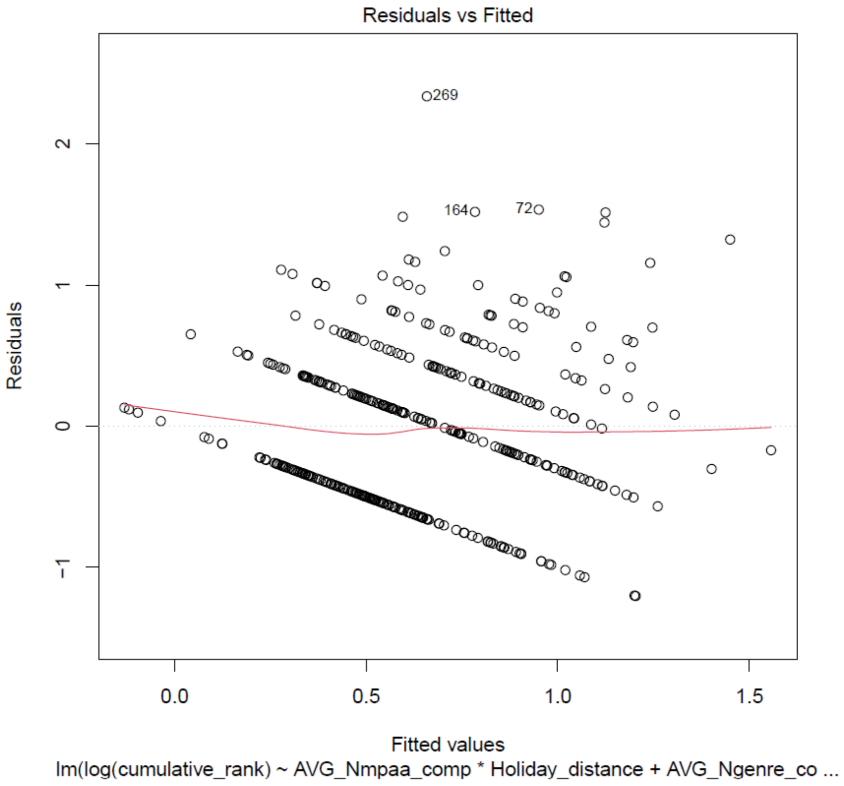


Figure B4: Q-Q plot for normality assumption

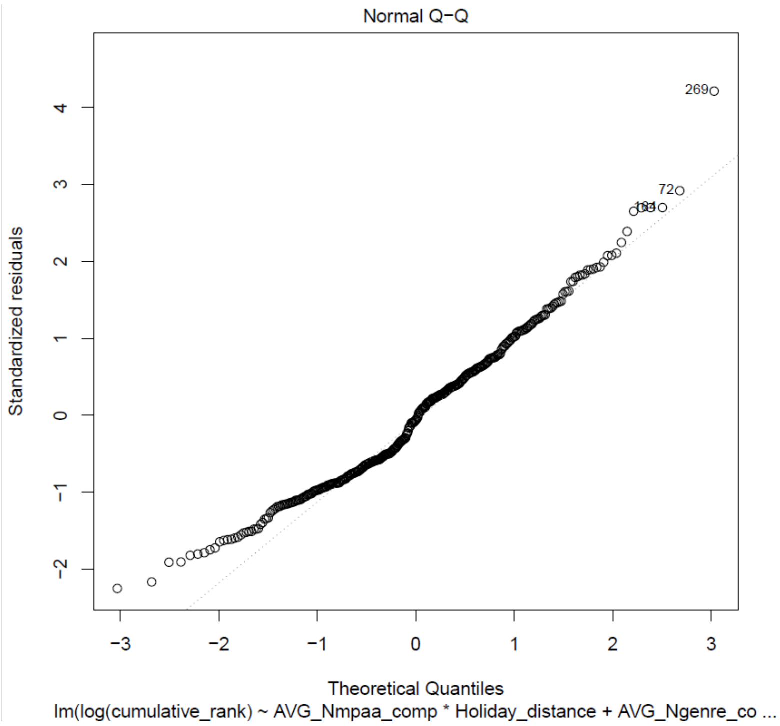


Figure B5: Most influential observations

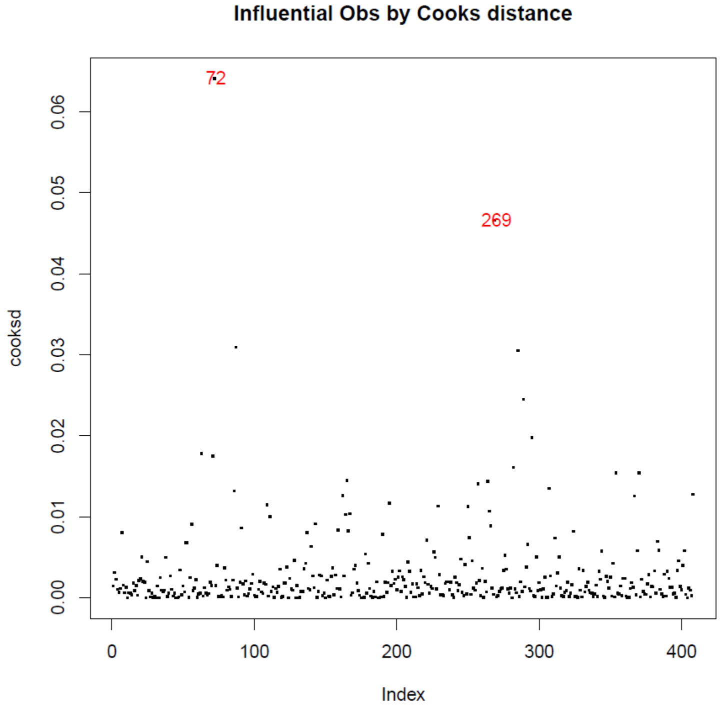


Figure B6: Outlier and leverage plot

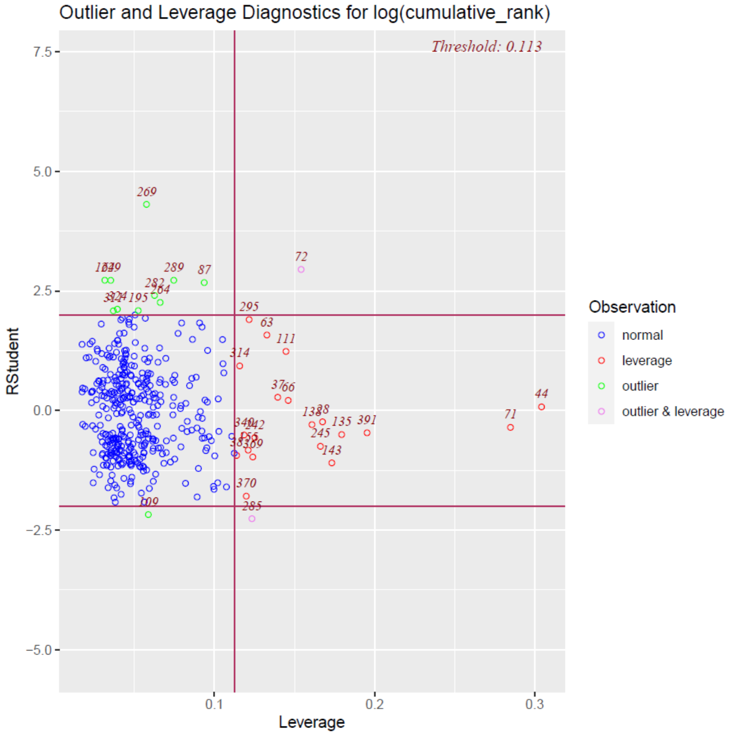


Figure B7: Influential observations according to Cook's Distance, with cut-off point (4/sample_size).

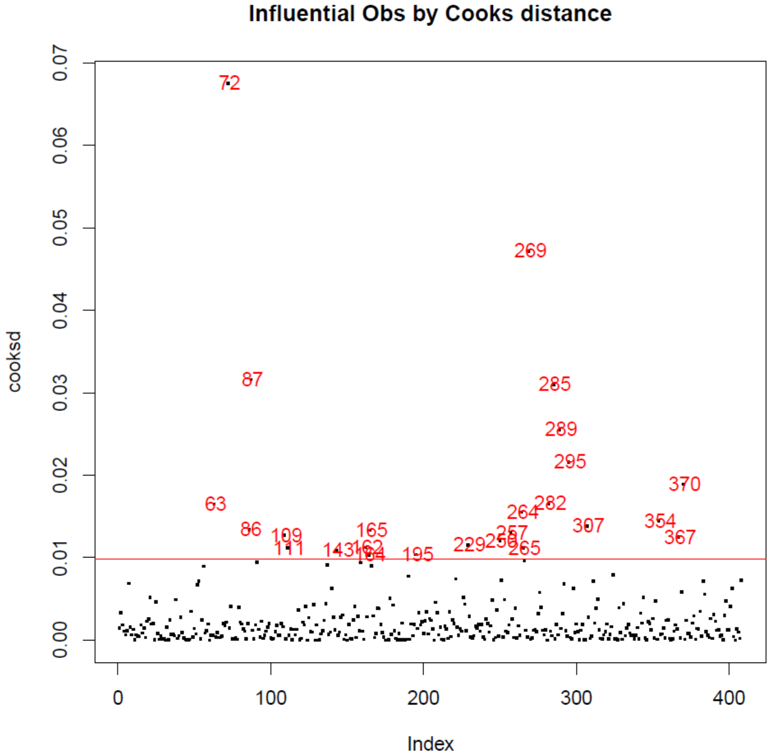
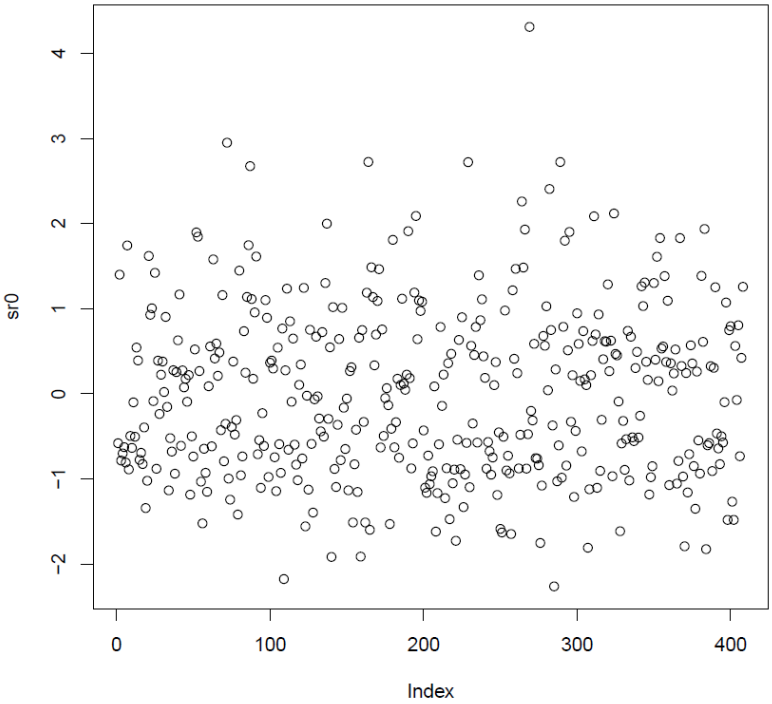


Figure B8: Outlier plot with studentized residuals. (Use a cut-off of 3/-3)



Appendix C

Table C1: VIF scores for multicollinearity assumption

Variables	VIF
AVG_Nmpaa_comp	**10.32**
Holiday_distance	4.63
AVG_Ngenre_comp	4.89
rating	1.36
exclusive	1.74
Type	1.92
sequel	1.23
Drama	1.56
Comedy	1.82
Suspense	1.84
Family	2.52
Action	1.51
Sci.Fi	1.22

Romance	1.67
Other	1.92
mpaa_R_Restricted	4.43
mpaa_G_General_audiences	1.60
mpaa_PG_Parental_guidance_suggested	9.36
mpaa_PG13_Parents_strongly_cautioned	**17.5**
mpaa_NC17_Adults_only	9.98
mpaa_Not_Rated	1.31
AVG_Nmpaa_comp:Holiday_distance	6.99
Holiday_distance:AVG_Ngenre_comp	6.37

Table C2a: Correlation matrix for multicollinearity assumption (part A)

	exclusive	cumulative_rank	Type	rating	AVG_Nmpaa_comp	AVG_Ngenre_comp	Holiday_distance	sequel	Drama	Comedy
exclusive	1	0.12	-0.43	-0.06	-0.2	-0.08	0.04	-0.05	0.03	-0.08
cumulative_rank	0.12	1	-0.21	0.2	0.01	-0.07	-0.05	0.23	0.04	0.05
Type	-0.43	-0.21	1	-0.28	0.05	-0.03	-0.03	-0.29	-0.03	0.23
rating	-0.06	0.2	-0.28	1	0.25	0.12	0.01	0.22	0.15	-0.07
AVG_Nmpaa_comp	-0.2	0.01	0.05	0.25	1	0.27	0.03	0.11	-0.07	0.16
AVG_Ngenre_comp	-0.08	-0.07	-0.03	0.12	0.27	1	0.11	0.05	-0.09	-0.14
Holiday_distance	0.04	-0.05	-0.03	0.01	0.03	0.11	1	-0.02	0.01	-0.08
sequel	-0.05	0.23	-0.29	0.22	0.11	0.05	-0.02	1	0	0
Drama	0.03	0.04	-0.03	0.15	-0.07	-0.09	0.01	0	1	-0.28
Comedy	-0.08	0.05	0.23	-0.07	0.16	-0.14	-0.08	0	-0.28	1
Family	-0.11	0.15	0.11	-0.01	0.07	-0.03	-0.05	0.05	-0.27	0.46
Action	-0.3	0	0.23	-0.03	0.04	-0.13	-0.03	0.04	-0.09	0.11
Sci.Fi	-0.09	0.13	-0.02	0.01	0.08	-0.1	-0.03	0.14	-0.05	0
Suspense	-0.04	0.01	0.01	0	-0.15	-0.21	0.05	-0.09	0.15	-0.31
Romance	-0.01	0.03	0.09	-0.16	0.08	-0.25	0.02	-0.02	0.1	0.21
Other	0.22	-0.07	-0.27	0.01	-0.04	0.27	0	0.01	-0.34	-0.22
mpaa_R_Restricted	-0.31	-0.11	0.33	-0.11	-0.27	0.05	0.01	-0.15	0.13	-0.05
mpaa_G_General_audiences	-0.09	0.13	0.04	-0.01	0.04	0.04	-0.02	0.06	-0.12	0.14
mpaa_PG_Parental_guidance_suggested	-0.14	0.08	0.11	0.09	0.36	0.15	-0.1	0.06	-0.18	0.34
mpaa_PG13_Parents_strongly_cautioned	-0.27	-0.03	0.18	0.17	0.65	0.06	-0.02	0.09	0.02	0.05
mpaa_NC17_Adults_only	0.41	0.03	-0.44	0.03	-0.08	0.02	0.14	0.04	0.03	-0.21
mpaa_Not_Rated	-0.11	-0.1	0.03	0.03	-0.05	0.18	-0.05	-0.05	-0.09	0.02

Table C2b: Correlation matrix for multicollinearity assumption (part B)

	Family	Action	Sci.Fi	Suspense	Romance	Other	mpaa_R	mpaa_G	mpaa_PG	mpaa_PG13
exclusive	-0.11	-0.3	-0.09	-0.04	-0.01	0.22	-0.31	-0.09	-0.14	-0.27
cumulative_rank	0.15	0	0.13	0.01	0.03	-0.07	-0.11	0.13	0.08	-0.03
Type	0.11	0.23	-0.02	0.01	0.09	-0.27	0.33	0.04	0.11	0.18
rating	-0.01	-0.03	0.01	0	-0.16	0.01	-0.11	-0.01	0.09	0.17
AVG_Nmpaa_comp	0.07	0.04	0.08	-0.15	0.08	-0.04	-0.27	0.04	0.36	0.65
AVG_Ngenre_comp	-0.03	-0.13	-0.1	-0.21	-0.25	0.27	0.05	0.04	0.15	0.06
Holiday_distance	-0.05	-0.03	-0.03	0.05	0.02	0	0.01	-0.02	-0.1	-0.02
sequel	0.05	0.04	0.14	-0.09	-0.02	0.01	-0.15	0.06	0.06	0.09
Drama	-0.27	-0.09	-0.05	0.15	0.1	-0.34	0.13	-0.12	-0.18	0.02
Comedy	0.46	0.11	0	-0.31	0.21	-0.22	-0.05	0.14	0.34	0.05
Family	1	0.27	0.08	-0.24	0.07	-0.12	-0.18	0.42	0.64	-0.17
Action	0.27	1	0.31	0.14	-0.22	-0.24	0.07	0.15	0.18	0.01
Sci.Fi	0.08	0.31	1	0.06	-0.13	-0.11	0	0.12	0.08	0.05
Suspense	-0.24	0.14	0.06	1	-0.33	-0.3	0.2	-0.07	-0.31	-0.06
Romance	0.07	-0.22	-0.13	-0.33	1	-0.15	-0.13	-0.03	0.11	0.09
Other	-0.12	-0.24	-0.11	-0.3	-0.15	1	-0.15	0.01	-0.02	-0.09
mpaa_R_Restricted	-0.18	0.07	0	0.2	-0.13	-0.15	1	-0.11	-0.23	-0.21
mpaa_G_General_audiences	0.42	0.15	0.12	-0.07	-0.03	0.01	-0.11	1	0.26	-0.1
mpaa_PG_Parental_guidance_suggested	0.64	0.18	0.08	-0.31	0.11	-0.02	-0.23	0.26	1	-0.09
mpaa_PG13_Parents_strongly_cautioned	-0.17	0.01	0.05	-0.06	0.09	-0.09	-0.21	-0.1	-0.09	1
mpaa_NC17_Adults_only	-0.22	-0.17	-0.05	0.14	-0.03	0.12	-0.27	-0.16	-0.31	-0.44
mpaa_Not_Rated	-0.03	-0.07	-0.03	-0.07	-0.09	0.05	-0.02	-0.04	-0.06	0

Table C3: Regression results for influential observations

Dependent variable: Log(Cumulative rank)	N = 383	N = 405	N = 408
Intercept	0.063	-0.238	-0.083
Amount competitor MPAA videos	0.009*	0.008*	0.004
Amount competitor genre videos	-0.047*	-0.040*	-0.034
Amount competitor MPAA videos * Holiday Distance	-0.002***	-0.002*	-0.001
Amount competitor genre videos * Holiday Distance	0.007*	0.005 ·	0.005
Rating	0.071***	0.106***	0.099***

Netflix Exclusive	0.261***	0.250***	0.232***
Movie	-0.309***	-0.254***	-0.239***
Sequel	0.097	0.092	0.152*
Drama	0.035	0.040	0.025
Comedy	0.063	0.072	0.047
Suspense	0.030	0.081	0.062
Family	0.076	0.128	0.187
Action	0.001	-0.020	-0.033
Sci.Fi	0.037	0.179 ·	0.117
Romance	0.090	0.159 ·	0.133
Other	-0.137	-0.085	-0.130
R - Restricted for children	0.109	0.096	0.073
G - General Audiences	0.081	0.142	0.215
PG - Parents guidance suggested	-0.010	0.013	-0.004
NC17 - Adults only	-0.054	-0.044	-0.044
No MPAA rating	-0.282 ·	-0.233	-0.262 ·
N	383	405	408
R^2	0.232	0.215	0.196
Adjusted R^2	0.186	0.170	0.150
F statistic	4.957***	4.765***	4.276***

Note: · p <0.1, *p<0.05, **p<0.01, ***p<0.001. Note: The dependent variable in all the models is the cumulative weeks spent in the Netflix top ten. The dependent variable is log-transformed.