A picture is worth a thousand words. Investigating the value of Instagram posts and Twitter posts in predicting movie box office revenues.

Master thesis: Master of Business Administration

UNIVERSITY OF TWENTE.

Author: Jeffrey Rouwenhorst

s1506986

j.h.j.rouwenhorst@student.utwente.nl

University of Twente

Supervisors: Dr. A.J.B.M. Wijnhoven

Dr. R. Effing

Date: 19-06-2018

Table of contents

Acknowledgements

This master thesis marks the end of my master of Business Administration and was internally executed at the University of Twente. During the past seven months, I have put a lot of effort and time into this master thesis assignment. For the execution of the study in this master thesis, I have read many scientific papers from different research field in which various research approaches were applied. These scientific papers were very useful and have significantly contributed to my personal knowledge with regard to social media predictions. On the one hand, gathering new knowledge is very interesting but sometimes I found it difficult to find a perfect balance between finalizing the chapters of this master thesis or reading more scientific material about the topic.

Personally, I would like to thank my supervisors Dr. Fons Wijnhoven and Dr. Robin Effing for spending their time, providing relevant feedback, the useful discussion and for taking my personal situation into consideration as a political leader during and after the local elections in Haaksbergen. I would like to thank them for giving me room and space to choose the subject and dimensions of this research. This master thesis helped me to extend my knowledge regarding data science. This knowledge will be of great importance during the remainder of my personal career. Without this master thesis, my future career would look very different.

Furthermore, I would also like to thank my brother for his mathematical thoughts about the data analysis. In addition, I would also like to thank my parents for offering personal support, for providing me the opportunity to successfully obtain a second master's degree together with being a local politician. Finally, I would also like to thank our dog, who have always surrounded me during the execution of my master thesis. Because the knowledge gathered during the execution of this master thesis would strongly impact the direction of my personal future career, I am very grateful to the people who supported me along the way.

I hope you enjoy reading this master thesis.

Jeffrey Rouwenhorst

Abstract

Purpose: Using Instagram and Twitter data for volume- and sentiment-related predictor variables, the purpose of this paper is to examine the predictive value of Instagram predictor variables compared to Twitter variables in predicting movie box office revenues in the same and the subsequent week.

Design/methodology/approach: A total number of 40.000 posts, containing textual and visual data were collected from Instagram and Twitter. Besides counting post volumes and post views, the Instagram and Twitter posts were used for performing textual sentiment and visual sentiment analysis.

Findings: The predictive value of Instagram predictor variables is on average higher than the predictive value of Twitter predictor variables in predicting movie box office revenues in the same and subsequent week.

Originality/value: Results of this research demonstrate that Instagram predictor variables are on average better predictors of movie box office revenues than Twitter predictor variables. Compared to textual sentiment, visual sentiment was found to be a more relevant predictor of movie box office revenues. With regard to the visual sentiment-related variables, negative visual sentiment-related on Instagram is identified as a significant predictor for movies after the moment of movie release.

1. Introduction

Internet usage has increased every year since the emergence of the digital world with internet connection. In 2017, the number of internet users worldwide was 3.58 billion up from 1 billion in 2005. Humans are currently generating 2.5 quintillion bytes of data every day. A multitude of sources ranging from social media to the increasing use of sensors gives rise to the ongoing increase with regard to the number of available data. The advent of internet results into people actively expressing opinions about products, services, events, political parties on digital platforms like social media (Dashtipour et al., 2016). More specifically, internet and especially social media, provide consumers and organizations with new forms of communications to share information (Cheung & Lee, 2012; Krishen, Berezan, Agarwal, & Kachroo, 2016).

With faster connection speed, internet users are now making social networks a huge reservoir of texts, images and video clips (Cai, Cao, Lin, & Ji, 2016). A large body of research have shown the potential of social media in predicting multiple phenomena of interest, like iPhone sales, stock prices, election outcomes, house prices and box office revenues. The majority of these studies found promising results with regard to the predictive value of social media. The vast majority of the research focusing on social media predictions uses Twitter as their data source (Asur & Huberman, 2010; Bollen, Mao, & Zeng, 2011; De Choudhury & Gamon, 2013; Elshendy, Colladon, Battistoni, & Gloor, 2017; Lassen, Madsen, & Vatrapu, 2014). The results of these studies have shown the potential of social media and in particular Twitter for predictive purposes (Lassen, la Cour, & Vatrapu, 2017). Compared to Twitter, other social media sources like Facebook (De Choudhury, Counts, Horvitz, & Hoff, 2014; Karabulut, 2012), Flickr (Jin, Gallagher, Cao, Luo, & Han, 2010), Instagram (Schmidbauer, Rösch, & Stieler, 2018) and Tumblr (Radosavljevic & Grbovic, 2014) were less often used for predicting multiple phenomena of interest. Despite that more and more visual content is generated and is going to play a more important role in the future. Thus, it can be said that research using image-based social media networking services for prediction purposes is quite scarce (Bashir, Wen, Kim, & Morris, 2018).

The importance of visual content is illustrated by Tiwari (2015). According to Tiwari (2015), images and image sequences (videos) represent a share of 80% of all corporate and public unstructured big data (Tiwari, 2015). Especially due to the fact that visual social networking services are easily accessible, simple to understand, to use and to enjoy (Zulli, 2017), it is expected that this number will further increase in the coming years (Tiwari, 2015). In the sharing economy, this results into increases in the importance of visual data and it will also lead to increases in terms of that information will be more often and easily gathered by visual data (Zulli, 2017). Marketing and Entrepreneurship (2018)[1] assume that an estimated 84 percent of all communications will be visual by 2018. A prominent phenomenon in social media communication is the rise of information exchanges in visual form (Lim & Childs, 2016). In today's digital world, wherein user personalized content such as text, video, photos and much more have become an integral part of people's daily lives, photo intensive social media applications have acquired enhanced adoption in social media users through Instagram (Mittal, Kaul, Gupta, & Arora, 2017). Currently, Instagram continues to evolve and grow rapidly (Mittal et al., 2017) and is the world's largest photo sharing platform (Zulli, 2017). The rise of Instagram is illustrated by the number of monthly active users which was 90 million in January 2013 and was already 800 million in September 2017, showing a more steadily increase in the number of users than social networking services like Facebook and Twitter. In comparison, in quartile 4 of 2017, the number of monthly active users on Instagram was already 3 times larger than the number of monthly active users on Twitter. The ongoing increasing number of users solidifies the statement that Instagram is the ''go-to platform for storytelling around the globe'' (Zulli, 2017).

Instagram has evolved into a unique social media platform where users can document their lives through visual content such as photos and videos (Bashir et al., 2018). Instagram is used by people as a platform to interact with each other, to share personal photos, videos, views and reviews on different topics of daily life, politics, sports, markets and much more (Mittal et al., 2017). Building further on this, Instagram is often used as a tool for self-promotion (Gibbs, Meese, Arnold, Nansen, & Carter, 2014; Marwick, 2015a). Instagram is seen as an image rich application that has the potential to influence consumers' behaviors and motivations differently than any previous social networking site (Lee, Lee, Moon, & Sung, 2015). Visual communication can elicit visceral responses and encourage emotions, ultimately impacting attitudes and influencing behaviors (Geise & Baden, 2015; Iyer, Webster, Hornsey, & Vanman, 2014). Shane-Simpson, Manago, Gaggi, & Gillespie-Lynch (2018) find that Instagram was most trusted of all social networking services because of that visual modalities like images and videos cue the ''realism heuristic'' (Pittman & Reich, 2016). Sundar (2008) argues that people generally have more trust in visual modalities than text. Compared to text-based networking services like for example Twitter, image-based social networking sites are particularly interesting because of that image-based social networking sites plays to the picture superiority effect (Childers & Houston, 1984; Turner & Lefevre, 2017).

[1] Marketing and Entrepreneurship (2018, March 9). 16 Visual Content Marketing Statitiscs That Will Wake You Up. Retrieved from https://medium.com/marketing-and-entrepreneurship/16-visual-content-marketing-statistics-that-will-wake-you-up-59c4c0b80465

Due to the increasing amount of visual content on social media networking services, social networking services emerge as channels for web users to express sentiments (Chen, Ji, Su, Cao, & Gao, 2018). The expression of sentiments is more and more tending to a multimodal way, which consists of images, videos short texts and emoticons (Chen et al., 2018).

1.1. Problem statement

According to an old saying, it can be said that an image is worth a thousand words (Yuan, You, & Luo, 2015). Nevertheless, unless the strong increases in digital visual content, the enormous potential of digital visual content in terms of influencing users and increases in the number of users of image-based platforms, research with regard to image-based platforms like Instagram and the use of visual communication is limited (Chua & Chang, 2016; Djafarova & Trofimenko, 2018; Sheldon & Bryant, 2016). While there is an comprehensive body of research focusing on other social networking sites, such as Facebook and Twitter, studies with regard to photo-sharing networking services are scarce (Bashir et al., 2018; Lee & Sin, 2016). Therefore, it can be said that knowledge with regard to Instagram, in particular in the context of prediction, is currently lacking.

Further, less is known with regard to the differential impact of text-based and image-based social media platforms. Some first research contributions were made by Pittman & Reich (2016), Turner & Lefevre (2017), Shane-Simpson et al. (2018), Chae (2018) and Arceneaux & Dinu (2018). First, Pittman & Reich (2016) study loneliness and find that image-based platforms have the potential to ameliorate loneliness due to the enhanced intimacy they offer. Second, Turner & Lefevre (2017) study orthorexia nervosa, which is an eating disorder. Turner & Lefevre (2017) find that higher Instagram use was associated with a greater tendency towards orthorexia nervosa. The findings of Turner & Lefevre (2017) highlight the implications Instagram can have on psychological well being and the influence social media ‘celebrities’ may have over hundreds of thousands of individuals. Due tot he fact that Instagram is an image-based platform, users may be more likely to follow advice or imitate diets of Instagram ‘celebrities’ because of that they feel a more personal connection than they would on a text-based platform (Turner & Lefevre, 2017). Third, Shane-Simpson et al. (2018) find hat people trust Instagram more than Twitter and that Instagram was most trusted of all social media platforms. Fourth, compared to text-based social network sites, Chae (2018) find that self-presentation on Instagram may be perceived as more realistic. Therefore, Chae (2018) assumes that Instagram is used more for social comparison and that it also leads to emotional contagion. Fourth, Arceneaux & Dinu (2018) use Twitter and Instagram to investigate how the presentation of textually-based and visually-based messages affect American college students’ recall of digital information. Arceneaux & Dinu (2018) find that information retention increases most by visually based information, and that Instagram was more effective than Twitter in terms of information recall.

Although some early advances were made by the aforementioned studies, more insight is needed with regard to the differential impact of textual-based and image-based social networking services. Because of the differences between text-based and image-based social platforms in terms of media appropriateness, media richness, information processing and recall, it is important that more insight is provided in the differential impact of text-based and image-based social platforms in predicting a particular phenomenon of interest. Additionally, more insight is needed with regard to the differential impact of image-based and textual-based social media platforms in other prediction contexts. Therefore, research should reveal whether the findings of Pittman & Reich (2016), Turner & Lefevre (2017), Shane-Simpson et al. (2018), Chae (2018) and Arceneaux & Dinu (2018) can be extended to other research fields. Therefore, knowledge with regard to the extent to which social media predictor variables from an image-based social platform (Instagram) or an textual-based social platform are better able to predict movie box office revenues should be produced.

The rise of information sharing on social media has also led to increases in the popularity of electronic word of mouth (eWOM) (Djafarova & Trofimenko, 2018). According to Djafarova & Trofimenko (2018), opinions formed from strong social connections on rich social media platforms like Instagram are viewed to be highly influential to users (Thoumrungroje, 2014; Wilcox & Stephen, 2013). Therefore, Chen & Chang (2018) argue that consumers and businesses should seek to transfer information through rich media formats. Lee et al. (2015) state that Instagram is seen as an image rich application that has the potential to influence consumers' behaviors and motivations differently than any previous social networking site. Especially the increases in visual content, increases in the number of users of image-based social networking sites and the potential of image-based social networking sites with regard to influencing users demands for more insight in the effect of visual eWOM.

With regard to sentiment analysis, most research retains on analyzing textual modalities alone, while analyzing sentiments from visual and other modalities retains as an open problem (Campos, Salvador, Jou, & Giró-i-Nieto, 2015; Chen, Gao, Cao, & Ji, 2015; Chen et al., 2018; Jingwen, Fu, Xu, & Mei, 2016; Wang, Wang, Tang, Liu, & Li, 2015; You, Luo, Jin, & Yang, 2016). Therefore, it can be said that there is a strong demand for insight with regard to the extent to which visual sentiment analysis is an accurate predictor of a phenomenon of interest, like for example movie box office revenues. Additionally, insight with regard to the extent to which Instagram volume-related and sentiment-related variables are able to predict weekly movie box office revenues is currently lacking. Therefore, it can be concluded that there is a strong demand for new insights with regard to whether textual and visual sentiment on Instagram and Twitter are able to predict movie box office revenues. In addition, by analyzing both sentiment in visual and textual posts, the predictive value of visual and textual posts can be identified for both platforms. In conclusion, the current knowledge with regard to the predictive value of visual and textual posts needs to be extended.

1.2. Research questions

The research goal is to identify the influence of Instagram and Twitter predictor variables in predicting movie box office-revenues. The reason for selecting this research goal is that knowledge with regard to the impact of social media predictor variables is currently lacking. In addition, knowledge with regard to whether textual and image-based social networking sites can be best used for predictive purposes is also lacking. More specifically, this study aims at revealing insight in the differential impact of textual and image based social media platforms in predicting movie box office revenues. To add knowledge to the existing literature, the following central research question and sub questions can be drawn:

- Central research question: To what extent can movie box office revenues be predicted by Instagram and Twitter predictor variables?

The following sub questions are formulated in order to completely answer the central research question:

Sub questions:

- What is the relationship between volume-related Instagram and Twitter variables and box office revenues of newly released movies?
- What is the relationship between textual sentiment-related variables on Instagram and Twitter and box office revenues of newly released movies?
- What is the relationship between visual sentiment-related variables on Instagram and Twitter and box office revenues of newly released movies?
- Which Instagram and Twitter predictor variables are significant predictors of movie box office revenues?

1.3. Academic relevance

Chen, Lu, & Wang (2017) argue that argue that especially newer social media data sources should be used for prediction purposes. Research using image-based social media networking services for prediction purposes is quite scarce (Bashir et al., 2018), whereas more and more visual content is generated and is going to play a more important role in the future. Further, research with regard to image-based platforms and visual communication is also limited (Chua & Chang, 2016; Djafarova & Trofimenko, 2018; Sheldon & Bryant, 2016). Zulli (2017) argues that research with regard to

Instagram is still in its infancy. From a scientific point of view, the academic relevance of this study is that extends the current knowledge with regard to Instagram in the context of predictions. In addition, this research contributes to the limited body of research with regard to image-based platforms and visual communication (Chua & Chang, 2016; Djafarova & Trofimenko, 2018; Sheldon & Bryant, 2016). Furthermore, this study sheds light on the differential impact of text-based and image-based social media platforms with regard to predictions.

Because of the increases in the use of visual communication, visual content and the number of users of image-based social network platforms, it is important to gain more insight in the differences between text-based and image-based social network platforms. From an academic point of view, this study is relevant because of that it sheds light on the differences between textual and image-based social media platforms in terms of information processing and information recall. More specifically, this study uses media appropriateness and media richness theory and extends the current knowledge with regard to media appropriateness and media richness theory. This research extends the current knowledge with regard to the extent to which image-based and text-based social media platforms can be used for prediction purposes in other contexts. Further, this study adds new knowledge with regard to the extent to which social media predictor variables from an image-based or an textual-based social platform are better able to predict movie box office revenues. This study also provides more insight in the effects of eWOM on textual and image-based social networking sites. Additionally, this study also sheds light on visual sentiment analysis. Especially with regard to the extent to which visual sentiment analysis is an accurate predictor of a phenomenon of interest, in this case movie box office revenues. Furthermore, this study also provides insight with regard to the extent to which Instagram volume-related and sentiment-related variables are able to predict weekly movie box office revenues. As a result, this study reveals the power of visual and textual content in predicting movie box office revenues.

Finally, the multilingual visual sentiment tool Complura is used in order to perform visual sentiment analysis. Complura is rarely used in scientific research and therefore this research also offers knowledge with regard to the extent Complura can be used to predict phenomena of interest, like for example movie box office revenues. Additionally, based on the findings of this research, it can be determined whether Complura is more suitable for detecting visual sentiment in Twitter or Instagram posts in predicting phenomena of interest in other research fields.

1.4. Practical relevance

Due to the global reach of Instagram, the strong increases in the number of users and the enormous potential of Instagram in terms of influencing users, there is a growing interest amongst marketers to use Instagram (Sheldon & Bryant, 2016). The industry has widely adopted Instagram, however academic research on Instagram and visual communication lags far behind (Chen, 2017; Chua & Chang, 2016; Djafarova & Trofimenko, 2018). Only a limited number of studies have been conducted to specifically examine Instagram usage and marketing (Chen, 2017). In this era of big data, it has been proven that integration of visual content can offer more reliable or complementary online social signals (You et al., 2016). According to Hubspot (2017),[2] 74% of social media marketers use visual assets in their social media marketing. Additionally, according to Buffer Social (2018),[3] images are the most shared content (95%) by business on social media, followed by links (85%) and written content (80%). Marketers agree that a picture is worth much more than a thousand words in this age (Lim & Childs, 2016). According to Abed (2018), Instagram is one of the best platforms to advertise on, because of that engagement with brands on Instagram is 10 times higher than Facebook, 54 times higher than Pinterest, and 84 times higher than Twitter. Instagram is seen as an image rich application that has the potential to influence consumers' behaviors and motivations differently than any previous social networking site (Lee et al., 2015). This findings of this study helps businesses in determining whether to invest in influencer marketing on Twitter or Instagram. Insight in the predictive value of both platforms provides marketers with opportunities to optimize their marketing communication strategy. Since that textual-based and image-based social media differ in terms of user characteristics, the findings of this research enable marketers to more efficiently spend their advertising budget. Likewise, this study also sheds light on which social media channels are most appropriate for reaching a particular target group. In addition, this research also reveals whether different user groups, for example gender and male and young and adult users, behave differently on textual and image-based social media platforms.

Despite the ongoing shift from text-based to image-based communication in the social web, little is known about image sharing practices (Thelwall & Vis, 2017). This study extends current knowledge with regard to image sharing practices. In addition, textual and image-based social media differ with regard to media richness and media appropriateness. This results into differences with regard to information processing, information recall and disclosures. For example, according to Digital

---

[2] Hubspot (2017, January 3). 42 Visual Content Marketing Statistics You Should Know in 2017. Retrieved from https://blog.hubspot.com/marketing/visual-content-marketing-strategy

[3] Buffer Social (2018, January 18). The State of Social 2018 Report: Your Guide to Latest Social Media Marketing Research. Retrieved from https://blog.bufferapp.com/state-of-social-2018

Marketing Institute (2017),[4] 90% of information transmitted to the brain is visual. Therefore, this study reveals the impact of the aforementioned differences between textual and image based social network platforms. The findings of this research are relevant for practitioners, because practitioners are better able to determine which social media is most appropriate for their business.

In the era of big data, practitioners more extensively rely on information collection from big data. Therefore, there is a strong demand for practitioners in gathering knowledge with regard to visual content and visual communication on social networks, such as Instagram. From a practical point of view, this study provides insight in the value of visual data in predicting movie box office revenues. The findings of this research reveal whether textual and visual data differ in their predictive value. In addition, this research provides practitioners with insights with regard textual and image-based social media platforms can be used for predictions in other contexts. From a practical point of view, this study is relevant in terms of that it sheds light on the extent to which Instagram and Twitter predictor variables are able to predict movie box office revenues. Furthermore, this study also reveals the impact of posts in the pre-release and the post-release stage of the movie. Therefore, this study provides insight in the extent to which posts on image-based or text-based social network platforms are relevant indicators of movie box office revenues in a particular week.

Finally, Complura, a multilingual visual sentiment ontology, is used for performing visual sentiment analysis. The findings of this study can be used by practitioners to investigate whether Complura is an appropriate visual sentiment analysis tool for their business. In addition, practitioners can also determine whether Complura can be best used for textual or image-based social media platforms. Furthermore, the findings of this study shed new light on the extent to which Complura can be used as a tool for predicting movie box office revenues.

Chapter outline

The remainder of this master thesis is organized as followed. Chapter 2 presents a literature review. In chapter 3 the methodology of this master thesis is discussed. In chapter 4 and 5 the results and the analysis are presented. In the end, a final conclusion is presented together with a discussion on the limitations and avenues for future research.

[4] Digital Marketing Institute (2017, n.d.). 2017: The Year of Visual Content. Retrieved from https://digitalmarketinginstitute.com/en-eu/blog/2017-7-18-2017-the-year-of-visual-content

2. Theory

In this chapter, the theory of this research is discussed. First, the strategy with regard to the literature search is discussed. Second, a general discussion on predictions and explanations is presented. Third, social media predictor variables are discussed. Fourth, visual social media communication, visual information processing and recall are discussed. Fifth, Instagram and Twitter are discussed. Finally, at the end of this chapter, media appropriateness theory and media richness theory are discussed.

2.1. Literature search

This section discusses the literature review. A systematic literature review was performed to gain an understanding of the predictive power of Instagram and Twitter. The literature review was performed according to the method of Wolfswinkel, Furtmueller & Wilderom (2013). In addition, the literature review also follows the principles as presented by Webster & Watson (2002). These approaches enable researchers to present a transparent and detailed literature review. The literature review was executed between November 2017 and April 2018. In order to check the available scientific publications in the field of predictions and social media, two commonly used search engines - Scopus and Google Scholar - for retrieving scientific publications were used. Additionally, snowballing was also used as a literature search method. Finally, the website of ScienceDirect, which owns scientific publications of leading journals (for example Computers in Human Behavior, Decision Support Systems and Internet Research) was used as a data source for finding publications. Additionally, the websites of MIS Quarterly and the Institute of Electrical and Electronics Engineer were also used as data sources for this research. The final sample of articles has been composed through abstract comparison, removing duplicates, analysis of the number of citations, forward and backward citations and by reading the articles (Wolfswinkel et al., 2013). The aforementioned process is described in more detail below.

One of the most widely cited studies with regard to social media and predictions is the study of Kalampokis et al. (2013). The literature review builds upon the social media predictor variables as identified by Kalampokis et al. (2013). The literature review of Kalampokis et al. (2013) was performed in 2012. Therefore, it can be said that this literature review needs to be updated with newer publications in the field of social media predictions, especially with regard to Twitter and Instagram. This master thesis focuses on movie box-office revenues. Therefore, a literature review regarding social media and box-office revenues is presented as well. The following sub-goals for the different parts of the literature review are as follows:

- Providing and update and extension to Kalampokis et al. (2013) with a specific focus on the use of Instagram and Twitter for predictions
- Gathering knowledge with regard to the use of social media for predicting movie-box office revenues

Different keywords were used in order to collect literature. The keywords can be categorized into two groups: the first group focuses on the social media predictor variables and the second group focuses on social media and movie box office revenue predictions. An overview of the categories, the topics and the keywords are presented in table 1.

Table 1 - Literature review categories, topics and keywords

Category	Topic	Keyword(s)
Social media predictor variables	Social media predictions	Social media AND predictions Social media AND predictive Social media AND predictive analytics
Social media predictor variables	Instagram	Instagram Instagram AND predictions Instagram AND predictive Instagram AND social media prediction
Social media predictor variables	Twitter	Twitter Twitter AND predictions Twitter AND predictive Twitter AND social media prediction
Social media and movie box office revenue	Movie box office revenue predictions	Movie box office AND predictions Box office AND predictions Movie success AND predictions Movie revenue AND predictions Box office revenue AND social media

In Scopus, the search terms social media AND predictions delivers 3402 results. On the other hand, the search terms Instagram AND predictions only leads to 31 results, which shows that there is only a small extent of research available with regard to Instagram and predictions. Another commonly used web search engine for scientific publications is Google Scholar. The search term social media prediction delivers 2020000 results. Furthermore, the search term Instagram prediction delivers 35700 results.

Social media predictions

The predictive power of social media has been researched by various scholars. This is illustrated by the overviews of social media predictions research as presented by Kalampokis et al. (2013) and Lassen et al. (2017). Kalampokis et al. (2013) present an literature study of social media and predictions and argue that there can be distinguished between three types of social media predictor variables: volume-related variables, sentiment-related variables and profile characteristics of online users. An overview of the literature is presented in appendix 1. The next sections describes predictions and explanations in general.

2.2. Predictions and explanations

First, it is important to describe the concepts of predictive and explanatory analytics in general. In predictive analytics, techniques like data mining, machine learning and statistical modeling are applied for predictive purposes (Lassen et al., 2017). Shmueli & Koppius (2011) state that predictive analytics consists of empirical methods that generate data predictions and methods that enable one to assess power in terms of prediction. Shmueli & Koppius (2011) further argue that predictive analytics play a key role besides explanatory modeling both with regard to theory building and testing. Additionally, there can be distinguished between six tasks of predictive analytics: generation of new theory, development of measurements, comparison of competing theories, improvement of existing models, relevance assessment, and assessment of the predictability of empirical phenomena (Shmueli & Koppius, 2011).

Shmueli & Koppius (2011) argue that predictive analytics and explanatory statistical modeling differ in terms of the various steps of the modeling process. Explanatory statistical models are useful for testing the causality of hypotheses, whereas predictive models are widely used for predictions (Shmueli & Koppius, 2011). Explanatory statistical models rely on the underlying causal relationships between theoretical constructs, whereas predictive models stress associations between variables (Shmueli & Koppius, 2011). Second, on the one hand explanatory modeling aims at decreasing model bias, like for example specification error to gather the most accurate representation of the underlying theoretical model. On the other hand, predictive modeling is aiming at increasing the combination of model bias and sampling variance (Shmueli & Koppius, 2011). Explanatory and predictive modeling further differ in terms of predictive power and prospective nature. More specifically, the predictive power of an explanatory model is restricted, while a predictive model based on the same data can be more powerful. Explanatory modeling builds a model for the purpose of testing already existing hypotheses (Shmueli & Koppius, 2011). As a consequence, in predictive modeling, all predicting variables must be available at the start of the prediction process. This is not a prerequisite in explanatory modeling (Shmueli & Koppius, 2011).

The traditional research approach and the predictive analytics approach

Hair (2007) compares the stages of the traditional research approach with the stages of predictive analytics. On the one hand, the traditional research approach starts with theory followed by the development and testing of hypotheses. On the other hand, the predictive analytics approach starts with data followed by identification of relationships in the data, the development of hypotheses, model building, testing of hypotheses and finally the validation of the model (Hair, 2007).

It can be said that the process of predictive analysis walks through various steps (Brynjolfsson, Geva, & Reichman, 2016). Brynjolfsson et al. (2016) present a typical prediction process based on the use of online crowd-based data. The process consists of the following six steps: 1) crowd-generated data (search trends & social media), 2) data selection, 3) relevant data extraction, 4) processing, summarizing and representation, 5) feature selection and 6) prediction (Brynjolfsson et al., 2016). The fifth step - feature selection - can be seen as an optional step. The first step refers to the fact that the crowd generates data, for example social media posts or search engine queries. The second step refers to identification of data by detecting keywords or entities. The third step refers to the extraction of relevant data. The fourth step refers to data processing and data summarization. This step typically results into a set of predictors (explanatory variables) (Brynjolfsson et al., 2016). The fifth step - feature selection - refers to the fact that statistical or machine-learning procedures can be applied to select features with high predictive power out of a larger set of features (Brynjolfsson et al., 2016). Finally, prediction refers to the predictions that are based on the final set of features (Brynjolfsson et al., 2016).

Predictive analytics is applied in different fields for various purposes like marketing, sales, business intelligence, healthcare, law enforcement, stock market management, financial management and governmental management (Hair, 2007). Predictive analytics can both have a single outcome (univariate) or multiple outcomes (multivariate) (Lassen et al., 2017). An example of an univariate outcome is the prediction of sales or the stock price. The multivariate approach refers to the specification of more than one relationship (Lassen et al., 2017).

2.3. Social media

This paragraph deals with social media. More specifically, social media predictor variables, social visual communication, social visual information processing and recall, and Instagram and Twitter are discussed in the next sections.

Researchers have extensively used social media to predict univariate or multivariate outcomes (Lassen et al., 2017). The results of the studies prove that social media can be successfully used to predict one or multiple phenomena of interest (Lassen et al., 2017). Lassen et al. (2017) argue that social media can provide businesses and researchers with insights in the wisdom of the crowds, because social media users build their social identity through the use of written and visual communication (Kim & Chock, 2015). Because of the enormous potential of social media in providing insight in the wisdom of the crowds, social media is selected as a data source for this research. A large body of research in the field of predictions is using social media networking sites like Facebook and Twitter (Lassen et al., 2017). Nevertheless, social media networking sites like Instagram and Snapchat are becoming more prevalent and will become more important in the future for predictive analytics (Jeong & Lee, 2017; Lassen et al., 2017). Therefore, both Twitter, as an textual-based social networking site, and Instagram, as an image-based social networking site, are selected as data sources for this research.

2.3.1. Social media predictor variables

This section discusses social media and predictions. First, a general introduction of research covering social media and predictions is presented. Social media, also known as consumer-generated content consists of platforms to create and exchange user-generated content (Yu & Kak, 2012). According to Pittman & Reich (2016), there can be distinguished between different types of social media: textual-based (like for example Twitter) and image-based social media (like for example Instagram). One of the most cited studies with regard to social media predictions is the study of Kalampokis et al. (2013). Kalampokis et al. (2013) present an literature study of social media predictions and argue that there can be distinguished between three types of social media predictor variables: volume-related variables, sentiment variables and profile characteristics of online users. In the remainder of this section, research with regard to volume-related and sentiment-related variables is discussed, followed by a discussion on Instagram and Twitter predictor variables and the role of the two types of predictor variables - volume-related and sentiment-related variables - regarding movie box office revenu predictions.

Volume-related variables

In general, most often volume-related variables were used in social media prediction research. One of the earliest examples is the study of Jin et al. (2010) who find that post volume on Flickr provided indications that Barack Obama would win the American Presidential Elections of 2008. In the field of health research, the study of Jashinsky et al. (2014) find strong correlations between a large number of risk tweets and real-time suicide data of the states in the United States. In addition, Young, Rivers, & Lewis (2014) find support for significant relationships between the volume of Tweets mentioning

HIV risk behavior and the actual number of HIV cases in the United States. Nevertheless, volume-related variables were not always found to be significant predictors of phenomena of interest, which was illustrated by Jungherr (2013) who found that the volume of Tweets was not a significant predictor of the 2009 German elections. It is important to mention that volume-related variables, in particular based on Twitter data, gained better results in predicting movie box office revenues than phenomena of interest in other contexts like health and politics.

Besides Twitter, Instagram was also used to analyze the relationship between volume-related variables and one or multiple outcomes. Kim, Cha & Kim (2016) apply an targeted ads experiment on Instagram and find that brands that were more frequently co-mentioned with the brand of interest had comparatively higher advertising effects than other brands on Instagram. Significant relationships were found between users who more jointly mentioned the brand name and more favorable responses to advertisements (Kim et al., 2016). In another study, Schmidbauer et al. (2018) studied the 2018 United States elections based on Instagram data. Schmidbauer et al. (2018) find that the number of Instagram media postings in favor of Trump was far higher than Clinton supporters and Trump opponents in the days before the presidential election. An overview of Twitter and Instagram predictions is presented in appendix 2. An overview of the literature with regard to volume-related variables and movie box office revenues can be found in appendix 3. Kalampokis et al. (2013) argue that sentiment-related variables are another type of social media predictor variables. Therefore, the next sections discusses literature with regard to sentiment-related variables.

Sentiment-related variables

Soleymani et al. (2017) state that sentiment comprises of a sentiment holder, an emotional disposition, a polarity and an object. Liu & Zhang (2012) argue that sentiment analysis should be able to capture sentiment holder, entity and aspect of entity in addition to the polarity. The most prolific domain for sentiment analysis is the classification of polarity of Tweets (Soleymani et al., 2017). With regard to sentiment analysis, there can be distinguished between four typical sentiment analysis approaches: textual (lexicon-based dictionaries, bag of words, word embedding and deep neural networks), speech (paralinguistic features), visual (adjective-noun pairs that carry strong sentiments through convolutional neural networks, facial expression, facial action units and visual aesthetics) and multimodal sentiment (multimodal fusion of text, facial expression and paralinguistic features) (Soleymani et al., 2017). In this research, textual and visual sentiment are used for performing analysis and therefore only these sentiment analysis approaches are discussed in more detail.

With regard to textual sentiment analysis, different tools like SentiStrength and SentiBank were used by researchers to perform analysis. Lipizzi, Iandoli, & Marquez (2016) use SentiStrength to predict movie box office revenues and find that sentiment is a weak predictor of movie box office revenues when it is used as the only predictor variable (Lipizzi et al., 2016). Liu, Ding, Chen, Chen, & Guo (2016) use a lexicon and find that sentiment is a significant predictor of movie box office revenues. In addition, White (2016) found that mean sentiment assigned by Twitter users to political candidates was the best predictor variable of their model. In another study, Schumaker, Jarmoszko, & Labedz (2016) use a tool similar to Sentistrength named OpinionFinder. Bly classifying text in positive or negative and subjective or objective, Schumaker et al. (2016) find that they were able to predict the results of Premier League soccer matches during the year 2014, and that large peaks in the sentiments were better able to predict the final results of matches than betting odds. Furthermore, machine learning is also often used as a tool for sentiment analysis. Machine learning algorithms are used to retrieve information from training and test sets. One of the most famous examples of applying machine learning for sentiment analysis is the study of Asur & Huberman (2010). Asur & Huberman (2010) find that sentiment has only a small influence on movie box office revenues and that the prediction accuracy slightly increases by including sentiment as a predictor variable. However, the volume of Tweets was the most important predictor variable (Asur & Huberman, 2010). Rui, Liu, & Whinston (2013) find that the ratio of positive Tweets was a positive and significant predictor of movie box office revenues, whereas the negative Tweets ratio was negatively and significantly associated with movie box office revenues (Rui et al., 2013). An overview of literature with regard to sentiment-related variables and movie box office revenues is presented in appendix 2.

Visual sentiment

More and more people use images in their social networking sites to share their opinions and ideas (Ji, Cao, Zhou, & Chen, 2016). Unless that there is an increase in image-based social network platforms and multimodal social web, such as images and vlogger opinion posts on YouTube, there is a small body of research with regard to visual and multimodal sentiment analysis (Soleymani et al., 2017). Although some texts provide clear sentiment information, an image could contain more information than text to understand the details of user sentiment (Ji et al., 2016). Islam & Zhang (2016) illustrate this by that images can overcome language boundaries and are more easily to understand.

Visual sentiment analysis refers to analyzing images and their associated tags posted on social media (Borth, Ji, Chen, Breuel, & Chang, 2013). The principal research tasks in visual sentiment analysis revolve around modeling, detecting and leveraging sentiment expressed by means of facial or bodily gestures or sentiment associated with visual multimedia (Soleymani et al., 2017). Just like textual sentiment is always towards an object of entity, likewise visual sentiment is defined towards an object, scene or event present in the visual content (Soleymani et al., 2017).

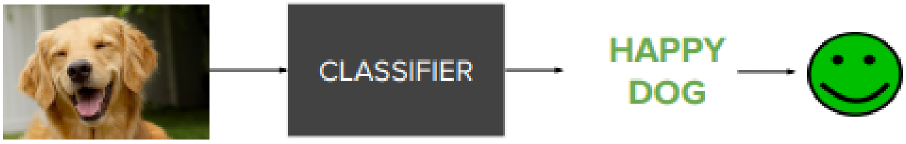
You et al. (2016) argue that there is a strong increase in the use of social media for expressing opinions and sharing experiences. Therefore, it is important to perform sentiment analysis on large scale visual content. Visual content sentiment is complementary to textual sentiment analysis (You et al., 2016) and can be used to extract user sentiment towards events or topics (You et al., 2016). You et al. (2016) use CNNs for Flickr and Twitter image sentiment analysis. The findings of You et al. (2016) show that CNNs are able to achieve better performance with regard to image sentiment analysis competing algorithms. Cai & Xia (2015) use CNNs for multimedia sentiment analysis and argue that a combination of text and image may be a better predictor of sentiment.

ObjectBank was one of the first tools being able to perform and handle attribute learning and mid-level feature representations. Sentibank is a library of trained concept detectors offering a mid-level-visual representation (Ji et al., 2016). Ji et al. (2016) argue that SentiBank is more abstract than low-level features, and experimental results reveal its potential in terms of representing the image sentiment. Almost 1200 detectors were used to construct SentiBank (Ji et al., 2016). The contribution of Borth et al. (2013) is that they construct a mid-level feature called adjective noun pairs (ANPs). Support vector machines were trained using a taxonomy of a semantic construct called adjective-noun pairs (ANPs) (Borth et al., 2013). These ANPs combine a ''noun'' for visually detectability and an ''adjective'' for sentiment modulation of the object described by noun semantics, resulting into pairs such as ''cute dog'' (Soleymani et al., 2017). Borth et al. (2013) proposed the ANPs based on seed keywords from Plutchik's Wheel of Emotion to query Flickr and YouTube's API for images and videos. This approach allows them to create a pool of ANP-candidates (Soleymani et al., 2017). The final set of ANPs was called the Visual Sentiment Ontology (VSO) (Borth et al., 2013).

The main advantage with regard to the VSO is that it included many categories in well-known visual ontologies, like for example LSCOM and ImageNet (Ji et al., 2016). Another advantage is that it can also be used in large sentiment prediction applications, such as microblog sentiment analysis (Ji et al., 2016). Jou et al. (2015) extended the Visual Sentiment Ontology by increasing the volume and breadth in a Multilingual Visual Sentiment Ontology (MVSO), including 15,630 ANPs from 12 major languages and 7,37 million images from over 235 countries (Dalmia, Liu, & Chang, 2016; Soleymani et al., 2017). The sentiment value of the ANPs is evaluated on a scale from 1 to 5 (Pappas et al., 2016), ranging from ''very negative'' (1), ''slightly negative'' (2), ''neutral'' (3), ''slightly positive'' (4) and ''very positive'' (5) (Pappas et al., 2016). According to Cai et al. (2016), ANPs are a state-of-the-art visual sentiment ontology for images.

With regard to the prediction process, object detectors are used to locate each object in a test image, extract features and apply ANP classifiers corresponding to detect nouns and the ANP predicted probability score and the corresponding noun detection score are fused to choose the best candidate (Ji et al., 2016). A simple overview of how visual sentiment analysis works is presented by (Fernández, Campos, Jou, Giró-I-Nieto, & Chang, 2016).

Figure 1 – Simple overview of the process of visual sentiment analysis



Retrieved from Fernández et al., (2016)

Visual sentiment analysis research
Unless that visual sentiment analysis is still in its infancy, some researchers already found promising results. Ahsan, De Choudhury, & Essa (2017) use visual content to capture sentiment information with regard to social event images and were able to demonstrate that the proposed event concept features could be mapped effectively to sentiments. Huber, Mcduff, Brockett, Galley, & Dolan (2018) use visual sentiment, facial expression and scene features from images to analyze the generated language. Data was used from a large Twitter data set. Huber et al. (2018) find that including image features increased how emotional, informative and relevant the generated dialogue was judged to be. In addition, Huber et al. (2018) find that visual sentiment and facial features in the images were the primary drivers of variations in sentiment. Gelli, Uricchio, Bertini, Del Bimbo, & Chang (2015) predict image-popularity of social media images and find that sentiment (as ANPs) are correlated to popularity. Gygli & Soleymani (2016) analyze animated GIFs and more specifically they study the interestingness of GIFs. Gygli & Soleymani (2016) compare different features and find that ANPs are the most informative features in predicting the interestingness of GIFs. Rayatdoost & Soleymani (2016) use mid-level semantic visual sentiment features and find similar results. Flaes, Rudinac, & Worring (2016) use visual sentiment (Flickr) and textual sentiment (Twitter) to study city livability and find significant relationships between extracted sentiment and the indicators of city livability.

Challenges of multimodal sentiment analysis
Multimodal sentiment analysis tools face several challenges. One of the challenges regarding sentiment detection is sarcasm in texts. Nevertheless, there is also a threat of people expressing sentiment for social reasons that are not related to their internal dispositions. Building further on this, a person might express like or dislike sentiments to conform with a certain cultural norm or to express and differentiate his or her identity (Soleymani et al., 2017). Another issue with regard to the use of multimodal sentiment analysis is that data is limited to certain demographics that are more

represented on a specific social media platform (Soleymani et al., 2017). In addition, based on existing work with regard to multimodal sentiment analysis, it is assumed that people are more likely to express positive or negative opinions. As a result, there is a lack of neutral opinions expressed online in all the reviewed multimodal sentiment analysis studies (Soleymani et al., 2017).

A prominent phenomenon in social media communication is the rise of information exchanges in imagery form (Lim & Childs, 2016). Therefore, the next section discusses visual social media communication.

2.3.2. Visual social media communication

In today's digital world, wherein user personalized content such as text, video, photos and much more have become an integral part of people's daily lives, photo intensive social media applications have acquired enhanced adoption in social media users through Instagram (Mittal et al., 2017).

Images can be viewed quickly with limited effort, which makes them attractive to individuals using a more peripheral oriented route (Filieri & McLeay, 2014). Goh, Ang, Chua, & Lee (2009) found that photographs are used by individuals to easily communicate emotions and feelings. Visual content shared between individuals is essential in developing interactions (Pittman & Reich, 2016). Images not only play a major role regarding the initial perception of webpages, but they are also seen as the feature most favored by users (Djafarova & Trofimenko, 2018). In addition, visual representation is more impactful than text for self-expression (Djafarova & Trofimenko, 2018; Zappavigna, 2016).

Social media photo-sharing services

The use of photo-sharing social networking services, like Instagram, Flickr, Pinterest and Snapchat is on the rise (Lee & Sin, 2016). These photo-sharing social networking services enable users to share and view photographs directly from mobile phones which empowers content shares and content seekers to interact through images (Lee & Sin, 2016). According to Goh et al. (2009), individuals are more likely to share photographs than text, audio or video in daily information sharing in social media. Viewing and sharing photographs in social media has become a routine activity for social media users (Bakhshi, Shamma, & Gilbert, 2014). Each day, millions of photographs are uploaded on social media sites (Bakhshi et al., 2014). Photo-sharing social networking services enable users to document their daily lives with photographs (Murray, 2008). This allows viewers to peek into the daily lives of ordinary people (Lee & Sin, 2016). Lee & Sin (2016) argue that photographs might act as a resource for information seekers to learn about social events and to find out personal and detailed information about other users on the network.

Posts with photos tend to receive more attention (Bakhshi et al., 2014). Visual elements in images and photographs are powerful and can affect viewers in emotional and cognitive ways (Lee & Sin, 2016). Increases in the use of image-based social networking sites can be explained by the desire for social connection, drivings users to share their views with other users (Zappavigna, 2016). Lee & Sin (2016) argue that social needs also revolve gaining a sense of belonging.

Self-disclosure

Different communication technologies, like text messaging, social networking sites, and mobile apps, have become the platforms through which many self-disclosures take place (Oeldorf-Hirsch & Nowak, 2018). Images have become an important means for self-presentation (Chua & Chang, 2016). According to Ellison, Vitak, Gray, & Lampe (2014), social media and other forms of mediated communication are able to make self-disclosure and frequent engagement easier.

Self-presentation is of key importance in the context of social media, because users extensively share information about themselves with other users (Djafarova & Trofimenko, 2018). Appropriate self-disclosure has been shown to encourage reciprocity and to increase feelings of closeness and relationship satisfaction, all of which can improve relationship quality (Oeldorf-Hirsch & Nowak, 2018). According to Chiang & Suen (2015), users gain peer acceptance through self-presentations that helps to build relationships with other individuals.

According to Kleemans, Daalmans, Carbaat, & Anschütz (2018), photos and videos are a very direct form of online self-presentation and have become an increasingly powerful form of social online currency. Prior research has revealed that men and women, both adolescents and adults, compare themselves more often to peers than to models or celebrities for social attributes (personality and intelligence) and physical attributes (like weight, height, body image) (Jones, 2001; Strahan, Wilson, Cressman, & Buote, 2006).

Personal self-disclosures vary in two dimensions: valence (positive or negative) and the degree of intimacy (Choi & Toma, 2014). Therefore, it can be said that valence and intimacy can influence the relationship between the disclosing individual and the individual being disclosed to (Taylor & Belgrave, 1986). On the one hand, it can be said that self-disclosure can enhance a particular relationship, but on the other hand disclosures must be kept with certain boundaries or they could have negative relational consequences (Derlega & Chaikin, 1977). Dibble (2014) states that negative information goes hand in hand with more face-saving concerns and results into a greater reluctance about disclosure. Therefore, it is more likely that stigmatized or negative information about the self is shared via private modalities, like face to face interaction (Choi & Toma, 2014).

Gender

With regard to gender, Thelwall & Vis (2017) argue that women were more likely to comment on their own images and images of others. Additionally, more female users were concerned for others to like or comment on their images (Thelwall & Vis, 2017). Men tend to have more individualistic and information communication needs or skills that are expressed visually (Thelwall & Vis, 2017). Women may have more relationship-based communication skills or needs that are expressed visually (Thelwall & Vis, 2017). This leads to the statement that women may be particularly drawn to platforms that provide images, like Instagram, due to the opportunities for intimate connections that Instagram may provide through images or related affordances (Shane-Simpson et al., 2018). Taddicken (2014) argues that women show higher tendencies regarding self-disclosure on social networking services than men. Shane-Simpson et al. (2018) argue that men are more likely to prefer Twitter than women.

Visual information processing

Powell, Boomgaarden, De Swert, & de Vreese (2015) state that visual communication captures viewers' attention. In addition, visual communication has an amplifying effect (Geise & Baden, 2015). Chaiken & Eagly (1976) and Chaiken (1980) argue that the heuristic-systematic processing model explains best the ability of visuals to attract attention. Brubaker & Wilson (2018) argue that visual information is more likely to be processed heuristically and requires little cognitive effort or involvement. On the other hand, systematic processing requires more cognitive involvement and scrutiny of the information presented (Brubaker & Wilson, 2018). Sparks, Areni, & Cox (1998) state that textual information requires more cognitive effort to interpret and that textual information is processed systematically. Referring to dual processing theory, visual stimuli evoke imagery information processing, while verbal stimuli induce discursive information processing (Demangeot & Broderick, 2010). With regard to consumers' decision making processing, Degeratu, Rangaswamy, & Wu (2000) suggest that there are two types of information relevant to consumers' decision making processes, namely sensory and non-sensory information. Nelson, Reed, & Walling (1976) argue that the picture superiority effect can be attributed to encoding distinctiveness at the sensory level of processing. Compared to words, pictures yield qualitatively superior sensory codes (Nelson et al., 1976). On the other hand, words do not lend themselves to sensory discrimination (Nelson et al., 1976). Pittman & Reich (2016) argue that a visual image is sensory-specific, because it is linked to the visual modality, whereas a mental model is not sensory-specific. Krauss, Chen, & Chawla (1996) argue that verbal and non-verbal cues involve distinct psychological mechanisms (Lim & Childs, 2016).

The MAIN model, as presented by Sundar (2008), takes a heuristic approach for understanding how digital technology has altered our perception of credibility. Sundar's MAIN model (2008) offers an theoretical explanation for the effect of presentation of the design elements of interface on information judgment. Sundar (2008) argues that there are four affordances, or mechanisms, for achieving gratification specific to digital media usage, referred to in Sundar's (2008) MAIN Model, that impact human information consumption: Modality, Agency, Interactivity and Navigability. Agency cue refers to the source of information (Sundar, 2008). More specifically, the author of the information influences the credibility, triggering the 'authority heuristic' (Sundar, 2008). Lu et al. (2014) state that interactivity tools trigger 'activity heuristic' by allowing users to act and interact with the system. Increases in interaction results into greater specificity of the resulting content. Finally, the navigability cue (for example hyperlinks) triggers the 'browsing heuristic' and allows users to search relevant information and to 'elaborate' the information (Lu et al., 2014). According to Sundar's MAIN model (2008), each interface serves as a cue to trigger cognitive heuristics for users' information judgment (Lu et al., 2014). According to Sundar (2008) and Zhou, Twitchell, Qin, Burgoon, & Nunamaker (2003), presenting information in rich form reduces the noise or deception that might occur when information is presented in a less rich form such as text.

Sundar's (2008) MAIN model supposes that our brains implicitly trust visual modalities like images and videos more than texts because those modalities cue the ''realism heuristic'' (Pittman & Reich, 2016). According to Kress & Van Leeuwen (2002), visuals become more truthful and more accurately depict reality because they can be seen by an individual. Kim & Lennon (2008) and MacInnis & Price (1987) argue that the heuristics processing model suggests that visual information processing can lead to holistics evaluations of brand and result in more favorable attitudes.

Electronic word of mouth and source credibility
The effect of eWOM is stronger than traditional word of mouth. This can be explained by that electronic word of mouth is quicker, more convenient, and that it affects a larger number of people (Cheung & Thadani, 2012; Godey et al., 2016). The rise of information sharing on social media has also led to increases in the popularity of eWOM (Djafarova & Trofimenko, 2018). According to Reichelt, Sievert, & Jacob (2014), the influence of eWOM on consumer behavior depends upon the credibility of the eWOM source. The credibility of eWOM communications, as perceived by other users, measures the extent to which a user views a recommendation as being a true reflection of reality (Cheung, Luo, Sia, & Chen, 2009; Fan, Miao, Fang, & Lin, 2013; Kareklas, Muehling, & Weber, 2015). High trust in online communicators decreases the level of scrutiny of messages, because readers presume there is validity (Kareklas et al., 2015).

During the first stage in judging online information credibility, users pass through the steps of evaluation surface credibility, such as considering surfaces characteristics like appearance/presentation and information recognition (Sundar, 2008, p. 76). Building further on this, Metzger (2007, p. 2083) argues that people rely most strongly on design and/or presentational elements for judging information credibility and quality (Cheung & Thadani, 2012). Yoon & Zhang (2018) argue that consumers are more likely to adopt arguments when online reviews contain images. This supports the fact that images increase people's ability to store and retrieve information (Yoon & Zhang, 2018).

Social comparison and emotional contagion

According to Kleemans et al. (2018), the effects of social comparison may be stronger when perceived similarity is high, which might be the case with exposure to images of peers in social media (Andsager et al., 2006; Montoya, Horton, & Kirchner, 2008). Fardouly, Diedrichs, Vartanian, & Halliwell (2015) argue that the appearance of peers in social media more directly triggers social comparison because they are seen as more attainable.

Chae (2018) find that the use of Instagram is positively associated with social comparison, whereas Twitter is negatively associated. According to Chae (2018), it seems that Twitter is more about communicating news rather than self-presentation. Stapleton, Luiz, & Chatwin (2017) and de Vries, Möller, Wieringa, Eigenraam, & Hamelink (2018) also find that Instagram and social comparison are associated. Self-presentation on Instagram based on photos and videos produces both social comparison and emotional contagion (de Vries et al., 2018). Individuals compare themselves and their lives to others based on information they receive about others (de Vries et al., 2018). De Vries et al. (2018) state that social comparison can both have negative and positive influences. Negative affect refers to the extent to which an individual experiences aversive emotions like hostility or fear (Watson, Clark, & Tellegen, 1988). On the other hand, positive affect refers to the extent to which one feels enthusiastic, active, and alert (Watson et al., 1988).

Social information encountered on Instagram may impact the viewer's affect through emotional contagion (Johnson & Knobloch-Westerwick, 2017). Emotional contagion refers to the process in which people adopt emotions expressed by others (Hatfield, Cacioppo, & Rapson, 1993). The detection of an emotion another person seems to be sufficient to transfer this emotion to the person who detected the emotion (Neumann & Strack, 2000). Ferrara & Yang (2015), Johnson & Knobloch-Westerwick (2017), Kramer, Guillory, & Hancock (2014) and Lin & Utz (2015) argue that emotional contagion can also happen as a result of viewing social media posts of other people. Despite the limited social cues, receivers of written messages are also able to successfully detect emotions

(Harris & Paradice, 2007). Therefore, individuals can also adopt emotions of others persons without directly viewing them (Neumann & Strack, 2000).

According to Sundar (2008), there is a strong demand for investigating the credibility of social media sources (Arceneaux & Dinu, 2018). Arceneaux & Dinu (2018) state that in general young social media users might view users as social peers. Therefore, they are seen as more credible information sources than professional users. Chen & Kim (2013) suppose that information consumers tend to believe information provided by social peers over national media outlets.

Information recall

According to Franklin & Mewhort (2015), the ability to recall information is dependent on three variables: the words used to convey a message, the associations made among the parts of the message, and the context in which the message was processed. With regard to the substantial influence of message structure on information recall, Sundar & Limperos (2013) found that the intellectual processing of text requires a larger amount of cognitive effort than does processing images. Building further on this, research of Bowman & Hodges (1999), Kumari & Bharadwaj (2017) and Yuan, Zhao, Luan, Wang, & Chua (2014) find support for the fact that the inclusion of visuals results into stronger degrees of cognitive association. According to Childers & Houston (1984), pictures are more memorable than their verbal counterparts. The stronger presence of pictures of memory can be explained by the superior ability to evoke the use of mental imagery - the process by which a previously experienced stimulus is recreated in one’s mind (Childers & Houston, 1984).

Turner & Lefevre (2017) argue that the strong influences of Instagram on users can be explained by different factors. First, Instagram is an image-based platform, which plays to the picture superiority effect - whereby images are more likely to be remembered than words (Childers & Houston, 1984; Turner & Lefevre, 2017). According to Chen (2017), since Instagram is an image-based social medium, by showing the specific scene or the environment when the events and activities actually happen, young consumers could better express and manage their identities and their social relationships in the present as well as retrieve their memories in the future. Thus it can be said that contextuality is a key advantage of Instagram over other types of social networking services (Chen, 2017). Arceneaux & Dinu (2018) investigate whether textually dominant or visually dominant communication was more effective with regard to information retentions. Arceneaux & Dinu (2018) find that information retention increased most by visually based information and that Instagram was more effective than Twitter in terms of information recall. Therefore, Arceneaux & Dinu (2018) suppose that visual-based social media platforms are more effective in producing memory recall than textual-based platforms. The findings of Arceneaux & Dinu (2018) support the findings of Sundar & Limperos (2013) who claim that visually dominant content is more effective in conveying information than textual content. Newhagen & Reeves (1992) and Powell et al. (2015) state that the superiority of visual communication can be explained by the fact that visuals provide more vivid and concrete representations of information than text, visuals are more accessible within a person's memory, making them both faster and easier to process and recall.

Instagram and Twitter

Instagram and Twitter are used as data sources for this master thesis. A general description with regard to Instagram, Instagram use, the content shared on Instagram and an overview of the literature with regard to Instagram and predictions is presented in appendix 4. Additionally, a general description with regard to Twitter, Twitter use, the content shared on Twitter and an overview of the literature with regard to Twitter and predictions is presented in appendix 5. A comparsion of Instagram and Twitter can be found in appendix 6.

Since image-based social media (Instagram) and text-based social media (Twitter) differ in terms of media appropriateness and media richness (Kaplan & Haenlein, 2010), both media appropriateness theory and media richness theory are discussed in the next sections of this chapter.

2.4. Media appropriateness theory

Media appropriateness theory predicts that people will select the modality they deem most appropriate (Rice, 1993) and the efficiency framework predicts that people select the most efficient, convenient, and easiest medium (Nowak, Watt, & Walther, 2009). According to Eden & Veksler (2016), in personal relationships, many motivations drive modality selection for communication, like convenience, intimacy, clarity of communication, and the ability to multitask during the interaction. Modality convenience is the perception that the modality is the easiest or most useful and suitable, which is a separate consideration from perceived appropriateness (Oeldorf-Hirsch & Nowak, 2018). Media appropriateness research has examined how various modalities are perceived and used for different interactions (Rice, 1993). Rice (1993) accounted for media availability, use, and organizational culture and found that the perception of the appropriateness of a medium for a given interaction was weakly associated with its actual use.

The different features for a social media platform seem to invite certain types of expressions and beliefs on what may be considered appropriate (Waterloo, Baumgartner, Peter, & Valkenburg, 2017). Oeldorf-Hirsch & Nowak (2018) argue that it is possible that perceived appropriateness is based on social norms and assumes ideal conditions. Individuals will generally choose the modality that is most efficient or convenient for disclosing information (Oeldorf-Hirsch & Nowak, 2018).

According to Oeldorf-Hirsch & Nowak (2018), the selection of modalities in personal disclosures is dependent upon the content and the form of the disclosure. Shane-Simpson et al. (2018) argue that people may prefer a particular social networking site because they feel they can trust the site to support them in self-expression. Further, with regard to less sensitive disclosures (not stigmatized or positive), efficiency of modality influence modality selection (Oeldorf-Hirsch & Nowak, 2018). Therefore, it is more likely that text messaging will be be prefered over face to face communication for less sensitive disclosures (Oeldorf-Hirsch & Nowak, 2018). Oeldorf-Hirsch & Nowak (2018) find that social networking sites are equally selected because of overall appropriateness and efficiency, but that they varied strongly from positive and public disclosures (posting that one is going to a basketball game) to negative and private disclosures (disclosures about sexually transmitted diseases). The findings of Oeldorf-Hirsch & Nowak (2018) support the assumption that face to face interaction was considered the most appropriate modality across disclosure types. Researchers have compared satisfaction with computer-mediated interactions to face to interactions (Oeldorf-Hirsch & Nowak, 2018). The fact that face to face communication is most appropriate and the contrasting view, nor that media are less appropriate seem to be reflected in actual media usage and interaction patterns (Hovick, Meyers, & Timmerman, 2003; Rice, 1993). The findings suggest that mediated interactions, especially text-based media, are viewed as less appropriate for many interaction goals (Caldwell, Uang, & Taha, 1995; Schmitz & Fulk, 1991).

2.5. Media richness theory

Media richness theory suggests that communication media differ in terms of richness (Daft & Lengel, 1984) - the ability to convey information and enable users to communicate and exchange understanding (Lu et al., 2014). According to Daft & Lengel (1986), media richness theory offers a conceptual framework for measuring a platform and its affordances, with regard to the ability to deliver ''rich information'' that can change understanding, often by reducing ambiguity (Tanupabrungsun & Hemsley, 2018). Therefore, it can be suggested that some media channels are more rich than other channels (Daft & Lengel, 1986). In other words, Carlson & Zmud (1999) state that media richness refers to the relative ability of communication channels to deliver messages containing rich information. Daft & Lengel (1986) state that organizations can select media channels that best solve their information problems.

Daft & Lengel (1984) suggest that various types of communication media convey different levels of information and the richness of information transferred over a period of time depends on the ability of the medium to convey various types of feedback (Chen & Chang, 2018). Tseng, Cheng, Li, & Teng (2017) find that media richness predicts customer's choice of communication media (Lee, Kozar, & Larsen, 2009), decision quality (Kahai & Cooper, 2003), user satisfaction and usage of instant messaging applications (Anandarajan, Zaman, Dai, & Arinze, 2010; Deng, Lu, Wei, & Zhang, 2010; Ogara, Koh, & Prybutok, 2014).

Dimensions of media richness

According to Tseng et al. (2017), media richness consists of four dimensions. The media richness theory dimensions are a) the availability of instant feedback, b) the capacity of the medium to transmit multiple cues such as body language, voice tone, and inflection, c) the use of natural language and d) the personal focus of the medium (Daft & Lengel, 1986).

Newberry (2001) find that richer media consist of more communication modes and social visual cues (for example gestures or immediate feedback) while less rich media have less cues or capacity to facilitate communication (Lu et al., 2014).

Brubaker & Wilson (2018) state that images and videos are part of visual communication. Visual communication can elicit visceral responses and encourage emotions, ultimately impacting attitudes and influencing behaviors (Geise & Baden, 2015; Iyer et al., 2014). Geise & Baden (2015), Iyer et al. (2014) and Powell et al. (2015) argue that visual communication is superior to text. Pittman & Reich (2016) find that image-based platforms like Instagram generate feelings of enhanced intimacy and connectedness relative to text-based platforms.

Media richness and trust

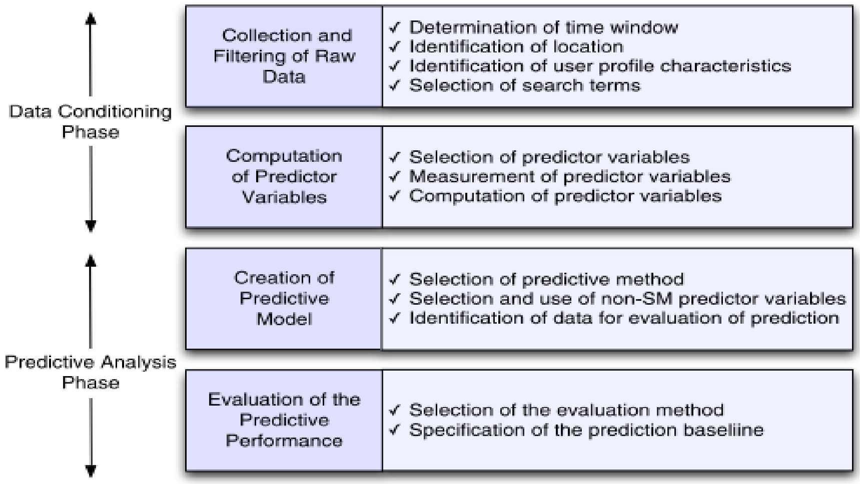
According to Kaplan & Haenlein (2010), text-based social media, like collaborative projects (for example Wikipedia) and blogs, score lowest with regard to social presence and media richness. Because users can share pictures and videos, content communities (for example YouTube) and social networking sites (for example Facebook, Instagram and Snapchat) score higher with regard to social presence and media richness (Kaplan & Haenlein, 2010). Kaplan & Haenlein (2010) argue that social networking sites allow for more self-disclosure than collaborative projects and content communities.

Lu et al. (2014) argue that richer information inspires greater trust. Sundar (2008) argues that technologies that make use of audio visual modalities cue a ''realism heuristic'', and are thus seen as more trustworthy and authentic because of their closer resemblance to the real world (Jeong & Lee, 2017). Therefore, it can be said that self-presentation based on photos and videos on Instagram may be perceived as more realistic (Chae, 2018). Newer social media like Instagram and Snapchat move away from the traditional models of Twitter (text-based) and Facebook (text and visual) to capitalize fully on the ''realism heuristic'' of visual modalities (Jeong & Lee, 2017). The appeal of visually-dominated social media is rooted in their mirroring of perceptual reality (Jeong & Lee, 2017). Jeong & Lee (2017) highlight that closeness to perceived actuality, especially with regard to face to face communication, is a highly desirable feature in computer-mediated communication. Visual social networking sites such as Instagram and Snapchat are platforms that tap into and harness underlying Theory of Mind and social presence mechanisms (Jeong & Lee, 2017). These mechanisms are crucial with regard to bringing computer-mediated communication closer to face to face communication (Jeong & Lee, 2017). Jeong & Lee (2017) state that Instagram and Snapchat successfully appeal to users by showcasing features that are fundamental to how we perceive and process information in face-to-face communication. In line with the findings of Jeong & Lee (2017), Shane-Simpson et al. (2018) find that people trust Instagram more than Twitter. More specifically, Instagram was most trusted of all social media platforms (Facebook, Instagram and Twitter) (Shane-Simpson et al., 2018). Choi & Toma (2014) find that easily accessible and nonintrusive media such as Twitter and texting are used more for sharing positive events, whereas richer, more intrusive media are used more for sharing negative events. In line with Daft & Lengel (1986), just as different information has different information needs, the same holds for different media environments (Tanupabrungsun & Hemsley, 2018).

3. Methodology

Introduction
This section discusses the hypotheses, the research model, the data collection and data analysis of this master thesis. Throughout the master thesis, the social media analysis framework as presented by Kalampokis et al. (2013) was followed. This framework consists of the data conditioning and the predictive analysis phase. The framework is presented below.

Figure 2 - Social media analysis framework



Retrieved from Kalampokis et al. (2013, p. 547)

3.1. Hypotheses

In order to answer the central research question and the sub questions of this master thesis, hypotheses are formulated. These hypotheses are based on the literature review and are tested in this research. Two types of the social media predictor variables as identified by Kalampokis et al. (2013) are included in the hypotheses and are operationalized in the next section. The hypotheses are categorized based on the social media predictor variable types as defined by Kalampokis et al. (2013), namely volume-related variables and sentiment-related variables.

Volume
Liu (2006) find that the volume of eWOM has an influence on movie box office revenue. In addition, Karniouchina (2011) finds that buzz with regard to a movie positively influences movie box office revenues. Baek, Ahn, & Oh (2014) find that the total volume of Tweets corresponds strongly with movie box office revenues. Asur & Huberman (2010) used both the volume and sentiment of Tweets in order to predict movie box office revenues. Their approach outperforms the Hollywood Stock

Exchange (HSE). Rui et al. (2013) find that the total number of Tweets has a significant and positive influence on movie box office revenue. Therefore, the following hypotheses can be drawn:

- *H1a: The volume of Instagram posts for each movie is positively associated with movie gross box office revenue.*
- *H1b: The volume of Twitter posts for each movie is positively associated with movie gross box office revenue.*

Bhavsar, Kumar, Kumar, Gaur, & Sheikh (2017) use social media features like the number of views and comments of movie trailers to predict movie popularity. In addition, Bhave, Kulkarni, Biramane, & Kosamkar (2015) find that the number of views on YouTube is a significant predictor of movie box office revenue. Furthermore, Mestyán, Yasseri, & Kertész (2013) find that the number of views on the movie page on Wikipedia corresponds with movie popularity. Therefore, the following hypotheses can be drawn:

- *H1c: The volume of Instagram post views is positively associated with movie gross box office revenue.*
- *H1d: The volume of Twitter post views is positively associated with movie gross box office revenue.*

Sentiment

Asur & Huberman (2010) find that sentiment in tweets can be used to improve predictions after the movie release. Rui et al. (2013) find that the ratio of Tweets with positive have a positive and significant influence on movie box office revenue. Liu et al. (2016) find that sentiment is a significant predictor of movie box office revenues. The model that achieves the best prediction accuracy consists of for example sentiment variables. The findings of Bhattacharjee, Sridhar, & Dutta (2017) reveal that the polarity of social media content and box-office revenues are causally associated. Further, Gaikar, Marakarkandy, & Dasgupta (2015) find that sentiment score corresponds strongly with the movie box office revenue of Indian movies. Additionally, Jain (2013) find that movie sentiment on social media can be used to predict movie box office success. Parimi & Caragea (2013) find that Tweet sentiment can be used to identify whether someone likes or dislikes a movie. Finally, Rui et al. (2013) find that positive Twitter sentiment is associated with higher movie sales, whereas negative sentiment is associated with lower movie box office revenues. Therefore, the following hypotheses can be drawn:

- *H2a: Positive sentiment in Instagram and Twitter textual posts is positively associated with movie box office revenue.*

- *H2b: Negative sentiment in Instagram and Twitter textual posts is negatively associated movie box office revenue.*

Giancristofaro & Panangadan (2016) find that including visual sentiment as a predictor variable increases the accuracy of predicting the sentiment towards the transport agency. Naumzik, Feuerriegel, & Neumann (2017) utilize images from real estate listings and predict the corresponding rent price. They show that visual sentiment detectors were able to accurately predict the rent prices. You et al. (2016) find that a combination of textual and visual sentiment analysis can achieve better performance than state-of-the-art textual and visual sentiment analysis approaches alone. This is in line with the findings of Cai & Xia (2015) who find that a combination of text and image is a better predictor of sentiment. Therefore, the following hypotheses can be drawn:

- *H3a: Positive sentiment in Instagram visual posts is positively associated with movie box office revenue.*
- *H3b: Positive sentiment in Twitter visual posts is positively associated with movie box office revenue.*
- *H3c: Negative sentiment in Instagram visual posts is negatively associated movie gross box office revenue.*
- *H3d: Negative sentiment in Twitter visual posts is negatively associated movie gross box office revenue.*

General hypothesis

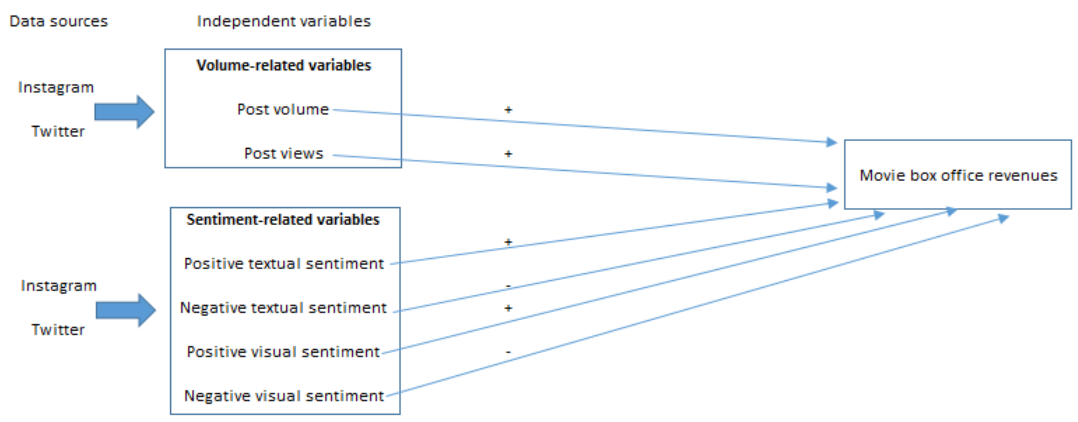
According to Bashir et al. (2018), Instagram provides a venue where communication can take place among its users with a richer graphic conversation via the exchanges of photos and videos. Instagram is seen as an image rich application that has the potential to influence consumers' behaviors and motivations differently than any previous social networking site (Lee et al., 2015). You et al. (2016) state that integration of visual content can offer more reliable or complementary online social signals. Marketers agree that a picture is worth much more than a thousand words in this age (Lim & Childs, 2016). Sundar's (2008) MAIN model supposes that our brains implicitly trust visual modalities like images and videos more than texts because those modalities cue the ''realism heuristic'' (Pittman & Reich, 2016). This heuristic immediately determines that a photograph of something is inherently more real than text written about the same thing; ''that is, we trust those things that we can see over those that we merely read about'' (Sundar, 2008) (Pittman & Reich, 2016). This heuristic underlies the general belief of people that pictures cannot lie (even in the age of digital manipulation) and the consequent trust in pictures over textual descriptions (Sundar, 2008, pp. 80-81). In other words, users trust visual information more than textual information (Lu et al., 2014).

Turner & Lefevre (2017), Shane-Simpson et al. (2018), Chae (2018) and Arceneaux & Dinu (2018) find that Instagram is because of the visual cues more influential than textual social networks such as Twitter, both in terms of information processing and information recall. In addition, compared to Twitter, Instagram is also more rich media and is therefore trusted more and is viewed as being more appropriate for self-disclosures (de Vries et al., 2018; Oeldorf-Hirsch, Birnholtz, & Hancock, 2017). Therefore, the following hypothesis can be presented:

- *H5: Instagram variables* are more accurate predictors *of movie box office revenue than Twitter variables.*

Based on the literature and the hypotheses, the following research model can be drawn for this research:

Figure 3 - Research model



The Instagram and Twitter predictor variables from the research model need to be operationalized. The Instagram and Twitter predictor variables of this research can be classified into two categories: volume-related variables, sentiment-related variables and profile characteristics of users (Kalampokis et al., 2013). Profile-characteristics of users are not included in this research, because user data could not be collected for Instagram users. Due to Instagram privacy restrictions, this data is not offered by Coosto for Instagram. Therefore, volume-related variables and sentiment-related are included in this research as predictor variables. This research consists of the following predictor variables:

- Volume-related variables:
    - Post volume: the volume of Instagram and Twitter posts mentioning a certain movie per week
    - Post views: The total number of post views a post mentioning a certain movie per week receives on Instagram of Twitter

- Sentiment-related variables:
  - Positive textual and visual sentiment ratio: the weekly ratio of Instagram and Twitter textual and visual posts related to a movie with positive sentiment
  - Negative textual and visual sentiment ratio: the weekly ratio of Instagram and Twitter posts related to a movie with negative sentiment in a week
- The dependent variable of this research is movie box office revenue per week.

Instagram and Twitter predictor variables

Based on the categories of predictor variables as presented by Kalampokis et al. (2013), the following predictor variables can be predicted for Instagram and Twitter:

Figure 4 - Instagram and Twitter predictor variables

Instagram

Predictor variable	Description
Movie posts	The total number of Instagram posts mentioning movie X from Friday to next Thursday
Post views	The total number of views of Instagram posts mentioning movie X from Friday to Thursday
Positive textual sentiment ratio	The ratio of textual Instagram posts expressing positive sentiment toward movie X per week
Negative textual sentiment ratio	The ratio of textual Instagram posts expressing negative sentiment toward movie X per week
Positive visual sentiment ratio	The ratio of visual Instagram posts expressing positive sentiment toward movie X per week
Negative visual sentiment ratio	The ratio of visual Instagram posts expressing negative sentiment toward movie X per week

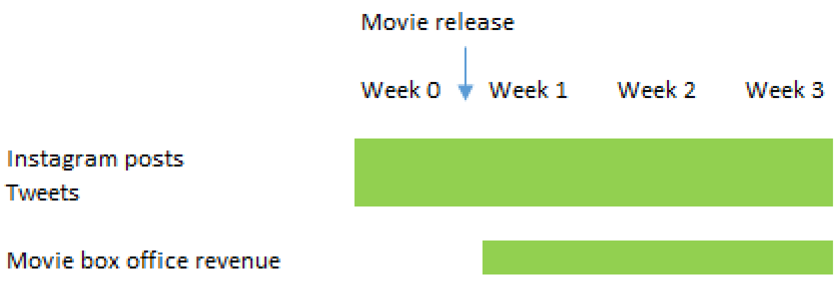
Twitter

Predictor variable	Description
Movie posts	The total number of Tweets mentioning movie X from Friday to next Thursday
Post views	The total number of views of Tweets mentioning movie X from Friday to Thursday
Positive textual sentiment ratio	The ratio of textual Tweets expressing positive sentiment toward movie X per week
Negative textual sentiment ratio	The ratio of textual Tweets expressing negative sentiment toward movie X per week
Positive visual sentiment ratio	The ratio of visual Tweets expressing positive sentiment toward movie X per week
Negative visual sentiment ratio	The ratio of visual Tweets expressing negative sentiment toward movie X per week

3.2. Data collection

For each movie, data from Instagram and Twitter is collected from week 0 till until week 3. Data was collected for movies, which were shown in Dutch cinemas with release dates between 4 January 2018 and 15 March 2018. For all movies, the movie box office revenues in week 1, 2 and 3 were collected. A list of the movie can be found in appendix 7. Instagram and Twitter posts were collected from the week before the movie release (week 0) till two weeks after the release (week 1, 2 to 3). This approach fits with the approach as presented by Baek et al. (2014) who distinguish between pre-consumption (number of Tweets before the movie release) and post-consumption (number of Tweets after the movie release). The Instagram posts and Tweets, one week before the release can be seen as pre-consumption posts, whereas the posts in the three weeks after the release can be seen as the post-consumption posts. Movie box office revenues were collected from the first week after the release (week 1 to 3). An example of the timeline of data collection for movies is presented in the figure below.

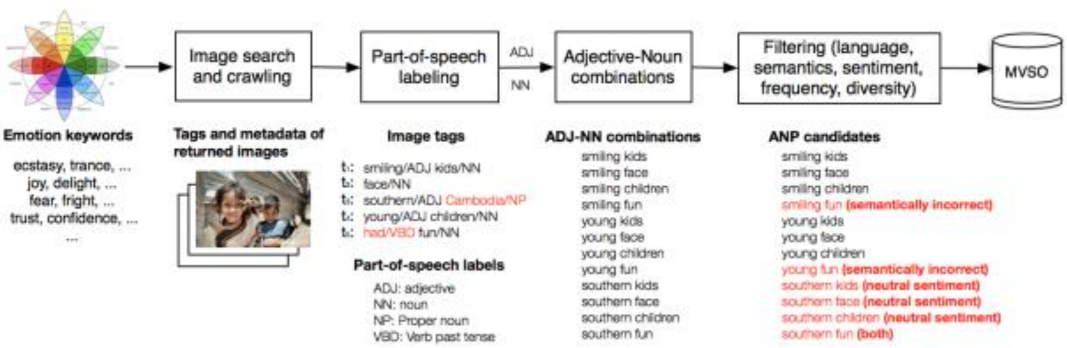
Figure 5 - Timeline of Instagram, Twitter and movie box office revenue data



As can be seen from figure 5, Instagram posts and Tweets were collected in the pre-release week (week 0) and the post-release weeks (week 1, 2 and 3). Movie box office revenues were collected for week 1, 2 and 3. Movie release and movie box office revenue data is collected from http://boxofficenl.net, which presents data from Film Distributeurs Nederland (the official association of Dutch movie distributors). The social media analysis framework for predictions as presented by Kalampokis et al. (2013) and the process of predictions based on crowd-based data model as presented by Brynjolfsson et al. (2016) were used as stepwise models for the predictive analysis of this master thesis.

Instagram posts and Tweets data were downloaded from Coosto. A total number of 2315 Instagram and 37722 Twitter posts were downloaded from Coosto for 42 movies. Both on Instagram and Twitter, most posts were collected during week 2 (the release week): 1,164 and 178.63 posts. In week 0 (the pre-release week) 628 and 10400 posts were collected, whereas 641 and 9459 posts were collected in week 3. The CSV files with Instagram and Twitter data were used for data analysis in SPSS. Coosto is used for sentiment classification. According to Team Nijhuis (2013) (in: Wijnhoven & Plant, 2017, p. 6), the sentiment classifier of Coosto reaches an accuracy of 80% which is quite accurate in comparison to other tools. On the one hand Coosto is used for textual sentiment classification, whereas Complura is used for visual sentiment classification. This is further explained in the section below. With regard to Complura, it can be said that ANPs are assigned to images, together with sentiment scores. After having assigned sentiment scores to the ANPs of images, the data collected for the Instagram and Twitter predictor variables was analyzed in SPSS. These images were stored and transformed into a JPEG file. The JPEG files were placed in Complura, which is a multilinguial visual sentiment ontology tool (Liu, Jou, et al., 2016). An overview of the process of MVSO is presented by Liu, Jou, et al. (2016). As already mentioned, a more detailed overview of how Complura works is presented.

Figure 6 – An overview of the MVSO process



Retrieved from Liu, Jou, et al. (2016, p.2)

The MVSO process

As can be seen from the figure, the MVSO construction process goes through different stages. First, the construction process starts with crawling images and metadata based on emotion keywords (Jou et al., 2015; Liu, Jou, et al., 2016). The emotion keywords are selected according to emotion ontologies from psychology such as Plutchik (1982) and Ekman (1993) (Jou et al., 2015; Liu, Jou, et al., 2016). Second, image tags (t1, t2, t3, t4 and t5 in figure 6) are labeled with part-of-speech tags and nouns are used to form candidate ANP combinations (Borth et al., 2013), while others are ignored (marked in red in figure 6) (Jou et al., 2015; Liu, Jou, et al., 2016). Jou et al. (2015) performed automatic part-of-speech labeling using pre-trained language-specific taggers which achieve high accuracy (more than 95% for the twelve languages: Arabic, Chinese, Dutch, English, French, German, Italian, Persian, Polish, Russian, Spanish, and Turkish), such as TreeTagger, Stanford tagger and HunPos tagger (Jou et al., 2015). Third, the discovered ANPs are filtered based on language, semantics, sentiment, frequency and diversity filters to ensure that the final set of ANPs have the following properties: (a) written in the target language, (b) do not refer to named entities, (c) reflect a non-neutral sentiment, (d) frequently used and (e) used by multiple speakers of the language (Jou et al., 2015; Liu et al., 2016). The aforementioned criteria help to remove incorrect pairs (marked in red in figure 6) (Jou et al., 2015; Liu, Jou, et al., 2016), forming a final MVSO with diversity and coverage. Neutral candidate ANPs were filtered out by scoring each ANP in sentiment using two sentiment ontologies: SentiStrength (Thelwall, Buckley, & Paltoglou, 2011) and SentiWordnet (Esuli & Sebastiani, 2006) (Jou et al., 2015; Liu, Jou, et al., 2016).

Jou et al. (2015) use CNNs to construct the bank of visual concept detectors of ANPS. More specifically, they adopt an AlexNet-styled architecture (Krizhevsky, Sutskever, & Hinton, 2012) because of its good performance on large-scale visual recognition and detection tasks (Jou et al., 2015). Six models were fine tuned in order to train the detector bank for each language. This was done by initializing networks with DeepSentiBank (Chen, Borth, Darrell, & Chang, 2014), an AlexNet model trained on the VSO dataset (Borth et al., 2013; Jou et al., 2015). This approach ensures that each network begins with weights that are already somewhat ‘’affectively’’’biased (Jou et al., 2015). CNN-based visual concept models trained for each language were used to extract image features (Jou et al., 2015). Additionally, sentiment scores of ANPs as supervised labels were used to learn sentiment prediction models (Jou et al., 2015). Furthermore, different layers of the CNN models were compared as image image features (Jou et al., 2015). Jou et al. (2015) binarized the ANP sentiment scores computed into positive and negative classes, and learned a binary classifier using linear support vector machines.

Jou et al. (2015) find that the softmax output features from the penultimate layer outputs of each language's CNN model performed best for all languages.

Graesser, Gupta, Sharma, & Bakhturina (2017) use Complura to classify sentiment by using images and label embeddings. Graesser et al. (2017) achieve reasonable results of 63,3% and beat the baseline accuracy of 61,6%. Pappas et al. (2017) use Complura to evaluate representations with regard to affective concept retrieval, concept clustering and concept predictions across different languages. Pappas et al. (2017) find that the examined ANP representations achieve a level of accuracy of 64-68% on the test set for ANP sentiment prediction. In comparison, Stanford’s deep learning models can achieve the highest accuracy of 85% classifying sentence-level sentiments of movie reviews (Socher, Perelygin, & Wu, 2013).

Limitations of Complura

The prediction accuracy of visual data analysis tools is lower than the prediction accuracy of textual data analytics tools. The relatively lower prediction accuracy of visual data analytics tools, in particular of Complura, can be seen as a limitation of this research. Chaturvedi, Cambria, Welsch, & Herrera (2018) state that Complura faces the limitation of that it only works well for detecting strong emotions. According to Jou et al. (2015), the discovery of affective visual concepts for these languages using ANPs goes hand in hand with challenges in lexical, structural and semantic ambiguities (Jou et al., 2015; Liu, Jou, et al., 2016). With regard to lexical ambiguity, it can be said that words can have multiple meanings which depend on the context, like for example ''sport jaguar'' or ''forest jaguar'' (Jou et al., 2015; Liu, Jou, et al., 2016). Structural ambiguity refers to the differential grammatical interpretation depending on the position in the context, like ambient light or light room (Jou et al., 2015; Liu, Jou, et al., 2016). Semantic ambiguity refers to a situation in which a combination of words with the same syntactic structure have different semantic interpretation, like ''big apple'' (Jou et al., 2015; Liu, Jou, et al., 2016).

3.3. Data analysis

In this section, the data analysis is described. Especially the data preprocessing and the data analysis methods are presented.

Data preprocessing
Tweets and Instagram posts were analyzed for a total number 42 movies. Originally, the total number was 49, but movie box office revenues could not be collected for all the 49 movies. In order to prevent against collection of general posts not related to the movie, keywords as movie, film and bios were added to the search queries in Coosto. The data for the predictor variables was downloaded from Coosto. This data was pre-processed by classifying the Instagram and Twitter posts into the corresponding weeks, like for example post X was posted in the pre-release week and is therefore categorized in week 0. Unless that sentiment scores were already assigned to the textual posts downloaded from Instagram and Twitter by Coosto, the positive and negative textual ratios for each movie were calculated in Excel. In addition, Complura assigns ANPs and corresponding sentiment scores to the images (Soleymani et al., 2017). These sentiment scores were summarized in Excel and assigned to the posts belonging to the different movies. The positive visual sentiment ratios and negative visual sentiment ratios were calculated for the posts about movies containing images. The data for all social media predictor variables and movie box office revenues were stored in one Excel file. This Excel file was used for data analysis in SPSS.

Data analysis methods
First, the mean values were summarized and discussed for the social media predictor variables and movie box office revenues. Second, the linearity of the phenomenon was checked by drawing charts presenting the number of posts on Instagram and Twitter and movie box office revenues during the different weeks. Third, a correlation analysis was performed. Pearson correlations was used because this research only deals with metric variables. Fourth, the model was tested. According to Hair, Black, Babin, & Anderson (2010), regression analysis is one of the most popular data analysis methods. In regression analysis, the linear dependency between a metric dependent variable and one or multiple metric independent variables is examined (Hair et al., 2010). Generally, regression analysis is

performed for predictions, analysis of causes, forecasting and time series analysis. The multiple decision diagram as presented by Hair et al. (2010) was used to perform multiple regression analysis. Multiple linear regression is performed for both the Instagram and Twitter variables in the same weeks and multiple linear regression was performed for prediction for the Instagram and Twitter variables using one-lagged-period variables. In the end, R^2 – the coefficient of determination - can be calculated in order to test the fit of the overall regression model. The R^2 helps to explain the proportion of variance in the dependent variables explained by the independent predictor variables. Compared to the R^2, the adjusted R^2 will not increase when adding extra variable to the linear model. So that is why the adjusted R^2 was used for this study. Finally, the results of the data analysis were used to test the hypotheses and to the research questions.

4. Results

The results of the data analysis are presented in this chapter. First, the dataset characteristics are discussed. Second, the results of the correlation analysis are presented. Third, results of the multiple linear regression analysis for the same week and the next week are presented for both the Instagram and Twitter predictor variables. Furthermore, additional variables are also tested and the results are presented at the end of this chapter.

4.1. Dataset characteristics

The dataset characteristics of this research are discussed. Summary statistics of the Instagram and Twitter variables are presented in table 2 and 3. These tables presents overviews of the mean values of the key variables per week. When inspecting the mean values of the Instagram and Twitter variables, it can be said that the number of posts mentioning the movie is higher on Twitter than on Instagram during week 0, 1 and 2. In week 1, the volume of posts mentioning a movie is highest both on Instagram and Twitter. In addition, in week 2, we see the smallest volume of posts mentioning a movie. This is the case for both Instagram and Twitter. The highest volume in week 1 could be explained by that it might be expected that most people would post about an movie shortly after it has been released. When comparing the number of posts views of Instagram and Twitter, we see that week 2 has the highest number of post views, followed by week 0 and week 1. The descriptive statistics present a similar view by showing that the textual sentiment on Instagram is highest in the release week (week 1). The highest number of textual sentiment is reached in the second post-release week (week 2) on Twitter. When comparing textual sentiment on Instagram and Twitter, we see that on the one hand the positive textual sentiment ratio is slightly higher on Instagram during week 0, 1 and 2. On the other hand, the negative textual sentiment ratio is also slightly higher on Instagram during week 0, 1 and 2. When inspecting the second category of sentiment predictor variables, visual sentiment, it can be said that positive visual sentiment in Instagram posts shows a peak in week 1. In addition, negative visual sentiment in Instagram is constant during the weeks, but the highest number of negative visual sentiment is reached in week 2. In the pre-release week the ratio of positive visual sentiment is higher than in week 2 and the ratio of negative visual sentiment is smaller in week 0 than in week 2. Positive visual sentiment in Twitter posts produces the highest ratio in week 2 and the lowest positive number in week 1, shortly after the movie release. In week 2, the negative visual sentiment ratio is also highest on Twitter.

Table 2 – Mean values of Instagram predictor variables

Predictor variable		
Post volume	Volume of posts on Instagram in week 0	12.07
	Volume of posts on Instagram in week 1	24.29
	Volume of posts on Instagram in week 2	11.93
Post views	Number of post views on Instagram in week 0	47598.91
	Number of post views on Instagram in week 1	2179336.91
	Number of post views on Instagram in week 2	1282179.62
Textual sentiment	Positive textual sentiment on Instagram in week 0	19.68
	Positive textual sentiment on Instagram in week 1	24.76
	Positive textual sentiment on Instagram in week 2	19.5
	Negative textual sentiment on Instagram in week 0	9.62
	Negative textual sentiment on Instagram in week 0	7.24
	Negative textual sentiment on Instagram in week 0	7.17
Visual sentiment	Positive visual sentiment on Instagram in week 0	30.23
	Positive visual sentiment on Instagram in week 1	50.36
	Positive visual sentiment on Instagram in week 2	18.12
	Negative visual sentiment on Instagram in week 0	0.52
	Negative visual sentiment on Instagram in week 1	1.57
	Negative visual sentiment on Instagram in week 2	2.97

Table 3 – Mean values of Twitter predictor variables

Predictor variable		
Post volume	Volume of posts on Twitter in week 0	242.05
	Volume of posts on Twitter in week 1	416.69
	Volume of posts on Twitter in week 2	219.45
Post views	Number of post views on Twitter in week 0	15719173.33
	Number of post views on Twitter in week 1	271842434.21
	Number of post views on Twitter in week 2	2066279.36
Textual sentiment	Positive textual sentiment on Twitter in week 0	18.38
	Positive textual sentiment on Twitter in week 1	20.71
	Positive textual sentiment on Twitter in week 2	23.70
	Negative textual sentiment on Twitter in week 0	6.09
	Negative textual sentiment on Twitter in week 0	5.17
	Negative textual sentiment on Twitter in week 0	6.05
Visual sentiment	Positive visual sentiment on Twitter in week 0	29,03
	Positive visual sentiment on Twitter in week 1	24.78
	Positive visual sentiment on Twitter in week 2	34.76
	Negative visual sentiment on Twitter in week 0	19.69
	Negative visual sentiment on Twitter in week 1	14.12
	Negative visual sentiment on Twitter in week 2	21.77

Table 4 presents the mean values of the weekly movie box office revenues. In the pre-release week (week 0), the movie has not been released, therefore week 0 is not included as a movie box office revenue week. The descriptive statistics show that the mean value of the movie box office revenues is highest in week 1, followed by the week 2 and 3. As expected, the mean values of movie box office revenues are highest during the first week of release (week 2) and steadily decreases with 23% and 30% in the second and the third week after the release.

Table 4 - Mean values of the dependent variable per week

Descriptive statistics - Movie box office revenues		
	N	Mean
Movie box office revenue in week 1	42	383948
Movie box office revenue in week 2	42	294264.81
Movie box office revenue in week 3	42	206094.71

4.2. Weekly trajectories of volume-related Instagram and Twitter variables and movie box office revenues

In order to investigate the weekly trajectory of the volume-related Instagram and Twitter variables and movie box office revenues, charts including the volume-related variables and movie box office revenues were drawn. The weekly trajectories of the volume-related Instagram and Twitter variables and movie box office revenues are presented in appendix 8.

Instagram posts

Only movies for which all data in all weeks was available were included in these charts. This results into a total number of 42 movies. When looking at the figure in appendix 8, it can be said that the weekly volume of Instagram posts peaks in week 1, which is the release week. In addition, the post volume decreases from week 1 to week 3. The movies ''Patser'', ''Fifty Shades Freed'', ''Maze Runner: The Death Cure'' and ''De Wilde Stad'' do not follow this trend. First, ''Patser'' and ''De Wilde Stad'' do not peak and remains constant across the weeks. Second, ''Fifty Shades Freed'' peaks during week 0, which is the pre-release week and the post volume decreases after this week. Third, ''Maze Runner: The Death Cure'' peaks in week 2, which is the second week after the release.

Twitter posts

When looking at figure in appendix 9, which presents the weekly post volume on Twitter, it can be said that there is a peak in posts during week 1, which is the first week after the release (post-release week 1). Additionally, the post volume decreases from week 1 to 2. This is in line with the weekly post volume of Instagram. The movies ''Bankier van het verzet'', ''All the Money in the World'' and ''24 Hours to Live'' do not follow the post volume trend on Twitter. These movies show stable post volumes during week 0 and 1 and the post volume will decrease after week 1.

Movie box office revenues

When looking at the figure in appendix 10, which presents the weekly movie box office revenues, it can be said that the movie box office revenue peaks in week 1 and that the movie box office revenue steadily decreases after week 1. The movies ''Bankier van het verzet'', ''Early Man'', ''Ted en het geheim van koning Midas'' and ''Black Panther'' do not follow this trend. Compared to the identified pattern, the movie box office revenue of ''Black Panther'' starts to decrease some later after week 1. ''Bankier van het verzet'' remains stable over time and slightly decreases during the weeks. The movie box office revenues of ''Early Man'' and ''Ted en het geheim van Koning Midas'' increase from week 1 to 2. This could explained by the fact that these movies are mainly visited by younger children. During week 1 and 2, in the Netherlands there were holidays, which could explain the increase in movie box office revenues for these movies. In the next section, the results of the correlation analysis are presented.

4.3. Correlation analysis

Before entering the variables into the multiple regression model, a correlation analysis has been performed. This study only consists of metric variables, which enables the use of Pearson correlation. Compared to Spearman's rho, Pearson correlation is only suitable for metric variables (Hair et al., 2010). By using Pearson correlation it is possible to analyze the extent to which a relationship is linear. This approach is able to reveal that the change in one variable is associated with proportional change in another variable (Goodwin & Leech, 2006). Pearson correlations were calculated using only complete observations (N= 42) for the Instagram and Twitter variables. A detailed description of the correlation analysis and the correlation matrix with two-tailed p-values can be found in appendix 11.

High positive correlations are observed between the volume of posts on Instagram and movie box office revenue. When comparing the correlations of Instagram and Twitter predictor variables, it can be said that several positive significant correlations have been identified. First, there is a positive significant correlation between Instagram post volume in week 1 with movie box office revenue in week 1, 2 and 3 (.693**, .675** and .660**, significant 0.01), Instagram post volume in week 1 with movie box office revenue in week 1,2 and 3 (.700**, .824** and .777**, significant 0.01) and Instagram post volume in week 2 with movie box office revenue in week 1, 2 and 3 (.772**, .856** and .819**, significant 0.01). Additionally, there is also a positive significant correlation between Twitter post volume in week 1 (.567**, .447** and .489**, significant 0.01), Twitter post volume in week 1 (.767**, .689** and .689**, significant at 0.01) and Twitter post volume in week 2 (.518**, .417** and .449**, significant at 0.01) with movie box office revenue in week 1, 2 and 3. Both the results for Instagram and Twitter with regard to the number of post volumes suggest that while the volume of posts on Instagram and Twitter increase, the movie box office revenues also increase. There were also positive significant correlations between post views in week 0 on Instagram (.565**, .610** and .598**, significant at 0.01) and post views in week 2 (.495**, .428** and .431**, significant at 0.01) and week 3 on Twitter (.812**, .816** and .812**, significant at 0.01) with movie box office revenue in week 1, 2 and 3. The results suggest that while the volume of post views on Instagram and Twitter during respectively week 0 and week 0 and 1 increase, the movie box office revenues also increase. Further, positive significant correlations were found between the positive ratio of visual sentiment in week 0 (.309*, significant at 0.05) and positive visual sentiment in Instagram posts in week 2 (.426**, .423** and .451**, significant at 0.01). These findings suggest that while the positive ratio of visual sentiment in week 2 on Instagram increases, the movie box office revenues will also increase. Witj regard to the positive ratio of visual sentiment on Twitter in week 0, this means that while the positive visual sentiment ratio increases, the movie box office revenue in week 0 also increases. A more detailed overview of the Pearson correlations, covering also the numerical values of the correlations, the strength and the sign of the correlation coefficient, and the statistical significance of the correlation can also be found in appendix 11.

4.4. Model testing

This section describes the results of the multiple regression analyses that was used to test the hypotheses. First, multiple regression analysis is performed on the Instagram and Twitter predictor variables and movie box office revenues for the same weeks to test the relationship between the variables. Second, Instagram and Twitter predictor variables are adopted to predict movie box office revenues in the subsequent week.

4.4.1. Multiple linear regression - same week

This section describes the test of the relationship between the Instagram and Twitter variables and the box office revenue in the same week. Both Instagram and Twitter variables and the box office revenues from week 0 and week 1 are included into a separate regression model. The independent variables are added into the model stepwise so that the effects of including each variable can be identified. First of all, the full model is explored and later on the separate steps are discussed.

Multiple linear regression was performed to test if the Instagram and Twitter predictor variables are able to significantly explain movie box office revenues in week 1 and week 2. Table 5 presents the results from the multiple linear regression with regard to Instagram. The standardized regression coefficient β is presented together with the significance levels for the predictor variables in week 1 and the box office revenue in week 1. The results of the multiple linear regression analysis indicate that the six Instagram variables explained 47,5% of the variance in movie box office revenue (adjusted $R^2 = .475$, $p < .001$). The analysis shows that post volume in week 1 was a significant predictor of movie box office revenue in week 2 (β = 0,711, p. <.001). When analyzing the stepwise additions in table 5 it remains clear that post volume is the significant predictor in each step. In case that post volume is only included in the model, it is able to explain 47,7% of the variance in movie box office revenue in week 1 (adjusted $R^2 = .477$, p. <.001). This suggests a strong relationship between the amount of times a movie is mentioned on Instagram and the amount of movie box office revenues a movie receives in the same week. Adding the positive textual sentiment ratio and the negative textual sentiment ratio to the model results into a slight increase of the adjusted R^2 to .511. Nevertheless, when only including volume of posts and positive sentiment in textual posts, positive sentiment in textual posts comes close to the significance level of p. <0.05 (β = 0,209, 0,063 ns > p. <0.05).

Table 6 presents the results from the multiple linear regression regarding Twitter. The standardized regression coefficient β is presented together with the significance levels for the predictor variables in week 1 and the box office revenue in week 1. The results of the multiple linear regression analysis indicate that the six Twitter variables explained 70,4% of the variance in movie box office revenue (adjusted R^2 = .704, p < .001). The analysis shows that the number of post views (β = 0,544, p. <.001) and post volume (β = 0,425, p. <.01) in week 1 were significant predictors of movie box office revenue in week 1. When analyzing the stepwise additions in table 6 it remains clear that the number of post views and post volume are significant predictors in each step. In case that the number post views is only included in the model, it is able to explain 65,5% of the variance in movie box office revenue in week 1 (adjusted R^2 = .650, p. <.001). This suggests a very strong relationship between the number of post views on Twitter and the amount of movie box office revenues a movie receives in the same week. Further, the results suggest a moderately strong relationship between post volume and the amount of movie box office revenues a movie receives in the same week. Adding positive textual sentiment ratio to the model results into a slight increase of the adjusted R^2 to .722.

Multiple linear regression was also performed on week 2 for the Instagram predictor variables and is presented in table 7. The results of the multiple linear regression analysis indicate that the Instagram predictor variables were able to strongly explain movie box office revenues in week 2 (adjusted R^2 = .719, p. <.001). Post volume is a strong and significant predictor of movie box office revenues in week 2. When adding the positive visual sentiment ratio to the model, together with post volume, the adjusted R^2 slightly increases to .730. The adjusted R^2 is much higher in week 3 (R^2 = .719) than in week 2 (R^2 = .475). This suggests that the release week could be more difficult to predict than other weeks. Furthermore, the results also show that the Instagram predictor variables explained 71,9% of the variance in movie box office revenues in week 2. The results for controlling on positive and negative textual and visual sentiment on Instagram in week 1 and week 2 show that positive and negative textual and visual sentiment on Instagram were no significant predictors in both weeks. Only positive sentiment in textual posts (on Instagram comes close to the significance level of p. <0.05: β = 0,209, 0,063 ns > p. <0.05). So in most cases there seems to be no significant relationship between the sentiment in Instagram posts and the amount of movie box office revenues. In most cases during both weeks adding positive and negative textual and visual sentiment to the model leads to slight increases in the adjusted R^2 measure. But still, the only p-value that was nearly significant was positive textual sentiment in week 2.

Table 5 - Results from stepwise multiple linear regression of Instagram predictor variables on movie box office revenues in week 1.

	Model					
	1	2	3	4	5	6
	β	β	β	β	β	β
Post volume	0,709***	0,711***	0,705***	0,710**	0,710***	0,700***
Number of post views	-0,032					
Positive textual sentiment ratio	0,186	0,193	0,186	0,207	0,209	
Negative textual sentiment ratio	-0,101	-0,102	-0,115			
Positive visual sentiment ratio	0,062	0,052	0,055			
Negative visual sentiment ratio	-0,038	-0,038				
Adjusted R-squared	0,475	0,488	0,501	0,511	0,510	0,477
Degrees of freedom	41	41	41	41	41	41
F	7,179	8,829	11,281	15,302	22,340	38,65
N	42	42	42	42	42	42

Significance levels
*** p <0.001
** p <0.01
* p <0.05

Multiple linear regression was also performed on week 2 for the Twitter predictor variables and is presented in table 8. The results of the multiple linear regression analysis indicate that the Twitter predictor variables were not strongly able to explain to movie box office revenues in week 2 (adjusted R^2 = .122, p. <.001). Only post volume is able to significantly explain movie box office revenues in week 2 (adjusted R^2 = .462, p. <.01). When adding the positive visual sentiment ratio to the model, together with post volume, the adjusted R^2 slightly increases to .197. The adjusted R^2 is much lower in week 2 (R^2 = .122) than in week 1 (R^2 = .704). This suggests that the second release week could be better predicted with the Twitter predictor variables than the release week. Furthermore, the results also show that the Twitter predictor variables explained 12,2% of the variance in movie box office revenues in week 2. In most cases during both weeks adding positive and negative textual and visual sentiment to the model leads to slight increases in the adjusted R^2 measure. But still, the only p-value that was nearly significant was positive visual sentiment in week 2.

Table 6 - Results from stepwise multiple linear regression of Twitter predictor variables on movie box office revenues in week 1.

	Model					
	1	2	3	4	5	6
	β	β	β	β	β	β
Post volume	0,425**	0,423**	0,409**	0,417**	0,386**	
Number of post views	0,544***	0,543***	0,543***	0,545***	0,539***	0,812***
Positive textual sentiment ratio	0,105	0,098	0,097	0,094		
Negative textual sentiment ratio	0,022					
Positive visual sentiment ratio	-0,042	-0,043				
Negative visual sentiment ratio	0,050	0,052	0,055			
Adjusted R-squared	0,704	0,711	0,71	0,722	0,720	0,650
Degrees of freedom	41	41	41	41	41	41
F	17,224	21,21	27,012	36,418	53,772	77,255
N	42	42	42	42	42	42
Significance levels						
*** p <0.001						
** p <0.01						
* p <0.05						

Table 7 - Results from stepwise multiple linear regression of Instagram predictor variables on movie box office revenue in week 2

	Model					
	1	2	3	4	5	6
	β	β	β	β	β	β
Post volume	0,814***	0,820***	0,824***	0,816***	0,815***	0,856***
Number of post views	0,052					
Positive textual sentiment ratio	-0,081	-0,080	-0,075	-0,077		
Negative textual sentiment ratio	-0,071	-0,076	-0,071			
Positive visual sentiment ratio	0,126	-0,130	0,105	0,122	0,106	
Negative visual sentiment ratio	-0,074	-0,073				
Adjusted R-squared	0,719	0,724	0,727	0,729	0,730	0,727
Degrees of freedom	41	41	41	41	41	41
F	18,526	22,54	28,234	37,68	56,336	110,031
N	42	42	42	42	42	42
Significance levels						
*** p <0.001						
** p <0.01						
* p <0.05						

Table 8 - Results from stepwise multiple linear regression of Twitter predictor variables on movie box office revenue in week 2

	Model					
	1	2	3	4	5	6
	β	B	B	β	β	β
Post volume	0,462**	0,461**	0,465**	0,470**	0,443**	0,417
Number of post views	0,017					
Positive textual sentiment ratio	0,066	0,064	0,066			
Negative textual sentiment ratio	-0,03	-0,03				
Positive visual sentiment ratio	0,216	0,213	0,21	0,218	0,251	
Negative visual sentiment ratio	-0,114	-0,117	-0,125	-0,108		
Adjusted R-squared	0,122	0,146	0,169	0,186	0,197	0,153
Degrees of freedom	41	41	41	41	41	41
F	1,953	2,407	3,079	4,125	6,03	8,412
N	42	42	42	42	42	42

Significance levels	
***	p <0.001
**	p <0.01
*	p <0.05

4.4.2. Multiple linear regression for prediction

In the previous section, the independent Instagram and Twitter variables and the dependent variable from the same weeks were analyzed. This section analyzes the predictive power of the independent Instagram and Twitter variables and the dependent variables in the subsequent week. This will be done by analyzing Instagram and Twitter one-period-lagged values. In other words, Instagram and Twitter variables of *week t* are used to predict movie box office revenues in *week t+1.*

Table 9 provides the results from the multiple linear regression analysis using one-period-lagged values. Results show that the Instagram predictor variables from week 0 were significant predictors of movie box office revenues in week 1 (adjusted R^2 = .475, p <.001). The significant predictor of movie box office revenues in week 1 is post volume on Instagram (β = .726, p <.001). Using Instagram predictor variables from week 0 the linear regression model is able to predict 47,5% of movie box office revenues in the next week. The Instagram predictor variables from week 1 were able to predict 68,8% of movie box office revenues in week 2 (adjusted R^2 = .688, p <.001). Like in the previous model, post volume (β = .914, p <.001) is the significant predictor. Furthermore, the Instagram predictor variables from week 1 were also able to significantly predict movie box office revenues in week 2 (adjusted R^2 = .688, p <.001). So it can be said that the Instagram predictor variables in week 1 are able to predict 68,8% of the movie box office revenue in the next week. In week 1, post volume (β = .814, $p < .0001$) is the significant predictor of movie box office revenues in week 3. Combined with post volume, the positive textual sentiment ratio (β = .197, $p < .001$) in week 1 is also a significant predictor of movie box office revenues in week 2. The Instagram predictor variables are able to significantly predict movie box office revenues in week 3 (adjusted R^2 = .666, $p < .0001$). Therefore, it can be said that the Instagram predictor variables in week 2 are able to predict 66,6% of the movie box office revenue in the next week.

In all weeks, Instagram post volume has an significant effect on movie box office revenue. In addition, in week 1, the positive textual sentiment ratio is also a significant predictor of movie box office revenues in the next week. For all the models, the number of post views, the negative textual sentiment ratio, the positive visual sentiment ratio and the negative visual sentiment ratio are not significant predictors of the movie box office revenues in the next week. In week 2, the positive sentiment ratio in visual posts on Instagram is almost nearly significant (β = .211, ns, $p < .05$). Comparing the adjusted R^2 measures of the three models reveals that the model that uses the Instagram predictor variables from week 1 to predict movie box office revenues in week 2 performs best with an adjusted R^2 measure of .688. While the model predicting week 1 and week 3 had an R^2 of respectively .475 and .666. The amount of received movie box office revenues in the release week (week 2) could be more difficult to predict because of other influences like for example promotional activities. Compared to week 3, the Instagram predictor variables perform only slightly better in week 2. Similar to the previous chapter all models show the relevance of the volume-related variable in predicting movie box office revenues.

Table 9 - Results from multiple linear regression between the Instagram variables and movie box office revenues using one-lagged-period values.

1. Instagram predictor variables in week 0 predicting box office revenues in week 1

	Model					
	1	2	3	4	5	6
	β	B	B	β	β	B
Post volume	0,726***	0,723***	0,788***	0,772***	0,761***	0,693***
Number of post views	0,082	0,083				
Positive textual sentiment ratio	-0,137	-0,137	-0,153	-0,14		
Negative textual sentiment ratio	-0,103	-0,105	-0,111			
Positive visual sentiment ratio	-0,213	-0,210	-0,211	-0,201	-0,202	
Negative visual sentiment ratio	0,023					
Adjusted R-squared	0,475	0,489	0,499	0,499	0,492	0,467
Degrees of freedom	41	41	41	41	41	41
F	7,177	8,842	11,199	14,621	20,817	36,938
N	42	42	42	42	42	42
Significance levels						
*** p <0.001						
** p <0.01						
* p <0.05						

2. Instagram predictor variables in week 1 predicting box office revenues in week 2

	Model					
	1	2	3	4	5	6
	B	B	B	β	β	B
Post volume	0,814***	0,842***	0,840***	0,833***	0,833***	0,824***
Number of post views	-0,014					
Positive textual sentiment ratio	0,193	0,197*	0,188*	0,182*	0,184*	
Negative textual sentiment ratio	-0,14	-0,141	-0,043	-0,136		
Positive visual sentiment ratio	-0,017	-0,022				
Negative visual sentiment ratio	-0,043	-0,043	-0,041			
Adjusted R-squared	0,688	0,696	0,704	0,710	0,698	0,671
Degrees of freedom	41	41	41	41	41	41
F	16,049	19,79	25,377	34,461	48,38	84,534
N	42	42	42	42	42	42
Significance levels						
*** p <0.001						
** p <0.01						
* p <0.05						

3. Instagram predictor variables in week 2 predicting box office revenues in week 3

	Model					
	1	2	3	4	5	6
	B	B	B	β	β	B
Post volume	0,763***	0,760***	0,760***	0,749***	0,759***	0,819***
Number of post views	0,093	0,095	0,101			
Positive textual sentiment ratio	-0,068	-0,07				
Negative textual sentiment ratio	-0,083	-0,086	-0,087			
Positive visual sentiment ratio	0,128	0,141	0,126	0,145	0,156	
Negative visual sentiment ratio	0,036					
Adjusted R-squared	0,666	0,674	0,678	0,679	0,676	0,663
Degrees of freedom	41	41	41	41	41	41
F	14,656	17,985	22,58	29,863	43,589	81,789
N	42	42	42	42	42	42

Significance levels
*** p <0.001
** p <0.01
* p <0.05

Table 10 provides the results from the multiple linear regression analysis using one-period-lagged values. Results show that the Twitter predictor variables from week 0 were significant predictors of movie box office revenues in week 2 (adjusted R^2 = .284, p <.001). The significant predictor of movie box office revenues in week 1 is post volume on Twitter (β = .524, p <.001). Using Twitter predictor variables from week 0 the linear regression model is able to predict 28,4% of movie box office revenues in the next week. The Instagram predictor variables from week 1 were able to predict 65,7% of movie box office revenues in week 2 (adjusted R^2 = .657, p <.001). Like in the previous model, post volume (β = .281, p <.05) is a significant predictor, but the number of Twitter post views (β = .657, p <.001) can be seen as a more significant predictor of movie box office revenue in week 2. Furthermore, the Twitter predictor variables from week 1 were also able to significantly predict movie box office revenues in week 1 (adjusted R^2 = .657, p <.001). So it can be said that the Instagram predictor variables in week 1 are able to predict 65,7% of the movie box office revenue in the next week. In week 1, number of post views (β = .656, p <.0001) is the significant predictor of movie box office revenues in week 2, combined with post volume (β = .226, p <.05). The Twitter predictor variables are not able to significantly predict movie box office revenues in week 3 (adjusted R^2 = .16, p <.0001). It can be said that the Instagram predictor variables in week 2 are only able to predict 16% of movie box office revenues in the next week. Post volume is the only significant predictor of box office revenues in week 2 (β = .486, p <.0001) and it should also be mentioned that positive sentiment in visual posts on Twitter is almost nearly significant (β = .253, p .075, ns >.05) in combination with the volume of post on Twitter in week 2.

In all weeks, post volume has an significant effect on movie box office revenue. In addition, in week 1, the number of post views can be seen as the significant predictor of movie box office revenues in the next week, followed by post volume. For all the models, the negative textual sentiment ratio, the positive visual sentiment ratio and the negative visual sentiment ratio are not significant predictors of the movie box office revenues in the next week. Comparing the adjusted R^2 measures of all three models reveals that the model that uses the Twitter predictor variables from week 1 to predict movie box office revenues in week 2 performs best with an adjusted R^2 measure of .657. While the model predicting week 1 and week 3 had an R^2 of respectively .284 and .160. The amount of received movie box office revenues in the release week (week 1) could be more difficult to predict because of other influences like for example promotional activities. Although it is unclear why the Twitter predictor variables perform much better in predicting week 2 than in week 3. This chapter also shows the importance of the volume-related variables on Instagram and Twitter in predicting movie box office revenues.

Table 10 - Results from multiple linear regression between the Twitter variables and movie box office revenues using one-lagged period value

1. Twitter predictor variables in week 0 predicting box office revenues in week 1

	Model					
	1	2	3	4	5	6
	B	B	β	β	β	β
Post volume	0,726***	0,723***	0,788***	0,772***	0,761***	0,693***
Number of post views	0,082	0,088				
Positive textual sentiment ratio	-0,137	-0,137	-0,153	-0,14		
Negative textual sentiment ratio	-0,103	-0,105	-0,111			
Positive visual sentiment ratio	-0,213	-0,210	-0,211	-0,201	-0,202	
Negative visual sentiment ratio	0,023					
Adjusted R-squared	0,475	0,489	0,499	0,499	0,492	0,467
Degrees of freedom	41	41	41	41	41	41
F	7,177	8,842	11,199	14,621	20,817	36,938
N	42	42	42	42	42	42

Significance levels	
***	p <0.001
**	p <0.01
*	p <0.05

2. *Twitter predictor variables in week 1 predicting box office revenues in week 2*

	Model					
	1	2	3	4	5	6
	Β	Β	β	β	Β	β
Post volume	0,814***	0,842***	0,840***	0,833***	0,833***	0,824***
Number of post views	-0,014					
Positive textual sentiment ratio	0,193	0,197*	0,188*	0,182*	0,184*	
Negative textual sentiment ratio	-0,14	-0,141	-0,043	-0,136		
Positive visual sentiment ratio	-0,017	-0,022				
Negative visual sentiment ratio	-0,043	-0,043	-0,041			
Adjusted R-squared	0,688	0,696	0,704	0,710	0,698	0,671
Degrees of freedom	41	41	41	41	41	41
F	16,049	19,79	25,377	34,461	48,38	84,534
N	42	42	42	42	42	42

Significance levels	
***	p <0.001
**	p <0.01
*	p <0.05

3. *Twitter predictor variables in week 2 predicting box office revenues in week 3*

	Model					
	1	2	3	4	5	6
	β	β	β	Β	β	Β
Post volume	0,763***	0,760***	0,760***	0,749***	0,759***	0,819***
Number of post views	0,093	0,095	0,094	0,101		
Positive textual sentiment ratio	-0,068	-0,07				
Negative textual sentiment ratio	-0,083	-0,086	-0,087			
Positive visual sentiment ratio	0,128	0,141	0,126	0,145	0,156	
Negative visual sentiment ratio	0,036					
Adjusted R-squared	0,666	0,674	0,678	0,679	0,676	0,663
Degrees of freedom	41	41	41	41	41	41
F	14,656	17,985	22,58	29,863	43,589	81,789
N	42	42	42	42	42	42

Significance levels	
***	p <0.001
**	p <0.01
*	p <0.05

4.5. Additional analyses

In this section, the additional analyses are discussed. Additional analyses are performed next to the main relationships between the predictor variables. The additional analysis provide insights in the associations between movie box office revenue and the additional variable. For the additional analyses, the classification of features for predicting box office revenue were used as a guideline. According to Lash & Zhao (2016), there can be distinguished between three types of features for predicting movie box office revenues: audience-based, release-based and movie-based features. First, with regard to audience-based features, gender and ages of the audiences are included as variables in our prediction model. The volume of discussions and sentiments are also part of audience-based features, but these variables are already studied and therefore, they were not included as an additional variable. Second, with regard to release-based features, the number of theaters in which a movie is shown is included as an additional variable in the model. Third, with regard to movie-based features, genre, sequel and movie budget have been added as additional variables.

Predictor variables

Predicting the same week

The analysis without the additional variables shows that post volume in week 0, 1, and 2 is a strong and significant predictor of box office revenues in week 1, 2 and 3 (week 0: β .7 and significant at <.005; week 1: β .8 and significant at <.005; week 2: β .8 and significant at <.005). With regard to Twitter, the results of the analysis show that the number of post views (β .8 and significant at <.005) is the strongest predictor of box office revenues, followed by post volume (β .58 and significant at <.01). In week 2, post volume is the strongest predictor (β .5 and significant at <.01) of movie box office revenues. When predicting the next week, in week 1 post volume is the strongest predictor of movie box office revenues (β .8 and significant at <.005) followed by the positive text ratio (β .2 and significant at <.01). In week 2, post volume is again the strongest predictor of movie box office revenues (β .8 and significant at <.005).

Predicting the subsequent week

When predicting the same week on Instagram, the adjusted R^2 is about .5 in week 1 and .7 in week 2. When predicting the next week, in week 0 the adjusted R^2 is .5, in week 1 the adjusted R^2 is .7 and in week 2 the adjusted R^2 is .7 for Instagram. When predicting the same week on Twitter, the adjusted R^2 is about .7 in week 1 and .2 in week 2. When predicting the next week by using data from week 0, the adjusted R^2 scores a value of .5, whereas it scores .7 when using Twitter data from week 1 and week.

Instagram

This section will continue by discussing the impact of adding the additional variables to the model. By adding the additional variables to the model, the predictive value of the model increases significantly. In week 1 the adjusted R^2 of the Instagram data ranges from .6 up to .8 in predicting box office revenues in the same week. The strong increase in the predictive value of the model is largely dependent upon adding movie budget (β .794, and significant at <.005) to the model and the number of theaters (β .658, and significant at <.005) in which a movie is shown. Especially, the impact of movie budget is higher in week 1 when movies are not a sequel (β .8, and significant at <.005). When looking at age, the predictive value of post volume increases strongly (β .982, and significant at <.005). Furthermore, negative visual sentiment ratio is a moderately strong predictor of movie box office revenue for movies with an audience age of 16 years and older (β .-484, and significant at <.05). Nevertheless, the number of theaters is the strongest predictor of movie box office revenue in week 1 for movies with an audience age of 16 years and older strongly (β .889, and significant at <.005). Especially the number of theatres is a significant and strong predictor of movie box office revenues of drama movies (β .77, and significant at <.01). In week 2, the adjusted R^2 of Instagram data ranges from .8 to .9. Therefore, it can be said that the predictive value of the model is strongly improved by including additional variables, in particular by including the number of theaters (β .4 and significant at <.005). The negative visual sentiment ratio is also a significant predictor of movie box office revenues in week 2 (β -.3 and significant at <.005). Movies not being part of a sequel slightly decrease the predictive value of the model. In week 2, negative visual sentiment ratio (β -.981, and significant at <.005 was found to be a strong and significantly negative predictor box office revenues of movies with an audience ranging from 6 to 9 years old in week 2. The number of post views (β 1,5, and significant at <.01), post volume (β .8, and significant at <.01) and positive visual sentiment ratio (β .803 ,and significant at <.0)1 were found to be strong and significant predictors of box office revenues for movies with audience between 12 and 16 years. For movies with an audience of 16 years and older, Instagram posts by women (β .8, and significant at <.005), were significant predictors of box office revenues of movies with an audience of 16 years and older. The number of theaters is a significant predictor of box office revenue of drama movies in week 2 (β .7, and significant at <.05).

When predicting box office revenue based on Instagram data from the previous week, post volume is a strong and significant predictor in week 1 for box office revenue in week 2 (β .79, and significant at <.005). The adjusted R^2 ranges between .617 and .863 when using Instagram posts for predicting box office revenues in week 2. When using Instagram data from week 2 for predicting box office revenues in week 3, the number of theatres is a significant predictor (β .416, and significant at <.005). Negative visual sentiment is also significant predictor, but its predictive value is much lower (β -.172 and significant at <.01). The adjusted R^2 ranges between .744 and .855 when using Instagram posts in week 2 for predicting box office revenues in week 3. An overview of the additional analyses can be found in appendix 6.

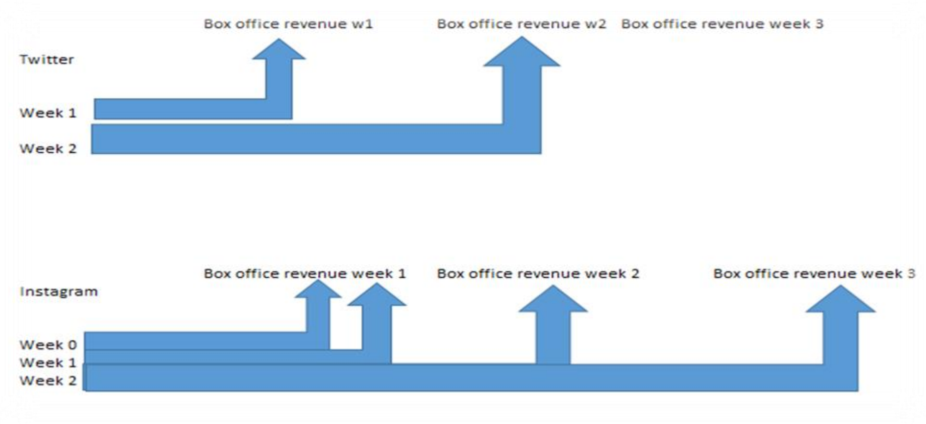
Twitter

When predicting movie box office revenues in week 1 based on Twitter data from the same week, it can be said that the number of post views is a strong and significant predictor (β .814 and significant at <.005). The adjusted R^2 ranges between .650 and .757 when using Twitter posts in week 1 for predicting box office revenues in week 2. Particularly interesting to see is that movie box office revenue in week 1 for drama movie is to a large extent predicted by the number of theaters (β .766 and significant at <.01) and that box office revenues of horror movies is to a large extent predicted by the negative visual sentiment ratio (β .999 and significant at <.01). With regard to the audience ages, post volume is a strong and significant predictor of movies with an audience of 12 to 16 years old (β .962 and significant at <.01), whereas the box office revenue in week 1 of movies with an audience of 6 to 9 years old is to a large extent predicted by the movie budget (β .97 and significant at <.05), whereas the number of theaters is able to predict to a large extent the box office revenue for movies of 16 years and older (β .889 and significant at <.005). This results into a high adjusted R^2 during week 1, ranging from .771 to .922. The box office revenue of drama movies are highly predicted by the number of theaters (β .7 and significant at <.05), followed by positive visual sentiment ratio (β .-543, ns.). When predicting week 2 with Twitter data from the same week, movie budget is the strongest and most significant predictor (β .848 and significant at <.005), followed by post volume (β .219 and significant at <.05). The adjusted R^2 ranges from .719 to .746. Box office revenues of movies with an audience of 6 to 9 years old are strongly and significantly predicted by movie budgets (β. 978 and significant at <.005), whereas movies with an audience of 16 years and older are strongly and significantly predicted by the number of post views (β .861 and significant at <.005) and the number of theaters (β. 562 and significant at <.01).

When predicting the next week based on Twitter data from week 1, movie budget is by far the strongest predictor (β .848 and significant at <.005) of movie box office revenues in week 2, followed by post volume (β. 436 and significant at <.05). This results into a an adjusted R^2 ranging from .708 to .761. Movie budget in week 2 is a strong and significant predictor (β. 845 and significant at <.005) of movie box office revenues in week 3. The adjusted R^2 ranges from .714 to .816. An overview of the additional analyses can be found in appendix 12.

Based on the findings of this chapter, the following model can be drawn for both Twitter and Instagram and their impact on movie box office revenues. The figure shows that Twitter posts in week 1 and week 2 are able to significantly predict the same and the subsequent week. In addition, Instagram posts in week 0, 1 and 2 are able to predict the same week and the subsequent weeks.

Figure 7 - Twitter and Instagram and movie box office revenues



As can be seen from table 11, Instagram data is able to predict box office revenues in the same and the subsequent weeks. In particular after the movie release, Instagram is able to predict box office revenues in the same and the subsequent weeks. On the other hand, Twitter is able to predict box office revenue in the same week (the first and the second week after the release of the movie).

Table 11- Adjusted R-squared for Instagram and Twitter predictions in the same and the subsequent week

Instagram prediction week	Adjusted R-squared	Twitter prediction week	Adjusted R-squared
Week 1 > week 1	.477	Week 1 > week 1	.720
Week 2 > week 2	.727	Week 2 > week 2	.153
Week 0 > week 1	.467	Week 0 > week 1	.304
Week 1 > week 2	.671	Week 1 > week 2	.657
Week 2 > week 3	.663	Week 2 > week 3	.189
Inclusive additional variables			
Week 1 > week 1	.803	Week 1 > week 1	.757
Week 2 > week 2	.902	Week 2 > week 2	.746
Week 0 > week 1	.863	Week 0 > week 1	.730
Week 1 > week 2	.875	Week 1 > week 2	.761
Week 2 > week 3	.855	Week 2 > week 3	.791

5. Analysis

After performing the multiple linear regression analysis in the previous chapter of this master thesis, this chapter presents a summary of the results and compares the results with the findings from the literature. The hypotheses are accepted or rejected based on the results from the data analysis part, in particular the multiple regression analyses. The sub-questions, as formulated in chapter 3 are also answered in this chapter.

First, the results with regard to the formulated hypotheses are discussed. The results with regard to the hypotheses are summarized in table 12.

H1a: The volume of Instagram posts is positively associated with movie gross box office revenue.

In every multiple linear regression analysis, the volume of Instagram posts was found to be a significant predictor of movie box office revenues, both in the same and the subsequent week. In the models in which movie box office revenue in the same week was predicted, the volume of Instagram posts was significant at the <.001 level in week 2 ($\beta = .709$, $p.<0.001$) and 3 ($\beta = .814$, $p.<0.001$). In week 1 ($\beta = .726$, $p.<0.001$), week 2 ($\beta = .814$, $p.<0.001$) and week 3 ($\beta = .763$, $p.<0.001$) post volume on Instagram was also identified as a significant predictor of movie box office revenues in the subsequent weeks. In all cases the relationship between post volume on Instagram and movie box office revenues was positive. These findings are in line with the findings of Asur & Huberman (2010) who were able to predict movie box office revenues with high accuracy by using volume-related variables. More specifically, Baek et al., (2014) find that the total volume of Tweets corresponds strongly with box office revenues. So based on the results, hypothesis 1a is accepted.

H1b: The volume of Twitter posts is positively associated with movie gross box office revenue.

Compared to the volume of Instagram posts, similar results were found for the volume of Twitter posts in relation to movie box office revenues. The volume of Twitter posts is a significant predictor of movie box office revenues both in the same weeks (week 2 ($\beta = .425$, $p.<0.01$) and week 3 ($\beta = .462$, $p.<0.01$) and the subsequent weeks (week 1 ($\beta = .567$, $p.<0.001$), week 2 ($\beta = .281$, $p.<0.05$) and week 3 ($\beta = .486$, $p.<0.01$). The relationship between the volume of Twitter posts and movie box office revenu was always positive. Again, the findings are in line with the findings of Asur & Huberman (2010) who were able to predict movie box office revenues with high accuracy by using volume-related variables. In addition, Liu (2006) and Karniouchina (2011) find that volume has an influence on movie box office revenue. So based on the results, hypothesis 1b is accepted.

H1c: The volume of Instagram post views is positively associated with movie gross box office revenue.

Based on the results of this study, it can be said the volume of Instagram posts views is not positively associated with movie box office revenues. This is in line with findings of Bhavsar et al., (2017) and

Bhave et al., (2015), who found that the number of views on YouTube is a not a strong significant predictor of movie box office revenues. When relating the findings of this research to the literature, it could be said that the number of Instagram post views is not a significant predictor of movie box office revenues, whereas the number of views of a trailer on YouTube and the number of Wikipedia page views are significant predictors of movie box office revenue. Instagram is mainly used for self-disclosure and sharing personal views (de Vries et al., 2018), whereas Wikipedia and YouTube are used for information seeking (Head & Eisenberg, 2009; Phua, Jin, & Kim, 2017). Due to the differences in user motives, it might be expected that the number of post views is a significant predictor of movie box office revenue in case that it is used for information seeking and that it is not a significant predictor of movie box office revenue in case that the social media is used for sociability and affective purposes.

H1d: The volume of Twitter post views is positively associated with movie gross box office revenue.
The multiple linear regression analysis in week 2 shows that the number of post views in week 2 (β = .544, p.<0.001) is the most significant predictor of movie box office revenue. When taking these variable alone in week 2, it shows a β of .812 (p.<0.001) and the number of post views is able to explain 65% of the movie box office revenues in the same weeks, which indicates that the number of post views is indeed a strong significant predictor of movie box office revenue. The number of post views was also able to significantly explain 65,7% of the movie box office revenues in the next week (β = .816, p.<0.001) (adjusted R^2 = .657, p <.001). In all regression models, Twitter post views had a positive relationship with movie box office revenue, but the number of post views was only a significant predictor in week 2. With regard to the previous hypothesis, it can be said that on the one hand Instagram is used for social activity, showing affection and sociability, whereas on the other hand Twitter users pursuing gratifications like information seeking (Lee & Ma, 2012; Pittman & Reich, 2016). The fact that the number of post views on Twitter is a significant predictor in the second week, which is the release week, shows that people collect information about a movie. Java, Song, Finin & Tseng (2007), Pittman & Reich (2016) and Lee & Ma (2012) argue that Twitter is used for information seeking purposes. The results of this research are consistent with the findings of the aforementioned studies. The significant relationships between the number of post views in week 2 and movie box office revenue in the same and the next week would suggest that people would visit directly or in the next week the movie after having collected information about the movie. Because the number of post views on Twitter is only a significant predictor in week 2, hypothesis 1d partially accepted.

H2a: Positive sentiment in Instagram textual posts is positively associated with movie gross box office revenue.

The results of the multiple regression analaysis reveal that positive textual sentiment is not a significant predictor of movie box office revenue in the same week. When looking at the multiple regression analysis, positive sentiment in textual posts on Instagram in week 2 (β = .184, ns, $p < 0.05$) is a significant predictor of movie box office revenue in the next week. The results of this research also show that sentiment alone was not a significant predictor of movie box office revenue, but that it improves the predictive power of the model when combining these variable with volume-related variables. Choi & Thoma (2014) argue that stigmatized or negative information is shared via private modalities, like face to face interaction. In addition, Choi & Toma (2014) suggest that negative information is disclosed more privately than positive information. Since Instagram is an visual rich social media platform and because of that it is based on visual cues, it might not be expected that Instagram is generally used for expressing positive sentiment. H2a is rejected, because positive sentiment in textual Instagram posts was not a significant predictor of movie box office revenue during the weeks.

H2b: Positive sentiment in Twitter textual posts is positively associated with movie gross box office revenue.

The results of the multiple regression analyses show that positive sentiment in Twitter posts is not a significant predictor of movie box office revenue in all weeks. In four out of five cases, inclusion of positive sentiment in Twitter textual posts leads to slight improvements of the predictive value (the adjusted R^2) of the model in predicting the movie box office revenue. This is consistent with the findings of Lassen et al. (2014) who find that the strong correlation between iPhone Tweets and iPhone sales becomes marginally stronger after incorporating sentiment in Tweets. In line with this, Asur & Huberman (2010) find that the addition of sentiment in the linear model only slightly improved the performance of the model. Other researchers like Ceron, Curini, & Iacus (2015), Ceron, Curini, Iacus, & Porro (2014) and Rui et al. (2013) suggest a positive association between positive sentiment in Tweets and positive outcomes. Lipizzi et al. (2016) are more skeptical about the predictive abilities of sentiment alone in predicting box office revenues. When only using sentiment, they find that sentiment was a weak predictor of box-office revenues. Liu et al. (2016) are also skeptical about the relationship between sentiment and movie box office revenues. Consistent with the findings of Lipizzi et al. (2016), the results of this research also show that sentiment alone was not a significant predictor of movie box office revenue, but that it improves the predictive power of the model when combining these variable with volume-related variables. Based on the findings of the research, hypothesis H2b is rejected.

H2c: Negative sentiment in Instagram textual posts is negatively associated movie gross box office revenue.

Negative sentiment in Instagram textual posts was not found to be a significant predictor of negative movie box office revenue. The relationship was very small and not significant in any of the regression models. This relationship might be expected because Instagram is mainly used for exchanging communication in imagery form (Lim & Childs, 2016). Instagram has evolved into a unique social media platform where users can document their lives through visual content such as photos and videos (Bashir et al., 2018). Therefore, it is not surprising that negative sentiment in textual posts on an image-based social media platform, which is non-sensory information, is not found to be a significant predictor of movie box office revenues. Therefore, based on the results of this study and theory, hypothesis H2c is rejected.

H2d: Negative sentiment in Twitter textual posts is negatively associated with movie gross box office revenue.

Based on the results of the multiple regression analyses, negative sentiment in Twitter textual posts in week 0 was only found to be a predictor of movie box office revenues in week 1, but still the strength of the association is very weak (β .2, significant at <.05). From a dual processing theory point of view, visual stimuli evoke imagery information processing, while verbal stimuli induce discursive information processing (Kim & Lennon, 2008). On the other hand, imagery processing is likely to evoke emotional responses more than discursive processing (MacInnis & Price, 1987). Therefore, it might be expected that textual posts will lead less to emotional responses than posts with visual content. Lipizzi et al. (2016) found that sentiment alone was not a significant predictor of movie box office revenue, but that it improves the predictive power of the model when combining these variable with volume-related variables. Based on the results of this study, hypothesis H2d is rejected.

H2e: Positive sentiment in Instagram visual posts is positively associated with movie gross box office revenue.

Positive sentiment in Instagram visual posts is found to be a strong and significant predictor of box office revenues for movies with an audience between 12 and 16 years (β .803, and significant at <.01). Instagram is the social media platform of choice for teenagers and young millennials (Len-Ríos, Hughes, McKee, & Young, 2016). Christofides, Muise, & Desmarais (2012)

and Waterloo et al. (2017) argue that younger people seem to disclose more to peers than older people online. Arceneaux & Dinu (2018) argue that young Instagram users might view other social media users as social peers. The finding that positive sentiment in Instagram visual posts is found to be a strong and significant predictor of box office revenues for movies with an audience between 12 and 16 years is therefore not surprising. Compared to Facebook and Twitter, Waterloo et al. (2017) found that Instagram was most appropriate for expressing positive emotions. Therefore, it seems logically that positive sentiment in Instagram visual posts is found to be a strong and significant predictor of box office revenues for movies with an audience between 12 and 16 years. Nevertheless, based on the findings of this study H2e is rejected.

H2f: Positive sentiment in Twitter visual posts is positively associated with movie gross box office revenue.

A moderate relationship and negative association was found between positive visual sentiment in Twitter posts (β .-543, ns.) and movie box office revenues for drama movies in week 1. Twitter is seen as a popular tool for short and immediate commentary on real-time happenings, including both personal and news events (Kaplan & Haenlein, 2011). Due to these characteristics, researchers like Naveed, Gottron, Kunegis, & Alhadi (2011) and Thelwall, Buckley & Paltoglou (2011) found that Twitter content is mainly negatively valenced even when it concerns positive events. Therefore, it is not surprising that the aforementioned association was found between positive visual sentiment in Twitter and movie box office revenues for drama movies in week 1. In addition, the main modality on Twitter is text and therefore it can be said that Twitter users generally share texts instead of images (Waterloo et al., 2017). Because significant associations and strong associations between positive visual sentiment in Twitter posts and movie box office revenue were generally not identified, hypothesis H2f is rejected.

H2g: Negative sentiment in Instagram visual posts is negatively associated movie gross box office revenue.

Negative sentiment in visual Instagram posts is identified as an significant predictor of movie box office revenues in week 1, 2 and 3. Interestingly, in week 0 the negative visual sentiment ratio was not found to be an significant predictor of movie box office revenues. This suggests that visual sentiment is shared on Instagram after the moment of release. Nevertheless, the association is relatively weak because negative visual sentiment in Instagram posts and movie box office revenues, because the β values are -.316, and -.297 in week 1 and 2. For movies with an audience of 16 years and older the negative visual sentiment ratio is a moderately strong predictor of movie box office revenue (β .-484), while a strong negative and significant association between negative visual sentiment ratio (β -.981, and significant at <.005) and movie box office revenues for movies with an audience ranging between 6 and 9 years was found in week 2. Movie visitors between 6 and 9 years old are typically accompanied by their parents when visiting a movie. Their parents use Instagram and post visual content about the movie. When inspecting the ratio of posts by women and men for these movies, it can be assumed that most posts about movies with an audience between 6 and 9 years comes from women (Vilnai-Yavetz & Tifferet, 2015). Tifferet & Vilnai-yavetz (2014) argue that women express more emotional content on social media and therefore this finding seems logically. When looking at the valence of the review, according to Ahluwalia (2002) and East, Hammond, &

Lomax (2008) negative reviews hurt a movie's box office performance more than positive reviews improve it. Chakravarty, Liu, & Mazumdar (2010) find that negative word of mouth is more influential, which can be explained by the diagnostics of information and prospect theory Hennig-Thurau, Wiertz, & Feldhaus (2015). According to Hennig-Thurau et al. (2015), negative information runs counter to consumers' expectations, such that negative messages have a higher diagnostic value for consumers (Chen, Wang, & Xie, 2011; Marchand, Hennig-Thurau, & Wiertz, 2017). This could explain why negative visual sentiment was identified for a limited amount of cases as a significant and strong predictor of movie box office revenues, whereas positive visual sentiment was not found to be a significant and strong predictor of movie box office revenues. Nevertheless, hypothesis H2g is rejected.

H2h: Negative sentiment in Twitter visual posts is negatively associated movie gross box office revenue.

Negative sentiment in visual Twitter posts was found to be a moderately strong and significant predictor of movie box office revenues of drama movies in week 1 (β .*-543* and significant at <.01). Naveed et al., 2011) and Thelwall et al. (2011) found that Twitter content is mainly negatively valenced even when it concerns positive events. Therefore, it is not surprising that the aforementioned association was found between negative visual sentiment in Twitter and movie box office revenues for drama movies in week 1. Because significant and strong associations between negative visual sentiment on Twitter and movie box office revenues were not found, hypothesis H2g is rejected. All hypotheses and the results are summarized in table 12.

After having discussed the literature, the methodology, pre-processing, reporting and performing data analysis and reporting on the hypothesis, it is now possible to answer the central research question. The central research question is answered in the next chapter.

Table 12 - Hypotheses and results

Hypothesis	***Result***
H1a: The volume of Instagram posts for each movie is positively associated with movie gross box office revenue.	*Accepted*
H1b: The volume of Twitter posts for each movie is positively associated with movie gross box office revenue.	*Accepted*
H1c: The volume of Instagram post views is positively associated with movie gross box office revenue.	*Rejected*
H1d: The volume of Twitter post views is positively associated with movie gross box office revenue.	*Partially accepted*
H2a: Positive sentiment in Instagram textual posts is positively associated with movie gross box office revenue.	*Rejected*
H2b: Positive sentiment in Twitter textual posts is positively associated with movie gross box office revenue.	*Rejected*
H2c: Negative sentiment in Instagram textual posts is negatively associated movie gross box office revenue.	*Rejected*
H2d: Negative sentiment in Twitter textual posts is negatively associated movie gross box office revenue.	*Rejected*
H2e: Positive sentiment in Instagram visual posts is positively associated with movie gross box office revenue.	*Rejected*
H2f: Positive sentiment in Twitter visual posts is positively associated with movie gross box office revenue.	*Rejected*
H2g: Negative sentiment in Instagram visual posts is negatively associated movie gross box office revenue.	*Rejected*
H2h: Negative sentiment in Twitter visual posts is negatively associated movie gross box office revenue.	*Rejected*

6. Conclusion

In this section, the main conclusion of this research is presented. Furthermore, this chapter also provides an answer to the central research question. After that, an discussion with regard to the limitations is presented, together with suggestions for future research.

6.1. Central research question

- The following central research question was drawn for this research: *To what extent can movie box office revenues be predicted by Instagram and Twitter predictor variables?*

The findings of this study show that there is a positive and significant relationship between volume-related predictor variables derived from Instagram and Twitter and movie box office revenues. Therefore, it can be said that movie box office revenues can be predicted during the first weeks of release. On average, the sentiment-related variables were not identified as significant predictors of movie box office revenues. However, including sentiment-predictor variables improves the extent to which the model is able to predict movie box-office revenues. Nevertheless, the findings of the additional analyses reveal that positive visual sentiment in Instagram posts was found to be a strong and significant predictor of movie box office revenues for movies with an audience between 12 and 16 years. This is consistent with the findings of prior research. Prior research found that Instagram is used by young people to express and manage their identities and social relationships in the present and in the future. Additionally, according to prior research younger people seem to disclose more to peers than older people online. Negative visual sentiment in Instagram posts was identified as a significant predictor of movie box office revenues in the release week and the two subsequent weeks. A strong and negative significant association was found between negative visual sentiment in Instagram posts and movie box office revenues for movies with an audience ranging between 6 and years. The additional analysis reveals that most of the posts for this movie categories were posted by women. Since children in the age between 6 and years will visit movies together with adults, in general their parents. These findings reveal that Instagram is used as an tool to share feelings and thoughts with other users through image sharing, like feedback with regard to a particular movie. With regard to Twitter visual sentiment, a moderate and negative association was found between positive visual sentiment in Twitter posts and movie box office revenues for drama movies in the release week. With regard to drama movies, negative visual sentiment in Twitter posts was identified as an significant predictor of movie box office revenues in the release week.

Based on the findings, it can be said that visual sentiment in Instagram posts were under certain circumstances predictors of movie-box office revenues. Especially for movies with younger movie audience ages and movies that are typically watched by women, visual sentiment in Instagram posts was identified as an significant predictor of movie box office revenues. Further, textual sentiment was not identified as an strong and significant predictor of movie box office revenues. In addition, the findings of this research reveal the differences between image-based and textual social media platforms. Compared to textual-based social media platforms, prior research argues image-based social media platforms are more rich, strong rely on visual communication and that self-presentation is perceived to be more realistic. From a media appropriateness theory perspective, the findings also show that image-based social media platforms are used more often for sharing negative information or feedback than textual-based social media platforms. The finding that often negative visual sentiment in Instagram was identified as an significant predictor of movie box office revenues is therefore not surprising, whereas similar effects were not found for Twitter.

What exactly do these results mean for the movie industry? The findings of this research show that Instagram predictor variables are more accurate predictors of movie box office revenue than Twitter predictor variables. In other words, image-based social media platforms are important platforms for the movie industry when predicting movie box office revenues. The results extend the findings of prior research by showing that image-based social media platforms are more influential than textual-based social media platforms. Twitter predictor variables are only able to predict movie box office revenues in the same week and more specifically in the release week and the week after the release, whereas Instagram predictor variables are able to predict movie box office revenues in the same and the subsequent weeks. This consistent with the findings of prior research, because it was found that information collected from more rich media was more easily processed and more easily and longer remembered by users. The findings of this research also reveal that visual sentiment can be used as a strong indicator to predict movie box office revenues for movies with younger adult ages and movies that typically viewed by womens. Based on visual sentiment data, estimations of the number of visitors can be made for the first release weeks of a movie.

The findings of this research show that Instagram is not only valuable in the field of health and for measuring college-student opinions, but that it also delivers valuable results in other contexts such as movie box office revenues. Furthermore, the findings of this research reveal that Instagram can be used for predictive purposes. Currently, research in the field of Instagram and predictive analytics is limited. Therefore, this study also shows the predictive value of Instagram predictor variables in predicting a particular phenomenon of interest, such as movie box office revenues.

6.2. Discussion

In this section a discussion with regard to the limitations of this research is presented. Furthermore, practical and scientific implications are discussed. Finally, avenues for future research are presented.

6.2.1. Limitations

The limitations of the research are discussed in this section. First, it can be said that this study focuses on movie box office revenues, which limits the scope of the research. Specific keywords were used to study the volume-related variables which contain the name of the movie. In order to extend this research to other fields, it is important to study different types of keywords. When studying movie box office revenues, mentioning the name of the movie in a Instagram post and/or a Tweet was enough to be included into the data set, but in other contexts like for example the music industry both Instagram posts and Tweets mentioning the name of the artist and the album should be collected for predictive purposes. Instagram posts and Tweets mentioning actors were not included in this research. Additionally, another factor like the rating on the Hollywood Stock Index was not not analyzed in this study. So therefore it is not known what the impact of this factor is on predicting movie box office revenues and to what extent this factor improves the prediction models. Further, in this research, Instagram and Twitter are used as data sources for predicting movie box office revenues. These social networking services differ in terms of that Instagram is an image-based platform, whereas Twitter is a text-based platform. Taking only one platform out of these categories limits the generalizability of results of this research to other platforms. More image-based and text-based platforms should be studied in order to able to present more generalizable conclusions about the predictive value of image-based and text-based platforms. In addition, the user demographics of Instagram and Twitter are different. This limitation can be seen as an important limitation of social media analytics, nevertheless this also holds for multimodal sentiment analysis. Data used for multimodal sentiment analysis is limited to certain demographics that are more represented on the internet and on a specific social media platform (Soleymani et al., 2017). So, the differences in user demographics should be considered in order to compare the social media platforms.

Coosto assigns sentiments to posts, but it is still unclear whether the sentiment scores represent the sentiment in the textual post. For example, sarcasm checks and context dependency checks were not performed. This could bias the results of this research. Additionally, different sentiment classifiers were not used, which limits the extent to which sentiment was optimally classified. This also limits the extent to which it is possible to state what sentiment classifiers deliver the most accurate results in classifying Instagram and Twitter posts. For example, it is also not possible to identify whether there is are differences in classification accuracy when classifying posts from text-based and image-based platforms with different sentiment classifiers. Additionally, due to time and budget restrictions, it was not possible to test sophisticated sentiment algorithms and it was because of the same reasons not possible to use machine learning algorithms. The sentiment limitations in this research are generally related to issues with regard to construct and content validity, because it is important to identify whether the data covers the underlying construct, to identify the accuracy of classifiers and the extent to which the data represents the right subjects. Additionally, it is also questionable whether social media is only source that is used for determining whether to visit a movie.

Complura was used for assigning sentiment to visual posts. Due to budget restrictions, the Instagram posts and Tweets were manually analyzed. Complura was also used to analyze the visual sentiment of Twitter, which is a text-based platform whereas Instagram is an image-based platform. No analysis is performed on the extent to which posted visual content differs on both types of platforms and the extent to which the results from a platform can be generalized to other platforms when taking into account user characteristics. Also fake accounts spamming on social media might limit the results. It was not possible to detect whether fake accounts or trolls were used and manipulation of data by companies (Couper, 2013). According to Chaturvedi et al., (2018), the explosion of deceptive activities declines the efficiency of sentiment prediction. With regard to multimodal sentiment analysis, this research is restricted by various limitations.In multimodal sentiment analysis, there is a threat of people expressing sentiment for social reasons that are not related to their internal dispositions. Additionally, it also remains unclear whether posts on Instagram and Tweets represent the real opinions of users (Couper, 2013). Building further on this, a person might express like or dislike sentiments to conform with a certain cultural norm or to express and differentiate his or her identity (Soleymani et al., 2017). In addition, based on existing work with regard to multimodal sentiment analysis, it is assumed that people are more likely to express positive or negative opinions. As a result, there is a lack of neutral opinions expressed online in all the reviewed multimodal sentiment analysis studies (Soleymani et al., 2017).

Sample biases are other important limitations of this research. Data was only collected for a period of four weeks and mainly focused on the release period of the movie. This time span limits the possibility to draw conclusions on causation between Instagram and Twitter and movie box office revenues. The goal of this research was to capture whether there was a relationship between Instagram and Twitter predictor variables and movie box office revenues. The results of this research show a relationship between the Instagram and Twitter predictor variables and movie box office revenues, but still it is difficult to prove causation. It could be possible that movie box office revenues result into Instagram posts and Tweets. Because of this, it is not possible to draw conclusion with regard to causality between Instagram posts and Tweets and movie box office revenues. Nevertheless, it should be mentioned that it was not intended to focus on this aspect in this research. Further, multiple tests need to be performed in order to improve the reliability of the findings of this research. Another limitation of this study is that the relationship between volume-related variables and movie box office revenue does not necessarily imply causation. Instagram posts or Twitter posts could lead to movie box office revenue, but movie box office revenue could also lead to more Instagram posts or Twitter posts. However, the volume of Instagram and Twitter posts can be used to predict movie box office revenue.

Another potential limitation of this research is that Coosto collects Instagram and Twitter posts, but there could be the risk that not all posts were publicly available. Because of that, not publicly available posts could not be found and were not included in the data set. Therefore, the impact of privacy issues regarding data collection should be considered. Further, Instagram posts, Tweets and movie box office revenues were collected during the months January, February and March. So the data was collected during a period of 12 weeks. This period limits the extent to which seasonality effects remain and it also remains unclear whether Instagram and Twitter are delivering the same predictive results for movie box office revenues in other times of the year. The sample size and the number of Instagram posts and Tweets of movies also limits the generalizability of the results of this research. In total, a number of 42 movies was included, because for all these movies box office revenues and Instagram posts and Tweets were available during the pre-release week, the first, the second and the third release week. A higher number of movies would definitely increase the generalizability. The number of posts comes close to a total number of 10,000, which is a small number when looking at larger data sets that were downloaded in other studies by using for example programs like R. After 2016, Instagram data could not be downloaded anymore from R due to that Instagram's API is not publicly accessible anymore. Therefore, posts were downloaded from Coosto. This limits the number of posts and it also limits the generalizability of the research. User-profile characteristics were not available for Instagram due to privacy issues and were therefore not included in this research because a comparison between Instagram and Twitter could not be made. This can be seen as another limitation of this study, because user-characteristics were identified as a social media predictor variable.

Microblogging data, for example Instagram and Twitter data, are often difficult to classify due to the lack of contextual information (Chaturvedi et al., 2018). This results into false classification of sentiment, such as classifying negative posts as positive posts (Chaturvedi et al., 2018). Building further on this, another limitation is context dependency. Context dependency refers to the different meaning of words in different contexts (Chaturvedi et al., 2018). Furthermore, a measurement checking for changes in the opinion of people about a movie over time is also not included in this research, which limits the results of this study.

6.2.2. Research implications

More and more people possess smart phones and other advances devices provided with technologies to produce and collect visual content. In the next coming years, it is expected that visual content will increase in number and that image-based social media platforms and applications will grow in number and use. Therefore, it is important to gather more insight in the predictive possibilities of visual sentiment. This research adds to the current knowledge with regard to whether an image-based social media platform like Instagram is different in its predictive ability than an text-based social media platform like Twitter. There remain enormous differences between for example the motives of using the social media platforms and the expressions of users. This research adds to the current understanding of the predictive value of Instagram, which is at the moment only ill-defined. The ongoing increasing number of Instagram users and the high influence during purchasing decisions of Instagram on users further demands for research with regard to the predictive value of Instagram. This research shows the potential of Instagram as an source for successfully predicting a phenomenon of interest. In the next coming years, not only the use of Instagram will increase, but also other image-based platforms will increase in use. Applications like Snapchat are popular, but it might be expected that the number of applications with visual possibilities will increase in the coming years.

But what do these results mean for the movie industry? First, it reveals the predictive potential of Twitter and especially Instagram in predicting movie box office revenues. More and more the internet is used for information seeking about movies and especially social media applications are used because information can be easily and quickly gathered. So therefore it is important to know where the box office revenues come from. This helps the movie industry to better target potential visitors, to better distribute marketing budgets over different (social media) channels and to invest more in pre-release promotional activities if it seems that a movie might underperform due to for example unfamiliarity by the crowd. Furthermore, this research is also interesting in terms of logistics and planning, because if a cinema knows what high or low values on a certain variable mean, this would enable them to accurately plan personnel, assigning rooms to the movie which fit with the number of visitors, the purchase of snacks and drinks, and other logistic issues. Besides for the movie distributors, it is also important for the cinemas to know where to spend their advertising budget on in order to attract movie visitors. Especially when people could be potentially interested in a movie, but are still in doubt whether to visit the movie. Furthermore, it is also important to know for investors and movie producers how comparable movies behave on social media with regard to post volumes and sentiment, because this helps them to better decide upon investing in for example new movie productions. Furthermore, they can also compare whether they are satisfied with the output in terms of movie box office revenue levels when investing certain sums of money on social media advertising or that the investment should be increased in order to gain higher movie box office revenues.

In addition, this research also contributes to the applicability of the Complura tool for predictive purposes. The Complura tool shows that it is able to produce visual sentiment scores suitable for research in different settings. Further, the generalizability of research towards other disciplines of social media is also important. This research builds on previous research with regard to volume-related and sentiment-related variables predicting movie box office revenues. This study found similar relationships between Twitter and movie box office revenues and extends the research field with knowledge with regard to Instagram. Particularly valuable are the results that in some cases visual sentiment is a significant and strong predictor. The findings revealed that audience age and genre matter regarding visual sentiment. This research could be extended to both text-based and/or image-based driven social media networking sites. Important issues that should be tackled are privacy issues regarding data for the research, which will come more in play due to the actual debates about data privacy. Nevertheless, it is interesting to find out whether the same results with regard to volume-related and sentiment-related variables can be observed on other platforms and whether different features on social media networking sites impact the outcomes.

Additionally, concepts of social media strategies are currently explored (Spil, Effing, & Both, 2016). A well-defined social media strategies consists of various elements. One of these elements is monitoring, but unless its importance companies seek to monitor until the later stages of implementing a social media strategy. This results into a lack with regard to social media monitoring. Hopefully the insights of this research could show social media managers that monitoring social media volume could provide essential real-time updates and that social media monitoring can be used as a powerful tool for prediction. Prediction models based on Instagram and Twitter empowers social media managers to monitor (the progress with regard to reaching) targets.

Possibly the most interesting field for predictive social media research will be the development of predictive models that are able to process real-time data from different social media network sites. Technological developments and advances in machine learning enables us to reach high degrees of prediction accuracy. Prediction algorithms can be improved in order to more accurately predict a phenomenon of interest. Also with regard to Big Data as an upcoming trend, this research could be seen as first steps in order to improve the predictive value out of large data sets. Optimized algorithms could enable better analysis of Big Data, which extends the possibilities with regard to output. Fine tuning specific features and parameters in machine learning algorithms will be an challenge in the future.

Also in other settings the findings of this research are relevant. Besides for companies, this research is also relevant for other fields like for example politics. A first step in understanding the predictive value of Instagram with regard to politics has been made by Schmidbauer et al. (2018), however, research in this field is quite limited. This research extends the current knowledge with regard to Instagram and Twitter by showing the predictive impact of volume-related and sentiment-related variables on a phenomenon of interest. More specifically, identifying significant variables as predictors could be relevant for politicians to improve their campaign by for example knowing when and how to invest their advertising budgets in order to reach the optimal effect and on which social media networking should be focused in order to influence public opinion in a beneficial way. Further, it also helps politicians to monitor how they are performing and to identify popular and/or relevant topics discussed by the crowd. This helps the politician to better understand what should be done and discussed in debates in order to receive support from citizens.

6.2.3. Future research

The findings of this research could be extended by future research. First, causality between the Instagram and Twitter variables and movie box office revenue should be analyze over a longer period of time. A panel data approach allowing researchers to test for Granger causality between variables needs to be employed in the future. Granger causality also makes it possible to identify causality and to analyze whether relationships between the social media predictor variables are different for weeks further away from the release week. This would help to answer questions like is sentiment more influential during the hype phase or is it of more influence during the phase(s) after the hype?

Further research should also deal with the influence of user-profile characteristics on Instagram. In this research, due to privacy issues user-profile characteristics were not available for Instagram. Therefore, they were not included in this research. Future research should study the impact of user-profile characteristics in predicting movie box office revenues. In addition, other factors that were identified as relevant factors should be included in future research, like for example the Hollywood Stock Index, the number of theatres in which a movie have been shown, the movie budget and the genre. With regard to social media research, more text-based and image-based social media platforms should be incorporated in future research. This would extend the current knowledge with regard to predictive value of these social media platforms and the impact of different features regarding predictions. When available, larger data sets should be downloaded from sources like R. Future research should also focus on the impact and the accuracy of sentiment classifiers. More specifically, the classification accuracy of different sentiment classifiers by comparing different social media platforms should be revealed in future research. For example, the question of which sentiment to use on what social media platform needs to be answered. Additionally, future research should also perform analysis on social media platforms by using sophisticated sentiment algorithms. This could for example be done for movie box office revenues in order to identify whether these algorithms were better able to predict movie box office revenues.

Another important direction for future research is collecting data for a larger time period and during different seasons. This approach enables to identify potential season effects and delivers more generalizable results. In addition, multiple tests should be performed in future research because this would generalize the research findings and results into a more complete overview of how movie box office revenue can be predicted. Larger sample sizes with regard to the number of posts and movies would also increase the results. Therefore, future research should work with larger sample sizes.

Further research should study the predictive value of Instagram in other research fields. Especially, it should be identified whether including different keywords categories, like for example posts mentioning artists, actors or albums, influence the predictive value of the models. Additionally, Complura should be used in other contexts for analyzing visual data on different social media platforms. Future research increase the understanding of the classifying accuracy of Complura for movie box office revenue and for other phenomena of interest, but also between different social media platforms. Furthermore, future research should also focus more on the impact of genres on visual sentiment scores. Future research should detect whether movie genres like horror and drama are more strongly producing negative sentiments, rather than movie genres like comedy and slapsticks. Future research should also improve tools for detecting and ruling out spam, fake and troll account and tools that are able to detect whether real opinions were expressed on social media, like for example whether there is a difference between Instagram and Twitter users in expressing real opinions or showing faked behavior.

Finally, another interesting avenue for future researe is data driven decision making. Every day, data increases both in terms of use and data production. Therefore, future research should study the impact of data driven based decisions. The insights of this study from the social media platforms Instagram and Twitter can be used as input for more detailed research with regard to the impact of data in decision making in both the private and public sector. More specifically, the impact of data with regard to skills, competences and behavior in terms of decision making is a topic that demands for further research. Additionally, more insight with regard to the implications regarding personal skills of managers is particulary relevant.

References

Abed, S. (2018). An empirical examination of Instagram as an s-commerce channel. *Journal of Advances in Management Research*, JAMR-05-2017-0057. https://doi.org/10.1108/JAMR-05-2017-0057

Ahluwalia, R. (2002). How Prevalent Is the Negativity Effect in Consumer Environments? *Journal of Consumer Research*, *29*(2), 270–279. https://doi.org/10.1086/341576

Ahsan, U., De Choudhury, M., & Essa, I. (2017). Towards using visual attributes to infer image sentiment of social events. *Proceedings of the International Joint Conference on Neural Networks*, *2017–May*(April 2016), 1372–1379. https://doi.org/10.1109/IJCNN.2017.7966013

Anandarajan, M., Zaman, M., Dai, Q., & Arinze, B. (2010). Generation Y adoption of instant messaging: An examination of the impact of social usefulness and media richness on use richness. *IEEE Transactions on Professional Communication*, *53*(2), 132–143. https://doi.org/10.1109/TPC.2010.2046082

Andsager, J. L., Bemker, V., Choi, H., & Torwel, V. (2006). Perceived Similarity of Effects on Message Evaluation. *Communication Research*, *33*(1), 3–18. https://doi.org/10.1177/0093650205283099

Apala, K. R., Jose, M., Motnam, S., Chan, C.-C., Liszka, K. J., & De Gregorio, F. (2013). Prediction of movies box office performance using social media. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013*, (July), 1209–1214. https://doi.org/10.1145/2492517.2500232

Araujo, C. S., Correa, L. P. D., Silva, A. P. C. Da, Prates, R. O., & Meira, W. (2014). It is not just a picture: Revealing some user practices in instagram. *Proceedings - 9th Latin American Web Congress, LA-WEB 2014*, (May), 19–23. https://doi.org/10.1109/LAWeb.2014.12

Arceneaux, P. C., & Dinu, L. F. (2018). The social mediated age of information: Twitter and Instagram as tools for information dissemination in higher education. *New Media & Society*, 146144481876825. https://doi.org/10.1177/1461444818768259

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*, 492–499. https://doi.org/10.1109/WI-IAT.2010.63

Ausserhofer, J., & Maireder, A. (2013). National politics on Twitter: Structures and topics of a networked public sphere. *Information Communication and Society*, *16*(3), 291–314. https://doi.org/10.1080/1369118X.2012.756050

Baek, H., Ahn, J., & Oh, S. (2014). Impact of tweets on box office revenue: Focusing on when tweets are written. *ETRI Journal*, *36*(4), 581–590. https://doi.org/10.4218/etrij.14.0113.0732

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. *Practice Report Yahoo*, 965–974. https://doi.org/10.1145/2556288.2557403

Bashir, A., Wen, J. T., Kim, E., & Morris, J. D. (2018). The Role of Consumer Affect on Visual Social Networking Sites: How Consumers Build Brand Relationships. *Journal of Current Issues and Research in Advertising*, *1734*, 1–14. https://doi.org/10.1080/10641734.2018.1428250

Bayer, J. B., Ellison, N. B., Schoenebeck, S. Y., & Falk, E. B. (2016). Sharing the small moments: ephemeral social interaction on Snapchat. *Information Communication and Society*, *19*(7), 956–977. https://doi.org/10.1080/1369118X.2015.1084349

Bhattacharjee, B., Sridhar, A. and, & Dutta, A. (2017). Identifying the causal relationship between social media content of a Bollywood movie and its box-office success - a text mining approach. *International Journal of Business Information Systems*, *24*(3), 344–368. https://doi.org/10.1504/IJBIS.2017.082039

Bhave, A., Kulkarni, H., Biramane, V., & Kosamkar, P. (2015). Role of different factors in predicting movie success. *2015 International Conference on Pervasive Computing (ICPC)*, *0*(c), 1–4. https://doi.org/10.1109/PERVASIVE.2015.7087152

Bhavsar, P., Kumar, A., Kumar, S., Gaur, A., & Sheikh, S. G. (2017). Outcome and Prediction of Popularity of Motion Picture Using Social Media, 3245–3248.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Borth, D., Ji, R., Chen, T., Breuel, T., & Chang, S.-F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. *Proceedings of the 21st ACM International Conference on Multimedia - MM '13*, 223–232. https://doi.org/10.1145/2502081.2502282

Bowman, & Hodges. (1999). Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *Journal of Visual Languages & Computing*, *10*(1), 37–53. https://doi.org/10.1006/jvlc.1998.0111

Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter Boyd, D., Golder, S., & Lotan, G. (2010, January). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. *43rd Hawaii International Conference on*, 1–10. https://doi.org/10.1109/HICSS.2010.412

Broekhuizen, T. L. J., Delre, S. A., & Torres, A. (2011). Simulating the cinema market: How cross-cultural differences in social influence explain box office distributions. *Journal of Product Innovation Management*, *28*(2), 204–217. https://doi.org/10.1111/j.1540-5885.2011.00792.x

Brubaker, P. J., & Wilson, C. (2018). Let's give them something to talk about: Global brands' use of visual content to drive engagement and build relationships. *Public Relations Review*, (June 2017), 0–1. https://doi.org/10.1016/j.pubrev.2018.04.010

Brynjolfsson, E., Geva, T., & Reichman, S. (2016). Crowd-Squared : Amplifying the Predictive Power of Large-Scale Crowd-Based Data. *MIS Quarterly*, *40*(4), 1–41. https://doi.org/10.2139/ssrn.2513559

Cai, Z., Cao, D., Lin, D., & Ji, R. (2016). A spatial-temporal visual mid-level ontology for GIF sentiment analysis. *2016 IEEE Congress on Evolutionary Computation, CEC 2016*, 4860–4865. https://doi.org/10.1109/CEC.2016.7744413

Cai, & Xia. (2015). Convolutional Neural Networks for Multimedia Sentiment Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9362*, 159–167. https://doi.org/10.1007/978-3-319-25207-0

Caldwell, B. S., Uang, S. T., & Taha, L. H. (1995). Appropriateness of communications media use in organizations: Situation requirements and media characteristics. *Behaviour and Information Technology*, *14*(4), 199–207. https://doi.org/10.1080/01449299508914633

Campos, V., Salvador, A., Jou, B., & Giró-i-Nieto, X. (2015). Diving Deep into Sentiment: Understanding Fine-tuned CNNs for Visual Sentiment Prediction. https://doi.org/10.1145/2813524.2813530

Carah, N., & Shaul, M. (2016). Brands and Instagram: Point, tap, swipe, glance. *Mobile Media and Communication*, *4*(1), 69–84. https://doi.org/10.1177/2050157915598180

Carlson, J. R., & Zmud, R. W. (1999). Channel Expansion Theory and the Experiential Nature of Media Richness Perceptions. *Academy of Management Journal*, *42*(2), 153–170. https://doi.org/10.2307/257090

Ceron, A., Curini, L., & Iacus, S. M. (2015). Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy. *Social Science Computer Review*, *33*(1), 3–20. https://doi.org/10.1177/0894439314521983

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media and Society*, *16*(2), 340–358. https://doi.org/10.1177/1461444813480466

Chae, J. (2018). Reexamining the relationship between social media and happiness: The effects of various social media platforms on reconceptualized happiness. *Telematics and Informatics*, (April), 0–1. https://doi.org/10.1016/j.tele.2018.04.011

Chaiken, S. (1980). Heuristic Versus Systemic Information Processing and the Use of Source Versus Message Clues in Persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037//0022-3514.39.5.752

Chaiken, S., & Eagly, A. H. (1976). Communication modality as a determinant of message persuasiveness and message comprehensibility. *Journal of Personality and Social Psychology*, *34*(4), 605–614. https://doi.org/10.1037/0022-3514.34.4.605

Chakravarty, A., Liu, Y., & Mazumdar, T. (2010). The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation. *Journal of Interactive Marketing*, *24*(3), 185–197. https://doi.org/10.1016/j.intmar.2010.04.001

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, *44*(December 2017), 65–77. https://doi.org/10.1016/j.inffus.2017.12.006

Chen. (2017). Can online social networks foster young adults' civic engagement? *Telematics and Informatics*, *34*(5), 487–497. https://doi.org/10.1016/j.tele.2016.09.013

Chen, F., Gao, Y., Cao, D., & Ji, R. (2015). Multimodal hypergraph learning for microblog sentiment prediction. *Proceedings - IEEE International Conference on Multimedia and Expo*, *2015–Augus*. https://doi.org/10.1109/ICME.2015.7177477

Chen, F., Ji, R., Su, J., Cao, D., & Gao, Y. (2018). Predicting Microblog Sentiments via Weakly Supervised Multimodal Deep Learning. *IEEE Transactions on Multimedia*, *20*(4), 997–1007. https://doi.org/10.1109/TMM.2017.2757769

Chen, Borth, Darrell, & Chang. (2014). DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks. Retrieved from http://arxiv.org/abs/1410.8586

Chen, & Chang. (2018). What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2018.03.019

Chen, & Kim. (2013). Problematic Use of Social Network Sites: The Interactive Relationship Between Gratifications Sought and Privacy Concerns. *Cyberpsychology, Behavior, and Social Networking*, *16*(11), 806–812. https://doi.org/10.1089/cyber.2011.0608

Chen, Lu, & Wang. (2017). Customers' purchase decision-making process in social commerce: A social learning perspective. *International Journal of Information Management*, *37*(6), 627–638. https://doi.org/10.1016/j.ijinfomgt.2017.05.001

Chen, Wang, & Xie. (2011). Online Social Interactions: A Natural Experiment on Word of Mouth versus Observational Learning. *Journal of Marketing Research*, *XLVIII*(April), 238–254. https://doi.org/10.1509/jmkr.48.2.238

Cheung, C. M. K., & Lee, M. K. O. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision Support Systems*, *53*(1), 218–225. https://doi.org/10.1016/j.dss.2012.01.015

Cheung, C. M. K., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, *54*(1), 461–470. https://doi.org/10.1016/j.dss.2012.06.008

Cheung, Luo, Sia, & Chen. (2009). Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations. *International Journal of Electronic Commerce*, *13*(4), 9–38. https://doi.org/10.2753/JEC1086-4415130402

Chiang, J. K. H., & Suen, H. Y. (2015). Self-presentation and hiring recommendations in online communities: Lessons from LinkedIn. *Computers in Human Behavior*, *48*, 516–524. https://doi.org/10.1016/j.chb.2015.02.017

Childers, T. L., & Houston, M. J. (1984). Conditions for a Picture-Superiority Effect on Consumer Memory. *Journal of Consumer Research*, *11*(2), 643. https://doi.org/10.1086/209001

Choi, M., & Toma, C. L. (2014). Social sharing through interpersonal media: Patterns and effects on emotional well-being. *Computers in Human Behavior*, *36*, 530–541. https://doi.org/10.1016/j.chb.2014.04.026

Christofides, E., Muise, A., & Desmarais, S. (2012). Hey mom, what's on your facebook? comparing facebook disclosure and privacy in adolescents and adults. *Social Psychological and Personality Science*, *3*(1), 48–54. https://doi.org/10.1177/1948550611408619

Chua, T. H. H., & Chang, L. (2016). Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Computers in Human Behavior*, *55*, 190–197. https://doi.org/10.1016/j.chb.2015.09.011

Coelho, R. L. F., Oliveira, D. S. de, & Almeida, M. I. S. de. (2016). Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics. *Online Information Review*, *40*(4), 458–471. https://doi.org/10.1108/OIR-06-2015-0176

Couper, M. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods*, *7*(3), 145–156. https://doi.org/10.18148/srm/2013.v7i3.5751

Daft, R. L., & Lengel, R. H. (1984). Information Richness: A New Approach to Managerial Behavior and Organization Design. *Research In Organizational Behavior*, *6*, 191–233. https://doi.org/N00014-83-C-0025

Daft, & Lengel. (1986). Organizational information requirements, media richness and structural design. *Management Science*, (May 1986), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Dalmia, V., Liu, H., & Chang, S.-F. (2016). Columbia MVSO Image Sentiment Dataset, 2–5. Retrieved from http://arxiv.org/abs/1611.04455

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y. A., Gelbukh, A., & Zhou, Q. (2016). Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. *Cognitive Computation*, *8*(4), 757–771. https://doi.org/10.1007/s12559-016-9415-7

De Choudhury, M., Counts, S., Horvitz, E. J., & Hoff, A. (2014). Characterizing and predicting postpartum depression from shared facebook data. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 626–638. https://doi.org/10.1145/2531602.2531675

De Choudhury, M., & Gamon, M. (2013). Predicting Depression via Social Media. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, *2*, 128–137. https://doi.org/10.1109/IRI.2012.6302998

de Vries, D. A., Möller, A. M., Wieringa, M. S., Eigenraam, A. W., & Hamelink, K. (2018). Social Comparison as the Thief of Joy: Emotional Consequences of Viewing Strangers' Instagram Posts. *Media Psychology*, *21*(2), 222–245. https://doi.org/10.1080/15213269.2016.1267647

Degeratu, A. M., Rangaswamy, A., & Wu, J. (2000). Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing*, *17*(1), 55–78. https://doi.org/10.1016/S0167-8116(00)00005-7

Dellarocas, C., Zhang, X., & 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

Demangeot, C., & Broderick, A. J. (2010). Consumer Perceptions of Online Shopping Environments. *Psychology & Marketing*, *30*(6), 461–469. https://doi.org/10.1002/mar

Deng, Z., Lu, Y., Wei, K. K., & Zhang, J. (2010). Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in China. *International Journal of Information Management*, *30*(4), 289–300. https://doi.org/10.1016/j.ijinfomgt.2009.10.001

Derlega, V. J., & Chaikin, A. L. (1977). Privacy and Self-Disclosure in Social Relationships. *Journal of Social Issues*, *33*(3), 102–115. https://doi.org/10.1111/j.1540-4560.1977.tb01885.x

DeWall, Buffardi, Bonser, & Campbell. (2011). Narcissism and implicit attention seeking: Evidence from linguistic analyses of social networking and online presentation. *Personality and Individual Differences*, *51*(1), 57–62. https://doi.org/10.1016/j.paid.2011.03.011

Dibble, J. L. (2014). Breaking Good and Bad News: Face-Implicating Concerns as Mediating the Relationship Between News Valence and Hesitation to Share the News. *Communication Studies*, *65*(3), 223–243. https://doi.org/10.1080/10510974.2013.811431

Ding, C., Cheng, H. K., Duan, Y., & Jin, Y. (2017). The power of the “like” button: The impact of social media on box office. *Decision Support Systems*, *94*, 77–84. https://doi.org/10.1016/j.dss.2016.11.002

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*, 12–28. https://doi.org/10.1016/j.intmar.2016.10.004

Djafarova, E., & Trofimenko, O. (2018). “Instafamous” – credibility and self-presentation of micro-celebrities on social media. *Information Communication and Society*, *0*(0), 1–15. https://doi.org/10.1080/1369118X.2018.1438491

East, R., Hammond, K., & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International Journal of Research in Marketing*, *25*(3), 215–224. https://doi.org/10.1016/j.ijresmar.2008.04.001

Eden, J., & Veksler, A. E. (2016). Relational Maintenance in the Digital Age: Implicit Rules and Multiple Modalities. *Communication Quarterly*, *64*(2), 119–144. https://doi.org/10.1080/01463373.2015.1103279

Ekman, P. (1993). Facial expression and emotion. *The American Psychologist*, *48*(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

Ellison, N. B., Vitak, J., Gray, R., & Lampe, C. (2014). Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication*, *19*(4), 855–870. https://doi.org/10.1111/jcc4.12078

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2017). Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, 16555151769829. https://doi.org/10.1177/0165551517698298

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of the 5th Conference on Language Resources and Evaluation*, 417–422. https://doi.org/10.1.1.61.7217

Fan, Y.-W., Miao, Y.-F., Fang, Y.-H., & Lin, R.-Y. (2013). Establishing the Adoption of Electronic Word-of-Mouth through Consumers' Perceived Credibility. *International Business Research*, *6*(3), 58–65. https://doi.org/10.5539/ibr.v6n3p58

Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: THE impact of Facebook on young women's body image concerns and mood. *Body Image*, *13*, 38–45. https://doi.org/10.1016/j.bodyim.2014.12.002

Fernández, D., Campos, V., Jou, B., Giró-I-Nieto, X., & Chang, S.-F. (2016). Is a ''happy dog'' image more ''happy'' than ''dog''? Analyzing Adjective and Noun Visual Contributions, 2016. Retrieved from https://imatge.upc.edu/web/sites/default/files/pub/xFernandeza.pdf

Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media. *PLoS ONE*, *10*(11), 1–10. https://doi.org/10.1371/journal.pone.0142390

Filieri, R., & McLeay, F. (2014). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal of Travel Research*, *53*(1), 44–57. https://doi.org/10.1177/0047287513481274

Flaes, B., Rudinac, & Worring. (2016). What Multimedia Sentiment Analysis Says About City Liveability. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9626*, 824–829. https://doi.org/10.1007/978-3-319-30671-1

Franklin, D. R. J., & Mewhort, D. J. K. (2015). Memory as a hologram: An analysis of learning and recall. *Canadian Journal of Experimental Psychology*, *69*(1), 115–135. https://doi.org/10.1037/cep0000035

Gaikar, D. D., Marakarkandy, B., & Dasgupta, C. (2015). Using Twitter data to predict the performance of Bollywood movies. *Industrial Management & Data Systems*, *115*(9), 1604–1621. https://doi.org/10.1108/IMDS-04-2015-0145

Geise, S., & Baden, C. (2015). Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory*, *25*(1), 46–69. https://doi.org/10.1111/comt.12048

Gelli, F., Uricchio, T., Bertini, M., Del Bimbo, A., & Chang, S.-F. (2015). Image Popularity Prediction in Social Media Using Sentiment and Context Features. *Proceedings of the 23rd ACM International Conference on Multimedia - MM '15*, 907–910. https://doi.org/10.1145/2733373.2806361

Geurin, A. N., & Burch, L. M. (2017). User-generated branding via social media: An examination of six running brands. *Sport Management Review*, *20*(3), 273–284. https://doi.org/10.1016/j.smr.2016.09.001

Giancristofaro, G. T., & Panangadan, A. (2016). Predicting Sentiment toward Transportation in Social Media using Visual and Textual Features. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2113–2118. https://doi.org/10.1109/ITSC.2016.7795898

Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2014). #Funeral and Instagram: death, social media, and platform vernacular. *Information, Communication & Society*, *0*(June). https://doi.org/10.1080/1369118X.2014.987152

Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R., & Singh, R. (2016). Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. *Journal of Business Research*, *69*(12), 5833–5841. https://doi.org/10.1016/j.jbusres.2016.04.181

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17486–90. https://doi.org/10.1073/pnas.1005962107

Goh, D. H. L., Ang, R. P., Chua, A. Y. K., & Lee, C. S. (2009). Why we share: A study of motivations for mobile media sharing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5820 LNCS*, 195–206. https://doi.org/10.1007/978-3-642-04875-3_23

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *Journal of Experimental Education*, *74*(3), 249–266. https://doi.org/10.3200/JEXE.74.3.249-266

Graesser, L., Gupta, A., Sharma, L., & Bakhturina, E. (2017). Sentiment Classification using Images and Label Embeddings. Retrieved from http://arxiv.org/abs/1712.00725

Guidry, J. P. D., Austin, L. L., Carlyle, K. E., Freberg, K., Cacciatore, M., Jin, Y., & Messner, M. (2018). Welcome or Not: Comparing #Refugee Posts on Instagram and Pinterest. *American Behavioral Scientist*. https://doi.org/10.1177/0002764218760369

Gygli, M., & Soleymani, M. (2016). Analyzing and Predicting GIF Interestingness. *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, 122–126. https://doi.org/10.1145/2964284.2967195

Ha, Y.-I., Kwon, S., Cha, M., & Joo, J. (2017). Fashion Conversation Data on Instagram. Retrieved from http://arxiv.org/abs/1704.04137

Hair. (2007). Knowledge creation in marketing: the role of predictive analytics. *European Business Review*, *19*(4), 303–315. https://doi.org/10.1108/09555340710760134

Hair, Black, Babin, & Anderson. (2010). *Multivariate Data Analysis*. Pearson Education. https://doi.org/doi.org/10.1016/j.ijpharm.2011.02.019

Han, K., Lee, S., Jang, J. Y., Jung, Y., & Lee, D. (2016). " Teens are from Mars , Adults are f rom Venus " : Analyzing and Predicting Age Groups with Behavioral Characteristics in Instagram. *ACM Web Science 2016*, 35–44. https://doi.org/10.1145/2908131.2908160

Harris, R. B., & Paradice, D. (2007). An Investigation of the Computer-mediated Communication of Emotions. *Journal of Applied Sciences Research*, *3*(12), 2081–2090.

Hatfield, Cacioppo, & Rapson. (1993). Emotional Contagion. *Current Directions in Psychological Science*, *2*(3), 96–99.

Head, & Eisenberg. (2009). What Today ’ s College Students Say about Conducting Research in the Digital Age. *Project Information Literacy Project Report*, (February), 1–18.

Hennig-Thurau, T., Marchand, A., & Hiller, B. (2012). The relationship between reviewer judgments and motion picture success: Re-analysis and extension. *Journal of Cultural Economics*, *36*(3), 249–283. https://doi.org/10.1007/s10824-012-9172-8

Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does Twitter matter? The impact of microblogging word of mouth on consumers’ adoption of new movies. *Journal of the Academy of Marketing Science*, *43*(3), 375–394. https://doi.org/10.1007/s11747-014-0388-3

Ho, J. Y. C., Dhar, T., & Weinberg, C. B. (2009). Playoff payoff: Super Bowl advertising for movies. *International Journal of Research in Marketing*, *26*(3), 168–179. https://doi.org/10.1016/j.ijresmar.2009.06.001

Houston, M. B., Kupfer, A. K., Hennig-Thurau, T., & Spann, M. (2018). Pre-release consumer buzz. *Journal of the Academy of Marketing Science*, 1–23. https://doi.org/10.1007/s11747-017-0572-3

Hovick, S. R. A., Meyers, R. A., & Timmerman, C. E. (2003). E-mail communication in workplace romantic relationships. *Communication Studies*, *54*(4), 468–482. https://doi.org/10.1080/10510970309363304

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram : a first analysis of Instagram photo content and user types. *Proceedings of the Eight International AAAI Conference on Weblogs and Social Media*, (McCune 2011), 595–598.

Huber, B., Mcduff, D., Brockett, C., Galley, M., & Dolan, B. (2018). Emotional Dialogue Generation using Image-Grounded Language Models. https://doi.org/10.1145/3173574.3173851

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior*, *28*(2), 561–569. https://doi.org/10.1016/j.chb.2011.11.001

Islam, J., & Zhang, Y. (2016). Visual Sentiment Analysis for Social Images Using Transfer Learning Approach. *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, 124–130. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.29

Iyer, A., Webster, J., Hornsey, M. J., & Vanman, E. J. (2014). Understanding the power of the picture: The effect of image content on emotional and political responses to terrorism. *Journal of Applied Social Psychology*, *44*(7), 511–521. https://doi.org/10.1111/jasp.12243

Jain, V. (2013). Prediction of Movie Success using Sentiment Analysis of Tweets. *International Conference on Soft Computing and Software Engineering 2013*, *3*(3), 308–313. https://doi.org/10.7321/jscse.v3.n3.46

Jang, J. Y., Han, K., Shih, P. C., & Lee, D. (2015). Generation Like: Comparative Characteristics in Instagram. *Acm Chi*, (October), 4039–4042. https://doi.org/10.1145/2702123.2702555

Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & Argyle, T. (2014). Tracking suicide risk factors through Twitter in the US. *Crisis*, *35*(1), 51–59. https://doi.org/10.1027/0227-5910/a000234

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Advances in Web Mining and Web Usage Analysis. In *Joint 9th WEBKDD and 1st SNA-KDDWorkshop ’07 , August 12, 2007 , San Jose, California , USA . Copyright*. https://doi.org/10.1007/978-3-642-00528-2

Jeong, D. C., & Lee, J. (2017). Snap back to reality: Examining the cognitive mechanisms underlying Snapchat. *Computers in Human Behavior*, *77*, 274–281. https://doi.org/10.1016/j.chb.2017.09.008

Ji, R., Cao, D., Zhou, Y., & Chen, F. (2016). Survey of visual sentiment prediction for social media analysis. *Frontiers of Computer Science*, *10*(4), 602–611. https://doi.org/10.1007/s11704-016-5453-2

Jin, X., Gallagher, A., Cao, L., Luo, J., & Han, J. (2010). The wisdom of social multimedia: using Flickr for prediction and forecast. *Proceedings of the International Conference on Multimedia*, 1235–1244. https://doi.org/10.1145/1873951.1874196

Jingwen, W., Fu, J., Xu, Y., & Mei, T. (2016). Beyond object recognition: Visual sentiment analysis with deep coupled adjective and noun neural networks. *IJCAI International Joint Conference on Artificial Intelligence*, *2016–Janua*, 3484–3490.

Johnson, & Knobloch-Westerwick, &. (2017). When misery avoids company: Selective social comparisons to photographic online. *Human Communication Research*, *43*, 54–75.

Jones, D. C. (2001). Social comparison and body image : Attractiveness comparisons to ... *Sex Roles*, *45*(November 2001), 645–664.

Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M., & Chang, S.-F. (2015). Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology, (1). https://doi.org/10.1145/2733373.2806246

Jungherr, A. (2013). Tweets and votes, a special relationship. *Proceedings of the 2nd Workshop on Politics, Elections and Data - PLEAD '13*, 5–14. https://doi.org/10.1145/2508436.2508437

Kahai, S. S., & Cooper, R. B. (2003). Exploring the Core Concepts of Media Richness Theory: The Impact of Cue Multiplicity and Feedback Immediacy on Decision Quality. *Journal of Management Information Systems*, *20*(1), 263–299. https://doi.org/10.1080/07421222.2003.11045754

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. *Internet Research*, *23*(5), 544–559. https://doi.org/10.1108/IntR-06-2012-0114

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, *54*(2), 105–113. https://doi.org/10.1016/j.bushor.2010.09.004

Karabulut, Y. (2012). Can Facebook Predict Stock Market Activity? *SSRN Electronic Journal*, *49*(0). https://doi.org/10.2139/ssrn.2017099

Kareklas, I., Muehling, D. D., & Weber, T. J. (2015). Reexamining health messages in the digital age: A fresh look at source credibility effects. *Journal of Advertising*, *44*(2), 88–104. https://doi.org/10.1080/00913367.2015.1018461

Karniouchina, E. V. (2011). Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, *28*(1), 62–74. https://doi.org/10.1016/j.ijresmar.2010.10.001

Kim, J. W., & Chock, T. M. (2015). Body image 2.0: Associations between social grooming on Facebook and body image concerns. *Computers in Human Behavior*, *48*, 331–339. https://doi.org/10.1016/j.chb.2015.01.009

Kim, Cha, & Kim. (2016). Targeted Ads Experiment on Instagram. In *Social informatics : 8th International Conference, SocInfo 2016* (Vol. 1, p. 545). https://doi.org/10.1007/978-3-319-47874-6

Kim, Hong, & Kang. (2017). Box Office Forecasting considering Competitive Environment and Word-of-Mouth in Social Networks: A Case Study of Korean Film Market. *Computational Intelligence and Neuroscience*, *2017*. https://doi.org/10.1155/2017/4315419

Kim, & Lennon. (2008). The Effects of Visual and Verbal Information on Attitudes and Purchase Intentions in Internet Shopping. *Psychology & Marketing*, *25*(2), 146–178. https://doi.org/10.1002/mar20204

Kleemans, M., Daalmans, S., Carbaat, I., & Anschütz, D. (2018). Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls. *Media Psychology*, *21*(1), 93–110. https://doi.org/10.1080/15213269.2016.1257392

Kramer, A., Guillory, J., & Hancock, J. T. (2014). Editorial Expression of Concern and Correction. *Proceedings of the National Academy of Sciences*, *111*(29), 8788–8790. https://doi.org/10.1073/pnas.1412469111

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal Behavior and Nonverbal Communication: What do Conversational Hand Gestures Tell Us? *Advances in Experimental Social Psychology*, *28*(C), 389–450. https://doi.org/10.1016/S0065-2601(08)60241-5

Kress, G., & Van Leeuwen, T. (2002). Colour as a Semiotic Mode:Notes for a Grammar of Colour. *Visual Communication, 1*(3), 343–368. https://doi.org/10.1177/147035720200100306

Krishen, A. S., Berezan, O., Agarwal, S., & Kachroo, P. (2016). The generation of virtual needs: Recipes for satisfaction in social media networking. *Journal of Business Research*, *69*(11), 5248–5254. https://doi.org/10.1016/j.jbusres.2016.04.120

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Kumari, & Bharadwaj. (2017). A Study and usage of Visual Features in Content Based Image Retrieval Systems. *International Journal of Advanced Research in Computer Science*, *8*(5), 2756–2581. Retrieved from http://ijarcs.info/index.php/Ijarcs/article/viewFile/3336/3766

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter , a Social Network or a News Media ? Categories and Subject Descriptors, 591–600.

Lash, M. T., & Zhao, K. (2016). Early Predictions of Movie Success: The Who, What, and When of Profitability. *Journal of Management Information Systems*, *33*(3), 874–903. https://doi.org/10.1080/07421222.2016.1243969

Lassen, N. B., la Cour, L., & Vatrapu, R. (2017). Predictive Analytics with Social Media Data. *The SAGE Handbook of Social Media Research Methods*, 328–341. Retrieved from http://www.inet.ox.ac.uk/library/view/809

Lassen, N. B., Madsen, R., & Vatrapu, R. (2014). Predicting iPhone Sales from iPhone Tweets. *Proceedings . IEEE 18th International Enterprise Distributed Object Computing Conference*, *2014–Decem*(December), 81–90. https://doi.org/10.1109/EDOC.2014.20

Lee, E., Lee, J.-A., Moon, J. H., & Sung, Y. (2015). Pictures Speak Louder than Words: Motivations for Using Instagram. *Cyberpsychology, Behavior, and Social Networking*, *18*(9), 552–556. https://doi.org/10.1089/cyber.2015.0157

Lee, Keeling, & Urbaczewski. (2017). The Economic Value of Online User Reviews with Ad Spending on Movie Box-Office Sales. *Information Systems Frontiers*, 1–16. https://doi.org/10.1007/s10796-017-9778-7

Lee, Kozar, & Larsen. (2009). Avatar e-mail versus traditional e-mail: Perceptual difference and media selection difference. *Decision Support Systems*, *46*(2), 451–467. https://doi.org/10.1016/j.dss.2007.11.008

Lee, & Ma. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, *28*(2), 331–339. https://doi.org/10.1016/j.chb.2011.10.002

Lee, & Sin. (2016). Why Do People View Photographs on Instagram? In *Digital Libraries: Knowledge, Information, and Data in an Open Access Society. ICADL 2016.* (Vol. 10075, pp. 339–350). https://doi.org/10.1007/978-3-319-49304-6

Len-Ríos, M. E., Hughes, H. E., McKee, L. G., & Young, H. N. (2016). Early adolescents as publics: A national survey of teens with social media accounts, their media use preferences, parental mediation, and perceived Internet literacy. *Public Relations Review*, *42*(1), 101–108. https://doi.org/10.1016/j.pubrev.2015.10.003

Lim, H., & Childs, M. (2016). Brand Storytelling on Instagram: How Do Pictures Travel to Millennial Consumers' Minds?, *12*, 0–2. Retrieved from http://lib.dr.iastate.edu/itaa_proceedings/2016/presentations/96

Lin, R., & Utz, S. (2015). The emotional responses of browsing Facebook: Happiness, envy, and the role of tie strength. *Computers in Human Behavior*, *52*, 29–38. https://doi.org/10.1016/j.chb.2015.04.064

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2016). Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies. *Technological Forecasting and Social Change*, *109*, 35–49. https://doi.org/10.1016/j.techfore.2016.05.013

Liu, Y. (2006). Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, *74*(1), 1–49. https://doi.org/10.1509/jmkg.70.3.74

Liu, Ding, Chen, Chen, & Guo. (2016). Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications*, *75*(3), 1509–1528. https://doi.org/10.1007/s11042-014-2270-1

Liu, Jou, Chen, Topkara, Pappas, Redi, & Chang. (2016). Complura: Exploring and Leveraging a Large-scale Multilingual Visual Sentiment Ontology, 417–420. https://doi.org/10.1145/2911996.2912030

Liu, & Zhang. (2012). *A Survey of Opinion Mining and Sentiment Analysis*. *Mining Text Data* (Vol. 9781461432). https://doi.org/10.1007/978-1-4614-3223-4

Lu, Y., Kim, Y., Dou, X., & Kumar, S. (2014). Promote physical activity among college students: Using media richness and interactivity in web design. *Computers in Human Behavior*, *41*, 40–50. https://doi.org/10.1016/j.chb.2014.08.012

Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed. *Cyberpsychology, Behavior, and Social Networking*, *18*(5), 247–252. https://doi.org/10.1089/cyber.2014.0560

MacInnis, D. J., & Price, L. L. (1987). The Role of Imagery in Information Processing: Review and Extensions. *Journal of Consumer Research*, *13*(4), 473. https://doi.org/10.1086/209082

Manikonda, L., Meduri, V. V., & Kambhampati, S. (2016). Tweeting the Mind and Instagramming the Heart: Exploring Differentiated Content Sharing on Social Media. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)* (pp. 639–642). Retrieved from http://arxiv.org/abs/1603.02718

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, *34*(2), 336–354. https://doi.org/10.1016/j.ijresmar.2016.09.003

Marwick, A. E. (2015). Instafame: Luxury Selfies in the Attention Economy. *Public Culture*, *27*(1 75), 137–160. https://doi.org/10.1215/08992363-2798379

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071226

Metzger. (2007). Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi

Mittal, V., Kaul, A., Gupta, S. Sen, & Arora, A. (2017). Multivariate Features Based Instagram Post Analysis to Enrich User Experience. *Procedia Computer Science*, *122*, 138–145. https://doi.org/10.1016/j.procs.2017.11.352

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). *Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity*. *Journal of Social and Personal Relationships* (Vol. 25). https://doi.org/10.1177/0265407508096700

Mulay, S. A., Joshi, S. J., & Shaha, M. R. (2016). Sentiment Analysis and Opinion Mining With Social Networking for Predicting Box Office Collection of Movie, *9359*(1), 74–79.

Murray, S. (2008). Digital images, photo-sharing, and our shifting notions of everyday aesthetics. *Journal of Visual Culture*, *7*(2), 147–163. https://doi.org/10.1177/1470412908091935

Naumzik, C., Feuerriegel, S., & Neumann, D. (2017). Understanding Consumer Behavior in Electronic Commerce with Image Sentiment. *Proceedings Der 13. Internationale Tagung Wirtschaftsinformatik (WI)*, 1264–1266. Retrieved from http://www.wi2017.ch/de/proceedings

Naveed, N., Gottron, T., Kunegis, J., & Alhadi, A. C. (2011). Bad news travel fast. *Proceedings of the 3rd International Web Science Conference on - WebSci '11*, (June 2011), 1–7. https://doi.org/10.1145/2527031.2527052

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology. Human Learning and Memory*, *2*(5), 523–528. https://doi.org/10.1037/0278-7393.2.5.523

Neumann, R., & Strack, F. (2000). “ Mood contagion ”: The automatic transfer of mood between persons “ Mood Contagion ”: The Automatic Transfer of Mood Between Persons. *Journal of Personality and Social Psychology*, *79*(2), 211–223. https://doi.org/10.1037//0022-3514.79.2.211

Newberry, B. (2001). Raising Student Social Presence in Online Classes. *World Conference on the WWW and Internet Proceedings*, *Orlando*, *F*, 7. Retrieved from http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED466611

Newhagen, J. E., & Reeves, B. (1992). The Evenings Bad-News - Effects of Compelling Negative Television-News Images on Memory. *Journal of Communication. 42*(2), 25–41. Retrieved from internal-pdf:/newhagen&reeves1992.pdf

Nowak, K. L., Watt, J., & Walther, J. B. (2009). Computer mediated teamwork and the efficiency framework: Exploring the influence of synchrony and cues on media satisfaction and outcome success. *Computers in Human Behavior*, *25*(5), 1108–1119. https://doi.org/10.1016/j.chb.2009.05.006

Oeldorf-Hirsch, A., Birnholtz, J., & Hancock, J. T. (2017). Your post is embarrassing me: Face threats, identity, and the audience on Facebook. *Computers in Human Behavior*, *73*, 92–99. https://doi.org/10.1016/j.chb.2017.03.030

Oeldorf-Hirsch, A., & Nowak, K. L. (2018). There is Something I Need to Tell You: Balancing Appropriateness and Efficiency in Modality Choice for Interpersonal Disclosures. *Communication Studies*, *69*(2), 125–144. https://doi.org/10.1080/10510974.2017.1417878

Ogara, S. O., Koh, C. E., & Prybutok, V. R. (2014). Investigating factors affecting social presence and user satisfaction with Mobile Instant Messaging. *Computers in Human Behavior*, *36*, 453–459. https://doi.org/10.1016/j.chb.2014.03.064

Oghina, A., Breuss, M., Tsagkias, M., & De Rijke, M. (2012). Predicting IMDB movie ratings using social media. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7224 LNCS*, 503–507. https://doi.org/10.1007/978-3-642-28997-2_51

Olszanowski, M. (2014). Feminist Self-Imaging and Instagram: Tactics of Circumventing Sensorship. *Visual Communication Quarterly*, *21*(2), 83–95. https://doi.org/10.1080/15551393.2014.928154

Pappas, N., Redi, M., Topkara, M., Jou, B., Liu, H., Chen, T., & Chang, S.-F. (2016). Multilingual Visual Sentiment Concept Matching. https://doi.org/10.1145/2911996.2912016

Pappas, N., Redi, M., Topkara, M., Liu, H., Jou, B., Chen, T., & Chang, S. F. (2017). Multilingual visual sentiment concept clustering and analysis. *International Journal of Multimedia Information Retrieval*, *6*(1), 51–70. https://doi.org/10.1007/s13735-017-0120-4

Parimi, R., & Caragea, D. (2013). Machine Learning and Data Mining in Pattern Recognition, *7988*(July). https://doi.org/10.1007/978-3-642-39712-7

Phua, J., Jin, S. V., & Kim, J. (Jay). (2017). Gratifications of using Facebook, Twitter, Instagram, or Snapchat to follow brands: The moderating effect of social comparison, trust, tie strength, and network homophily on brand identification, brand engagement, brand commitment, and membership intentio. *Telematics and Informatics*, *34*(1), 412–424. https://doi.org/10.1016/j.tele.2016.06.004

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155–167. https://doi.org/10.1016/j.chb.2016.03.084

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Social Science Information*, *21*(4–5), 529–553. https://doi.org/10.1177/053901882021004003

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2015). A Clearer Picture: The Contribution of Visuals and Text to Framing Effects. *Journal of Communication*, *65*(6), 997–1017. https://doi.org/10.1111/jcom.12184

Qin, L. (2011). Word-of-Blog for Movies: a Predictor and an Outcome of Box Office Revenue? *Journal of Electronic Commerce Research*, *12*(3), 187–198.

Radosavljevic, V., & Grbovic, M. (2014). Large-scale World Cup 2014 outcome prediction based on Tumblr posts. *KDD Workshop on …*, (August), 2–5. Retrieved from http://astro.temple.edu/~tua95067/radosavljevic2014sportskdd.pdf

Rayatdoost, S., & Soleymani, M. (2016). Ranking images and videos on visual interestingness by visual sentiment features. *CEUR Workshop Proceedings*, *1739*.

Reichelt, J., Sievert, J., & Jacob, F. (2014). How credibility affects eWOM reading: The influences of expertise, trustworthiness, and similarity on utilitarian and social functions. *Journal of Marketing Communications*. Taylor & Francis. https://doi.org/10.1080/13527266.2013.797758

Rice, R. E. (1993). Media Appropriateness - Using Social Presence Theory to Compare Traditional and New Organizational Media. *Human Communication Research*, *19*(4), 451–484. https://doi.org/10.1111/j.1468-2958.1993.tb00309.x

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, *55*(4), 863–870. https://doi.org/10.1016/j.dss.2012.12.022

Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: Who was in the lead? *Computers in Human Behavior*, *81*, 148–160. https://doi.org/10.1016/j.chb.2017.11.021

Schmitz, & Fulk. (1991). Organizational Colleagues, Media Richness, and Electronic Mail. *Communication Research*, *18*(4), 487–523. https://doi.org/10.1177/07399863870092005

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Shane-Simpson, C., Manago, A., Gaggi, N., & Gillespie-Lynch, K. (2018). Why Do College Students Prefer Facebook, Twitter, or Instagram? Site Affordances, Tensions Between Privacy and Self-Expression, and Implications for Social Capital. *Computers in Human Behavior*, *86*, 276–288. https://doi.org/10.1016/j.chb.2018.04.041

Sheldon, P., & Bryant, K. (2016). Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior*, *58*, 89–97. https://doi.org/10.1016/j.chb.2015.12.059

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, *35*(3), 553–572.

Socher, R., Perelygin, A., & Wu, J. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the …*, 1631–1642. https://doi.org/10.1371/journal.pone.0073791

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3–14. https://doi.org/10.1016/j.imavis.2017.08.003

Sparks, J. R., Areni, C. S., & Cox, K. C. (1998). An investigation of the effects of language style and communication modality on persuasion. *Communication Monographs*, *65*(2), 108–125. https://doi.org/10.1080/03637759809376440

Spil, T. A. M., Effing, R., & Both, M. P. (2016). Enable, engage and evaluate: Introducing the 3e social media strategy canvas based on the european airline industry. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9844 LNCS*, 15–30. https://doi.org/10.1007/978-3-319-45234-0_2

Stapleton, P., Luiz, G., & Chatwin, H. (2017). Generation Validation: The Role of Social Comparison in Use of Instagram Among Emerging Adults. *Cyberpsychology, Behavior, and Social Networking*, *20*(3), 142–149. https://doi.org/10.1089/cyber.2016.0444

Strahan, E. J., Wilson, A. E., Cressman, K. E., & Buote, V. M. (2006). Comparing to perfection: How cultural norms for appearance affect social comparisons and self-image. *Body Image*, *3*(3), 211–227. https://doi.org/10.1016/j.bodyim.2006.07.004

Sundar. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. *Digital Media, Youth, and Credibility*, 73–100. https://doi.org/10.1162/dmal.9780262562324.073

Sundar, S. S., & Limperos, A. M. (2013). Uses and Grats 2.0: New Gratifications for New Media. *Journal of Broadcasting and Electronic Media*, *57*(4), 504–525. https://doi.org/10.1080/08838151.2013.845827

Taddicken, M. (2014). The “Privacy Paradox” in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure1. *Journal of Computer-Mediated Communication*, *19*(2), 248–273. https://doi.org/10.1111/jcc4.12052

Tanupabrungsun, S., & Hemsley, J. (2018). Studying Celebrity Practices on Twitter Using a Framework for Measuring Media Richness. *Social Media and Society*, *4*(1). https://doi.org/10.1177/2056305118763365

Taylor, D. & Belgrave, F. (1986). The Effects of Perceived Intimacy and Valence on Self-Disclosure Reciprocity. *Personality and Social Psychology Bulletin*, *12*(2), 247–255. https://doi.org/0803973233

Thelwall, M., & Vis, F. (2017). Gender and image sharing on Facebook, Twitter, Instagram, Snapchat and WhatsApp in the UK. *Aslib Journal of Information Management*, *69*(6), 702–720. https://doi.org/10.1108/AJIM-04-2017-0098

Thelwall, Buckley, & Paltoglou. (2011). Sentiment inTwitter Events. *Journal of the American Society for Information Science and Technology*, *2*(62), 406–418. https://doi.org/10.1002/asi.21462

Thoumrungroje, A. (2014). The Influence of Social Media Intensity and EWOM on Conspicuous Consumption. *Procedia - Social and Behavioral Sciences*, *148*(November 2012), 7–15. https://doi.org/10.1016/j.sbspro.2014.07.009

Tifferet, S., & Vilnai-Yavetz, I. (2014). Computers in Human Behavior Gender differences in Facebook self-presentation : An international randomized study. *Computers in Human Behavior*, *35*, 388–399. https://doi.org/10.1016/j.chb.2014.03.016

Ting, & De Run. (2015). Beliefs about the Use of Instagram : An Exploratory Study. *International Journal of Business and Innovations*, *2*(2).

Tiwari, A. K. (2015). A Review on Big Data Integration. *International Journal of Computer Applications*, *1*(1), 1–5.

Tobin, S. J., & Chulpaiboon, P. (2016). The role of social connection in satisfaction with Instagram photographs. *Translational Issues in Psychological Science*, *2*(3), 303–312. https://doi.org/10.1037/tps0000084

Tseng, F.-C., Cheng, T. C. E., Li, K., & Teng, C.-I. (2017). How does media richness contribute to customer loyalty to mobile instant messaging? *Internet Research*, *27*(3), 520–537. https://doi.org/10.1108/IntR-06-2016-0181

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–185. https://doi.org/10.1074/jbc.M501708200

Turner, P. G., & Lefevre, C. E. (2017). Instagram use is linked to increased symptoms of orthorexia nervosa. *Eating and Weight Disorders*, *22*(2), 277–284. https://doi.org/10.1007/s40519-017-0364-2

Vilnai-Yavetz, I., & Tifferet, S. (2015). A Picture Is Worth a Thousand Words: Segmenting Consumers by Facebook Profile Images. *Journal of Interactive Marketing*, *32*, 53–69. https://doi.org/10.1016/j.intmar.2015.05.002

Wakefield, L. T., & Bennett, G. (2016). Sports fan experience: Electronic word-of-mouth in ephemeral social media. *Sport Management Review*. https://doi.org/10.1016/j.smr.2017.06.003

Wang, Y., Wang, S., Tang, J., Liu, H., & Li, B. (2015). Unsupervised Sentiment Analysis for Social Media Images. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, (Ijcai), 2378–2379. https://doi.org/10.1109/ICDMW.2015.142

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2017). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media and Society*. https://doi.org/10.1177/1461444817707349

Watson. (2016). Is Twitter an Alternative Medium? Comparing Gulf Coast Twitter and Newspaper Coverage of the 2010 BP Oil Spill. *Communication Research*, *43*(5), 647–671. https://doi.org/10.1177/0093650214565896

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. https://doi.org/10.1.1.104.6570

White. (2016). Forecasting Canadian Elections Using Twitter. *Advances in Artificial Intelligence*, *9673*, 1–15. https://doi.org/10.1007/978-3-319-34111-8

Wijnhoven, F., & Plant, O. (2017). Sentiment Analysis and Google Trends Data for Predicting Car Sales. *Icis*, (2014), 1–16.

Wilcox, K., & Stephen, A. T. (2013). Are Close Friends the Enemy? Online Social Networks, Self-Esteem, and Self-Control. *Journal of Consumer Research*, *40*(1), 90–103. https://doi.org/10.1086/668794

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45–55. https://doi.org/10.1057/ejis.2011.51

Yao, R., & Chen, J. (2013). Predicting movie sales revenue using online reviews. *Proceedings - 2013 IEEE International Conference on Granular Computing, GrC 2013*, 396–401. https://doi.org/10.1109/GrC.2013.6740443

Yoon, S., & Zhang, D. (2018). Social media, information presentation, consumer involvement, and cross-border adoption of pop culture products. *Electronic Commerce Research and Applications*, *27*, 129–138. https://doi.org/10.1016/j.elerap.2017.12.005

You, Q., Luo, J., Jin, H., & Yang, J. (2016). Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining - WSDM ’16*, 13–22. https://doi.org/10.1145/2835776.2835779

Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive Medicine*, *63*, 112–115. https://doi.org/10.1016/j.ypmed.2014.01.024

Yu, S., & Kak, S. (2012). A survey of prediction using social media. *arXiv Preprint arXiv:1203.1647*, 1–20. https://doi.org/10.1186/s13174-014-0008-y

Yuan, J., You, Q., & Luo, J. (2015). Multimedia Data Mining and Analytics. In A.K. Baughman et al. (eds.) (Ed.), *Multimedia Data Mining and Analytics* (pp. 31–59). Springer International Publishing Switzerland 2015. https://doi.org/10.1007/978-3-319-14998-1

Yuan, Zhao, Luan, Wang, & Chua. (2014). Memory recall based video search: Finding videos you have seen before based on your memory. *ACM Trans. Multimedia Comput. Commun. Appl.*, *10*(2), 1–21. https://doi.org/10.1145/2534409

Zappavigna, M. (2016). Social media photography: construing subjectivity in Instagram images. *Visual Communication*, *15*(3), 271–292. https://doi.org/10.1177/1470357216643220

Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., & Nunamaker, J. F. (2003). An exploratory study into deception detection in text-based computer-mediated communication. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003*, (February 2003). https://doi.org/10.1109/HICSS.2003.1173793

Zulli, D. (2017). Capitalizing on the look: insights into the glance, attention economy, and Instagram. *Critical Studies in Media Communication*, *5036*, 1–14. https://doi.org/10.1080/15295036.2017.1394582

Appendices

List of appendices

Appendix 1 – Overview of literature with regard to social media predictor variables

Volume-related variables

Article	Context								Type of SM					Type of SM predictor variable		
	Movies	Business	Politics	Music	Health	Disasters	News	Other	Twitter	Instagram	Facebook	Flickr	Other	Volume	Sentiment	User-profile
Volume-related variables																
Jin & Gallagher (2010)		x	x									x		x		
Jahinsky et al. (2014)					x				x					x		
Young et al. (2014)					x				x					x		
Jungherr (2013)			x						x					x	x	
Kim et al. (2016)								x		x				x		
Schmidbauer et al. (2018)			x											x	x	
Asur & Huberman (2010)	x								x					x	x	
Liu (2006)	x												x	x		
Karnouchina et al. (2011)	x												x	x		
Baek et al. (2014)	x								x					x	x	
Rui et al. (2013)	x								x					x	x	
Oghina et al. (2012)	x								x				x	x	x	
Liu et al. (2016)	x								x					x	x	
Bhavsar et al. (2017))	x								x		x			x	x	
Ding et al. (2017)	x										x			x		
Goel et al. (2010)	x			x				x					x	x		
Houston et al. (2018)	x								x		x			x		
Divakaran et al. (2017)	x												x	x		
Ho et al. (2009)	x												x	x		
Hennig-Thurau et al (2012)	x								x					x		
Broekhuizen et al. (2011)	x												x	x		
Kim et al. (2017)	x								x		x			x		
Lee et al. (2017)	x												x	x		
Lipizzi et al. (2016)	x								x					x	x	
Mestyán et al. (2013)	x												x	x		
Dellarocas et al. (2007)	x												x	x		
Qin (2011)	x												x	x		

Literature:

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*, 492–499. https://doi.org/10.1109/WI-IAT.2010.63

Baek, H., Ahn, J., & Oh, S. (2014). Impact of tweets on box office revenue: Focusing on when tweets are written. *ETRI Journal*, *36*(4), 581–590. https://doi.org/10.4218/etrij.14.0113.0732

Bhavsar, P., Kumar, A., Kumar, S., Gaur, A., & Sheikh, S. G. (2017). Outcome and Prediction of Popularity of Motion Picture Using Social Media, 3245–3248.

Broekhuizen, T. L. J., Delre, S. A., & Torres, A. (2011). Simulating the cinema market: How cross-cultural differences in social influence explain box office distributions. *Journal of Product Innovation Management*, *28*(2), 204–217. https://doi.org/10.1111/j.1540-5885.2011.00792.x

Dellarocas, C., Zhang, X., & 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

Ding, C., Cheng, H. K., Duan, Y., & Jin, Y. (2017). The power of the “like” button: The impact of social media on box office. *Decision Support Systems*, *94*, 77–84. https://doi.org/10.1016/j.dss.2016.11.002

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*, 12–28. https://doi.org/10.1016/j.intmar.2016.10.004

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17486–90. https://doi.org/10.1073/pnas.1005962107

Hennig-Thurau, T., Marchand, A., & Hiller, B. (2012). The relationship between reviewer judgments and motion picture success: Re-analysis and extension. *Journal of Cultural Economics*, *36*(3), 249–283. https://doi.org/10.1007/s10824-012-9172-8

Ho, J. Y. C., Dhar, T., & Weinberg, C. B. (2009). Playoff payoff: Super Bowl advertising for movies. *International Journal of Research in Marketing*, *26*(3), 168–179. https://doi.org/10.1016/j.ijresmar.2009.06.001

Houston, M. B., Kupfer, A. K., Hennig-Thurau, T., & Spann, M. (2018). Pre-release consumer buzz. *Journal of the Academy of Marketing Science*, 1–23. https://doi.org/10.1007/s11747-017-0572-3

Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & Argyle, T. (2014). Tracking suicide risk factors through Twitter in the US. *Crisis*, *35*(1), 51–59. https://doi.org/10.1027/0227-5910/a000234

Jin, X., Gallagher, A., Cao, L., Luo, J., & Han, J. (2010). The wisdom of social multimedia: using Flickr for prediction and forecast. *Proceedings of the International Conference on Multimedia*, 1235–1244. https://doi.org/10.1145/1873951.1874196

Jungherr, A. (2013). Tweets and votes, a special relationship. *Proceedings of the 2nd Workshop on Politics, Elections and Data - PLEAD '13*, 5–14. https://doi.org/10.1145/2508436.2508437

Karniouchina, E. V. (2011). Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, *28*(1), 62–74. https://doi.org/10.1016/j.ijresmar.2010.10.001

Kim, Cha, & Kim. (2016). Targeted Ads Experiment on Instagram. In *Social informatics : 8th International Conference, SocInfo 2016* (Vol. 1, p. 545). https://doi.org/10.1007/978-3-319-47874-6

Kim, Hong, & Kang. (2017). Box Office Forecasting considering Competitive Environment and Word-of-Mouth in Social Networks: A Case Study of Korean Film Market. *Computational Intelligence and Neuroscience*, *2017*. https://doi.org/10.1155/2017/4315419

Lee, Keeling, & Urbaczewski. (2017). The Economic Value of Online User Reviews with Ad Spending on Movie Box-Office Sales. *Information Systems Frontiers*, 1–16. https://doi.org/10.1007/s10796-017-9778-7

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2016). Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies. *Technological Forecasting and Social Change*, *109*, 35–49. https://doi.org/10.1016/j.techfore.2016.05.013

Liu, T., Ding, X., Chen, Y., Chen, H., & Guo, M. (2016). Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications*, *75*(3), 1509–1528. https://doi.org/10.1007/s11042-014-2270-1

Liu, Y. (2006). Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, *74*(1), 1–49. https://doi.org/10.1509/jmkg.70.3.74

Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071226

Oghina, A., Breuss, M., Tsagkias, M., & De Rijke, M. (2012). Predicting IMDB movie ratings using social media. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7224 LNCS*, 503–507. https://doi.org/10.1007/978-3-642-28997-2_51

Qin, L. (2011). Word-of-Blog for Movies: a Predictor and an Outcome of Box Office Revenue? *Journal of Electronic Commerce Research*, *12*(3), 187–198.

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, *55*(4), 863–870. https://doi.org/10.1016/j.dss.2012.12.022

Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: Who was in the lead? *Computers in Human Behavior*, *81*, 148–160. https://doi.org/10.1016/j.chb.2017.11.021

Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive Medicine*, *63*, 112–115. https://doi.org/10.1016/j.ypmed.2014.01.024

Sentiment-related variables

Article	Context								Type of SM					Type of SM predictor variable		
	Movies	Business	Politics	Music	Health	Disasters	News	Other	Twitter	Instagram	Facebook	Flickr	Other	Volume	Sentiment	User-profile
Sentiment-related variables																
Soleymani et al. (2017)								x					x		x	
Lipizzi et al. (2016)	x								x					x	x	
Liu et al. (2016)	x								x					x	x	
White (2016)			x						x						x	
Schumaker et al. (2016)								x	x						x	
Asur & Huberman (2010)	x								x					x	x	
Rui et al. (2013)	x								x					x	x	
Borth et al. (2013)								x				x	x		x	
Ji et al. (2016)								x	x				x		x	
Jou et al. (2015)								x				x	x		x	
Pappas et al. (2016)								x				x	x		x	
Islam & Zhang (2016)								x	x						x	
Graesser et al. (2017)								x				x	x		x	
Ahsan et al. (2017)								x					x		x	
Huber et al. (2018)								x	x						x	
Gelli et al. (2015)								x				x			x	
Flaes et al. (2016)								x	x			x			x	
You et al. (2016)								x		x		x			x	
Cai & Xia (2015)								x	x						x	
Apala et al. (2013)	x								x				x		x	x

Literature:

Ahsan, U., De Choudhury, M., & Essa, I. (2017). Towards using visual attributes to infer image sentiment of social events. *Proceedings of the International Joint Conference on Neural Networks*, *2017–May*(April 2016), 1372–1379. https://doi.org/10.1109/IJCNN.2017.7966013

Apala, K. R., Jose, M., Motnam, S., Chan, C.-C., Liszka, K. J., & De Gregorio, F. (2013). Prediction of movies box office performance using social media. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013*, (July), 1209–1214. https://doi.org/10.1145/2492517.2500232

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*, 492–499. https://doi.org/10.1109/WI-IAT.2010.63

Borth, D., Ji, R., Chen, T., Breuel, T., & Chang, S.-F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. *Proceedings of the 21st ACM International Conference on Multimedia - MM '13*, 223–232. https://doi.org/10.1145/2502081.2502282

Cai, & Xia. (2015). Convolutional Neural Networks for Multimedia Sentiment Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9362*, 159–167. https://doi.org/10.1007/978-3-319-25207-0

Flaes, B., Rudinac, & Worring. (2016). What Multimedia Sentiment Analysis Says About City Liveability. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9626*, 824–829. https://doi.org/10.1007/978-3-319-30671-1

Gelli, F., Uricchio, T., Bertini, M., Del Bimbo, A., & Chang, S.-F. (2015). Image Popularity Prediction in Social Media Using Sentiment and Context Features. *Proceedings of the 23rd ACM International Conference on Multimedia - MM '15*, 907–910. https://doi.org/10.1145/2733373.2806361

Graesser, L., Gupta, A., Sharma, L., & Bakhturina, E. (2017). Sentiment Classification using Images and Label Embeddings. Retrieved from http://arxiv.org/abs/1712.00725

Huber, B., Mcduff, D., Brockett, C., Galley, M., & Dolan, B. (2018). Emotional Dialogue Generation using Image-Grounded Language Models. https://doi.org/10.1145/3173574.3173851

Islam, J., & Zhang, Y. (2016). Visual Sentiment Analysis for Social Images Using Transfer Learning Approach. *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, 124–130. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.29

Ji, R., Cao, D., Zhou, Y., & Chen, F. (2016). Survey of visual sentiment prediction for social media analysis. *Frontiers of Computer Science*, *10*(4), 602–611. https://doi.org/10.1007/s11704-016-5453-2

Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M., & Chang, S.-F. (2015). Visual Affect Around the World: A Large-scale Multilingual Visual Sentiment Ontology, (1). https://doi.org/10.1145/2733373.2806246

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2016). Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies. *Technological Forecasting and Social Change*, *109*, 35–49. https://doi.org/10.1016/j.techfore.2016.05.013

Liu, T., Ding, X., Chen, Y., Chen, H., & Guo, M. (2016). Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications*, *75*(3), 1509–1528. https://doi.org/10.1007/s11042-014-2270-1

Pappas, N., Redi, M., Topkara, M., Jou, B., Liu, H., Chen, T., & Chang, S.-F. (2016). Multilingual Visual Sentiment Concept Matching. https://doi.org/10.1145/2911996.2912016

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, *55*(4), 863–870. https://doi.org/10.1016/j.dss.2012.12.022

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3–14. https://doi.org/10.1016/j.imavis.2017.08.003

White. (2016). Forecasting Canadian Elections Using Twitter. *Advances in Artificial Intelligence*, *9673*, 1–15. https://doi.org/10.1007/978-3-319-34111-8

You, Q., Luo, J., Jin, H., & Yang, J. (2016). Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining - WSDM '16*, 13–22. https://doi.org/10.1145/2835776.2835779

Appendix 2 – Instagram and Twitter predictions literature

Twitter

Article	Context								Type of SM					Type of SM predictor variable		
	Movies	Business	Politics	Music	Health	Disasters	News	Other	Twitter	Instagram	Facebook	Flickr	Other	Volume	Sentiment	User-profile
Twitter and predictions									x							
Asur & Huberman (2010)	x								x					x	x	
Bothos et al. (2010)	x								x						x	
Lassen et al. (2014)								x	x						x	
Bollen et al. (2011)		x							x						x	
Elshendy et al. (2017)		x							x				x	x	x	
Mohd Sharriff et al. (2017)								x	x							x
Ritterman et al. (2009)					x				x					x		
De Choudhury & Gamon (2013)					x				x					x		x
Culotta (2010)					x				x					x		
Eichstaedt et al. (2015)					x				x						x	
Li & Cardie (2013)					x				x					x		
Schumaker et al. (2016)								x	x						x	
Chung & Mustafaraj (2011)			x						x						x	
Conover et al. (2011)			x						x					x		
Sang & Bos (2012)			x						x					x	x	
Tumasjan et al. (2010)			x						x					x	x	
Eysenbach (2011)								x	x							

Literature:

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*, 492–499. https://doi.org/10.1109/WI-IAT.2010.63

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bothos, E., Apostolou, D., & Mentzas, G. (2010). Using Social Media to Predict Future Events with Agent-Based Markets. *IEEE Intelligent Systems*, *25*(6), 50–58. https://doi.org/10.1109/MIS.2010.152

Chung, J., & Mustafaraj, E. (2011). Can collective sentiment expressed on twitter predict political elections? *Aaai*, 1770–1771. https://doi.org/10.1007/s00247-002-0848-7

Conover, M., Ratkiewicz, J., & Francisco, M. (2011). Political polarization on twitter. *Icwsm*, *133*(26), 89–96. https://doi.org/10.1021/ja202932e

Culotta, A. (2010). Detecting influenza outbreaks by analyzing Twitter messages, (May), 1–11. https://doi.org/10.1145/1964858.1964874

De Choudhury, M., & Gamon, M. (2013). Predicting Depression via Social Media. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, *2*, 128–137. https://doi.org/10.1109/IRI.2012.6302998

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., … Seligman, M. E. P. (2015). Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychological Science*, *26*(2), 159–169. https://doi.org/10.1177/0956797614557867

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2017). Using four different online media sources to forecast the crude oil price. *Journal of Information Science*, 16555151769829. https://doi.org/10.1177/0165551517698298

Eysenbach G. (2011). Can tweets predict citations? Metrics of social. *Journal of Medical Internet Research*, *13*(4), e123. https://doi.org/10.2196/jmir.2012

Lassen, N. B., Madsen, R., & Vatrapu, R. (2014). Predicting iPhone Sales from iPhone Tweets. *Proceedings . IEEE 18th International Enterprise Distributed Object Computing Conference*, *2014–Decem*(December), 81–90. https://doi.org/10.1109/EDOC.2014.20

Li, J., & Cardie, C. (2013). Early Stage Influenza Detection from Twitter. Retrieved from http://arxiv.org/abs/1309.7340

Ritterman, J., Osborne, M., & Klein, E. (2009). Using prediction markets and Twitter to predict a swine flu pandemic. *1St International Workshop on {…}*, (2004).

Sang, E., & Bos, J. (2012). Predicting the 2011 dutch senate election results with twitter. *Proceedings of the Workshop on Semantic Analysis in …*, (53), 53–60. Retrieved from http://dl.acm.org/citation.cfm?id=2389969.2389976%5Cnhttp://dl.acm.org/citation.cfm?id=2389976

Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, *88*, 76–84. https://doi.org/10.1016/j.dss.2016.05.010

Shariff, S. M., Zhang, X., & Sanderson, M. (2017). On the credibility perception of news on Twitter: Readers, topics and features. *Computers in Human Behavior*, *75*, 785–796. https://doi.org/10.1016/j.chb.2017.06.026

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–185. https://doi.org/10.1074/jbc.M501708200

Instagram

Article	Context								Type of SM					Type of SM predictor variable		
	Movies	Business	Politics	Music	Health	Disasters	News	Other	Twitter	Instagram	Facebook	Flickr	Other	Volume	Sentiment	User-profile
Instagram and predictions																
Almgren et al. (2016)								x		x					x	x
Amencherla & Varshney (2017)								x		x					x	
Arcenaux & Dinu (2018)								x		x						
Bashir et al. (2018)		x								x						x
Chae (2018)					x					x	x		x			x
Chen (2018)								x		x						x
Cheung et al. (2015)								x		x		x				
Chu et al. (2017)								x		x						
De et al. (2017)								x		x				x		
De Vries et al. (2018)					x					x						x
Djafarova & Rushworth (2017)								x		x						x
Djafarova & Trofimenko (2018)								x		x						x
Ferguson & Greer (2018)		x								x						
Giancristofaro & Panangadan (2016)		x								x					x	
Goel & Prakash (2016)		x							x	x					x	
Han et al. (2016)								x		x						
Hosseinmardi et al. (2015)								x		x					x	
Hu et al. (2016)												x				
Jin & Gallagher (2010)		x	x									x		x		
Jang et al. (2016)								x		x						x
Katsurai & Satoh (2016)								x		x					x	
Kim et al. (2016)								x		x				x		
Kim et al. (2017)								x		x						x
Kleemans et al. (2018)								x		x						x
Lee & Sin (2016)								x		x						x
Li et al. (2015)								x		x					x	x
Mazloom et al. (2015)		x								x					x	
Overgoor et al (2017)		x								x					x	
Pang & Zhang (2016)								x		x						x
Pang & Zhang (2017)		x								x						x
Peng et al. (2016)					x					x					x	
Park et al. (2016)		x								x					x	
Pittman & Reich (2016)					x				x	x			x			
Phua et al. (2018)					x					x						x
Schmidbauer et al. (2018)			x											x	x	
Shane-Simpson et al. (2018)								x	x	x	x					x
Stapleton et al. (2017)								x		x						x
Thelwall & Vis (2017)								x	x	x	x		x			x
Turner & Lefevre (2017)					x					x						x
Zappavigna (2016)								x		x						

Literature:

Almgren, K., Lee, J., & kim, M. (2016). Predicting the Future Popularity of Images on Social Networks. *Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016 - MISNC, SI, DS 2016*, 1–6. https://doi.org/10.1145/2955129.2955154

Amencherla, M., & Varshney, L. R. (2017). Color-based visual sentiment for social communication. *2017 15th Canadian Workshop on Information Theory, CWIT 2017*, 1–5. https://doi.org/10.1109/CWIT.2017.7994829

Arceneaux, P. C., & Dinu, L. F. (2018). The social mediated age of information: Twitter and Instagram as tools for information dissemination in higher education. *New Media & Society*, 146144481876825. https://doi.org/10.1177/1461444818768259

Bashir, A., Wen, J. T., Kim, E., & Morris, J. D. (2018). The Role of Consumer Affect on Visual Social Networking Sites: How Consumers Build Brand Relationships. *Journal of Current Issues and Research in Advertising*, *1734*, 1–14. https://doi.org/10.1080/10641734.2018.1428250

Chae, J. (2018). Reexamining the relationship between social media and happiness: The effects of various social media platforms on reconceptualized happiness. *Telematics and Informatics*, (April), 0–1. https://doi.org/10.1016/j.tele.2018.04.011

Chen, & Chang. (2018). What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2018.03.019

Cheung, M., She, J., & Li, X. (2015). Non-user Generated Annotation on User Shared Images for Connection Discovery. *Proceedings - 2015 IEEE International Conference on Data Science and Data Intensive Systems; 8th IEEE International Conference Cyber, Physical and Social Computing; 11th IEEE International Conference on Green Computing and Communications and 8th IEEE Inte*, *1*, 204–209. https://doi.org/10.1109/DSDIS.2015.113

Chu, D., Shen, Z., Zhang, Y., B, S. Y., & Lin, X. (2017). Real-Time Popularity Prediction on Instagram. In *Australian Database Conference* (Vol. 10538, pp. 275–279). https://doi.org/10.1007/978-3-319-68155-9

De, S., Maity, A., Goel, V., Shitole, S., & Bhattacharya, A. (2017). Predicting the popularity of instagram posts for a lifestyle magazine using deep learning. *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*, 174–177. https://doi.org/10.1109/CSCITA.2017.8066548

de Vries, D. A., Möller, A. M., Wieringa, M. S., Eigenraam, A. W., & Hamelink, K. (2018). Social Comparison as the Thief of Joy: Emotional Consequences of Viewing Strangers' Instagram Posts. *Media Psychology*, *21*(2), 222–245. https://doi.org/10.1080/15213269.2016.1267647

Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, *68*, 1–7. https://doi.org/10.1016/j.chb.2016.11.009

Djafarova, E., & Trofimenko, O. (2018). “Instafamous” – credibility and self-presentation of micro-celebrities on social media. *Information Communication and Society*, *0*(0), 1–15. https://doi.org/10.1080/1369118X.2018.1438491

Ferguson, D. A., & Greer, C. F. (2018). Visualizing a Non-Visual Medium through Social Media : The Semiotics of Radio Station Posts on Instagram Visualizing a Non-Visual Medium through Social Media : The Semiotics of Radio Station Posts on Instagram. *Journal of Radio & Audio Media*, *25*(1), 126–141. https://doi.org/10.1080/19376529.2017.1385617

Giancristofaro, G. T., & Panangadan, A. (2016). Predicting Sentiment toward Transportation in Social Media using Visual and Textual Features. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2113–2118. https://doi.org/10.1109/ITSC.2016.7795898

Goel, L., & Prakash, A. (2016). Sentiment Analysis of Online Communities Using Swarm Intelligence Algorithms. *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*, 330–335. https://doi.org/10.1109/CICN.2016.71

Han, K., Lee, S., Jang, J. Y., Jung, Y., & Lee, D. (2016). “ Teens are from Mars , Adults are f rom Venus ” : Analyzing and Predicting Age Groups with Behavioral Characteristics in Instagram. *ACM Web Science 2016*, 35–44. https://doi.org/10.1145/2908131.2908160

Hosseinmardi, H., Li, S., Yang, Z., Lv, Q., Rafiq, R. I., Han, R., & Mishra, S. (2015). A comparison of common users across instagram and Ask.fm to better understand cyberbullying. *Proceedings - 4th IEEE International Conference on Big Data and Cloud Computing, BDCloud 2014 with the 7th IEEE International Conference on Social Computing and Networking, SocialCom 2014 and the 4th International Conference on Sustainable Computing and C*, 355–362. https://doi.org/10.1109/BDCloud.2014.87

Hu, Yamasaki, & Aizawa. (2016). Multimodal Learning for Image Popularity Prediction on Social Media. *2016 International Conference on Consumer Electronics-Taiwan*, 0–1. https://doi.org/10.1145/2812802

Jang, J. Y., Han, K., Lee, D., Jia, H., & Shih, P. C. (2016). Teens Engage More with Fewer Photos. *Proceedings of the 27th ACM Conference on Hypertext and Social Media - HT ’16*, 71–81. https://doi.org/10.1145/2914586.2914602

Jin, X., Gallagher, A., Cao, L., Luo, J., & Han, J. (2010). The wisdom of social multimedia: using Flickr for prediction and forecast. *Proceedings of the International Conference on Multimedia*, 1235–1244. https://doi.org/10.1145/1873951.1874196

Katsurai, & Satoh. (2016). Image sentiment analysis using latent correlations among visual, textual and sentiment views. *IEEE, ICASSP*, 2837–2841.

Kim, D. H., Seely, N. K., & Jung, J. H. (2017). Do you prefer, Pinterest or Instagram? The role of image-sharing SNSs and self-monitoring in enhancing ad effectiveness. *Computers in Human Behavior*, *70*, 535–543. https://doi.org/10.1016/j.chb.2017.01.022

Kim, Cha, & Kim. (2016). Targeted Ads Experiment on Instagram. In *Social informatics : 8th International Conference, SocInfo 2016* (Vol. 1, p. 545). https://doi.org/10.1007/978-3-319-47874-6

Kleemans, M., Daalmans, S., Carbaat, I., & Anschütz, D. (2018). Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls. *Media Psychology*, *21*(1), 93–110. https://doi.org/10.1080/15213269.2016.1257392

Lee, & Sin. (2016). Why Do People View Photographs on Instagram? In *Digital Libraries: Knowledge, Information, and Data in an Open Access Society. ICADL 2016.* (Vol. 10075, pp. 339–350). https://doi.org/10.1007/978-3-319-49304-6

Li, X., Nguyen Pham, T.-A., Cong, G., Yuan, Q., Li, X.-L., & Krishnaswamy, S. (2015). Where you Instagram? Associating Your Instagram Photos with Points of Interest. *Cikm*, 1231–1240. https://doi.org/10.1145/2806416.2806463

Mazloom, M., Rietveld, R., Rudinac, S., Worring, M., & van Dolen, W. (2015). Multimodal Popularity Prediction of Brand-related Social Media Posts. *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, 197–201. https://doi.org/10.1145/2964284.2967210

Overgoor, G., Mazloom, M., Worring, M., Rietveld, R., & van Dolen, W. (2017). A Spatio-Temporal Category Representation for Brand Popularity Prediction. *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval - ICMR '17*, 233–241. https://doi.org/10.1145/3078971.3078998

Pang, J., & Zhang, Y. (2017). Quantifying Location Sociality. *Proceedings of the 28th ACM Conference on Hypertext and Social Media - HT '17*, 145–154. https://doi.org/10.1145/3078714.3078729

Pang, & Zhang. (2016). DeepCity: A Feature Learning Framework for Mining Location Check-ins. *CEUR Workshop Proceedings*, *1828*, 53–59. https://doi.org/10.1145/1235

Park, J., Ciampaglia, G. L., & Ferrara, E. (2016). Style in the Age of Instagram: Predicting Success within the Fashion Industry using Social Media. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 64–73. https://doi.org/10.1145/2818048.2820065

Peng, X., Luo, J., Glenn, C., Zhan, J., & Liu, Y. (2016). Large-scale sleep condition analysis using selfies from social media. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10354 LNCS*, 151–161. https://doi.org/10.1007/978-3-319-60240-0_19

Phua, J., Lin, J. S. (Elaine), & Lim, D. J. (2018). Understanding consumer engagement with celebrity-endorsed E-Cigarette advertising on instagram. *Computers in Human Behavior*, *84*, 93–102. https://doi.org/10.1016/j.chb.2018.02.031

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*(September 2017), 155–167. https://doi.org/10.1016/j.chb.2016.03.084

Schmidbauer, H., Rösch, A., & Stieler, F. (2018). The 2016 US presidential election and media on Instagram: Who was in the lead? *Computers in Human Behavior*, *81*, 148–160. https://doi.org/10.1016/j.chb.2017.11.021

Shane-Simpson, C., Manago, A., Gaggi, N., & Gillespie-Lynch, K. (2018). Why Do College Students Prefer Facebook, Twitter, or Instagram? Site Affordances, Tensions Between Privacy and Self-Expression, and Implications for Social Capital. *Computers in Human Behavior*, *86*, 276–288. https://doi.org/10.1016/j.chb.2018.04.041

Stapleton, P., Luiz, G., & Chatwin, H. (2017). Generation Validation: The Role of Social Comparison in Use of Instagram Among Emerging Adults. *Cyberpsychology, Behavior, and Social Networking*, *20*(3), 142–149. https://doi.org/10.1089/cyber.2016.0444

Thelwall, M., & Vis, F. (2017). Gender and image sharing on Facebook, Twitter, Instagram, Snapchat and WhatsApp in the UK. *Aslib Journal of Information Management*, *69*(6), 702–720. https://doi.org/10.1108/AJIM-04-2017-0098

Turner, P. G., & Lefevre, C. E. (2017). Instagram use is linked to increased symptoms of orthorexia nervosa. *Eating and Weight Disorders*, *22*(2), 277–284. https://doi.org/10.1007/s40519-017-0364-2

Zappavigna, M. (2016). Social media photography: construing subjectivity in Instagram images. *Visual Communication*, *15*(3), 271–292. https://doi.org/10.1177/1470357216643220

Appendix 3 – Movie box office revenue and social media predictions literature

Article	Context								Type of SM					Type of SM predictor variable		
	Movies	Business	Politics	Music	Health	Disasters	News	Other	Twitter	Instagram	Facebook	Flickr	Other	Volume	Sentiment	User-profile
Social media and movie box office revenues																
Baek et al. (2014)	x								x					x	x	
Apala et al. (2013)	x								x				x		x	x
Asur & Huberman (2010)	x								x					x	x	
Bhattacharjee et al. (2017)	x								x		x		x		x	
Bhavsar et al. (2017)	x								x				x	x	x	
Liu et al. (2016)	x								x					x	x	
Gaikar et al. (2015)	x								x						x	
Jain et al. (2013)	x								x						x	
Mulay et al. (2016)	x								x					x	x	
Yao & Chen (2013)	x												x		x	
Karnouchina et al. (2011)	x												x	x		
Rui et al. (2013)	x								x					x	x	
Oghina et al. (2012)	x								x				x	x	x	
Ding et al. (2017)	x										x			x		
Goel et al. (2010)	x			x				x					x	x		
Houston et al. (2018)	x								x		x			x		
Divakaran et al. (2017)	x												x	x		
Ho et al. (2009)	x												x	x		
Hennig-Thurau et al (2012)	x								x					x		
Broekhuizen et al. (2011)	x												x	x		
Kim et al. (2017)	x								x		x			x		
Lee et al. (2017)	x												x	x		
Lipizzi et al. (2016)	x								x					x	x	
Mestyán et al. (2013)	x												x	x		
Dellarocas et al. (2007)	x												x	x		
Qin (2011)	x												x	x		

Literature:

Apala, K. R., Jose, M., Motnam, S., Chan, C.-C., Liszka, K. J., & De Gregorio, F. (2013). Prediction of movies box office performance using social media. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013*, (July), 1209–1214. https://doi.org/10.1145/2492517.2500232

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *1*, 492–499. https://doi.org/10.1109/WI-IAT.2010.63

Baek, H., Ahn, J., & Oh, S. (2014). Impact of tweets on box office revenue: Focusing on when tweets are written. *ETRI Journal*, *36*(4), 581–590. https://doi.org/10.4218/etrij.14.0113.0732

Bhattacharjee, B., Sridhar, A. and, & Dutta, A. (2017). Identifying the causal relationship between social media content of a Bollywood movie and its box-office success - a text mining approach. *International Journal of Business Information Systems*, *24*(3), 344–368. https://doi.org/10.1504/IJBIS.2017.082039

Bhavsar, P., Kumar, A., Kumar, S., Gaur, A., & Sheikh, S. G. (2017). Outcome and Prediction of Popularity of Motion Picture Using Social Media, 3245–3248.

Broekhuizen, T. L. J., Delre, S. A., & Torres, A. (2011). Simulating the cinema market: How cross-cultural differences in social influence explain box office distributions. *Journal of Product Innovation Management*, *28*(2), 204–217. https://doi.org/10.1111/j.1540-5885.2011.00792.x

Dellarocas, C., Zhang, X., & 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

Ding, C., Cheng, H. K., Duan, Y., & Jin, Y. (2017). The power of the “like” button: The impact of social media on box office. *Decision Support Systems*, *94*, 77–84. https://doi.org/10.1016/j.dss.2016.11.002

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*, 12–28. https://doi.org/10.1016/j.intmar.2016.10.004

Gaikar, D. D., Marakarkandy, B., & Dasgupta, C. (2015). Using Twitter data to predict the performance of Bollywood movies. *Industrial Management & Data Systems*, *115*(9), 1604–1621. https://doi.org/10.1108/IMDS-04-2015-0145

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17486–90. https://doi.org/10.1073/pnas.1005962107

Hennig-Thurau, T., Marchand, A., & Hiller, B. (2012). The relationship between reviewer judgments and motion picture success: Re-analysis and extension. *Journal of Cultural Economics*, *36*(3), 249–283. https://doi.org/10.1007/s10824-012-9172-8

Ho, J. Y. C., Dhar, T., & Weinberg, C. B. (2009). Playoff payoff: Super Bowl advertising for movies. *International Journal of Research in Marketing*, *26*(3), 168–179. https://doi.org/10.1016/j.ijresmar.2009.06.001

Houston, M. B., Kupfer, A. K., Hennig-Thurau, T., & Spann, M. (2018). Pre-release consumer buzz. *Journal of the Academy of Marketing Science*, 1–23. https://doi.org/10.1007/s11747-017-0572-3

Jain, V. (2013). Prediction of Movie Success using Sentiment Analysis of Tweets. *International Conference on Soft Computing and Software Engineering 2013*, *3*(3), 308–313. https://doi.org/10.7321/jscse.v3.n3.46

Karniouchina, E. V. (2011). Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, *28*(1), 62–74. https://doi.org/10.1016/j.ijresmar.2010.10.001

Kim, Hong, & Kang. (2017). Box Office Forecasting considering Competitive Environment and Word-of-Mouth in Social Networks: A Case Study of Korean Film Market. *Computational Intelligence and Neuroscience*, *2017*. https://doi.org/10.1155/2017/4315419

Lee, Keeling, & Urbaczewski. (2017). The Economic Value of Online User Reviews with Ad Spending on Movie Box-Office Sales. *Information Systems Frontiers*, 1–16. https://doi.org/10.1007/s10796-017-9778-7

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2016). Combining structure, content and meaning in online social networks: The analysis of public's early reaction in social media to newly launched movies. *Technological Forecasting and Social Change*, *109*, 35–49. https://doi.org/10.1016/j.techfore.2016.05.013

Liu, T., Ding, X., Chen, Y., Chen, H., & Guo, M. (2016). Predicting movie Box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications*, *75*(3), 1509–1528. https://doi.org/10.1007/s11042-014-2270-1

Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071226

Mulay, S. A., Joshi, S. J., & Shaha, M. R. (2016). Sentiment Analysis and Opinion Mining With Social Networking for Predicting Box Office Collection of Movie, *9359*(1), 74–79.

Oghina, A., Breuss, M., Tsagkias, M., & De Rijke, M. (2012). Predicting IMDB movie ratings using social media. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7224 LNCS*, 503–507. https://doi.org/10.1007/978-3-642-28997-2_51

Qin, L. (2011). Word-of-Blog for Movies: a Predictor and an Outcome of Box Office Revenue? *Journal of Electronic Commerce Research*, *12*(3), 187–198.

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, *55*(4), 863–870. https://doi.org/10.1016/j.dss.2012.12.022

Yao, R., & Chen, J. (2013). Predicting movie sales revenue using online reviews. *Proceedings - 2013 IEEE International Conference on Granular Computing, GrC 2013*, 396–401. https://doi.org/10.1109/GrC.2013.6740443

Volume-related variables and movie box office revenues

Volume-related variables were also used to predict movie box-office revenues. With regard to movie-box office revenue, Asur & Huberman (2010) were able to predict movie box office revenues with high accuracy by using volume-related variables. In addition, Liu (2006) and Karniouchina (2011) find that volume has an influence on movie box office revenue. In line with these findings, Baek, Ahn, & Oh (2014) and Rui, Liu, & Whinston (2013) find that the volume of Tweets is a positive and significant predictor of movie box office revenues. Further, Oghina, Breuss, Tsagkias, & De Rijke (2012) prove that IMDB scores can be predicted with high accuracy by the number of likes and dislikes on YouTube. Liu et al. (2016) find that including Tweet volume in the prediction model leads to highly accurate predictions. Bhavsar et al. (2017) use the number of views and comments on the YouTube trailer of a movie and the number of fan followers on Twitter to predict movie box office revenues. Bhavsar et al. (2017) find that the number of views and comments on the YouTube trailer of a movie and the number of fan followers on Twitter are strong and significant predictors of movie box office revenues. Ding, Cheng, Duan, & Jin (2017) study the number of likes on Facebook and find that increases in the number of likes in the prerelease are significant predictors of movie box office revenues. Goel, Hofman, Lahaie, Pennock, & Watts (2010) find that the volume of search counts corresponds highly with movie box office revenues. Ho, Dhar, & Weinberg (2009) study weekly search behavior on IMDB and find that the volume of search is a strong and significant predictor of movie box office revenues (Broekhuizen, Delre, & Torres, 2011; Hennig-Thurau, Marchand, & Hiller, 2012). Lipizzi et al. (2016) find that traffic metrics are significant predictors of movie box office revenues. Liu et al. (2016) find that the unique number of users who Tweeted about a movie and the number of Tweets during the two weeks before the movie's release are significant of movie box office revenues. Houston, Kupfer, Hennig-Thurau, & Spann (2018) study pre-release consumer buzz and find that the volume of pre-release consumer buzz is a significant predictor of movie box office revenues. Divakaran, Palmer, Søndergaard, & Matkovskyy (2017) use word-of-mouth volume as an variable to predict a movie's box office revenue in the opening week and find that word-of-mouth volume is a significant predictor of movie box office revenue in the opening week. Kim, Hong, & Kang (2017) find that the number mentions on social networks is a significant predictor of movie box office revenues. Further, Lee, Keeling, & Urbaczewski (2017) find that the volume of online ratings significantly predicts movie box office revenues. Mestyán et al. (2013) use data from Wikipedia and find that movie box office revenues can be significantly predicted by the number of viewers of a Wikipedia movie page. Dellarocas, Zhang, & Awad (2007) find that early volume of online reviews on Yahoo!Movies is a proxy for early movie box office revenues. Finally, Qin (2011) uses data from BlogPulse and finds significant relationships between the word of blog volume and movie box office revenues. Therefore, it can be said that multiple studies find support that volume-related variables are significant predictors of movie box office revenue (Liu et al., 2016).

Sentiment-related variables and movie box office revenues

Sentiment-related variables were studied in relation to movie box office revenues. Asur & Huberman (2010) use Twitter to forecast box-office revenues for movies. They find that topic related Tweets can outperform market-based predictors, like for example the Hollywood Stock Exchange. Additionally, they show how social media sentiment can be further developed in order to increase the forecasting power of social media. Their findings show that sentiment in Tweets can be used to improve predictions after the movie releases. Asur & Huberman (2010) use linear regression for predicting box-office revenues of movies before the release. More specifically, they find a strong correlation between the degree of attention of a particular topic and the future ranking. Bhattacharjee et al., (2017) study the extent to which the polarity of social media sentiment can be used to predict box office revenues of Bollywood movies. They study sentiment both in the period after and before the movie release by applying linear regression models. The findings of Bhattacharjee et al., (2017) support the notion that the polarity of social media content and box-office revenues are causally associated. In other words, Bhattacharjee et al., (2017) argue that comments, likes and Tweets definitively impact the public opinion regarding movies. In addition, it can be said that social media influences the extent to which a movie-viewer is going to visit a theatre. Furthermore, Bhattacharjee et al., (2017) argue that the polarity of comments expresses the tendency of social media users to view a movie in theatre. In general, Bhattacharjee et al., (2017) state that their study shows the importance of social media content analytics as a promotional tool. Baek et al. (2014) study the influence of Tweets on box office revenue. More specifically, they studied the moments Tweets were posted by analyzing the influence of pre- and post-consumption Tweets on box office revenue (Baek et al., 2014). Additionally, they study the influence of subjective, intention and negative Tweets on box office revenue. Tweets were collected two weeks before the release till the last moment the movie was shown. Baek et al. (2014) find that disconfirmation in pre-consumption Tweets negatively influences box office revenues. In addition, Baek et al. (2014) find that intention Tweets more strongly impacts box office revenues than subjective Tweets. More specifically, the impact of intention Tweets is three times more than the impact of subjective Tweets. Further, with regard to valence, Baek et al. (2014) find that a higher proportion of negative Tweets leads to stronger decreases in box office revenue. In addition, Baek et al. (2014) find that the total volume of Tweets corresponds strongly with box office revenues.

Bhavsar et al. (2017) use the sentiment score of Tweets related to particular movies to predict movie popularity. Bhavsar et al. (2017) were able to predict the popularity of movies and to categorize them as a hype or not. Gaikar et al. (2015) use Twitter data to predict the performance of Bollywood movies. Based on social media data, they use sentiment analysis and prediction algorithms to investigate Indian movie performance. Gaikar et al. (2015) find that actor rating and sentiment score were able to deliver accurate prediction results for the performance of Indian movies. Jain (2013) predict movie box office success based on sentiment analysis. The findings of Jain (2013) prove that movie sentiment on social media can be used to predict movie box office success.

Liu et al. (2016) use a visual-analytics toolkit and extracts data from Twitter and Bitly to predict movie ratings and movie revenue. More specifically, they study Tweet volume and sentiment in Tweets with regard to movies. Liu et al. (2016) identify the following useful metrics: three-day opening-weekend revenue, movie budget (according to the IMDB), movie genre, number of unique users who Tweeted about a movie, the average daily number of Tweets during the two weeks before the movie's release, Tweet sentiment score (summation of polarity of word's sentiments), movie sentiment score (movie overall sentiment score) and movie star power (summation of the number of followers on Twitter of the three highest-billed movie stars according to IMDB. Mulay, Joshi, & Shaha (2016) use sentiment analysis and opinion mining on social network websites for predicting movie box office revenues. Mulay et al. (2016) use a system assigning weights to Tweets based upon criteria such as followers of actors, directors, producers, sentiments and the number of Tweets containing the movie hashtags were used for predicting movie box office revenues. Yao & Chen (2013) apply sentiment analysis and machine learning methods to study the relationship between online reviews for a movie and movie's box office revenue performance. They find that a simplified version of the sentiment-aware autoregressive model is able to accurately predict box office sales by using online review data (Yao & Chen, 2013). Yao & Chen (2013) create a classification model using a Naive Bayes classifier for predicting box office revenue trends from review sentiment data. In addition, Yao & Chen (2013) argue that online reviews are good indicators for predicting a movie's box office revenue. Contrary to the other research findings, according to Apala et al. (2013), the sentiment score of YouTube viewers' comments was not found to be a relevant predictor of movie box office revenue.

Appendix 4 – Instagram

Introduction

This section discusses the social networking service Instagram in general. Instagram was launched in October 2010 (Araujo, Correa, Silva, Prates, & Meira, 2014). Instagram is generally viewed as a social photo sharing service and is classified as an image-based social media platform (Pittman & Reich, 2016) and as an ephemeral social media application (Wakefield & Bennett, 2016). Ha, Kwon, Cha, & Joo (2017) view Instagram as an virtual-centric social media. Carah & Shaul (2016) define Instagram in terms of being an ʻʻimage machine that captures and calibrates attentionʼʼ (p. 69). More specifically, according to Araujo et al. (2014), Instagram ʻʻincludes dedicated mobile applications that allows users to take and manipulate photographs by adding filters and frames, and to share them online where other users can react through comments and likesʼʼ (p. 19). On average, 60 million photographs are posted per day. So based on this, it can be said that Instagram is one of the most popular applications for photo sharing and interacting with friends, acquaintances and businesses (Araujo et al., 2014).

Currently, Instagram continues to evolve and grow rapidly (Mittal et al., 2017) and is the world's largest photo sharing platform (Zulli, 2017). This is further encouraged by the fact that visual social media networking services are easily accessible, simple to understand, to use and to enjoy (Zulli, 2017). The rise of Instagram is illustrated by the number of monthly active users which was 90 million in January 2013 and was already 800 million in September 2017,[5] showing a more steadily increase in the number of users than social networking services like Facebook and Twitter. In comparison, in quartile 4 of 2017, the number of monthly active users on Instagram was already 3 times larger than the number of monthly active users on Twitter.[6] According to Zulli (2017), the ongoing increasing number of users solidifies the statement that Instagram is the ʻʻgo-to platform for storytelling around the globeʼʼ (p. 140).

[5] Statista (2018). Number of monthly active Instagram users from January 2013 to September 2017 (in millions). Retrieved from https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/

[6] Statista (2018). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2018 (in millions). Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Instagram is seen as an image rich application that has the potential to influence consumers' behaviors and motivations differently than any previous social networking site (Lee et al., 2015). In a recent study, Retaildive (2017)[7] found that Instagram influences almost 75% of user purchase decisions. Additionally, 50% of the respondents argue that Instagram influenced their shopping habits most strongly, followed by Facebook (23%), Pinterest (22%), Twitter (3%) and Snapchat (1%). Chen (2017) argues that young consumers are more likely to consider marketing information that their friends posted (Chen, 2017). In addition, Instagram users are more like to pay more attention to the products recommended by friends (Chen, 2017). According to Bloglovin (2016)[8], micro-influencers state that Instagram is the best social media platform for engaging audiences. This is due to Instagram's focus on images, its simplicity, and the combination of large audiences and the tight (relative to Twitter) personal networks it provides (Shane-Simpson et al., 2018). Pittman & Reich (2016) find that image-based platforms like Instagram conform to the need of individuals for communicating thoughts and feelings quicker and more effectively than text-based media, like for example Twitter.

According to Pew Research Center (2018)[9] and Lee et al. (2017), Instagram is one of the best platforms to advertise on, because their findings show that engagement with brands on Instagram is 10 times higher than Facebook, 54 times higher than Pinterest, and 84 times higher than Twitter.

Compared to Facebook, Twitter and Snapchat, the main differences between the features of Instagram are that you can post filtered photographs and short videos, like and comment on photos and short videos, add captions and use hashtags in order to search topics (Phua et al., 2017). According to Coelho, Oliveira, & Almeida (2016), specific characteristics are that postings can exclusively be posted with smartphones and tablets.

[7] Retaildive (2017, August 23). Study: Instagram influences almost 75% of user purchase decisions. Retrieved from https://www.retaildive.com/news/study-instagram-influences-almost-75-of-user-purchase-decisions/503336/

[8] Bloglovin (2016). We asked, they answered: How micro-influencers really want to work with brands. Retrieved from: https://www.adweek.com/digital/influencers-instagram-is-the-most-engaging-platform-report/

[9] Pew Research Center (2018). Social media use in 2018. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

Instagram use
Instagram is one of the fastest-growing online photo social web services and compared to similar applications (Sheldon & Bryant, 2016). Instagram, as a social media platform, is used by individuals, companies, vendors and interest groups, where people can join in and upload photos or pictures (Zulli, 2017). This results into increases in the spreading of information and that businesses use Instagram as a social media platform to more efficiently reach and communicate with actual and potential consumers (Guidry et al., 2018).

Pew Research Center (2018)[10] found that 32% of online adults use Instagram of which female internet users are more likely to use Instagram than men (Mittal et al., 2017). According to data from Pew Research Center (2018)[11], Instagram is used by 26% men and 38% women. Instagram is most used by people in the age between 18 to 29 years old (59%), followed by 30 to 49 years old (33%), 50 to 64 years (18%) and 65 years and older (8%) (Pew Research Center, 2018).[12] Besides that younger people are strongly represented on Instagram, Shane-Simpson et al. (2018) find that emerging adults are also attracted by Instagram, because of its relative novelty and its focus on visual communication. Today, young people may rely more on visual communication than older people do (Shane-Simpson et al., 2018). Instagram is especially popular among young people because it enables immediate sharing of images (Stapleton et al., 2017), which provides information for social comparison.

Participants in the study of Shane-Simpson et al. (2018) most often mentioned the modes of communication and types of interactions available as the reasons for site preference. People who preferred Instagram identified the visual imagery available on Instagram as a primary reason for their preference, whereas those who preferred Twitter or Facebook were less likely to indicate that visual images were a reason for their preference (Shane-Simpson et al., 2018). Compared to Facebook and Twitter, Chen (2017) argues that Instagram is used because of the opportunities with regard to visual engagement. Convenience is another key advantage of Instagram over other types of social networking sites (Chen, 2017).

[10] Pew Research Center (2018). Social media use in 2018. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

[11] Pew Research Center (2018). Social media use in 2018. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

[12] Pew Research Center (2018). Social media use in 2018. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

More specifically, convenience refers to both the content and format (Chen, 2017). Furthermore, Chen (2017) argues that Instagram is able to create intimate co-presence. Co-presence refers to the perception of the presence of other social actors in a particular contextual environment through multiple paths of connections (Kim & Chock, 2015).

People who seek attention from others online, for example people with heightened narcissistic traits may be drawn to social networking sites like Instagram (DeWall, Buffardi, Bonser, & Campbell, 2011). Instagram enables them to broadcast images of themselves to large audiences (Shane-Simpson et al., 2018). Due to its visual properties, Instagram is often used for self-promotional content (Sheldon & Bryant, 2016).

With regard to the motivations of using Instagram, Gibbs et al. (2014) and (Marwick, 2015) found that Instagram is often used for self-promotion or self-branding. Besides self-promotion, motivations of using Instagram are primarily grounded in documentary and posting images in the here and now (Gibbs et al., 2014; Marwick, 2015). Or as Mittal et al. (2017) put it, Instagram is a way to capture and share the moment application. Instagram is becoming a popular tool to search for friends and to maintain relationships (Lee & Sin, 2016). Jang, Han, Shih, & Lee (2015) compared teenager and adult user in Instagram and find that teenage users posted selfies and engaged in self-representation activities to develop and maintain social connection with friends highlighting a strong need for personal identity.

Lee et al. (2015) identify five social and psychological motives for using Instagram: social interaction, archiving, self-expression, escapism and peeking. Furthermore, Han, Lee, Jang, Jung, & Lee (2016) argue that there can be distinguished between three types of activities on Instagram, namely content-based, interaction-based and relation-based activities. First, content-based activities refer to posts including photos, tags and locations. Second, interaction-based activities refer to actions like for example sharing content, likes or comments. Users can ‘’like’ or ‘’comment’’ on Instagram posts or short videos (Turner & Lefevre, 2017). Third, relation-based activities refer to actions resulting from following users (Han et al., 2016). Connections on Instagram are nonreciprocal, which means that users can follow users without that the other person follows the user (Hu, Manikonda, & Kambhampati, 2014). Instagram users can follow an unlimited number of accounts and can view content posted by the users they follow (Turner & Lefevre, 2017). Further, Instagram is able to propose new accounts to follow based on the content the user is already exposed to (Turner & Lefevre, 2017). Ting & De Run (2015) argue that Instagram is used because of personal gratification, features usefulness, the socializing role, collecting product information and for entertainment purposes. Chen (2017) states that Instagram is used for seeking information about products, companies, and services through online social interaction. Compared to Facebook, Twitter and Snapchat users, Phua et al. (2017) argue that Instagram users scored highest for showing affection, following fashion (using Instagram as a style guide) and demonstrating sociability. In addition, Phua et al. (2017) state that Instagram users had the highest brand community engagement and commitment. Furthermore, Instagram users further show the highest degrees of belongingness and pride with regard to the brand pages. Pittman & Reich (2016) argue that Instagram is generally used for social activities. Marwick (2015) argues that Instagram allows average individuals to achieve social recognition.

Content

Instagram provides a venue where communication can take place among its users with a richer graphic conversation via the exchanges of photos and videos (Bashir et al., 2018). Shane-Simpson et al. (2018) find that Instagram was generally selected as the favorite social media platform, because of its visual affordances. Instagram can be viewed as a media where content sharing is persistent (Bayer, Ellison, Schoenebeck, & Falk, 2016). Users can organize, use, document and remember in a persistent manner (Bayer et al., 2016). According to Turner & Lefevre (2017), Instagram posts are dominated by the content of the images. Hu et al. (2014) state that the pictures that are shared on Instagram can be classified into eight different photo categories, namely friends, food, gadgets, captioned photos, pets, activity, selfies and fashion. Araujo et al. (2014) find that people show tendencies that they endorse pictures with a lot of likes and comments, which results into the so-called 'the rich get richer' phenomenon. Bakhshi et al. (2014) argue that photos with faces attract more likes and comments on Instagram. Tobin & Chulpaiboon (2016) argue that comments on Instagram posts are satisfying to the user and result into user happiness.

Instagram has evolved into a unique social media platform where users can document their lives through visual content such as photos and videos (Bashir et al., 2018). According to Geurin & Burch (2017), Instagram serves as a kind of a journal of everyday life. Instagram is used by people as a platform to interact with each other, to share personal photos, videos, views and reviews on different topics of daily life, politics, sports, markets and much more (Mittal et al., 2017). In addition, Instagram is also used for encouraging social and political activism (Zulli, 2017). Olszanowski (2014) argues that Instagram is particularly useful for encouraging social and political activism because followers are guaranteed to see the images that are posted. Lee & Sin (2016) find that viewing photographs on Instagram can help users to acquire news and new information. In addition, viewing photographs on Instagram helps to satisfy the needs of users to find out what other users are doing (Lee & Sin, 2016). Photographs posted on Instagram typically disclose detailed information about the user, like social activities, social circles and hobbies (Lee & Sin, 2016). The content from photos viewed is also valuable to provide more contextual information about the social network environment (like events, activities and latests happenings) and creating a sense of belonging (Lee & Sin, 2016). Instagram is a platform that enable viewers to peek into the personal lives of other users to satisfy their curiosities (Lee & Sin, 2016).

Social information encountered on Instagram differs from social information on other social networking services with regard to the centrality of images and that users are especially likely to view posts from strangers (Lup, Trub, & Rosenthal, 2015; Pittman & Reich, 2016). Compared to text-central social media such as Twitter, image-based social networking sites have a differential impact on the mood of viewers (Johnson & Knobloch-Westerwick, 2017). Further, posts on Instagram are likely to be positively biased (Lup et al., 2015), which means that individuals engage in self-presentation and tend to select and emphasize the most positive aspects of themselves and their lives (de Vries et al., 2018). Turner & Lefevre (2017) find that higher Instagram use was associated with a greater tendency towards orthorexia nervosa[13]. Other social media channels were not able to produce similar effects. The findings of Turner & Lefevre (2017) highlight the implications Instagram can have on psychological well being and the influence social media 'celebrities' may have over hundreds of thousands of individuals. Because of the fact that Instagram is an image-based platform, users may be more likely to follow advice or imitate diets of Instagram 'celebrities' because of that they feel a more personal connection than they would on a text-based platform (Turner & Lefevre, 2017).

[13] Orthorexia nervosa is defined as an unhealthy obsession with eating healthy food (Turner & Lefevre, 2017).

Appendix 5 – Twitter

Introduction

Twitter was launched in 2006, and broke into the mainstream in 2008 and 2009 (Marwick & Boyd, 2011). According to Kwak, Lee, Park, & Moon (2010), due to its low level of reciprocal connections and its focus on information sharing, Twitter is classified as a micro-blogging site rather than a social network site. Twitter emphasizes on text-based information sharing through short Tweets (Shane-Simpson et al., 2018). Tweets can be posted and read on desktop computers, smartphones and other devices (Marwick & Boyd, 2011). These methods allow for instant postings of photos, reports and quick replies to other users (Marwick & Boyd, 2011).

Twitter is a text-based social medium (Pittman & Reich, 2016), nevertheless users can attach photos and videos to their posts (Chae, 2018). According to Chae (2018), with the 140 characters that Twitter allows, it is hard to illustrate one's life. Marwick & Boyd (2011) argue that Twitter creates a constantly updated timeline, or stream, of short messages that range from humor and musings on life to links and breaking news. Besides Tweets, private direct messages can be sent to other individuals on Twitter (Marwick & Boyd, 2011). Tweets can also be retweeted, which means that they are spread further when users repost Tweets through their accounts (Marwick & Boyd, 2011). In addition, users can also cite the author in a Tweet by using the @username function (Marwick & Boyd, 2011). Just like blogs, Twitter shows user' Tweets in reverse chronological order (Boyd, Golder, & Lotan, 2010).

Twitter use

Twitter allows users to follow Twitter accounts in their streams and other users can also be followed (Marwick & Boyd, 2011). Technical requirements of reciprocity are absent, which allows Twitter users to maintain non-reciprocal relationships (Marwick & Boyd, 2011). Shane-Simpson et al. (2018) find that respondents in their study who preferred Twitter seemed to be more individualistic in terms of values for self-expression and choice.

According to Pew Research Center (2018)[14], 24% of the users on Twitter is a male user, whereas 25% of the users are female (Pew Research Center, 2018). [15] Furthermore, 36% of the Twitter users is 18 to 29 years old, followed by 23% of the users being 30 to 49 years old, 21% of the users being 50 to

[14] Pew Research Center (2018). Social media use in 2018. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

[15] Pew Research Center (2018). Social media use in 2018. Retrieved from http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

64 yours and 10% of the users beings 65 years and older. Java, Song, Finin, & Tseng (2007) identified three main categories of Twitter users: information sources, friends, and information seekers.

Lee & Ma (2012) and Pittman & Reich (2016) find that Twitter users pursuing gratifications like information seeking. Phua et al. (2017) find that Twitter users had the highest brand community identification and membership intention. According to Tumasjan, Sprenger, Sandner, & Welpe (2010) and Watson (2016), Twitter is used more for communicating news about politics, sports, and natural disasters than presenting themselves. Additionally, Phua et al. (2017) argue that Twitter is also used for following news. Furthermore, Ausserhofer & Maireder (2013) argue that Twitter is used by politicians and journalists to spread information. Hughes, Rowe, Batey, & Lee (2012) argue that Twitter may appeal to people seeking intellectual stimulation.

Appendix 6 – Twitter and Instagram

Instagram and Twitter share some similaries, but there are also some differences. A discussion with regard to the similarities and differences is presented in this section.

The affordances of Instagram and Twitter encourage a one-to-many form of social interaction, and may facilitate the formation and maintenance of looser, distal ties, or a mix of close and loose ties wherein reciprocity is more arbitrary (Shane-Simpson et al. 2018). Both on Instagram and Twitter, an option to to 'friend' people is lacking. More specifically, Instagram and Twitter provide users with opportunities 'follow' and 'follow' back other users (Shane-Simpson et al. 2018). Shane-Simpson et al. (2018) argue that reciprocal relationships are more common on Instagram than Twitter. According to Manikonda, Meduri, & Kambhampati (2016), Twitter and Instagram differ in terms of their image. Manikonda et al. (2016) argue that Twitter was viewed more as having informational content, while Instagram was seen as being more personal and social in nature. Furthermore, Pittman & Reich (2016) state that Twitter is an text-based social networking site, whereas Instagram is a image-based social networking site. Turner & Lefevre (2017) assume that their findings - that Instagram has an association with orthorexia nervosa and that Twitter has a small positive association with orthorexia nervosa - can be attributed to the fact that Twitter is an text-focused social networking site and the strictly character-limited nature of Twitter. Turner & Lefevre (2017) argue that the strong influences of Instagram on users can be explained by the different factors. First, Instagram is an image-based platform, which plays to the picture superiority effect - whereby images are more likely to be remembered than words (Childers & Houston, 1984; Turner & Lefevre, 2017). Second, users choose which account they wish to follow, and so are then continually exposed to the type of content these accounts produce (Turner & Lefevre, 2017). Due to this limited exposure, this may result into users believing a behavior is more prevalent or normal than is actually the case, and this may result into perceived social pressures to conform to such behaviors (Turner & Lefevre, 2017). Further, Chae (2018) find that the use of Instagram is positively associated with social comparison with friends, whereas Twitter is negatively associated with friends. The findings of Chae (2018) suggest that Instagram is used for social comparison with friends, whereas Twitter is generally not used for social comparison with friends. Stapleton et al. (2017) and de Vries et al. (2018) also found that Instagram and social comparison are associated. Chae (2018) find a negative association between Twitter and social comparison. According to Chae (2018), it seems that Twitter is more about communicating news rather than self-presentation. In addition, observing others' self-presentation is less likely to be the main goal of Twitter users (Chae, 2018).

Appendix 2 provides a detailed overview of literature with regard to Instagram and Twitter and predictions

Appendix 7 – Movies and release dates

Movie	Release date
12 strong	1 februari 2018
Black Panther	14 februari 2018
Bob de Bouwer	15 februari 2018
Den of thieves	22 februari 2018
Diep in de zee	24 januari 2018
Downsizing	25 januari 2018
Early man	7 februari 2018
Every day	22 februari 2018
Fifty shades Freed	8 februari 2018
Game night	22 februari 2018
I Tonya	22 februari 2018
Maze Runner The Death Cure	25 januari 2018
Médecin de Campagne	25 januari 2018
Patser	1 februari 2018
Shape of water	15 februari 2018
Ted en het geheim van koning Midas	14 februari 2018
The Florida Project	8 februari 2018
Winchester The House that ghosts built	8 februari 2018
24 hours to live	18 januari 2018
All the money in the world	11 januari 2018
Call me by your name	11 januari 2018
Darkest hour	18 januari 2018
Father figures	18 januari 2018
Insidious: The last key	11 januari 2018
Taal is zegmaar echt mijn ding	18 januari 2018
The commuter	11 januari 2018
The greatest showman	4 januari 2018

The leisure Seeker	4 januari 2018
Three Billboards outside Ebbing Missouri	11 januari 2018
You were never really here	4 januari 2018
Wild	1 februari 2018
The Post	1 februari 2018
Death Wish	1 maart 2018
Phantom Thread	1 maart 2018
Red Sparrow	1 maart 2018
De Wilde Stad	1 maart 2018
Bankier van het verzet	8 maart 2018
Film stars don't die in Liverpool	8 maart 2018
Mannen van Mars	8 maart 2018
The Strangers 2: Prey at Night	8 maart 2018
Sweet Country	15 maart 2018
Tomb Raider	15 maart 2018

Coosto search queries (keywords) per movie

The Greatest Showman

The Greatest Showman

Greatest Showman

Greatest Showman film

The Greatest Showman movie

Diep in de zee

Diep in de zee

Diep in de zee film

Diep in de zee bios

Diep in de zee movie

Downsizing

Downsizing movie

Downsizing

Downsizing film

Downsizing bios

Médecin de campagne

Medecin de campagne

Médecin de campagne

Médecin de campagne bios

Médecin de campagne movie

Médecin de campagne film

Patser

Patser

Patser film

Pater bios

Patser movie

Early Man

Early Man

Early Man movie

Early Man film

Early Man bios

Fifty Shades Freed

Fifty shades freed

Fifty shades

Fifty shades movie

Mr Grey

Fifty Shades film

Fifty Shades bios

The Florida Project

The Florida Project

Florida Project movie

Florida Project bios

Florida Project film

Winchester: The House that Ghosts built

Winchester film

Winchester movie

Winchester bios

House that ghosts built

Ghosts built film

Ghosts built movie

Ghosts built bios

Winchester ghosts movie

Winchester ghosts bios

Winchester ghosts film

Bob de Bouwer - Mega Machines

Bob de bouwer

Bob de bouwer mega machines

Bob de bouwer film

Bob de bouwer movie

Bob de bouwer bios

Mega machines film

Mega machines movie

Mega machines film

Den of Thieves

Den of thieves

Den of thieves film

Den of thieves movie

Den of thieves bios

Every Day

Every day film

Every day movie

Every day bios

Wild

Wild film

Wild bios

Wild movie

The Post

The Post film

The Post bios

The Post movie

Game Night

Game night film

Game night movie

Game night bios

I Tonya

I Tonya

Tonya film

Tonya bios

Tonya movie

The Shape of Water

Shape of water

Shape of water movie

Shape of water bios

Shape of water film

Ted en het geheim van Koning Midas

Ted geheim koning midas

Ted midas film

Ted koning film

Ted koning bios

Ted koning movie

Ted geheim movie

Ted geheim bios

Ted koning midas film

Ted koning midas movie

Ted koning midas bios

Koning midas

Koning midas film

Koning midas bios

Koning midas movie

Black Panther

Black panther

Black panther film

Black panther movie

Black panther bios

12 strong

12 strong bios

12 strong film

12 strong movie

12 strong

Maze Runner: The Death Cure

Death cure film

Death cure movie

Death cure

Maze runner film

Maze runner movie

Maze runner the death cure

Maze runner

The death cure

The Leisure Seeker

The Leisure Seeker

Leisure Seeker

Leisure Seeker film

Leisure Seeker movie

You Were Never Really Here

You Were Never Really Here

You Were Never Really Here film

You Were Never Really Here movie

All the Money in the World

All the Money in the World

All Money in the World

All the Money in the World movie

All the Money in the World film

Call Me By Your Name

Call Me By Your Name

Call Me By Your Name movie

Call me By Your Name movie

Insidious: The Last Key

Insidious The Last Key

Insidious Last Key

Insidious The Last Key movie

Insidious The Last Key film

The Commuter

The Commuter

The Commuter film

The Commuter movie

Three Billboards Outside Ebbing, Missouri

Three Billboards Outside Ebbing

Three Billboards Outside Ebbing Missouri

Three Billboards Outside Ebbing film

Three Billboards Outside Ebbing movie

Billboards Missouri

24 hours to live

24 hours to live

Hours to live movie

Hours to live bios

Hours to live film

Darkest hour

Darkest hour movie

Darkest hour

Darkest hour bios

Darkest hour film

Father figures

Father figures

Father figures film

Father figures movie

Father figures bios

Taal is zeg maar echt mijn ding

Taal is zeg maar echt mijn ding taal is mijn ding film

Taal is mijn ding movie

Taal is mijn ding

Taal is zeg maar echt mijn ding

Death wish

Death Wish film

Death Wish movie

Death wish bios

Death Wish

Phantom Thread

Phantom Thread

Phantom Thread film

Phantom Thread bios

Phantom Thread movie

Red Sparrow

Red Sparrow

Red Sparrow film

Red Sparrow bios

Red Sparrow movie

De Wilde Stad

De Wilde Stad

De Wilde Stad film

De Wilde Stad bios

De Wilde Stad movie

Bankier van het verzet

Bankier van het verzet

Bankier van het verzet film

Bankier van het verzet movie

Bankier van het verzet bios

Film Stars Don't Die in Liverpool

Film stars don't die in liverpool

Film stars dont' die film

Film stars don't die movie

Film stars don't die bios

Mannen van Mars

Mannen van mars

Mannen van mars bios

Mannen van mars film

Mannen van mars movie

The Strangers 2: Prey at Night

The strangers 2 prey at night

The strangers prey at night

The strangers film

The strangers movie

The strangers bios

Prey at night film

Prey at night movie

Prey at night bios

Sweet Country

Sweet Country

Sweet Country film

Sweet Country bios

Sweet Country movie

Tomb Raider

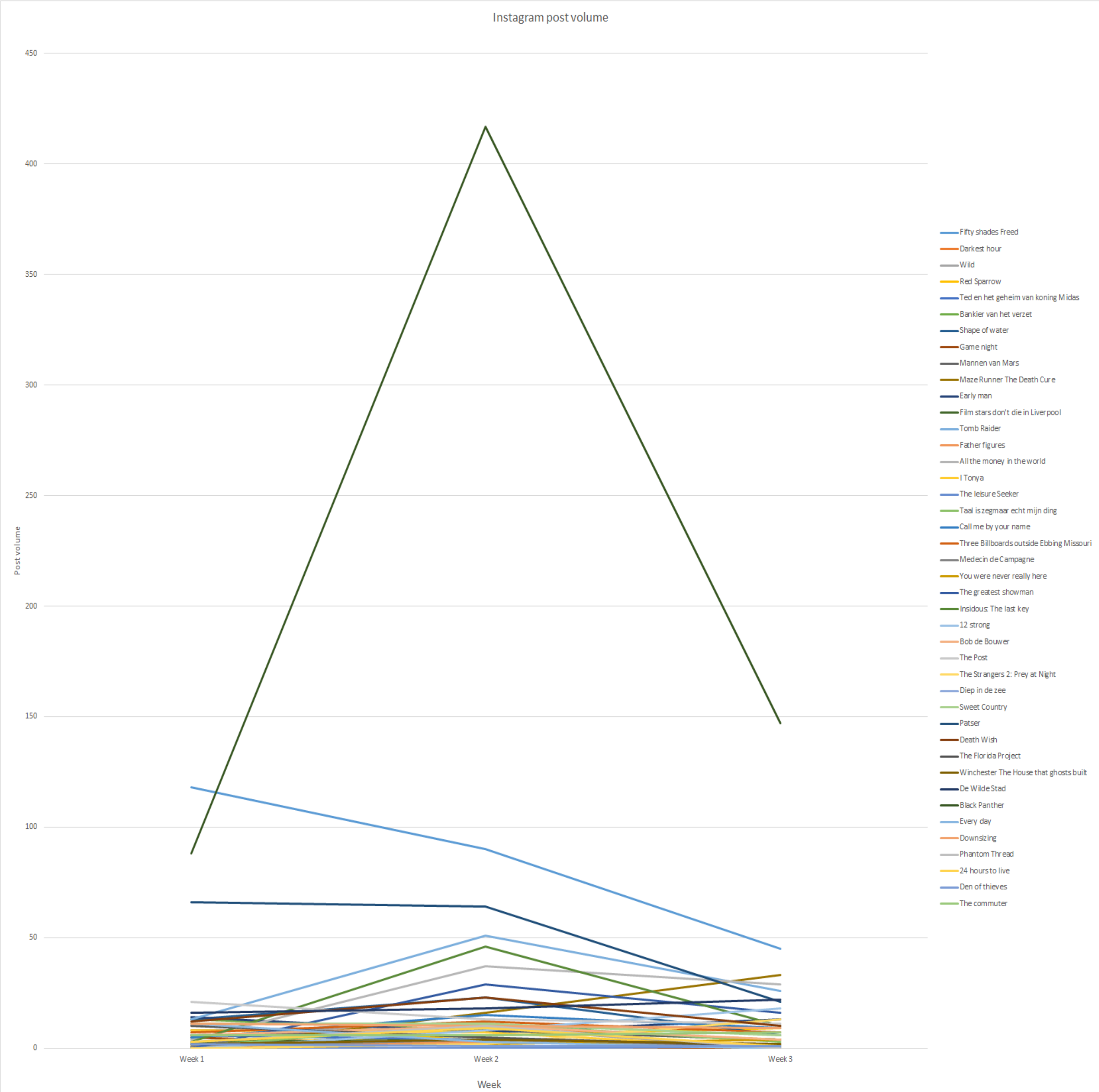
Tomb Raider

Tomb Raider film

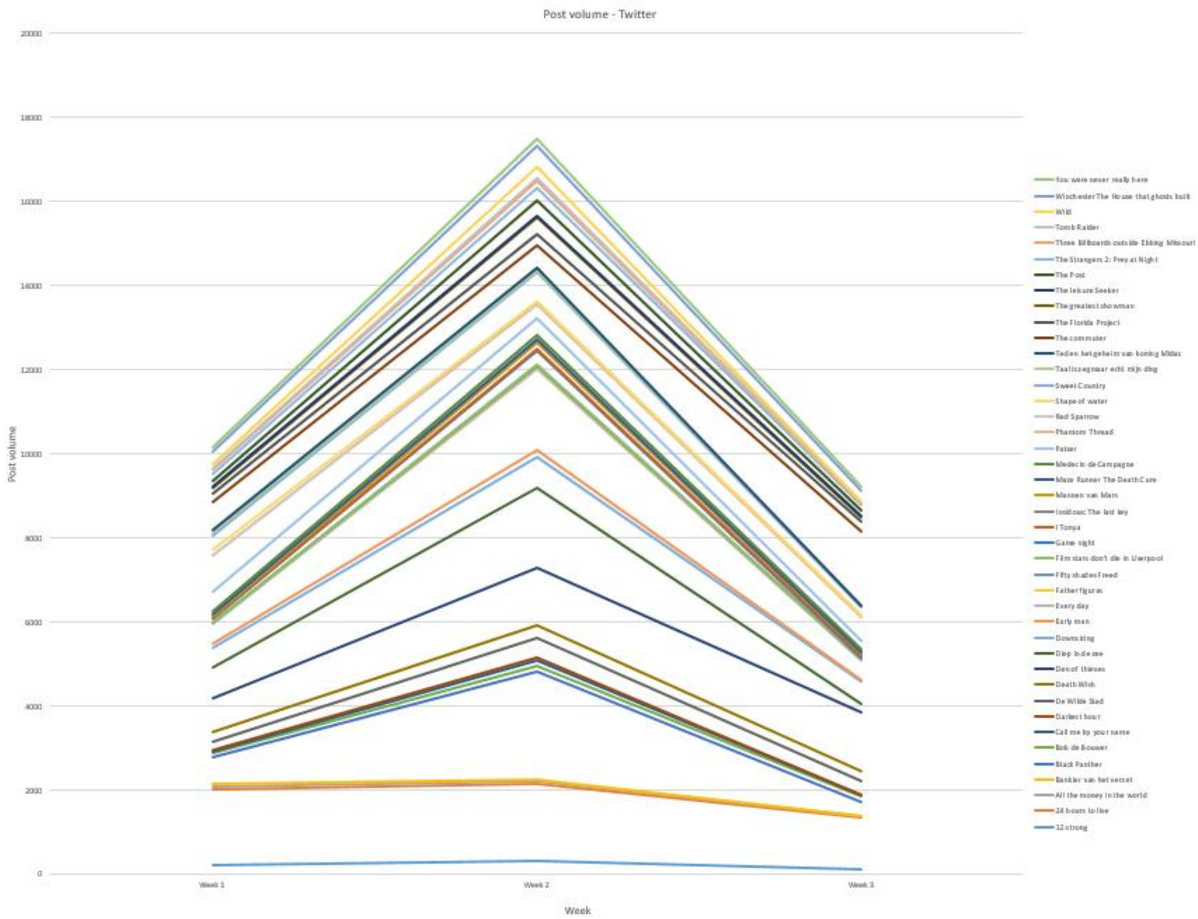
Tomb Raider movie

Tomb Raider bios

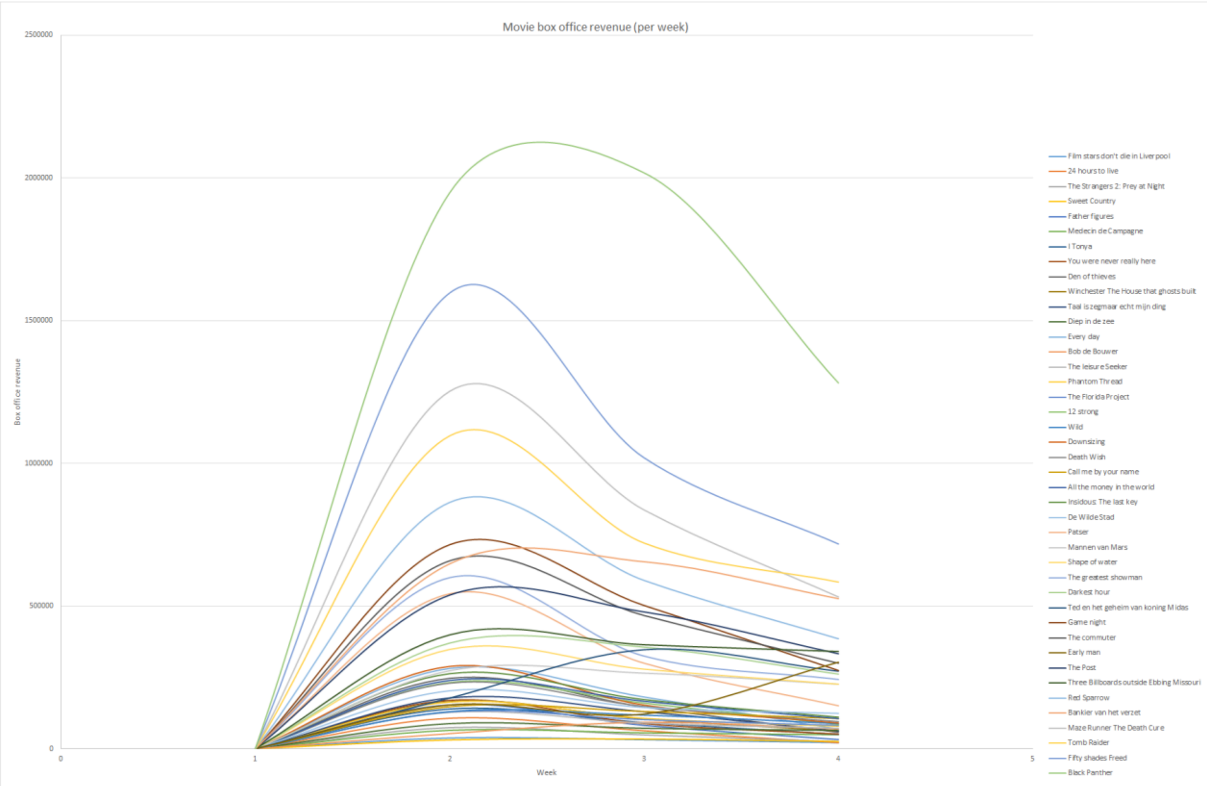
Appendix 8 – Weekly trajectories of volume-related Instagram variables



Appendix 9 – Weekly trajectories of volume-related Twitter variables



Appendix 10 – Weekly movie box office revenues



Appendix 11 – Correlation analysis

it can be said that post volume on Instagram in week 1, 2 and 3 shows strong positive and significant correlations at a .01 level with movie box office revenue in week 2 (.693**, .700** and .772**), week 3 (.675**, .824** and . 856**) and week 4 (.660**, .777** and .819**). The r's of .824** and .856** can be seen as a very strong positive and significant correlation, the other reported r's can be seen as strong correlation (Hair et al., 2010). These r scores suggests that an increase in the volume of posts on Instagram results into strong to very strong increases in the weekly movie box office revenue. Twitter post volume shows strong and moderate positive correlations at a .01 level with movie box office revenue in week 2 (.567**, .447** and .489**), week 3 (.767**, .689** and .698**) and week 4 (.518**, .417** and .449**). These r scores suggest that an increase in the volume of posts on Twitter results into a moderate to strong increase the weekly movie box office revenue. The number of Instagram post views in week 1, 2 and 3 show negative and very weak (-.035, -.036 and -.057), positive weak (.227, .179 and .224) and significant moderate (.586**, significant at .01 level) and strong correlations (.610** and .598**, significant at .01 level) with movie box office revenues in week 2,3 and 4. This suggests that increases in the number of posts on Instagram results into respectively almost no increases (very weak correlations and weak correlations) and moderately increases in the weekly movie box office revenues. More specifically, only the number of post views in week 1 on Instagram ensures moderately increases in the weekly movie box office revenue. The number of post views on Twitter in week 1, 2 and 3 show respectively positive moderate correlations (.495**, .428** and .431**, significant at .01 level), positive very strong correlations (.812**, .816** and .812**, significant at .01 level) and negative very weak correlations (-.054, -.056 and -.077) with movie box office revenues in week 2, 3 and 4. This suggests that increases in the number of posts views on Twitter results into respectively moderate, very strong increases and very weak decreases in the weekly box office revenues. Positive sentiments in textual posts on Instagram in week 1, 2 and 3 show respectively very weak negative (-.086, -.056 and -.032) and very weak positive correlations (.176, .145, .148, .037, .019 and .030) with movie box office in week 2, 3 and 4. This suggests that increases in positive sentiment in texts on Instagram in week 1, 2 and 3 results into almost no impact on movie box office in week 2, 3 and 4. Negative sentiment in Instagram posts in week 1, 2 and correlates negatively and very weak (-.002, -.058, -.109, -.116, -1.36, -.160, -.042, -0.65 and -.098) with box office revenues. This suggest that increases in the negative sentiment in textual posts on Instagram results into very weak decreases in the weekly box office revenues. Positive sentiment in textual posts on Twitter in week 1, 2 and 3 shows positive weak (.079, .074, .172, .158, .107 and .151) and negative very weak correlations (.-126, -.103 and -.076) with movie box office revenue in week

1,2 and 3. These r scores suggest that increases or decreases in the number of positive sentiment in textual posts on Twitter does not really result into changes in the weekly movie box office revenues. Negative sentiment in textual posts on Twitter correlates weak to very weak and positive (.246, .153 and .098), and very weakly and negative (-.050, -.062, -.105, -.083, -.083 and -.084) with movie box office revenue in week 2, 3 and 4. This suggests that decreases and increases in the number of negative sentiment in textual posts on Twitter does not significantly result into changes in the weekly movie box office revenues. Negative sentiment in visual posts on Instagram correlates negatively weak and very weak in week 1 (-.002, , -.058, -.109), negatively and very weak in week 2 (-.116, -.136, -.160) and negatively and very weak in week 3 (-.042, -.065 and -.098) with movie box office revenue in week 2, 3 and 4. This suggests that if the ratio of negative sentiments becomes higher than this will almost not impact the amount of weekly box office revenue. Negative sentiment in visual posts on Twitter correlates positively weak and very weak in week 1 (.246, .153, .098), negatively and very weak in week 2 (-.126, -.103 and -.076) and positively strong in week 3 (.603, .602 and .598) with the weekly movie box office revenues. This suggests that in the third week if the negative sentiment ratio on visual increases than the weekly amount of movie box office revenue will increase. Positive visual sentiment on Instagram correlates positively and very weak in week 1 (.054, .083 and .128), positively weak and very weak in week 2 (.212, .161 and .155) and significantly positively moderate in week 3 (.426**, .423** and 451**, significant at 0.01) with movie box office revenue. This suggests that an increase in the positive ratio of visual sentiment in week 3 (the second post-release) results into moderately higher scores regarding the weekly movie box office revenues. Positive visual sentiment on Twitter correlates significantly positively weak and positively very weak in week 1 (.309*, .231 and .188, significant at 0.05), positively weak and very weak in week 2 (.207, .143 and .197), positively and weak and very weak in week 3 (.148, .205 and .204) with the movie box office revenues. The results suggest that higher scores with regard to positive visual sentiment on Twitter do not go hand in hand with significantly higher scores with regard to weekly movie box office revenues. Negative visual sentiment on Instagram correlates positively very weak in week 1 (.095, .243 and .246), positively very weak in week 2 (0.098, .243 and .246) and negatively very weak and positively very weak in week 3 (.-002, .038 and .151.) with weekly movie box office revenues. This suggests that higher scores with regard to negative visual sentiment on Instagram does not significantly change the amount of movie box office revenues. Negative visual sentiment on Twitter correlates negatively very weak and positively very weak in week 1 (-.011, .034 and .081), positively very weak in week 2 (.052, .013 and .040) and negatively very weak in week 3 (-.034, -.046 and -.023) with weekly box office revenues. This suggests that higher scores on negative visual sentiment on Twitter does not significantly change the amount of weekly box office revenues. The table below presents an overview of the numerical value of the correlation, the sign of the correlation, the strength of the correlation and thet statistical significance of the correlation for the independent variables.

Pearson correlations with two-tailed p values for Instagram and Twitter

Correlations Instagram

		Movie box office revenue in week 2	Movie box office revenue in week 3	Movie box office revenue in week 4
Volume of posts on Instagram in week 1	Pearson Correlation	,693**	,675**	,660**
	Sig. (2-tailed)	,000	,000	,000
	N	42	42	42
Volume of posts on Instagram in week 2	Pearson Correlation	,700**	,824**	,777**
	Sig. (2-tailed)	,000	,000	,000
	N	42	42	42
Volume of posts on Instagram in week 3	Pearson Correlation	,772**	,856**	,819**
	Sig. (2-tailed)	,000	,000	,000
	N	42	42	42
Number of post views on Instagram in week 1	Pearson Correlation	,565**	,610**	,598**
	Sig. (2-tailed)	,000	,000	,000
	N	42	42	42
Number of post views on Instagram in week 2	Pearson Correlation	-,035	-,036	-,057
	Sig. (2-tailed)	,823	,823	,718
	N	42	42	42
Number of post views on Instagram in week 3	Pearson Correlation	,227	,179	,224
	Sig. (2-tailed)	,148	,257	,155
	N	42	42	42
Positive sentiment in texts on Instagram in week 1	Pearson Correlation	-,086	-,056	-,032
	Sig. (2-tailed)	,590	,726	,841
	N	42	42	42
Positive sentiment in textual posts on Instagram in week 2	Pearson Correlation	,176	,145	,148
	Sig. (2-tailed)	,265	,359	,351
	N	42	42	42
Positive sentiment in textual posts on Instagram in week 3	Pearson Correlation	,037	,019	,030
	Sig. (2-tailed)	,818	,904	,849
	N	42	42	42
Negative sentiment in textual posts on Instagram in week 1	Pearson Correlation	-,002	-,058	-,109
	Sig. (2-tailed)	,989	,717	,491
	N	42	42	42
Negative sentiment in textual posts on Instagram in week 2	Pearson Correlation	-,116	-,136	-,160
	Sig. (2-tailed)	,466	,391	,313
	N	42	42	42
Negative sentiment in textual posts on Instagram in week 3	Pearson Correlation	-,042	-,065	-,098
	Sig. (2-tailed)	,792	,681	,535
	N	42	42	42

Pearson correlations with two-tailed p values for Instagram and Twitter

Positive sentiment in visual posts on Instagram in week1	Pearson Correlation	,054	,083	,128
	Sig. (2-tailed)	,735	,603	,418
	N	42	42	42
Positive sentiment in visual posts on Instagram in week 2	Pearson Correlation	,212	,161	,155
	Sig. (2-tailed)	,177	,310	,326
	N	42	42	42
Positive sentiment in visual posts on Instagram in week 3	Pearson Correlation	,426**	,423**	,451**
	Sig. (2-tailed)	,005	,005	,003
	N	42	42	42
Negative sentiment in visual posts on Instagram in week 1	Pearson Correlation	,095	,243	,246
	Sig. (2-tailed)	,549	,121	,116
	N	42	42	42
Negative sentiment in visual posts on Instagram in week 2	Pearson Correlation	,098	,112	,216
	Sig. (2-tailed)	,536	,481	,170
	N	42	42	42
Negative sentiment in visual posts on Instagram in week 3	Pearson Correlation	-,002	,038	,151
	Sig. (2-tailed)	,992	,810	,340
	N	42	42	42

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations Twitter

		Movie box office revenue in week 2	Movie box office revenue in week 3	Movie box office revenue in week 4
Volume of posts on Twitter in week 1	Pearson Correlation	,567**	,447**	,489**
	Sig. (2-tailed)	,000	,003	,001
	N	42	42	42
Volume of posts on Twitter in week 2	Pearson Correlation	,767**	,689**	,698**
	Sig. (2-tailed)	,000	,000	,000
	N	42	42	42
Volume of posts on Twitter in week 3	Pearson Correlation	,518**	,417**	,449**
	Sig. (2-tailed)	,000	,006	,003
	N	42	42	42
Number of post views on Twitter in week 1	Pearson Correlation	,495**	,428**	,431**
	Sig. (2-tailed)	,001	,005	,004
	N	42	42	42
Number of post views on Twitter in week 2	Pearson Correlation	,812**	,816**	,812**
	Sig. (2-tailed)	,000	,000	,000
	N	42	42	42
Number of post views on Twitter in week 3	Pearson Correlation	-,054	-,056	-,077
	Sig. (2-tailed)	,733	,722	,630
	N	42	42	42
Positive sentiment in textual posts on Twitter in week 1	Pearson Correlation	,079	,074	,172
	Sig. (2-tailed)	,618	,643	,277
	N	42	42	42
Positive sentiment in textual posts on Twitter in week 2	Pearson Correlation	-,126	-,103	-,076
	Sig. (2-tailed)	,426	,517	,634
	N	42	42	42
Positive sentiment in textual posts on Twitter in week 3	Pearson Correlation	,158	,107	,151
	Sig. (2-tailed)	,316	,500	,339
	N	42	42	42
Negative sentiment in textual posts on Twitter in week 1	Pearson Correlation	,246	,153	,098
	Sig. (2-tailed)	,116	,332	,536
	N	42	42	42
Negative sentiment in textual posts on Twitter in week 2	Pearson Correlation	-,050	-,062	-,105
	Sig. (2-tailed)	,755	,696	,510
	N	42	42	42
	Pearson Correlation	-,083	-,083	-,084

Negative sentiment in textual posts on Twitter in week 3	Sig. (2-tailed)	,603	,602	,598
	N	42	42	42
Positive sentiment in visual posts on Twitter in week 1	Pearson Correlation	,309	,231	,188
	Sig. (2-tailed)	,046	,142	,234
	N	42	42	42
Positive sentiment in visual posts on Twitter in week 2	Pearson Correlation	,207	,143	,197
	Sig. (2-tailed)	,189	,365	,211
	N	42	42	42
Positive sentiment in visual posts on Twitter in week 3	Pearson Correlation	,148	,205	,204
	Sig. (2-tailed)	,350	,192	,196
	N	42	42	42
Negative sentiment in visual posts on Twitter in week 1	Pearson Correlation	-,011	,034	,081
	Sig. (2-tailed)	,945	,831	,611
	N	42	42	42
Negative sentiment in visual posts on Twitter in week 2	Pearson Correlation	,052	,013	,040
	Sig. (2-tailed)	,741	,935	,803
	N	42	42	42
Negative sentiment in visual posts on Twitter in week 3	Pearson Correlation	-,034	-,046	-,023
	Sig. (2-tailed)	,832	,773	,884
	N	42	42	42

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Independent variable	Dependent variable	Source	Numerical value of the correlation	Sign of the correlation coefficient	Strength of the correlation (Evans, 1996)	Statistical significance of the correlation
Post volume week 1	Movie box office revenue week 2	Instagram	.693**	Positive	Strong	.01
Post volume week 1	Movie box office revenue week 3	Instagram	.675**	Positive	Strong	.01
Post volume week 1	Movie box office revenue week 4	Instagram	.660**	Positive	Strong	.01
Post volume week 1	Movie box office revenue week 2	Twitter	.567**	Positive	Moderate	.01
Post volume week 1	Movie box office revenue week 3	Twitter	.447*	Positive	Moderate	.01
Post volume week 1	Movie box office revenue week 4	Twitter	.489**	Positive	Moderate	.01
Post volume week 2	Movie box office revenue week 2	Instagram	.700**	Positive	Strong	.01
Post volume week 2	Movie box office revenue week 3	Instagram	.824**	Positive	Strong	.01
Post volume week 2	Movie box office revenue week 4	Instagram	.777**	Positive	Strong	.01

Post volume week 2	Movie box office revenue week 2	Twitter	.767**	Positive	Strong	.01
Post volume week 2	Movie box office revenue week 3	Twitter	.689**	Positive	Strong	.01
Post volume week 2	Movie box office revenue week 4	Twitter	.698**	Positive	Strong	.01
Post volume week 3	Movie box office revenue week 2	Instagram	.772**	Positive	Strong	.01
Post volume week 3	Movie box office revenue week 3	Instagram	.856**	Positive	Strong	.01
Post volume week 3	Movie box office revenue week 4	Instagram	.819**	Positive	Strong	.01
Post volume week 3	Movie box office revenue week 2	Twitter	.518**	Positive	Moderate	.01
Post volume week 3	Movie box office revenue week 3	Twitter	.417**	Positive	Moderate	.01
Post volume week 3	Movie box office revenue week 4	Twitter	.449**	Positive	Moderate	.01
Post views week 1	Movie box office revenue week 2	Instagram	.565**	Positive	Moderate	.01

Post views week 1	Movie box office revenue week 3	Instagram	.610**	Positive	Strong	.01
Post views week 1	Movie box office revenue week 4	Instagram	.598**	Positive	Strong	.01
Post views week 1	Movie box office revenue week 2	Twitter	.495**	Positive	Moderate	.01
Post views week 1	Movie box office revenue week 3	Twitter	.428**	Positive	Moderate	.01
Post views week 1	Movie box office revenue week 4	Twitter	.431**	Positive	Moderate	.01
Post views week 2	Movie box office revenue week 2	Instagram	-.035	Negative	Very weak	
Post views week 2	Movie box office revenue week 3	Instagram	-.036	Negative	Very weak	
Post views week 2	Movie box office revenue week 4	Instagram	-.057	Negative	Very weak	
Post views week 2	Movie box office revenue week 2	Twitter	.812	Positive	Very strong	.01
Post views week 2	Movie box office revenue week 3	Twitter	.816	Positive	Very strong	.01

Post views week 2	Movie box office revenue week 4	Twitter	.812	Positive	Very strong	.01
Post views week 3	Movie box office revenue week 2	Instagram	.227	Positive	Weak	
Post views week 3	Movie box office revenue week 3	Instagram	.179	Positive	Very weak	
Post views week 3	Movie box office revenue week 4	Instagram	.224	Positive	Weak	
Post views week 3	Movie box office revenue week 2	Twitter	-.054	Negative	Very weak	
Post views week 3	Movie box office revenue week 3	Twitter	-.056	Negative	Very Weak	
Post views week 3	Movie box office revenue week 4	Twitter	-.077	Negative	Very Weak	
Positive ratio textual sentiment week 1	Movie box office revenue week 2	Instagram	-.086	Negative	Very weak	
Positive ratio textual sentiment week 1	Movie box office revenue week 3	Instagram	-.056	Negative	Very Weak	
Positive ratio textual sentiment week 1	Movie box office revenue week 4	Instagram	-.032	Negative	Very Weak	

Positive ratio textual sentiment week 1	Movie box office revenue week 2	Twitter	.079	Positive	Very weak	
Positive ratio textual sentiment week 1	Movie box office revenue week 3	Twitter	.074	Positive	Very weak	
Positive ratio textual sentiment week 1	Movie box office revenue week 4	Twitter	.172	Positive	Very weak	
Positive ratio textual sentiment week 2	Movie box office revenue week 2	Instagram	.176	Positive	Very weak	
Positive ratio textual sentiment week 2	Movie box office revenue week 3	Instagram	.145	Positive	Very weak	
Positive ratio textual sentiment week 2	Movie box office revenue week 4	Instagram	1.48	Positive	Very weak	
Positive ratio textual sentiment week 2	Movie box office revenue week 2	Twitter	-.126	Negative	Very weak	
Positive ratio textual sentiment week 2	Movie box office revenue week 3	Twitter	-.103	Negative	Very weak	
Positive ratio textual sentiment week 2	Movie box office revenue week 4	Twitter	-.076	Negative	Very weak	

Positive ratio textual sentiment week 3	Movie box office revenue week 2	Instagram	.037	Positive	Very weak	
Positive ratio textual sentiment week 3	Movie box office revenue week 3	Instagram	.019	Positive	Very weak	
Positive ratio textual sentiment week 3	Movie box office revenue week 4	Instagram	.030	Positive	Very weak	
Positive ratio textual sentiment week 3	Movie box office revenue week 2	Twitter	.158	Positive	Very weak	
Positive ratio textual sentiment week 3	Movie box office revenue week 3	Twitter	.107	Positive	Very weak	
Positive ratio textual sentiment week 3	Movie box office revenue week 4	Twitter	.151	Positive	Very weak	
Negative ratio textual sentiment week 1	Movie box office revenue week 2	Instagram	-.002	Negative	Very weak	
Negative ratio textual sentiment week 1	Movie box office revenue week 3	Instagram	-.058	Negative	Very weak	
Negative ratio textual sentiment week 1	Movie box office revenue week 4	Instagram	-.109	Negative	Very weak	

Negative ratio textual sentiment week 1	Movie box office revenue week 2	Twitter	.246	Positive	Weak	
Negative ratio textual sentiment week 1	Movie box office revenue week 3	Twitter	.153	Positive	Very weak	
Negative ratio textual sentiment week 1	Movie box office revenue week 4	Twitter	.098	Positive	Very weak	
Negative ratio textual sentiment week 2	Movie box office revenue week 2	Instagram	-.116	Negative	Very weak	
Negative ratio textual sentiment week 2	Movie box office revenue week 3	Instagram	-.136	Negative	Very weak	
Negative ratio textual sentiment week 2	Movie box office revenue week 4	Instagram	-.160	Negative	Very weak	
Negative ratio textual sentiment week 2	Movie box office revenue week 2	Twitter	-.126	Negative	Very weak	
Negative ratio textual sentiment week 2	Movie box office revenue week 3	Twitter	-.103	Negative	Very weak	
Negative ratio textual sentiment week 2	Movie box office revenue week 4	Twitter	-.076	Negative	Very weak	

Negative ratio textual sentiment week 3	Movie box office revenue week 2	Instagram	-.042	Negative	Very weak	
Negative ratio textual sentiment week 3	Movie box office revenue week 3	Instagram	-.065	Negative	Very weak	
Negative ratio textual sentiment week 3	Movie box office revenue week 4	Instagram	-.098	Negative	Very weak	
Negative ratio textual sentiment week 3	Movie box office revenue week 2	Twitter	.603	Positive	Strong	
Negative ratio textual sentiment week 3	Movie box office revenue week 3	Twitter	.602	Positive	Strong	
Negative ratio textual sentiment week 3	Movie box office revenue week 4	Twitter	.598	Positive	Strong	
Positive ratio visual sentiment week 1	Movie box office revenue week 2	Instagram	.054	Positive	Very weak	
Positive ratio visual sentiment week 1	Movie box office revenue week 3	Instagram	.083	Positive	Very weak	
Positive ratio visual sentiment week 1	Movie box office revenue week 4	Instagram	.128	Positive	Very weak	

Positive ratio visual sentiment week 1	Movie box office revenue week 2	Twitter	.309*	Positive	Weak	0.05
Positive ratio visual sentiment week 1	Movie box office revenue week 3	Twitter	.231	Positive	Weak	
Positive ratio visual sentiment week 1	Movie box office revenue week 4	Twitter	.188	Positive	Very weak	
Positive ratio visual sentiment week 2	Movie box office revenue week 2	Instagram	.212	Positive	Weak	
Positive ratio visual sentiment week 2	Movie box office revenue week 3	Instagram	.161	Positive	Very weak	
Positive ratio visual sentiment week 2	Movie box office revenue week 4	Instagram	.155	Positive	Very weak	
Positive ratio visual sentiment week 2	Movie box office revenue week 2	Twitter	.207	Positive	Weak	
Positive ratio visual sentiment week 2	Movie box office revenue week 3	Twitter	.143	Positive	Very weak	
Positive ratio visual sentiment week 2	Movie box office revenue week 4	Twitter	.197	Positive	Weak	

Positive ratio visual sentiment week 3	Movie box office revenue week 2	Instagram	.426**	Positive	Moderate	0.01
Positive ratio visual sentiment week 3	Movie box office revenue week 3	Instagram	.423**	Positive	Moderate	0.01
Positive ratio visual sentiment week 3	Movie box office revenue week 4	Instagram	.451**	Positive	Moderate	0.01
Positive ratio visual sentiment week 3	Movie box office revenue week 2	Twitter	.148	Positive	Very weak	
Positive ratio visual sentiment week 3	Movie box office revenue week 3	Twitter	.205	Positive	Weak	
Positive ratio visual sentiment week 3	Movie box office revenue week 4	Twitter	.204	Positive	Weak	
Negative ratio visual sentiment week 1	Movie box office revenue week 2	Instagram	.095	Positive	Very weak	
Negative ratio visual sentiment week 1	Movie box office revenue week 3	Instagram	.243	Positive	Weak	
Negative ratio visual sentiment week 1	Movie box office revenue week 4	Instagram	.246	Positive	Weak	

Negative ratio visual sentiment week 1	Movie box office revenue week 2	Twitter	-.011	Negative	Very weak	
Negative ratio visual sentiment week 1	Movie box office revenue week 3	Twitter	.034	Positive	Very weak	
Negative ratio visual sentiment week 1	Movie box office revenue week 4	Twitter	.081	Positive	Very weak	
Negative ratio visual sentiment week 2	Movie box office revenue week 2	Instagram	.098	Positive	Weak	
Negative ratio visual sentiment week 2	Movie box office revenue week 3	Instagram	.243	Positive	Weak	
Negative ratio visual sentiment week 2	Movie box office revenue week 4	Instagram	.246	Positive	Weak	
Negative ratio visual sentiment week 2	Movie box office revenue week 2	Twitter	.052	Positive	Very weak	
Negative ratio visual sentiment week 2	Movie box office revenue week 3	Twitter	.013	Positive	Very weak	
Negative ratio visual sentiment week 2	Movie box office revenue week 4	Twitter	.040	Positive	Very weak	

Negative ratio visual sentiment week 3	Movie box office revenue week 2	Instagram	-.002	Negative	Very weak	
Negative ratio visual sentiment week 3	Movie box office revenue week 3	Instagram	.038	Positive	Very weak	
Negative ratio visual sentiment week 3	Movie box office revenue week 4	Instagram	.151	Positive	Very weak	
Negative ratio visual sentiment week 3	Movie box office revenue week 2	Twitter	-.034	Negative	Very weak	
Negative ratio visual sentiment week 3	Movie box office revenue week 3	Twitter	-.046	Negative	Very weak	
Negative ratio visual sentiment week 3	Movie box office revenue week 4	Twitter	-.023	Negative	Very weak	

Appendix 12 – Additional analyses

Analyses of additional variables for both Instagram and Twitter

1. *Instagram predictor variables in week 1 predicting movie box office revenues in week 1*

Instagram	Model				
Week 1 > week 1	1	2	3	4	5
	β	β	β	β	β
Post volume					.592***
Number of post views					
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio					-0,316*
Additional variables					
Posts by men in week 1					
Posts by men in week 1					
Number of theaters		.306*	.459*	.515***	.658***
Movie budget	.794***	.612***	.140	.568***	
Adjusted R^2	.617	.667	.721	.728	.803
Degrees of freedom	27	27	27	27	27
F	44,475	28,029	24,203	37,077	37,743
N	42	42	42	42	42

2. *Instagram predictor variables - Sequel*

Sequel (no sequel movies)	
Week 1 > week 1	1
	β
Post volume	
Number of post views	
Positive textual sentiment ratio	
Negative textual sentiment ratio	
Positive visual sentiment ratio	
Negative visual sentiment ratio	
Additional variables	
Posts by men in week 1	
Posts by men in week 1	
Number of theaters	
Movie budget	.800***
Adjusted R^2	.625
Degrees of freedom	26
F	44,406
N	42

3. *Age*

Age				
	6-9 years	12-16 years	16 +	16+
Week 1 > week 1	1	2	3	4
	β	β	β	β
Post volume		.982***		
Number of post views				
Positive textual sentiment ratio				
Negative textual sentiment ratio				
Positive visual sentiment ratio				
Negative visual sentiment ratio				.484**
Additional variables				
Posts by men in week 1				
Posts by men in week 1				
Number of theaters			.889***	.547**
Movie budget	.970			
Adjusted R^2	.921	.956	.771	.889
Degrees of freedom	4	5	12	12
F	47,333	110,675	41,328	48,817
N	42	42	42	42

4. *Movie genre*

Movie genre	
	Drama
Week 1 > week 1	1
	β
Post volume	
Number of post views	
Positive textual sentiment ratio	
Negative textual sentiment ratio	
Positive visual sentiment ratio	
Negative visual sentiment ratio	
Additional variables	
Posts by men in week 1	
Posts by men in week 1	
Number of theaters	.766*
Movie budget	
Adjusted R^2	.518
Degrees of freedom	7
F	8,536
N	42

5. *Instagram predictor variables in week 1 predicting movie box office revenues in week 2*

Instagram			
	Model		
Week 2 > week 2	1	2	3
	β	β	β
Post volume	.890***	.795***	.758***
Number of post views			
Positive textual sentiment ratio			
Negative textual sentiment ratio			
Positive visual sentiment ratio			
Negative visual sentiment ratio			.-297 ***
Additional variables			
Posts by men in week 2			
Posts by men in week 2			
Number of theaters		.245**	.390***
Movie budget			
Adjusted R^2	.784	.830	.902
Degrees of freedom	27	27	27
F	98,98	67,13	83,927
N	42	42	42

6. *Sequel*

Sequel (no sequel movies)		
Week 2 > 2	1	2
	β	β
Post volume	.905***	.850***
Number of post views		
Positive textual sentiment ratio		
Negative textual sentiment ratio		
Positive visual sentiment ratio		
Negative visual sentiment ratio		
Additional variables		
Posts by men in week 2		
Posts by women in week 2		.204*
Number of theaters		
Movie budget		
Adjusted R^2	.812	.845
Degrees of freedom	26	26
F	112,98	72,125
N	42	42

7. *Age*

Age					
	6-9 years	12-16 years	12-16 years	16+	16+
Week 2 > week 2	1	2	3	4	5
	β	β	β	β	β
Post volume		.862*			
Number of post views			1,501*		
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio			.803**		.304*
Negative visual sentiment ratio	.981***				
Additional variables					
Posts by men in week 2					
Posts by women in week 2				.820***	.854***
Number of theaters					
Movie budget					
Adjusted R^2	.949	.678	.966	.704	.784
Degrees of freedom	4	5	5	12	12
F	75,251	11,542	71,978	29,54	22,796
N	42	42	42	42	42

8. Movie genre

Movie genre	
	Drama
Week 2 > week 2	1
	β
Post volume	
Number of post views	
Positive textual sentiment ratio	
Negative textual sentiment ratio	
Positive visual sentiment ratio	
Negative visual sentiment ratio	
Additional variables	
Posts by men in week 2	
Posts by women in week 2	
Number of theaters	.717*
Movie budget	
Adjusted R^2	.433
Degrees of freedom	7
F	6,354
N	42

9. Instagram predictor variables in 1 predicting movie box office revenues in week 2

Instagram	Model				
Week 1 > week 2	1	2	3	4	5
	β	β	β	β	β
Post volume	.794***	.430**	.411***		
Number of post views					
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio					
Additional variables					
Posts by men in week 1			-0,325		
Posts by men in week 1					
Number of theaters			.298*		
Movie budget		.567***	.534***		
Adjusted R^2	.617	.746	.808	.863	
Degrees of freedom	27	27	27	27	
F	44,475	40,655	38,836	43,465	
N	42	42	42	42	42

10. Instagram predictor variables in week 2 predicting movie box office revenues in week 3

Instagram	Model				
Week 2 > week 3	1	2	3	4	5
	β	β	β	β	β
Post volume	.868***	.739***	.718***		
Number of post views					
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio			-0,172 *		
Additional variables					
Posts by men in week 2					
Posts by men in week 2					
Number of theaters		.332***	.416***		
Movie budget					
Adjusted R^2	.744	.835	.855		
Degrees of freedom	27	27	27		
F	79,472	69,276	53,817		
N	42	42	42	42	42

11. Twitter predictor variables in week 1 predicting movie box office revenues in week 1

Twitter	Model	
Week 1 > week 1	1	2
	β	β
Post volume		
Number of post views	.814***	.642***
Positive textual sentiment ratio		
Negative textual sentiment ratio		
Positive visual sentiment ratio		
Negative visual sentiment ratio		
Additional variables		
Posts by men in week 1		
Posts by men in week 1		
Number of theaters		.376**
Movie budget		
Adjusted R^2	.650	.757
Degrees of freedom	27	27
F	51,201	43,048
N	42	42

12. *Sequel*

Sequel (no sequel movies)

Week 1 > week 1	1	2	3	4	5
	β	β	β	β	β
Post volume					
Number of post views	.841***	677***			
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio					
Additional variables					
Posts by men in week 1					
Posts by women in week 1					
Number of theaters		.329**			
Movie budget					
Adjusted R^2	.696	.771			
Degrees of freedom	26	26			
F	60,518	44,777			
N	42				

13. Movie genre

	Drama	Horror			
Week 1 > week 1	1	2	3	4	5
	β	β	β	β	β
Post volume					
Number of post views					
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio		.999*			
Additional variables					
Posts by men in week 1					
Posts by women in week 1					
Number of theaters	.766*				
Movie budget					
Adjusted R^2	.518	.996			
Degrees of freedom	7				
F	8,536				
N	42				

14. Age

	6-9 years	12-16 years	16 +	16+
Week 1 > week 1	1	2	3	4
	β	β	β	β
Post volume		.962**		.466**
Number of post views				
Positive textual sentiment ratio				
Negative textual sentiment ratio				
Positive visual sentiment ratio				
Negative visual sentiment ratio				
Additional variables				
Posts by men in week 1				
Posts by women in week 1				
Number of theaters			.889***	.620***
Movie budget	.970**			
Adjusted R^2	.921	.906	.771	.922
Degrees of freedom	4	5	12	12
F	47,333	49,46	41,328	71,575
N	42	42	42	42

15. Twitter predictor variables in week 1 predicting movie box office revenues in week 2

Week 1 > week 2	1	2	3	4	5
	β	β	β	β	β
Post volume					
Number of post views		.436*			
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio					
Additional variables					
Posts by men in week 1					
Posts by men in week 1					
Number of theaters					
Movie budget	.848***	.486**			
Adjusted R^2	.708	.761			
Degrees of freedom	27	27			
F	66,525	43,942			
N	42	42	42	42	42

16. Twitter predictor variables in week 2 predicting movie box office revenues in week 2

Twitter	Model	
Week 2 > 2	1	2
	β	β
Post volume		.219*
Number of post views		
Positive textual sentiment ratio		
Negative textual sentiment ratio		
Positive visual sentiment ratio		
Negative visual sentiment ratio		
Additional variables		
Posts by men in week 2		
Posts by men in week 2		
Number of theaters		
Movie budget	.848***	.807***
Adjusted R^2	.719	.746
Degrees of freedom	27	27
F	66,525	40,719
N	42	42

17. Sequel

Sequel (no sequel movies)

Week 2 > 2	1	2	3	4
	β	β	β	β
Post volume		.231*	.247*	.445**
Number of post views				
Positive textual sentiment ratio				
Negative textual sentiment ratio				
Positive visual sentiment ratio				
Negative visual sentiment ratio				-0,269*
Additional variables				
Posts by men in week 2				
Posts by women in week 2			.206*	.197*
Number of theaters				
Movie budget	.851***	.808***	.765***	.715***
Adjusted R^2	.714	.757	.792	.823
Degrees of freedom	26	26	26	26
F	65,788	41,556	34,091	31,268
N	42	42	42	42

18. Age

Week 2 > week 2	6-9 years 1	16+ 2	16 + 3	16+ 4	5
	β	β	β	β	β
Post volume					
Number of post views		.861***	.507**	.360	
Positive textual sentiment ratio				.260*	
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio					
Additional variables					
Posts by men in week 2					
Posts by women in week 2					
Number of theaters			.475*	.562**	
Movie budget	.978**				
Adjusted R^2	.942	.719	.810	.866	
Degrees of freedom	4	12	12	12	
F	65,911	31,638	26,635	26,865	
N	42	42	42	42	

19. Genre

Movie genre					
	Drama				
Week 1 > week 1	1	2	3	4	5
	β	β	β	β	β
Post volume					
Number of post views					
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio		-0,543			
Negative visual sentiment ratio					
Additional variables					
Posts by men in week 2					
Posts by women in week 2					
Number of theaters	.717*	.737*			
Movie budget					
Adjusted R^2	.433	.733			
Degrees of freedom	7	7			
F	6,354	10,593			
N	42	42			

20. *Twitter predictor variables in week 2 predicting movie box office revenue in week 3*

Twitter	Model				
Week 2 > week 3	1	2	3	4	5
	β	β	β	β	β
Post volume		.261*	.272**		
Number of post views					
Positive textual sentiment ratio					
Negative textual sentiment ratio					
Positive visual sentiment ratio					
Negative visual sentiment ratio					
Additional variables					
Posts by men in week 2					
Posts by men in week 2			.193*		
Number of theaters					
Movie budget	.845***	.796***	.757***		
Adjusted R^2	.714	.780	.816		
Degrees of freedom					
F	64,959	44.301	35,412		