The Predictive Power of Social Media:

Using Twitter to predict Spotify streams for newly released music albums

MASTER THESIS BUSINESS ADMINISTRATION
Name: Rutger Ruizendaal
Student Number: S1225898
E-Mail: R.Ruizendaal@Student.utwente.nl
Supervisors: DR. IR. A.A.M. Spil & Dr. R. Effing

UNIVERSITY
OF TWENTE.

Acknowledgements

This master thesis marks the end of my studies in Business Administration and was performed internally at the University of Twente. During the past sixth months I have worked on this master thesis with great passion. On one hand it is a relief to finish such a big project, while on the other hand it is difficult to let go. However, this thesis does not mark the end of my academic studies. After graduating in Business Administration I will continue my masters in Communication studies.

I would like to thank my supervisors Ton Spil and Robin Effing for always providing relevant feedback and for the useful and interesting discussions we had on the subject. It is very motivating to see your supervisors be genuinely interested and invested in your research. I would also like to thank them for giving me the freedom in choosing the subject and dimensions of this research. Throughout this research I got drawn into the world of data science. I followed many online courses in order to supplement me with the knowledge needed to analyze big data like the data used in this thesis. I really want to thank my supervisors for supporting me in trying something new. Currently, I am still following courses in data science and machine learning and cannot wait to also implement this new knowledge in my graduation thesis for Communication studies.

Additionally I would like to thank my brother for implementing the database in which the tweets were originally stored. Finally, I would like to thank my family, friends and flat mates for always supporting me during these busy and sometimes difficult times. This graduation project taught me many things and introduced me to data science and machine learning. Without this master thesis my future career would look very different. Therefore, I will be forever grateful to the people who supported me along the way.

I hope you enjoy reading this master thesis.

Rutger Ruizendaal

Enschede, 2016

Abstract

Kalampokis, Tambouris and Tarabanis (2013) categorized social media predictor variables in three categories: volume-related variables, sentiment-related variables and profile characteristics of online users. Previous research has shown the predictive power of Twitter when using a single type of predictor variable. However, few studies have focused on the combination of multiple types of social media predictor variables. Moreover, research on the predictive power of social media in music often solely focused on volume-related variables and focused on blogs and Myspace. Nonetheless, Twitter plays an increasingly important role in the music industry and in academic research. This study tests the predictive power of all three categories of social media predictor variables on Spotify streams of newly released music albums. Based on an extensive literature review the research model in figure 1 has been designed.

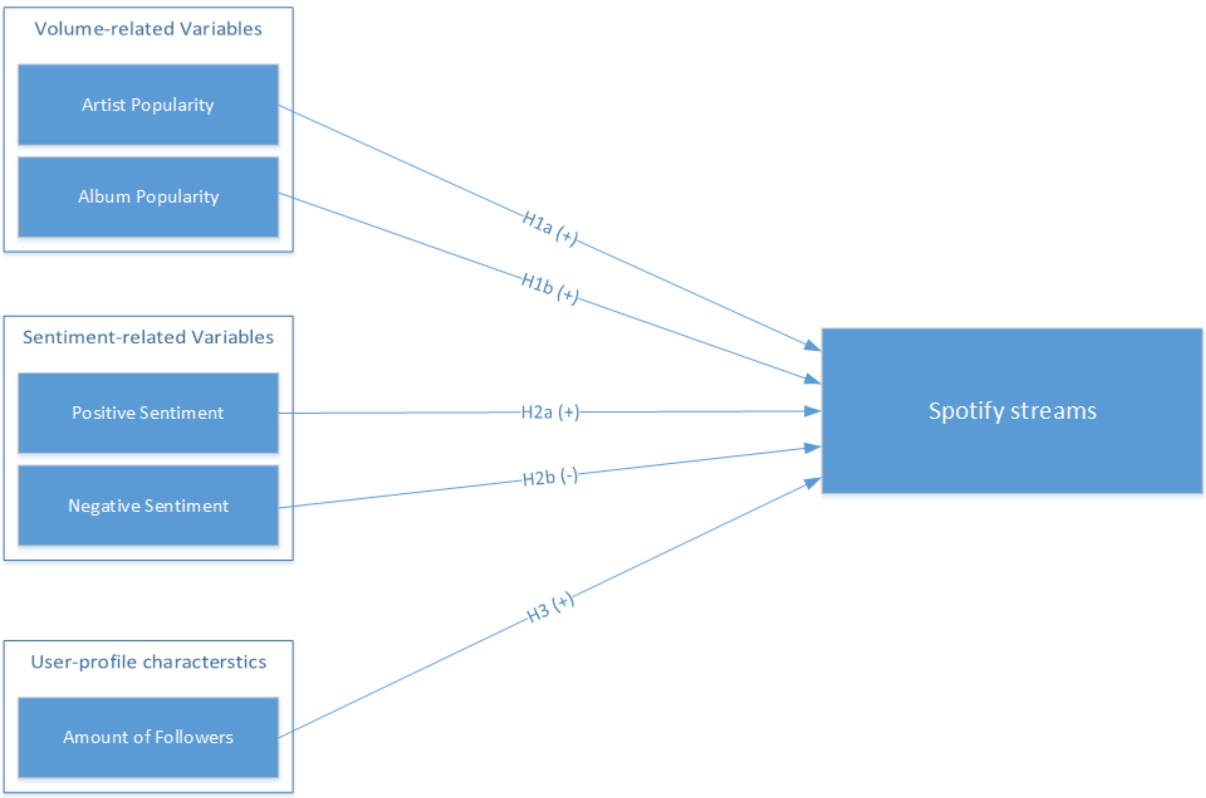

Figure 1: Research model

Over 2.4 million tweets were collected over a period of five weeks using keywords related to the artist and the album title. Multiple regression analyses were performed in order to test the relationship between Twitter predictor variables and Spotify streams in the same week. Additionally, multiple regression analyses were performed using one-period-lagged values in order to predict Spotify streams using Twitter variables from the previous week. Furthermore, additional analyses were performed including a daily time series analysis, exploratory analysis into age and gender differences and the inclusion of additional variables.

Hypotheses	Results
H1a: The volume of tweets for each album is positively associated with Spotify streams.	Accepted
H1b: The volume of tweets for each artist is positively associated with Spotify streams.	Accepted
H2a: Positive sentiment in tweets is positively associated with Spotify streams.	Rejected
H2b: Negative sentiment in tweets is negatively associated with Spotify streams.	Rejected
H3: Amount of followers is positively associated with Spotify streams.	Rejected

Table 1: Results on hypotheses

Results show the importance of volume-related variables in predicting Spotify streams. The sentiment-related variables and profile characteristics of online users were found to have no significant predictive power on Spotify streams. Results on the hypotheses formulated according to the research model can be found in table 1. The volume of tweets related to the album performed better at predicting Spotify streams than the volume of tweets related to the artist. A daily time series of the volume of tweets containing the album title was able to predict first week streams with high accuracy. Exploratory analysis suggests that the relationship between volume-related variables and streams could be influenced by the age of the musician. For younger musicians the Twitter predictor variables were able to significantly predict Spotify streams, while for older musicians the overall regression model was not significant. Furthermore, additional analyses suggest that the type of performer (male vs. female vs. band) could also influence the predictive power of Twitter on Spotify streams.

Keywords = Twitter, Spotify, streaming, predictive analysis, music, social media

Table of Contents

1. Introduction

From the top five most followed accounts on Twitter, four are musicians and the other is Barack Obama. The first Twitter account to exceed five million followers was a musician's. Furthermore, some of the highest rates of tweets per second are observed during televised music events. These cases all stress the importance of Twitter in the music industry, which is rapidly changing. In 2015 streaming became the biggest source of revenues (Friedlander, 2016). However, few academic studies have explored the relationship between Twitter and music revenues. On the other hand, Twitter has been used to predict box-office revenues (Asur & Huberman, 2010; Rui, Liu & Whinston, 2013), political elections (Ceron, Curini & Lacos, 2015) and the spread of diseases (Kim, Seok, Oh, Lee & Kim, 2013). Kalampokis, Tambouris and Tarabanis (2013) defined three types of social media predictor variables: volume-related variables, sentiment-related variables and profile characteristics of online users. Despite their framework many studies only focus on one social media predictor variable which often is volume-related. In this study all three social media predictor variables are used and the interaction between them is explored. Using social media predictor variables derived from Twitter this study tries to predict the amount of Spotify streams received by newly released albums.

Some early studies in the field of social media prediction in the music industry have explored relations between blog buzz and album sales (Dhar & Chang, 2009; Dewan & Ramaprasad, 2014), and Myspace broadcasting and music sales (Chen, De & Hu, 2015). Kim, Suh and Lee (2014) were the first to explore the option of using Twitter to predict music sales. However, Kim et al. (2014) only included volume-related variables in their analysis. This study expands upon that study by testing how effective the use of Twitter is in predicting Spotify streams by including multiple social media predictor variables. The rest of the introduction is organized as follows. First, Twitter and its use in academic research is discussed. Second, recent developments in the music industry are explored. Third, the central research question and sub-questions are formulated.

1.1 Twitter

The type of social media that this study focuses on is Twitter. Twitter is a popular social network with 313 million monthly active users (Twitter, 2016). Although the amount of active users on Twitter seems to slightly reach a plateau, it still has its highest amount of monthly active users ever. Twitter is the fourth most popular social networking site worldwide after Facebook, YouTube and Instagram (Kallas, 2016). Another reason that makes Twitter so attractive to social media research is its provision of the Twitter streaming application programming interface (API). This streaming API is widely used in academic research and allows for the real-time collection of tweets (Burnap et al., 2014). Additionally, the API also allows for the collection of additional data like the amount of followers a user has. In comparison to other social networking sites like Facebook and Instagram, many posts on Twitter are publicly available. Most post from Facebook, for example, are private which means that they cannot be used in academic research. Twitter is also highly relevant for the music industry. The most popular accounts on Twitter have often been musicians. For example, Britney Spears became the first account to exceed five million followers in 2010 and four out of the five currently most followed Twitter accounts are musicians. The predictive power of Twitter has been shown multiple times in previous research. For example, Asur and Huberman (2010) were able to accurately predict the box-office revenues of newly released movies using Twitter. Ceron et al. (2015) found that using the sentiment expressed in Twitter messages clearly predicted Obama to be the winner of the 2012 US presidential election, even though traditional polling considered the race too close to call. A high amount of tweets being public and the availability of the streaming API make Twitter a popular choice for academic research (Burnap et al., 2014; Ceron et al., 2014).

1.2 Music Industry

The music industry is a multibillion-dollar industry which has been experiencing an on-going decline since 1995 as a result of multiple difficulties including digitalization and piracy. In 2014 global revenues dropped below US$15 billion for the first time in decades, as revenues decreased by 0.4% compared to 2013 (IFPI, 2015). The International Federation of the Phonographic Industry (IFPI) recently reported that 2015 marks a change. For the first time in two decades global recorded music industry revenues saw a significant growth with 3.2% compared to 2014 (IFPI, 2016). This development led to positive responses within the industry. Besides, 2015 was the first year where streaming surpassed digital downloads as the biggest component of revenues in the United States, the largest music industry in the world (Friedlander, 2016). Streaming is rapidly growing and is already the main source of revenue in multiple markets. The disruption streaming causes is bigger than digital downloading did, because streaming changes the way consumers have access to music. PASM (2015) predicts that streaming will be able to double the value of the music business by 2020. In July of 2016 Nielsen published their mid-year report analyzing music data from the first half of 2016. This report contains the most updated information on the current state of the music industry in the United States. For the first time, audio streaming (54%) surpassed video streaming (46%) as the biggest streaming format (Nielsen, 2016). Despite the constant popularity of YouTube and Vevo, streaming services like Spotify and Apple Music contributed to audio streams increasing by 97.4% compared to the same half-year period last year. In comparison, video streaming only grew by 28.6%. These are not the only statistics that point to the growing importance of streaming in the music industry. Despite the resurgence of vinyl sales, overall album sales were down 13.6% with digital album sales dropping 18.4%. Digital track sales were hit even harder, decreasing by 23.9% (Nielsen, 2016). Forecasters now predict that if similar trends continue, Apple might have to turn off the iTunes store in 2020. Mark Mulligan, analysist at Midia wrote: “By 2020 download business would be tracking to be 10 times smaller than streaming revenue but, crucially, streaming revenue would nearly have reached the 2012 iTunes Store download revenue peak.” (Music Industry Blog, 2016). Next to Nielsen, market monitors report similar results (BuzzAngle, 2016). All these statistics point to the indisputable growth of streaming and the downfall of digital downloads. Spotify is the biggest service provider with over 40 million paid subscribers and more than 100 million active users. Since March, Spotify increased its paid subscriber base with 10 million. In comparison, Apple Music has 17 million subscribers (Schneider, 2016). Therefore, this research focuses on Spotify.

Besides streaming, social media has also had an impact on the music industry. Social media has made it easier than ever for artists to reach people, communicate with them in a meaningful way, and build a fan base. In the current music climate, social media is important for young artists (Leenders, Farrell, Zwaan & Ter Bogt, 2015). Currently, music consumption is at an all-time high (IFPI, 2016). The downside for artists is that there is also more music being produced than ever, and all these artists have access to social media. This can actually make it much harder for artists to get noticed from the crowd. Social media is the place where most fans go to follow their favorite artists. New releases like singles, albums and music videos are often launched on social media. This presents record labels with an opportunity to track the buzz and sentiment regarding the release (Franklin, 2013). Social media can also be used to obtain user data, find new artist and predict future hits (Shubber, 2014). For example, Asur and Huberman (2010) used social media data to predict box-office revenues around the time of release. Using data from Twitter, their predictions were more accurate than those of the Hollywood Stock Exchange, which is seen as the golden standard in the industry.

This study will focus on albums, because their release dates are often known in advance and their releases are big events and focal points in the social media campaign of artists. Singles generally receive less attention. Since 2015 albums have a global release day, on Fridays, which makes the buzz surrounding the album release easier to track. Albums can also be compared to movies, which have often been the focus of research into the predictive power of Twitter. Both are released following a comprehensive release strategy and both are part of the entertainment industry.

1.3 Research Questions

Previous research has discussed the predictive power of social media in general and has taken steps into testing this phenomenon in the music industry. Despite Twitter being a popular type of social media in academic research, only one study (Kim et al., 2014) explored the predictive power of Twitter in the music industry. Additionally, all studies researching the link between social media and music success have focused on sales data. Recent developments in the music industry have shifted the main attention to streaming. However, academic research has not yet caught up with this development. To the best of our knowledge, this is the first study that uses social media data in order to predict and explain the amount of streams received by music albums. Because Spotify is the most important provider of audio streaming at this point, the focus of this study is on Spotify streams. Based on the introduction the following research goal and central research question have been formulated.

Research goal: *To capture the relationship between Twitter predictor variables and Spotify streams received by newly released albums.*

***Central research* question**: *What is the relationship between volume-related, sentiment-related and profile-related Twitter variables and the volume of Spotify streams of newly released music albums?*

The following sub-questions have been formulated based on the central research question:

- *What Twitter predictor variables has previous literature identified as having significant predictive power on predicting both real-world and online outcomes?*
- *To what extent has the predictive power of social media in the music industry been demonstrated before?*
- *How should data from Twitter be collected and preprocessed for further analysis?*
- *Which Twitter predictor variables are significant predictors of album streams on Spotify?*

The remainder of this research is organized as follows. Section 2 describes how the literature review search has been organized. Section 3 presents the literature review with relevant literature on the predictive power of Twitter and the predictive power of social media in the music industry. Section 4 discusses the methodology used in the research design. Section 5 presents the results from statistical analyses. Section 6 provides the analysis by presenting the results on the hypotheses and answering the sub-questions. Finally, section 7 discusses the conclusions of the study, its limitations and presents suggestions for future research.

2. Literature Search

In order to gain an understanding of the predictive power of Twitter a systematic literature review was performed according to the method described by Wolfswinkel, Furtmueller and Wilderom (2013). The literature review also follows the principles of Webster and Watson (2002). The main idea behind these approaches is that the method and approach of the literature review is transparent and described in detail. A computer search was conducted at the end of July 2016 on the international research databases of Scopus and Web of Science. This search was extended in September 2016 to include a new set of keywords in the search. The final sample of papers has been constructed through a comparison of abstracts, removing duplicates, number of citations, forward and backward citations and finally reading the full texts. This process is described in more detail below.

The starting point of this literature review is a previous literature study. Kalampokis et al. (2013) wrote a literature review on the predictive power of social media. Their literature review provided a comprehensive review of the literature in this field. However, it could use an update in 2016. Kalampokis et al. (2013) performed the search in 2012 using the search term '(predict OR forecast) AND social media' on Google Scholar with the following inclusion and exclusion criteria:

- Excluded qualitative or purely theoretical articles.
- Included only studies aiming at making predictions.
- Included only studies that attempt to predict real world outcomes. For example, studies that focused on online features such as tie strength, volume of comments on online news or movie rating on IMDB were excluded.

While the literature review by Kalampokis et al. (2013) is a great starting point there are certain things that the literature review in this paper would like to add:

- The published literature on social media and predicting after 2012. The literature search by Kalampokis et al. (2013) was performed in 2012. Since their study there has not been another comprehensive literature review on this subject. This might be caused by how many papers on the subject are published by computer scientists. Most articles on the topics of computer science are presented at conferences and therefore only include a short overview of past research with more focus on the methods and results. This study will provide a comprehensive overview of the literature published between 2013 and the first half of 2016.
- Literature that focuses on online outcomes. The study by Kalampokis et al. (2013) excluded articles that attempted to predict online outcomes. This thesis focuses on an online outcome, online music streaming. These excluded articles from Kalampokis et al. (2013) are included in this review, because they might describe differences between offline and online outcomes. Therefore it is important to include these studies for the literature review in this paper.

Besides, this study focuses on the music industry. Therefore, next to updating and expanding Kalampokis et al. (2013) this review will also pay close attention to the music industry. First, studies related to how social media is changing the industry will be included in order to understand the extent of social media effects. Second, studies specially focusing on predicting or monitoring social media in the music industry will be discussed in detail. After analyzing the papers found in the first search in July 2016 results showed that many studies in the review focused on Twitter. Therefore, Twitter was planned to be the type of social media used in the empirical part of the research. Accordingly, it was decided to expand to literature search to focus specifically on Twitter. Extra searches with these new keywords were conducted in September 2016.

The goal of this literature review is to gain an understanding of what types of social media, methods, variables and results have been recorded in previous research on the predictive power of social media. The sub-goals for each part of the review are as follows:

- Providing an update and extension to the work of Kalampokis et al. (2013) with a specific focus on the use of Twitter for predicting.
- Gathering knowledge on the specifics related to 'online outcomes' that were omitted in Kalampokis et al. (2013).
- Capturing the previous ways that predictions using social media have been studied in the music industry.

A basic search on Scopus into social media was performed. This search showed that academic research into social media, did not start until 2006. Therefore, only papers published in 2006 or later are included in the literature review. Computer searches using keywords were performed on the online databases of Scopus and Web of Science. These are multidisciplinary databases that can cover a broad subject like social media. Table 1 displays the keywords used in the search and provides detailed information on how many papers were found and selected in each phase. The set of keywords can be split up into three groups, where the first group focuses on expanding the work of Kalampokis et al. (2013), the second group narrows its focus to the music industry and the third group focuses on the use of Twitter for predictions. Because of the big amount of results for certain search terms a restriction was placed of having a minimum number of ten references to be included. This means that papers found using these search terms were only included in the review if they had at least ten references on the databases they were found at. This criterion allows for the inclusion of only highly relevant papers in the review. The use of broad search terms allows us to review all the different aspects of the social media spectrum. Further narrowing down these terms could lead to the exclusion of key papers. On the other side, the keywords that specifically focused on the predictive power of social media in the music industry resulted in few papers. Therefore, no minimum requirement of citations was set for these papers. During the extra search in September 2016 it was also decided to include recent papers since these could include state-of-the-art knowledge related to Twitter. For these papers published in 2016 the minimum requirement of 10 citations was also dropped. A detailed description regarding the combination of search terms used can be found in Appendix A.

The selection process was organized as follows. Firstly, articles were selected based on their title and abstract. Secondly, duplicates were removed. Finally, full texts were read and a concept matrix was updated after reading each paper (Webster & Watson, 2002). This concept matrix can be found in Appendix B. Forward and backward citations were also performed during this stage. Afterwards, papers were analyzed using the grounded theory approach which is responsible for how the literature review has been organized (Wolfswinkel et al., 2013). This literature search resulted in the inclusion of 48 papers (see table 1). Table 2 presents the key papers regarding the use of social media prediction in the music industry found during the literature search. These are 12 out of the 48 papers that were essential in writing the literature review.

Search Terms	Years	Ref. Min	Scopus Results	Scopus Selected	WOS Results	WOS Selected	Total Selected	Actual Selected
“social media” AND (forecast OR predict)	2013-2016	≥10	40	9	28	6	15	12
“Social media” AND “monitoring”	2013-2016	≥10	64	15	19	7	22	6
“Social media” AND (forecast OR predict)	2006-2012	≥10	42	9	27	4	13	7
“Social media” AND “music”	2006-2016	≥10	31	8	7	3	11	4
“Social media” AND “music” AND (forecast OR predict)	2006-2016	-	7	1	9	1	2	11
“Social media” AND (music sales OR album sales)	2006-2016	-	9	9	7	5	12	
“Twitter” AND (forecast OR predict)	2013-2016	≥10	36	7	18	1	8	8
“Twitter” AND “monitoring”	2013-2016	≥10	34	4	14	0	4	
“Twitter” AND (forecast OR predict)	2016	-	95	7	59	5	12	
“Twitter” AND “monitoring”	2016	-	53	2	24	0	2	
Total selected after reading abstracts							101	-
After removing duplicates							83	-
Final sample after reading full texts & forward/backward citations							48	48

Table 1: Literature review overview

Abbreviations: Ref. Min. = Reference minimum; amount of references needed in order to be selected. WOS = Web of Science.

Authors	**Year**	**Journal/Proceedings**
Asur & Huberman	2010	International Conference on Web Intelligence and Intelligent Agent Technology
Bischoff et al.	2009	International Conference on Advanced Data Mining and Applications
Ceron et al.	2014	New Media & Society
Chen et al.	2015	Information Systems Research
Dewan & Ramaprasad	2009	Pacific Asia Conference on Information Systems
Dewan & Ramaprasad	2012	Information Systems Research
Dewan & Ramaprasad	2014	Management Information Systems (MIS) Quarterly
Dhar & Chang	2009	Journal of Interactive Marketing
Kim et al.	2014	Proceedings of the first international workshop on Social media retrieval and analysis
Rui et al.	2013	Decision Support Systems
Sharma et al.	2012	Journal of Information, Information Technology, and Organizations
Saboo et al.	2015	International Journal of Research in Marketing

Table 2: Key papers

3. Literature Review

The literature review is split-up based on the type of social media predictor variable studied. Table 3 features an overview of the academic literature and what type of social media predictor variables are included within these studies. Volume-related variables are the most studied type of social media predictor variable for the Twitter studies, as well as for the music specific studies. For sentiment-related variables table 3 shows that only two studies focus on social media in music using sentiment-related variables. Research into user-profile characteristics is relatively scarce in both cases.

This research focuses on both Twitter and the music context. In this literature review one paper was found that also focused on the use of Twitter to predict a music-specific phenomenon. Kim et al. (2014) researched whether it was possible to predict album sales using volume-related Twitter variables. This paper is further explored in chapter 3.1. Because there is only one paper that includes both Twitter and music it is important that in the literature review both the predictive power of Twitter and the predictive power of other social media in music are represented. Therefore, for each type of social media predictor variable the findings from the Twitter prediction studies are discussed first and then we zoom in on studies that applied social media prediction in the music industry. Research into sentiment-related variables and user profile characteristics is relatively scarce in the music context. It is therefore especially important to explore how these variables were studied in different contexts. The literature review is organized as follows. First, volume-related variables are discussed. Second, sentiment-related variables and sentiment analysis is elaborated upon. Third, user-profile characteristics of online users are discussed. Fourth, additional variables found in the literature are presented. Finally, the hypotheses and research model are formulated and designed.

	Predictive Power of Twitter	**Predictive Power of Social Media in Music**
Volume-related (3.1)	Asur & Huberman (2010), Burnap et al. (2014), Caldarelli et al. (2014), Ceron et al. (2014), de Choudhury et al. (2013), Denecke et al. (2013), Hanson et al. (2013), Hong et al. (2011), Jashinsky et al. (2014), Jungherr (2013), Kim et al. (2013), Lipizzi et al. (2016), Liu et al. (2016), Oghina et al. (2012), Rui et al. (2013), Schumaker et al. (2016), Suh et al. (2010), Young et al. (2014), Zhu et al. (2011) *Total = 19*	Bischoff et al. (2009), Chen et al. (2015), Dewan & Ramaprasad (2012). Dewan & Ramaprasad (2014). Dewan & Ramaprasad (2009), Dhar & Chang (2009), Kim et al. (2014), Maecker et al. (2013), Salganik et al. (2006), Sharma et al. (2012), Saboo et al. (2015) *Total = 11*
Sentiment-related (3.2)	Asur & Huberman (2010), Burnap et al. (2014), Burnap et al. (2016), Ceron et al. (2014), Ceron et al. (2015), Jungherr (2013), Li et al. (2016), Lipizzi et al. (2016), Liu et al. (2016), Oghina et al. (2012), Rui et al. (2013), White (2016) *Total = 12*	Dewan & Ramaprasad (2012), Dhar & Chang (2009), *Total = 2*
User-profile characteristics (3.3)	Burnap et al. (2014), de Choudhury et al. (2013), Rui et al. (2013), Zhu et al. (2011), Suh et al. (2010) *Total = 5*	Bischoff et al. (2009), Dewan & Ramaprasad (2012), Dhar & Chang (2009), Saboo et al. (2015) *Total = 4*

Table 3: Academic literature organized by variables studied and context. The amount of papers and the respective paragraph where they are discussed are included in each cell.

3.1 Volume-related Variables

Volume-related variables are the main type of variable used in social media prediction research. For example, Luo & Zhang (2013) found that the volume of consumer reviews from an electronic product review website had a significant positive effect on firm value. Jin, Gallagher, Cao, Luo and Han (2010) showed that Flickr provided hints that Obama would win the presidential election in 2008. Next to the US election, Flickr also showed the geographical distribution of iPod and iPhone sales and was able to predict quarterly sales with relatively low error (Jin et al., 2010). Also, Won et al. (2013) built a prediction model that tried to predict the national number of suicides in South Korea using counts of blogs that mentioned words related to suicide or dysphoria. These previous studies used various forms of social media. However, many studies that research the predictive power of social media use Twitter. The research in this literature review that used volume-related variables derived from Twitter can be split-up in three categories: health, politics and movies. In this section the main conclusions from these contexts will be discussed. For a detailed overview of the findings of literature in health, politics and movies see appendix C. After discussing the use of volume-related variables from Twitter the effect of volume-related variables on predicting music outcomes are discussed.

In the context of health the main focus of many papers is not strictly on predicting, but rather on monitoring. A common practice is the gathering of Twitter data by using health related keywords and then matching the volume of tweets to official disease reports per area. In these cases, it is important that tweets are geolocated or can be traced to a certain area. Hanson et al. (2013) monitored Adderall use among college students while Jashinsky et al. (2014) found strong correlations between high risk tweets and actual suicide data per US state. Another study found a significant relationship between the volume of tweets containing HIV risk behavior and actual HIV cases in the Unites States (Young, Rivers & Lewis, 2014). Kim et al. (2013) used linear regression to build a model with good predictive power on the spread of influenza by using around thirty influenza-related keywords. While the use of volume-related metrics has received good results in a health context, there are some limitations related to data availability that pop-up in almost every health-related study. The low amount of geotagged tweets (Hanson et al., 2013), outdated regional disease information (Young et al., 2014) and shame possibly restricting Twitter users are often mentioned (Stoové & Pedrana, 2014).

Gayo-Avello (2013) provides a critical review of literature that only uses volume-related social media data to predict political outcomes. While trying to predict the outcome of the German 2009 federal election Jungherr (2013) found that the volume of tweets was a bad measurement to solely use when predicting election outcomes. Also, the amount of hashtags used to refer to a party alone has a similar fate. The main conclusion is that, in a political context, only counting the mentions of a particular candidate or party is not reliable enough to use for predicting electoral results. Another option could be the relative support parameter introduced by Caldarelli et al. (2014). Next to the volume of tweets, these authors include the ratio of time variation between the volume of tweets in the comparison of two parties. However, results show that this approach is still in its infancy and does not achieve significant results. In a political context there is a more critical opinion on the sole use of volume-related variables. People have a tendency to not always openly express what party they are actually planning to vote for and some elections deal with controversial parties who are popular online but do not receive many actual votes.

These studies already show the importance of context in social media prediction. The third and final context found in most Twitter prediction studies is the movie industry. From health, politics and movies, this is the context that shows the most similarities to the context in which this master thesis is performed; the music industry. Studies in health and politics mainly focused on one type of social media predictor variable. Research that focuses on social media prediction in a movie context also focuses on combining multiple types of social media predictor variables. One of the studies that initiated the research trend into the predictive power of social media was a study related to the movie industry. Asur and Huberman (2010) used the volume and sentiment of tweets to predict box-office scores with a higher accuracy than the Hollywood Stock Exchange (HSE). Their goal is to not monitor all movies mentioned on Twitter, but to predict box-office scores in the opening weekends of newly released movies. Over 2.8 million tweets for 24 movies were collected over a period of three months. By only using the tweet-rate in a linear regression model the authors were able to significantly predict box-office scores ($R^2 = 0.80$). Next, the authors constructed a time series from seven days before the release of a movie from the tweet-rates and also include an additional variable; the amount of theatres a movie was released in. This new model is able to predict 97.3% of the variance in box-office scores. Rui et al. (2013) extend on this research by extending the data collection and performing a panel data analysis using one-period lagged values. Rui et al. (2013) determine whether a tweet contains the clear intention to go see a movie and create the variable ‘intention tweets’. The dynamic panel shows that the total number of tweets, tweets from users with a high following, intention tweet ratio and the ratio of tweets with a positive sentiment all have a significant and positive influence on box-office scores. Using textual features of tweets and the likes-dislikes ratio from YouTube Oghina, Breuss, Tsagkias and De Rijke (2012) were able to predict IMDB scores with high accuracy. Liu, Ding, Chen, Chen and Guo (2016) focused on three Twitter metrics when predicting box-office scores: purchase intention, tweet volume and sentiment. The model that performs best in predicting box-office revenues uses a combination of purchase intention, sentiment, the amount of theatres the movie was released in, and the popularity of the movie’s director (Liu et al., 2016). With regard to using volume-related Twitter variables for predictions the movie context achieved much better results than the health and politics context when using only volume-related variables. Like the music industry, the movie industry is part of the entertainment industry. When tweeting about what movies or music people are interested in they generally do not face the shame that comes with disease symptoms or the controversial nature of politics. Also, data on box-office revenues is always updated and limitations related to geolocation also play no role in these studies.

After focusing on the predictive power of Twitter the focus is now switched to the music industry. Social media plays a role in almost every part of the music industry. For example, fans share their experiences of live shows on social media and even ‘tweet seats’ have been introduced at some concerts (Bennett, 2012). The same goes for live television shows. The MTV Video Music Awards and the Black Entertainment Television Awards created some of the highest rates of tweets per second in 2011 (Highfield, Harrington & Bruns, 2013). Furthermore, Twitter also plays an important role in album campaigns (Kaplan & Haenlein, 2012). For a detailed description of ways that social media plays a role in the music industry see appendix D. Compared to previously discussed contexts, the literature that focuses on predicting in a music context using Twitter is extremely scarce. Actually, only one study that fits these criteria was found during the literature review (Kim et al, 2014). Therefore, other types of social media have the forefront in this next part.

Similar to the Twitter prediction studies, volume-related variables are the most popular type of social media predictor variable used when applying social media prediction to the music industry. These volume-related variables can range from the amount of blog posts, number of tweets, number of Myspace comment to the number of reviews and more. The dependent variable is often related to music success like song sales or album sales. In order to fully understand the relationship between social media and music, we first take a look at social influence in a music context. Salganik, Dodds and Watts (2006) created an artificial music market to tests the effect of social influence on music downloads. Participants were shown a list of unknown bands. They then had the possibility to listen to tracks, rate them on a 1-5 scale and download the songs. The independent condition gave no extra information, while the social influence condition showed how many times each song was downloaded by previous participants. The social influence condition also consisted of eight different virtual worlds that developed separately from each other. The results suggest that social influence increases the inequality of outcomes, which results in popular songs becoming more popular and less popular songs becoming even less popular. Maecker, Grabenströer, Clement and Heitmann (2013) re-created the study by Salganik et al. (2006) and found similar results for music, but also for movies and fashion products. The authors also find that the popularity of products is related to social information for both positive and negative types of information. Social information regarding the purchases or popularity of music is often found on social media. Therefore, when social media sales ranks become available companies can expect a shift in market shares (Maecker et al., 2013).

The first set of researchers that studied the predictive power of social media in music focused on the amount of blog posts, also defined as 'blog buzz'. Dewan and Ramaprasad (2009) research the simultaneous relationship between blog buzz and album sales. Using a time-series cross-section dataset with the weekly blog buzz and weekly album sales of 2694 albums the authors find results that indicate strong bi-directional causality. This relationship seems to be stronger for albums that were recently released. Other variables like type of record label, artist reputation and the number of Amazon reviews were also included in the analysis. These variables will be discussed in their respective parts below. Dhar and Chang (2009) also find support that the amount of blog posts is a good indicator of album sales. Blog buzz seems to be a consistent variable in predicting album sales, one, two and three weeks ahead. After Sharma, Morales-Arroyo and Pandey (2012) included latency effects in their analysis, blog posts were able to predict album sales 80% of the time. In 2014, Dewan and Ramaprasad performed a similar research to their 2009 research. However, this research had pretty different results. This research extended the scope to also include traditional media and song sales. Data was collected for a 24 week period with Google Blog Search and Nielsen Soundscan and was supplemented with additional variables like type of record label and artist reputation. After performing a panel vector auto regression the authors find that the relationship between blog buzz and album sales is insignificant. Impulse response functions show that the reaction of album sales to a shock in album buzz hovers around zero. The relation between blog buzz and song sales turns out to be negative, showing that an increase in blog posts about songs actually leads to a decrease in sales for these songs. The authors attribute this negative relationship to sampling and argue that free sampling displaces sales. Music blogs often offer an opportunity to sample the songs or albums they blog about (Dhar & Chang, 2009). Dewan and Ramaprasad (2012) previously studied music sampling and found that it is positively associated with blog popularity and music popularity. This relationship seems to be stronger for recently released albums and for niche music (versus mainstream music). Although admitting that the conclusion is speculative, Dewan and Ramaprasad (2012) suggest that sampling drives music sales. This suggestion is later debunked by Dewan and Ramaprasad (2014) who conclude that sampling is the reason that a higher amount of blog posts leads to lower song sales. Sharma et al. (2012) and Saboo, Kumar and Ramani (2014) study music sampling from social media websites. The number of times an album was sampled on Myspace predicted album sales in four out of eleven weeks (Sharma et al., 2012). Saboo et al. (2014) study social sampling over various social media websites like Facebook, Twitter, Myspace, YouTube and Last.fm and conclude that social sampling has a significant negative influence on music sales. Therefore, it is important to keep the effects of sampling in mind when researching social media types that offer the opportunity to sample music, like blogs for example.

Research on volume-related variables later switched to other social media sites like Myspace. In a study by Sharma et al. (2012) the number of posts on Myspace was a good predictor of album sales, because it was significantly related to album sales in seven out of eleven weeks. However, this number dropped significantly when latency effects were included. Chen et al. (2015) found that there is a significant relationship between the amounts of posts made by an artist on Myspace on music sales. Nonetheless, this effect is mainly caused by bulletins and not by friend updates. Bulletins are messages posted by the artist themselves and are therefore seen as more personal than friend updates. This distinction is specific to Myspace and therefore difficult to apply to other types of social media. The authors also find that around the time of new music releases the effect of personal messages is about six times larger than otherwise. Myspace was once the most popular social media for artists to use, but it has experienced a serious decline in recent years. Therefore, it is important to also research other types of social media (Chen et al., 2015). As described by Saboo et al. (2014) comments on social media have a significantly positive influence on music sales. The interesting aspect is that this effect keeps increasing as the amount of comments increases. It is important for managers to encourage participation on social media as “social comments are like testimonies, create a positive buzz about the brand and bring new consumers into fold” (Saboo et al., 2014, p. 9). Sharma et al. (2012) found no relationship between YouTube comments and album sales. YouTube views and the number of uploaded videos seemed to have a small predictive effect on album sales. Bischoff, Firan, Georgescu, Nejdl and Paiu (2009) used social media data gathered from Last.fm and build an algorithm for hit prediction. Their algorithm improved performance by 28% on a similar algorithm that did not include social media data.

Twitter is a popular social network that has recently seen an increased importance in academic research due to its streaming API and the high level of public posts. However, only one study researched the predictive power of Twitter in a music context. Kim et al. (2014) collected tweets regarding the listening behavior of users with the music-related hashtags #nowplaying, #np and #iTunes. The hashtag #nowplaying is a very popular hashtag on Twitter, for example, in the dataset collected by Suh, Hong, Pirolli and Chi. (2010) is was the most used hashtag. After collecting data over ten weeks with the Twitter streaming API, Kim et al. (2014) used this data to predict Billboard rankings and hit songs. Mining these tweets for ten weeks led to the collection of more than 31 million tweets. The amount of tweets associated with a song was defined as the ‘song play-count’ and the amount of tweets associated with an artist became ‘artist popularity’. Another variable was taken into account as well, the number of weeks an album had been on the Billboard chart. The authors then build three regression models and compare the results. Compared to linear regression and quadratic linear regression, support vector regression performs best. Therefore, the authors suggest that linear models may not fit the distribution of data. When using the support vector machine the model with just Twitter information achieves an $R^2 = .57$. The combined model of Twitter information and the amount of weeks an album has been on the Billboard charts improved the performance to 0.75. An algorithm for hit prediction was constructed and used to predict hits across various ranges. The algorithm, using Twitter information and weeks on the chart, was able to predict whether a song would be in the 1-10 range on the Billboard chart with 92% accuracy. The authors conclude that despite the combined model having the highest accuracy “music listening behavior available in Twitter can generate an outstanding predictive model” (Kim et al., 2014, p. 55).

The final volume-related variable is the number of reviews. Dewan and Ramaprasad (2009, 2012) take the number of reviews on Amazon into account in their analysis. Although both variables appear to be significant predictors of album sales, the authors do not further elaborate on this. Chen et al. (2015) include the number of new Amazon reviews in their research. They find that the volume of new reviews seems to have some association with the sales rank of albums. Also, an artist’s posts are not influenced by customer reviews, but friend updates by artists seems to weakly predict the amount of new customer reviews.

3.2 Sentiment-related Variables

Jungherr (2013) and Luo and Zhang (2013) used sentiment-related variables in their approach without actually performing a sentiment analysis. Jungherr (2013) analyses the German election of 2009. Throughout this election people have been encouraged to express support or disapproval by tweeting a party name preceded by a hashtag and followed by a + or -. These mentions have been collected by the German website Wahlgetwitter and are later used by Jungherr (2013) for analysis. The overall sentiment of tweets was more negative than positive, and sentiments were a valuable addition to purely using tweet volume in predicting the election outcome. Luo and Zhang (2013) found a significant positive relationship between the average score of consumer reviews and firm value metrics.

When performing a sentiment analysis there generally are two main approaches. The first approach is lexicon-based. In this method the sentiment of a text is defined by matching words in the text to words in a pre-defined lexicon. The second approach is based in machine learning. Here, a classifier model is trained to classify texts into sentiment categories. In the next part studies using the lexicon-based approach are discussed first and studies using machine learning are discussed next.

The lexicon-based approach in sentiment analysis focuses on the matching of words in a string of text to words in a pre-defined lexicon. Using such a lexicon, Chen et al. (2014) summed the fraction of negative words used in articles and comments on one of the most popular investment-related social networking sites. Their results showed that the amount of negative words used in the articles and comments was able to significantly predict stock prices over the next three months. All other studies discussed in this part on sentiment analysis used Twitter data. Burnap et al. (2014, 2016) and Lipizzi, Landoli and Marquez (2016) used the SentiStrength tool. This tool has been developed by Thelwal, Buckley, Paltoglou, Cai and Kappas (2010) and calculates a negative and positive score for each text, ranging from -5 to +5 respectively. Burnap et al. (2014) found that sentiment was statistically significant in predicting both the size and survival of tweets, measured by its amount of retweets.

Burnap et al. (2016) also used the SentiStrength tool when calculating the sentiment of tweets in trying to predict the UK 2015 general election. The authors take into account the earlier critics by Gayo-Avello (2013) for predicting political elections with social media. Mainly the authors analyse sentiment instead of volume, take into account parliamentary representation and test their prediction model on a new set of data. The authors combine volume and sentiment by counting tweets with a positive sentiment as a vote for a specific party. Their results highlight the difficulty in predicting political elections when there is a certain party that is very big in specific regions of the country. The authors suggest new methods for geolocation in future similar research. Lipizzi et al. (2016) also used the SentiStrength tool and were critical of the results they received regarding the predictive power of sentiment. When only using sentiment, it was found to be a weak predictor of box-office revenues. Liu et al. (2016) use a sentiment lexicon constructed by the Harbin Institute of Technology in China. The ratio of positive to negative tweets is used for measuring sentiment. Liu et al. (2016) find that sentiment is a good predictor of box-office revenues and it is included in their model that achieves the best performance. In an attempt to predict political elections using Twitter, White (2016) found that the mean sentiment that each Twitter user gave to each candidate was the best predictor out of the variables they tested. Sentiment was calculated using word polarity from the qdap R package, which uses a lexicon of positive and negative words. The Twitter forecast model using the mean sentiment score was able to forecast the overall Canadian election as well as provincial results (White, 2016). Li, Zhou and Liu (2016) used a lexicon that searched for frequency of happy, sad, anger, fear, disgust and surprise words. The authors only use correlational measures and find a high correlation between the frequency of happy, sad and angry words and stock behaviour. Finally, Schumaker, Jarmoszko and Labedz (2016) used a similar tool to SentiStrength called OpinionFinder. This tool classifies texts on two axes, positive or negative and subjective or objective. The sentiment of tweets was used to predict the outcomes of matches in the English Premier League in 2014. Results showed that positive surges of the sentiment in tweets can lead to a better prediction of match outcomes than traditional betting odds. However, the authors also mention the high volume of tweets from opposing teams inserting negative sentiment about the other team (Schumaker et al., 2016).

The second method commonly used for sentiment analysis is machine learning. When using machine learning for sentiment classification the dataset is usually split-up in a training set and a test set. The sentiment labels for the training set are manually coded. A machine learning algorithm then uses information retrieved from the training set in order to predict the sentiment classification of the test set. After Asur and Huberman (2010) reached high accuracy in predicting movie box-office revenues using volume-related variables they decided to also include sentiment in their linear model. A language model classifier was trained and taught to classify tweets as either negative, neutral or positive. The inclusion of sentiment was found to have a small effect on box-office revenues and showed a slight increase in the prediction accuracy of the model. However, the tweet-rate was still the most important variable in the model. Ceron et al. (2014, 2015) use a method described by Hopkins and King (2010) and released in the readme package for R (Hopkins, King, Knowles & Mendelez, 2010). Ceron et al. (2014) perform a sentiment analysis on Twitter in three case studies; popularity of Italian party leaders, the 2012 French presidential ballot and the 2012 French legislative election. The method of Hopkins and King (2010) uses a supervised machine learning approach where a manually coded subsample is used by an algorithm to classify the remaining set of the data. The classifier had to be trained separately for each case studied. Results in the Italian case study show that the Italian leaders are generally scored as less positive on social media than in traditional polls. For some leaders the poll and social media results are very similar, while this is not the case for other leaders. The Mean Absolute Error (MAE) of the predictions seems to decrease as we get closer to the election date. For both of the French case studies, the results from the social media analysis were in line with the actual election results and with published surveys. Twitter allows for a proper analysis of day-to-day reactions from the general public. Concluding, while internet users might not always be representative of the full population, this analysis shows that there is a consistent correlation between social media results and that they are also able to forecast electoral results. Also, social media sentiment reacts to day-to-day exogenous factors (Ceron et al., 2014). In 2015, Ceron et al. (2015) used similar methods and received similar results in the 2012 US presidential election and in predicting the centre-left coalition leader in Italy. Ceron et al. (2015) again report on the ability of social media to monitor any momentum gained in a campaign. In the 2012 US presidential election the authors report that online sentiment clearly predicted Obama to be the winner although the race was considered too close to call by traditional polling. On the other hand, Rui et al. (2013) built a Naive Bayesian classifier. This classifier was trained on a corpus of 3000 tweets in order to classify tweets as either positive, negative or neutral. Through different robustness checks the positive tweets ratio always had a significant and positive effect on box-office revenues. Similarly, the negative tweets ratio always had a negative impact on box-office revenues. Although this effect was not significant in all cases.

None of the articles related to the predictive power of social media in music performed a sentiment analysis where a set of text is classified as positive or negative. It certainly seems relevant to include the sentiment of social media content when researching its effect on sales. As shown before sentiment analysis is often found in different topics when performing this type of research like politics (Ceron et al., 2014; Ceron et al., 2015) and movies (Asur & Huberman, 2010; Rui et al., 2013). Although this absence of sentiment analysis on social media prediction in music is often mentioned as a limitation or suggestion for future research (Dhar & Chang, 2009; Dewan & Ramaprasad, 2014), no studies performing a sentiment analysis were found in the search for this literature review. On the other hand, there were some studies that used sentiment-related variables in their analysis. The previously discussed studies that included the volume of reviews, also included the average score of these reviews. Dewan and Ramaprasad (2012) find that albums in the top 5000 ranking on Amazon have a higher amount of customer reviews than albums below the 5000 ranking. However, the valence of the reviews is similar for both groups. Regression analysis showed that review valence was not significant in predicting music sampling. Dhar and Chang (2009) find that in their model three variables are significant in predicting album sales; average customer reviews, blog buzz and type of record label.

3.3 Profile Characteristics of Online Users

From the three types of social media predictor variables identified by Kalampokis et al. (2013), profile characteristics of online users is the least popular one in academic literature. Only a small sample of the literature in this review included these type of variables. Zhang and Pennacchiotti (2013) collected demographic information like age and gender, and Facebook likes from Facebook when trying to predict eBay sales. Results show that there is a difference in the categories that men and women buy products from. Most sales in eBay categories could be predicted by the Facebook likes per user. In a study on predicting postpartum depression through Twitter, de Choudhury, Counts and Horvitz (2013) included the number of followers and followees from recent mothers. The number of followees was a better predictor than the number of followers in predicting extreme behavioural change. However, in this case this is probably caused by mothers who are faced with postpartum depression attempting to minimize the amount of people they follow. Next to the number of followers and followees, Burnap et al. (2014) also included the number of previous tweets and a variable 'reach'. This variable was calculated by summing the number of followers of each user that retweeted a certain tweet. Only the amount of followers had a positive effect on the amount of retweets. Rui et al. (2013) also include the amount of followers in their model that tried to predict the box-office revenues of movies. Tweets from users with a following above 400 are categorized as type-2 tweets and tweets from users with a following below 400 as type-1 tweets. The dynamic panel shows that the percentage of tweets from users with a high following has a positive and significant effect on box-office scores. A robustness check for the cut-off point of 400 followers for type-2 tweets shows similar results (Rui et al., 2013).

Dhar and Chang (2009) include a variable related to the profile characteristics of online users in their model by calculating the weekly change of Myspace friends. However, this turns out to not be a significant predictor of album sales. When analysing the link between blog buzz and music sales, it could be argued that blog popularity is a profile characteristic. In this case, the user is seen as the blog who posted the blog article. Blog popularity has a significant and positive effect of music sampling, especially for albums with a lower Amazon ranking (Dewan & Ramaprasad, 2012). Saboo et al. (2014) find that social following has a positive effect on a song's Billboard hot 100 ranking in a decreasing way. Followers on social media have a positive effect on music success. However, as the follower count increases, its effect decreases.

Volume-related variables, sentiment-related variables and user profile characteristics in the literature have been discussed until now. As an addition to Kalampokis et al. (2013) their literature review, papers that focused on online outcomes were also analysed. Only two of these papers explicitly focused on the prediction of online outcomes (that were not social media related like the amount of retweets) using social media. Therefore it seems logical that these papers were excluded from the review of Kalampokis et al. (2013). The methods and models used in these papers regarding online outcomes are similar as the ones used in the previously described articles that focus on 'real-world outcomes'. Therefore, it seems that in our later analysis of online streams there are no specificalities related to online outcomes that need to be taken into account. A more detailed description of these papers can be found in appendix E.

3.4 Additional Variables

Additional variables are often exclusive to the industry or context the predictive study is applied in. For example, Won et al. (2013) collected economical and meteorological measures as well as celebrity suicide data when predicting national suicides. The number of theaters a movie was released in has been included in various studies that used Twitter data to predict box-office revenues of movies (Asur & Huberman, 2010; Liu et al., 2016). This section discusses variables outside of the social media predictor variables that were included in studies that explored the predictive power of social media in music. These variables are specific to the music industry and could provide additional explanations when exploring the predictive power of social media in music.

3.4.1 Artist Reputation

Many studies include some form of artist reputation in their statistical analysis. Dewan and Ramaprasad (2009, 2012 & 2014) include a dummy variable for artist reputation as a control variable. This variable is set to 1 if an artist appeared on the "top Artists of the Year" chart by Billboard between 2002 and 2006, or if the artist appeared on the "Billboard All-Time Hot 100 Artists". Artist reputation did not have a significant effect on music sampling (Dewan & Ramaprasad, 2012). Dewan and Ramaprasad (2014) split up their dataset based on artist reputation. The analysis showed that the negative relationship between song buzz and song sales is mainly caused by artist with a low reputation. It seems like artist who have not yet established a big reputation are mostly impacted by free sampling on blogs displacing sales. Chen et al. (2015) found that a decrease in sales from an artist can lead to a decrease in artist reputation, defined as google searches, in the following week.

3.4.2 Record Label

Similar to artist reputation, Dewan and Ramaprasad (2009, 2012 & 2014) include a dummy variable for the type of record label as a control variable in their model. The dummy variable is set to 1 if the music has been released by a major record label and to 0 if music is released by an independent record label. A record label is classified as a major label when it is part of the Recording Industry Association of America. The relationship between type of record label and music sampling is significant, indicating that songs released by independent labels are sampled more (Dewan & Ramaprasad, 2012). Dewan and Ramaprasad (2014) split their sample based on type of record label. From the results it appears that the negative relationship between song buzz and song sales is mainly caused by music released by independent labels. For albums released by major record labels the relationship between buzz and sales is insignificant. Results by Dhar and Chang (2009) indicate that albums released by a major label generally receive higher album sales.

3.4.3 Traditional Media

Chen et al. (2015) included expenses on traditional media in their analysis, which they obtained from an advertising intelligence company. Traditional media includes TV, radio, newspapers and magazines among others. Only 108 out of the 616 artists in their dataset used traditional forms of promotion. Their results imply that traditional media might have a benefit for artists who post more automated messages compared to personal messages on Myspace. In 2014, Dewan and Ramaprasad found that radio play has a positive and significant effect on sales, for both song and album sales. This is a short-term effect.

3.5 Research Model

From previous research it can be conclude that volume-related variables, sentiment-related variables and user profile characteristics on Twitter have all seen significant results when used as predictor variables. However, the importance of each type of social media predictor variable depends on the context it is used in. Based on previous research, a significant relationship is expected between all three types of social media predictor variables in the music industry. A schematic representation of the research model can be found below in figure 1. It should be noted that this model is a simplification of reality. There are many different factors influencing the amount of streams an album will receive. However, in this study the focus is on using metrics received from Twitter. The effects of additional variables related to the music industry will be explored in the additional analyses in chapter 5.3. Based on the literature review the following hypotheses have been formulated.

H1a: The volume of tweets for each album is positively associated with Spotify streams.
H1b: The volume of tweets for each artist is positively associated with Spotify streams.

H2a: Positive sentiment in tweets is positively associated with Spotify streams.
H2b: Negative sentiment in tweets is negatively associated with Spotify streams.

H3: Amount of followers is positively associated with Spotify streams.

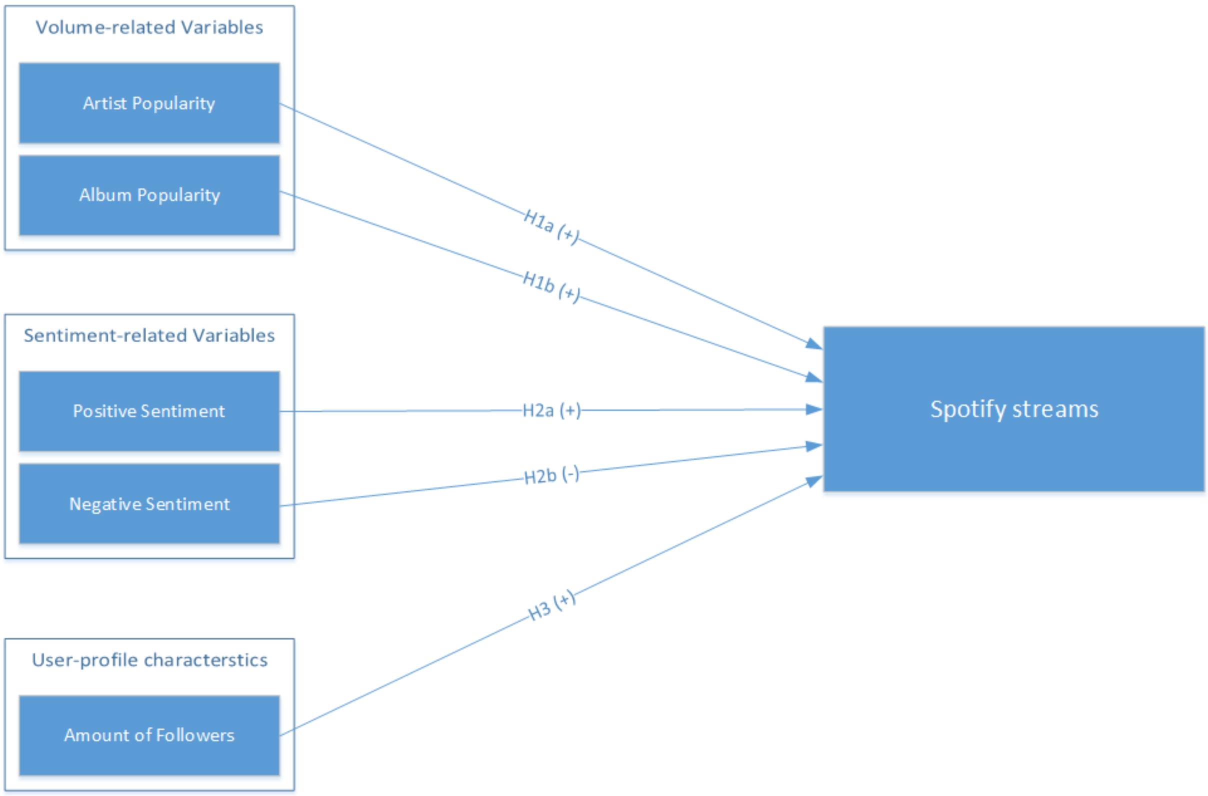

Figure 1: Research Model

4. Methodology

4.1 Data Collection

Twitter data has been collected for five weeks between the 26th of August and the 30th of September. Tweets were collected from the week before the album release till two weeks after release (week 0 to week 2). In order to perform predictive analysis Spotify streams were collected from the first week of release to three weeks after the release (week 1 to week 3). Figure 2 provides more detail into the timeline of the data collection. The specific dates in figure 2 refer to the data collection process for albums released on the 2nd of September. Week 0 refers to the week before the album release, also referred to as the pre-release week. Twitter data has been collected in this week in order to test the predictive power of Twitter (see chapter 5.3). However, albums cannot receive any streams before they are released and therefore there is no streaming data for week 0. Week 1 is the album release week, the first day of week 1 is the release day of the album. Week 2 and week 3 are the week after the release week and the second week after the release week respectively. Album release dates were retrieved from www.hitsdailydouble.com and www.pauseandplay.com. The amount of albums released per week differs strongly. Albums are selected based on how likely a proper set of search query can be constructed in order to retrieve relevant tweets. If an album uses a very commonly used word as a title it is therefore less likely to be used in this research because many irrelevant tweets would be collected. The focus is on newly released albums, therefore EPs, live performances or tribute albums are excluded from the dataset.

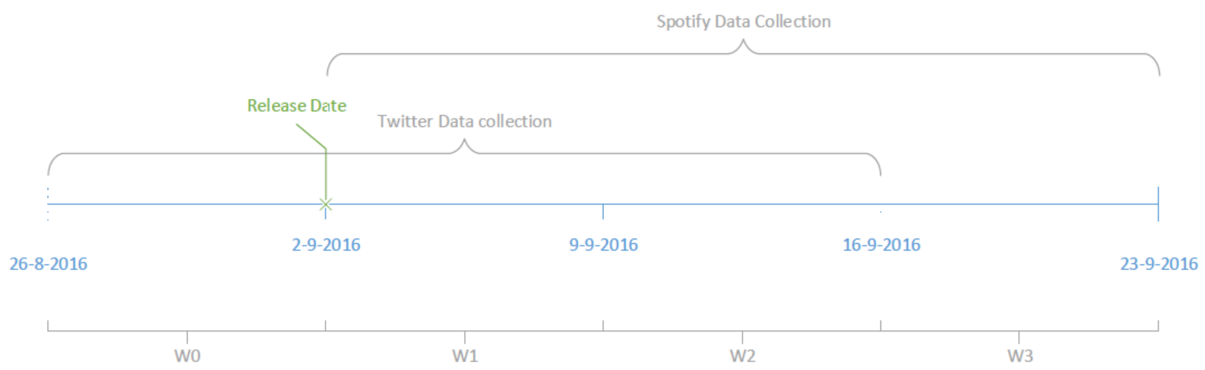

Figure 2: Timeline of the data collection; example of albums released on 02-09-2016

4.1.1 Twitter

The Twitter streaming application programming interface (API) was used to collect data from Twitter, which allows for the real-time collection of tweets. The Twitter streaming API was used to query the server in order to deliver tweets that contain selected keywords. Tweets were collected if they contained either the name of the artist or the name of the album. These tweets were then stored in a MySQL database. The week between the 19th of August and the 26th of August was used as a test week to see what type of keywords resulted in the collection of relevant tweets. As previous research showed album titles that consist of a general phrase result in the collection of many irrelevant tweets (Rui et al., 2013). For example, the album 'Amnesty' by Crystal Castles was mentioned in 94,906 tweets during this test week. However, many of these tweets were about Amnesty International and not about the Crystal Castles album. This test week showed the importance of using correct keywords and not just using the album title without consideration. Afterwards a new feature was implemented in the streaming API program that allowed for the use of the Boolean operator AND. Now, keywords can be more carefully chosen to collect more relevant tweets. Next to the content of the tweet, the account name, amount of followers of the account and the time stamp of the tweet were also collected using the streaming API. Our program collected tweets for five weeks between the 26th of August and the 30th of September. This resulted in the collection of 2,469,574 tweets from 898,837 unique users.

4.1.2 Spotify

The number of weekly streams per album were manually collected from Spotify. Because Spotify only publishes the daily streams for the 200 most streamed tracks per day, album streams had to be collected manually. As with Twitter, a short test of data collection was done to see if the streaming number of songs is updated regularly and correctly. Data was collected for a sample of songs including very popular songs and less popular songs. Song data and album data showed that streaming numbers were all updated regularly. For the tracks that were in the top 200 most streamed daily songs we could calculate the difference between the chart numbers and the manually collected numbers. The average difference between these two was 1.26%, which is probably caused by the official chart numbers being collected in a different 24 hour timeframe than the manually collected numbers. For the albums in our dataset, the number of streams received by singles and promotional songs released before the album release date were collected on Thursday night right before the release of the album. Afterwards, streaming data was collected weekly and previously released streams were subtracted to calculate the amount of new streams received per week.

4.1.3 Additional Data

The data was supplemented with additional variables next to data from Twitter and Spotify. The focus of this research is on the relationship between Twitter predictor variables and Spotify streams. However, research identified some additional variables that could play a role in this relationship. Therefore, additional variables were collected which will be used in the additional analyses in chapter 5.3. For each artist in the dataset the amount of Twitter followers and Spotify monthly listeners were manually collected on the night before release of the album. The amount of tracks on each album and whether the album was released by a major or independent label were also collected. More details regarding these additional variables can be found in table 4.

Additional Variables	
Twitter Followers	Amount of followers of the artist's official verified Twitter account collected the night before the album release.
Spotify Monthly Listeners	Amount of Spotify monthly listeners of the artist's official Spotify page collected the night before the album release.
Tracks	Amount of tracks on the album.
Record Label	A dummy variable indicating whether the album was released by a major or independent label collected from www.amazon.com. If an album was released by a major label = 1, by an independent label = 0. A record is classified as a major label when it is part of the Recording Industry Association of America (RIAA) (Dewan & Ramaprasad, 2014).

Table 4: Additional Variables

4.2 Data Analysis

4.2.1 Preprocessing

A small sample of the albums that tweets were collected for were not released on Spotify (for example, *Jason Aldean – They Don't Know*). Therefore, keywords for these albums were removed from the database. One extreme outlier (*Shawn Mendes – Illuminate*) was removed from the dataset. This album had extremely high values for both tweet volume and streams compared to the rest of the dataset. For example, the inclusion of this album resulted in the mean of the tweets containing the album name per week becoming ten times as big as its median. Finally, random samples per keyword were checked to test if these keywords actually captured tweets about the artist or album, or if these keywords were also used to discuss other things. Twenty-eight albums were used for the analysis. These albums and their release dates can be found in appendix F.

Before sentiment analysis can be performed on a text it is important that the text is properly preprocessed. Parts of the text that are not influencing the sentiment expressed in the tweet should be removed or replaced. For example, if a tweet is addressed to a user called @horrible than that tweet could be wrongly classified as negative despite the real message of the tweet being positive. All preprocessing and statistical analyses have been performed in R. The tweets in our dataset have been preprocessed using the following steps:

- Replace all capitalization with lower case
- Remove hashtags
- Remove @mentions
- Remove URLs
- Remove stop words (top 25)
- Album title/artist name filtered out. Before calculating the sentiment the name of the respective artist or album for which the tweet was collected is removed. These names are only removed for the sentiment analysis and not from the real dataset. This is to prevent tweets that mention album titles like 'Bad Vibrations' and 'False Readings On' to be classified as a false negative.

4.2.2 Sentiment Analysis

Various methods for performing the sentiment analysis have been explored before settling on the methods used in this study. First, the polarity function from the 'qdap' package in R was tested (White, 2016). This function calculates the average sentiment of each tweet by using a lexicon of positive and negative words. However, the results of this function were not satisfying because the average scores given did not correspond to our tweets. Second, the sentiment function from the 'sentiment140' R package was tested. Here, each tweet is classified as either negative, neutral or positive. This data was then used to calculate the ratios of positive and negative tweets. However, this function only estimated a very small proportion of the tweets to contain a sentiment, less than one percent in most cases. As this was much smaller than the sentiment ratios observed in previous papers other options were explored. Ceron et al. (2014, 2015) used the readme R package, which uses a machine learning approach. Unfortunately, the package has not been updated since 2013 and its mailing list for questions and discussions is not active anymore. After running into various errors with the readme method it was decided to go for a different approach. We explored the option of building our own classifier using machine learning (Asur & Huberman, 2010; Rui et al., 2013).

However, machine learning models require a large manually coded test set. Due to time and budget constraints the decision was made to use a lexicon-based approach. The recently released 'syuzhet' package in R was used to calculate the sentiment. This package offers four different lexicons to detect sentiment that all have roots in academic research. These lexicons are taken from: syuzhet (sentiment dictionary from the Nebraska Literary Lab), Liu, Hu and Cheng (2005), Nielsen (2011) and Mohammad and Turney (2010).

Sentiment analysis is still an immature topic. Experiments with previous methods showed the sensitivity of various methods. Therefore, a combination of all lexicons is used which only decides on the classification of sentiment when multiple lexicons agree. Each lexicon provides an average polarity score per tweet. A function was written that rounds up the value of sentiment classification for each method per tweet, where positive values become +1, negative values become -1 and zeros stay as 0. Each method produces an output with different ranges, so by rounding each individual score to +1, -1 or 0 they can be properly compared. Then, these scores are summed for each tweet and the tweet is only classified as positive if the sum of rounded sentiments is at least two, and negative if that sum is at least minus two. If the score is in between these values, the tweet is classified as neutral since it does not contain a clear sentiment. Combining the scores of these four methods and checking for congruency between them allows us to better classify the sentiment of tweets. Table 5 presents an example of the process. This table already features the rounded scores to -1, +1 or 0 for ten random tweets in the dataset.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Syuzhet	0	1	1	1	0	-1	1	1	1	0
Liu et al. (2005)	0	1	0	1	0	-1	1	1	1	0
Nielsen (2011)	0	0	-1	1	0	0	1	0	1	0
Mohammad & Turney (2010)	0	0	1	0	0	0	1	1	1	0
Total score	0	2	1	3	0	-2	4	3	4	0
Sentiment	Neu	Pos	Neu	Pos	Neu	Neg	Pos	Pos	Pos	Neu

Table 5: Example of the sentiment classification using the four lexicons. T1-T10 are ten random tweets from our dataset. Abbreviations: neu = neutral, neg = negative, pos = positive.

4.3 Model Design

The Twitter predictor variables used in this research can be split up in three categories:

- *Volume-related variables:* The volume of tweets mentioning a certain album per week: album popularity. And the volume of tweets mentioning a certain artist (that released said album) per week: artist popularity.
- *Sentiment-related variables:* Positive ratio, negative ratio.
- *Profile characteristics of online users*: Ratio of type 2 followers.

These Twitter predictor variables are the independent variables used in this study. Spotify streams is the dependent variable. The key variables are further explained in table 6.

Key Variables	**Description**
Album Popularity	The volume of tweets mentioning album *x* per week from Friday to next Thursday.
Artist Popularity	The volume of tweets mentioning artist *y, who released album x,* per week from Friday to next Thursday.
Positive Ratio	Ratio of tweets expressing positive sentiment toward album x per week.
Negative Ratio	Ratio of tweets expressing negative sentiment toward album x per week.
Type 2 Follower Ratio	Ratio of tweets from users with more than 1500 followers who tweeted about album x per week.
Spotify Streams	The volume of streams on Spotify received by album x per week from Friday to next Thursday.

Table 6: Key variables

Before performing linear regression, correlations were calculated first. Because the data in our dataset is not normally distributed, Spearman's rho was chosen instead of Pearson's r. Spearman's correlation coefficient is denoted as r_s and is restricted by $-1 \leq r_s \leq 1$, where the closer r_s is to 1 the stronger the positive correlation. Similarly, the closer r_s is to -1, the stronger the negative correlation. Spearman's rho is calculated according to equation 1 presented below, where r_s is the correlation coefficient (Spearman's rho), d is the difference in ranks between a pair of scores and n is the number of ranks.

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \tag{1}$$

After performing the correlational analysis the relationship between the key variables will be tested using multiple linear regression analysis. The multiple linear regression equation is presented below in equation 2. Let Y denote the amount of Spotify streams and ϵ the error. Then β_i refers to the regression coefficient for each parameter. The parameters are:

- X1: Album Popularity
- X2: Artist Popularity
- X3: Positive Ratio
- X4: Negative Ratio
- X5: Type 2 Follower Ratio

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon \tag{2}$$

In order to explore the relationship between the Twitter predictor variables and Spotify streams, multiple linear regression is first performed on these variables in the same week. Using stepwise regression the independent variables are gradually entered into the regression model. After focusing on the relationships between these variables in the same week, the predictive power of the Twitter predictor variables will be tested. Once more multiple linear regression will be performed. However, in this case one-period-lagged values will be used (Rui et al., 2013; Sharma et al., 2012). Thus, the Twitter variables of $week_t$ will be used to predict the Spotify streams in $week_{t+1}$. For example, the

Twitter predictor variables from the week before release (week 0) will be used to predict the Spotify streams received in the release week (week 1).

The main measure for testing the fit of the overall regression model is the coefficient of determination, the R^2. The R^2 measures the proportion of variance in the dependent variable that is explained by the independent variables. While the R^2 always increases when more variables are added to the linear model, the adjusted R^2 takes the addition of extra variables into account. Therefore, the adjusted R^2 is used in this study.

Kalampokis et al. (2013) provided a framework to analyze current research and design future research into the predictive power of social media. This framework can be found in figure 3. This framework was followed throughout the design of this research. A structured overview of the methodology used in this paper can be found in the completed framework in Appendix G.

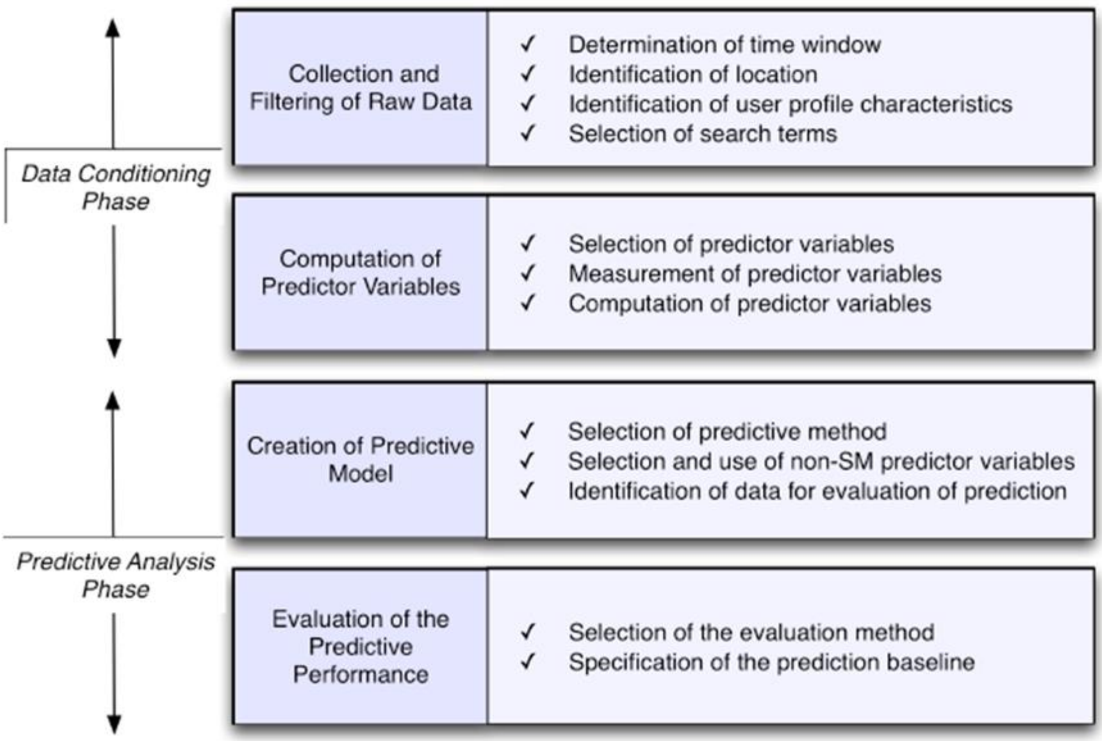

Figure 3: Social Media Analysis Framework for Predictions by Kalampokis et al. (2013)

5. Results

5.1 Dataset Characteristics

Summary statistics for all variables can be found in table 7. For albums released close to the 30th of September it was impossible to collect Twitter data for two weeks after the release date since the data collection had already been closed. Table 7 provides more detail into how much data is included in the analysis per week. When inspecting the mean values in table 7 a big peak of the volume of tweets containing the album name (album popularity) is observed in the release week (5151). As one might expect, most people would tweet about an album shortly after it has been released. The volume of tweets containing the album name (album popularity) is also higher in week zero than in week two, suggesting that pre-release promotional activities increase the volume of tweets. From now on, variables are referred to by their variable names as described in table 6. Therefore, instead of ‘the volume of tweets mentioning the artist’, we will now refer to this variable as ‘artist popularity’. For artist popularity we see a similar peak during release week, although the difference in volume between the week before release and week two is much smaller than for album popularity. Asur and Huberman (2010) observed an increase in the amount of sentiments after a movie release. In our dataset the ratios of positive and negative sentiment per week are constant across all weeks. As expected the amount of Spotify streams peak during the first week of release and steadily decrease with 25.7% and 29.2% in the second and third week.

	Week 0	**Week 1**	**Week 2**	**Week 3**
Album Popularity	1,803	5,151	1,329	
Artist Popularity	9,889	17,487	8,552	
Positive Ratio (%)	29.72	32.66	31.32	
Negative Ratio (%)	8.71	7.219	11.85	
Type 2 Follower Ratio (%)	25.09	22.46	25.36	
n	*28*	*24*	*19*	
Spotify Streams		2,415,820	1,983,524	1,405,063
n		*28*	*24*	*19*

Table 7: Mean values of key variables per week

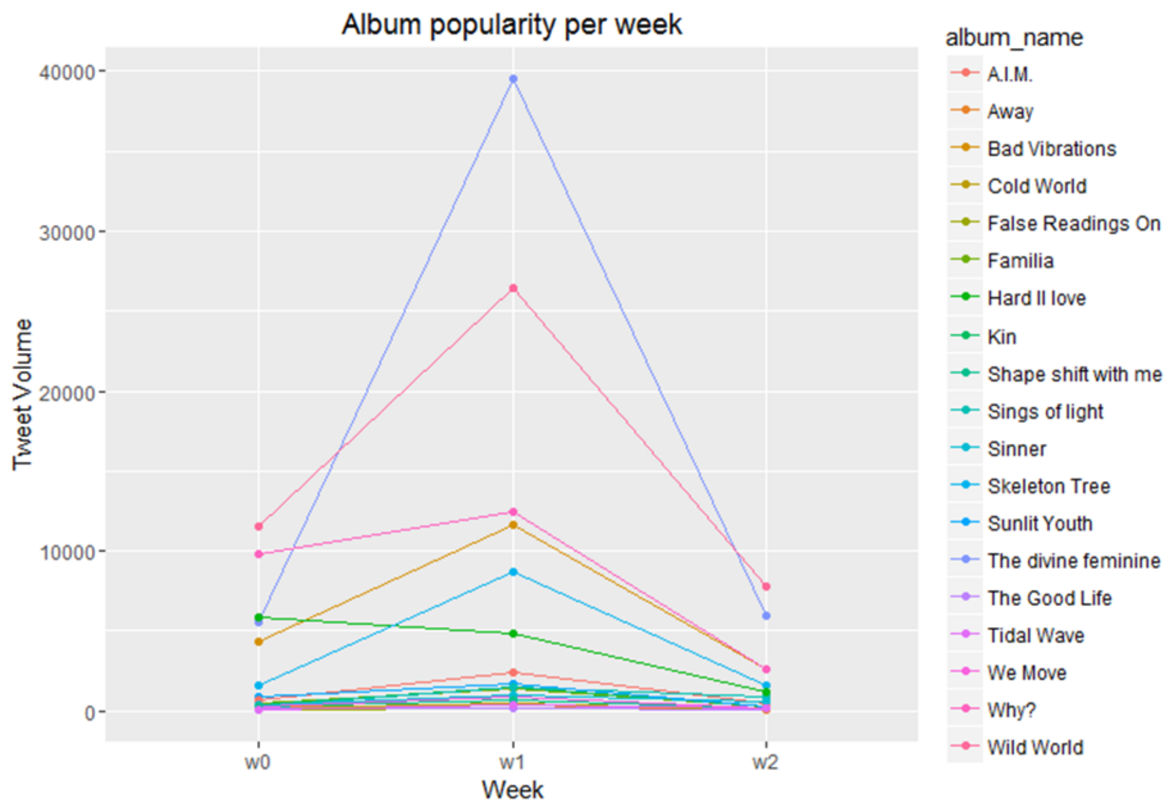

Figure 4: Volume of tweets containing the album name per week (album popularity).

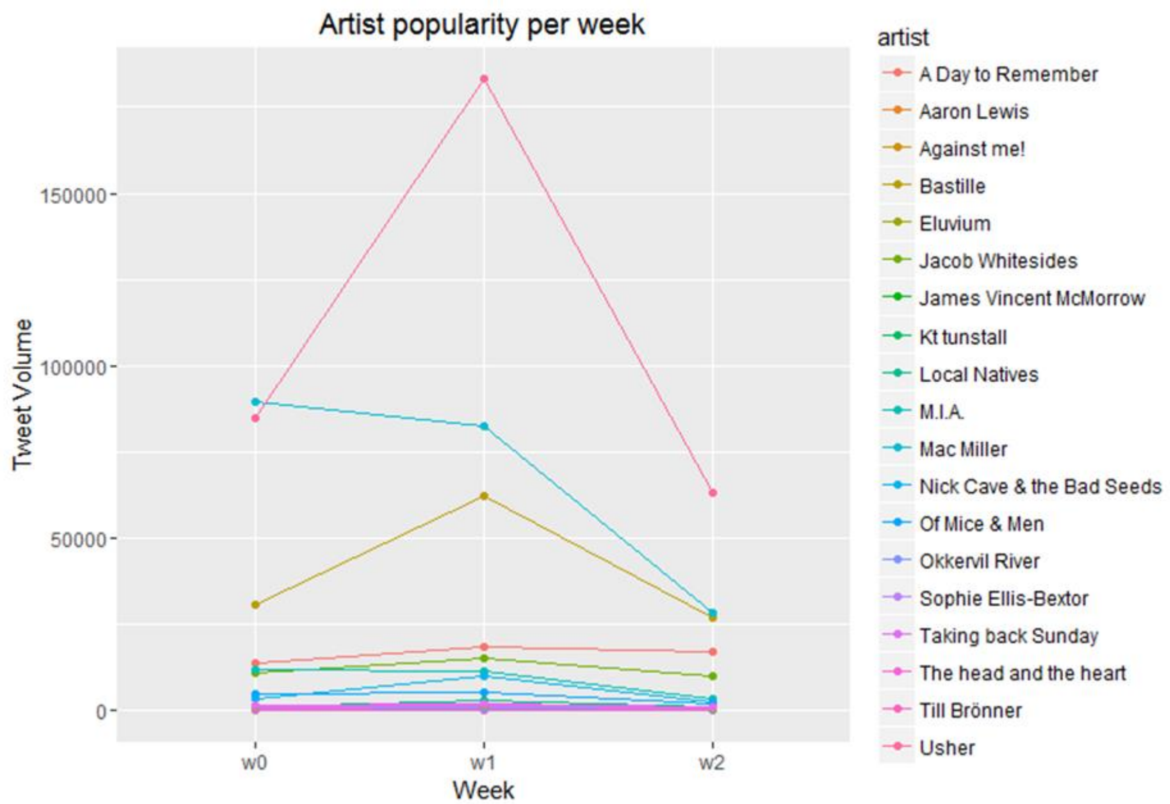

Figure 5: Volume of tweets containing the artist per week (artist popularity).

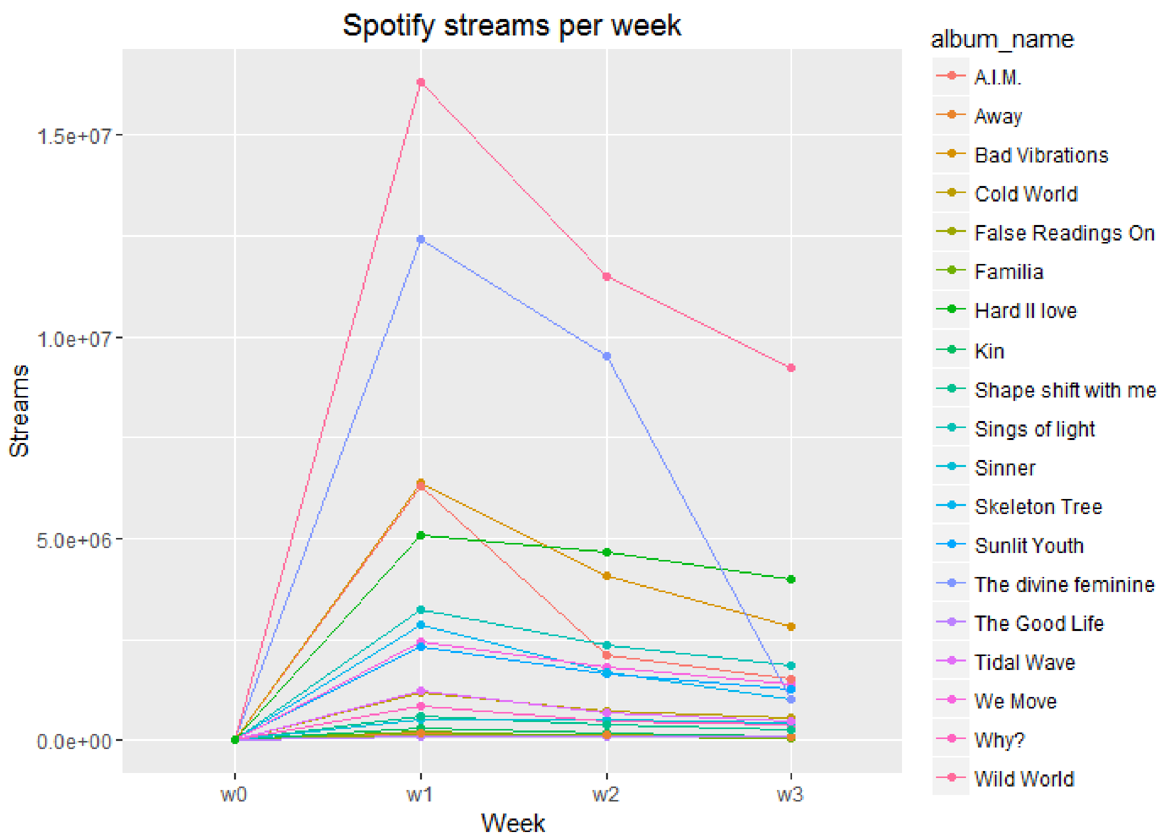

Figure 6: Amount of Spotify streams per week. W0 denotes Week 0, which is the week before the release of the album. Therefore, all albums received zero streams in week 0.

In order to further explore the weekly trajectory of the volume-related variables and streams, charts were made which can be found in figures 4-6. Figures 4-6 present weekly data which is also used in further analysis. Only albums for which data in all weeks was available are included in these charts, which means that nineteen albums have been included. For album popularity a clear peak during the release week can be observed in figure 4. The only album that does not follow this trend is 'Hard II Love' which shows a decrease from week 0 to week 2. Generally, tweet volume in week 0 seems to be slightly higher than the tweet volume in week 2. Figure 5 presents the artist popularity per week, which shows a similar peak in week 1 like figure 4. However, the tweet volume containing the artist 'Mac Miller' shows a different pattern by peaking in week 0. The amount of Spotify streams are shown in figure 6. The amount of streams peaks in the release week and steadily decreases afterwards.

With regard to the peak of album popularity in week 1 an additional chart was created which can be found in figure 7. For this chart the album popularity has been calculated per day instead of per week. The chart in figure 7 shows a big peak on release day (day 8) for most albums, which is represented by the black dotted line. The build-up of tweet volume containing the album name starts a couple of days before the album release and can be clearly seen in this chart. After about two to three days after the album release most hype seems to settle down. Because the lower half of this chart is very crowded another version of this chart that explores the lower half of the chart (max tweet volume = 1500) is included in Appendix H. For albums that received a lower average amount of tweets and reside in the lower half of figure 7 the same trends are observed.

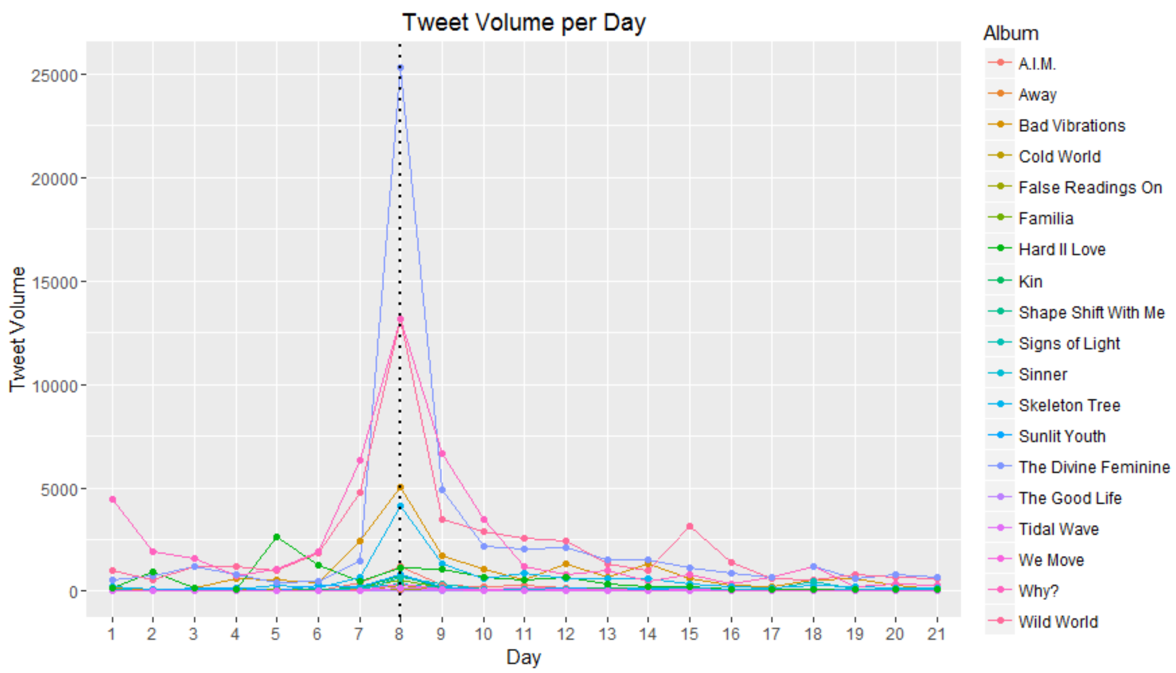

Figure 7: Tweet volume of tweets containing the album name (album popularity) per day. The dotted line on day 8 represents the album release.

Before entering the variables into the multiple linear regression model, correlations were calculated. The correlation matrix, including two-tailed p-values, can be found in table 8. Spearman's rho correlations were calculated using only complete observations (n = 43). A high positive correlation is observed between album popularity and artist popularity, r_s = .83, p < .01. This suggests that the volume of tweets about album and artist are closely related. The positive sentiment ratio of tweets is positively correlated with both album popularity (r_s = .35, p < 0.05) and artist popularity (r_s = .45, p < .01). It seems like the ratio of positive tweets increases along the volume of tweets. Between Spotify streams and the independent variables, multiple significant correlations can be observed. Both album popularity (r_s = .69, p < .01) and artist popularity (r_s = .82, p < 0.01) show a significant positive correlation with Spotify streams. The type 2 follower ratio displays a significant negative correlation with Spotify streams (r_s = -.30, p < .05). These correlations suggest that while the volume of tweets about the album and artist increase, Spotify streams also increase. On the other hand, when the amount of tweets from users with a high following increases, Spotify streams seem to decrease.

	Album Popularity	Artist Popularity	Positive Ratio	Negative Ratio	Type 2 Follower Ratio	Spotify Streams
1. Album Popularity	-					
2. Artist Popularity	0.83**	-				
3. Positive Ratio	0.35*	0.45**	-			
4. Negative Ratio	0.07	0.15	-0.08	-		
5. Type 2 Follower Ratio	-0.25	-0.19	-0.18	0.21	-	
6. Spotify Streams	0.69**	0.82**	0.29	0.13	-0.30*	-

*Table 8: Correlation matrix. Significance levels = ** p < 0.01, * p < 0.05*

5.2 Model Testing

This section will present the results of the multiple regression analyses that will be used to test our hypotheses. First, multiple regression analysis is performed on the Twitter predictor variables and Spotify streams from the same week in order to test the relationship between these variables. Next, Twitter predictor variables are used in order to predict Spotify streams in the subsequent week.

5.2.1 Multiple Linear Regression for Same Week

In this section the relationship between the Twitter predictor variables and Spotify streams within the same week are tested. Both Twitter variables and the volume of streams from week 1 and week 2 are entered into a separate regression model. The independent variables are added into the model stepwise in order to further explore the effects of including each variable. First, the full model is explored and later the separate steps are discussed.

Multiple linear regression was used to test if Twitter predictor variables significantly explained Spotify streams in week 1. Table 9 presents the results from multiple linear regression and includes the standardized regression coefficient ß as well as the significance levels. Results of the regression analysis indicated that the five Twitter predictors explained 71.1% of the variance in Spotify streams (adj. $R^2 = .711$, $F(5, 18) = 12.34$, $p < .001$). The analysis shows that album popularity was a significant predictor of Spotify streams in week 1 (ß = .751, $p < .01$). However, artist popularity (ß = .262, ns), positive ratio (ß = -.103, ns), negative ratio (ß = -.012, ns) and type 2 follower ratio (ß = -.175, ns) were not significant predictors of Spotify streams. When exploring the stepwise additions in table 9 it becomes clear that album popularity was the significant predictor in each step. Even when only album popularity is included in the regression this predictor variable is able to explain 69.9% of variance in Spotify streams (adj. $R^2 = .699$, $F(1, 22) = 54.45$, $p < .001$). This suggests a very strong relationship between the amount of times an album name is mentioned on Twitter and the amount of Spotify streams said album receives in the same week. Adding artist popularity to the model creates a slight increase of the adjusted R^2 to .715.

Week 1

	1	2	3	4	5
	ß	ß	ß	ß	ß
Album Popularity	0.844***	0.753***	0.776***	0.783***	0.751***
Artist Popularity		0.190	0.233	0.235	0.262
Positive Ratio			-0.085	-0.095	-0.103
Negative Ratio				-0.023	-0.012
Type 2 Follower Ratio					-0.175
Adjusted R^2	0.699	0.715	0.706	0.691	0.711
Degrees of Freedom	22	21	20	19	18
F	54.45	29.88	19.36	13.83	12.34
n	24	24	24	24	24

*Table 9: Results from stepwise multiple linear regression of Twitter predictor variables on Spotify streams in week 1. Significance levels = *** p < 0.001, ** p < 0.01, * p < 0.05*

Similarly, multiple linear regression was also performed for week 2 and is presented in table 10. Results of the regression analysis indicate that the Twitter predictor variables were able to significantly explain Spotify streams in week 2 (adj. R^2 = .919, $F(5, 13) = 41.76$, $p < .001$). Further analysis shows that significant predictors of Spotify streams were album popularity (ß = .750, $p < .001$), artist popularity (ß = .277, $p < .01$) and type 2 follower ratio (ß = -.210, $p < .01$). The sentiment-related variables, positive ratio (ß = -.031, ns) and negative ratio (ß = -.031, ns) were both not significant. While artist popularity and the type 2 follower ratio were not significant in week 1, they are both significant in week 2. Additionally, the adjusted R^2 is much higher in week 2 ($R^{2\,=}$ .919) than in week 1 (R^2 = .711). These results suggest that the release week could be more difficult to predict than other weeks. This also shows that the Twitter predictor variables explained 91.9% of the variance of Spotify streams in week 2.

The results for controlling on sentiment in both weeks was very weak. There is no regression model where either of the two sentiment-related variables was a significant predictor of Spotify streams. In most cases the positive ratio even has a negative relationship with Spotify streams, although this relationship is far from significant. There seems to be no significant relationship between the sentiment of tweets and the amount of Spotify streams. In three out of four cases adding sentiment to the model actually decreased the adjusted R^2 measure. The only exception is the addition of the positive ratio in week 2 where the R^2 slightly increases, although its p-value was very far from significance. In all cases the p-values of the sentiment-related variables hovered around 0.8 which indicates that the sentiment-related variables show no relationship with Spotify streams.

Week 2

	1	2	3	4	5
	ß	ß	ß	ß	ß
Album Popularity	0.924***	0.794***	0.794***	0.808***	0.750***
Artist Popularity		0.249*	0.246*	0.242*	0.277**
Positive Ratio			0.007	-0.006	-0.031
Negative Ratio				-0.039	-0.031
Type 2 Follower Ratio					-0.210**
Adjusted R^2	0.845	0.886	0.879	0.872	0.919
Degrees of Freedom	17	16	15	14	13
F	98.91	71.05	44.42	31.54	41.76
n	19	19	19	19	19

*Table 10: Results from stepwise multiple linear regression of Twitter predictor variables on Spotify streams in week 2. Significance levels = *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$*

5.2.2 Multipe Linear Regression for Prediction

The previous section used both independent and the dependent variables from the same week. In this section the predictive power of the Twitter variables on Spotify streams is analyzed. In order to test the predictive power of Twitter one-period-lagged values are used. This means that the Twitter variables of $week_t$ will be used to predict the Spotify streams in $week_{t+1}$.

In this chapter the sentiment-related variables are kept out of the analyses. The sentiment-related variables showed to have no relation with Spotify streams in all models in the previous chapter. In order to effectively test the predictive power of the significant Twitter predictor variables the sentiment-related variables are not included in the regression models. The adjusted R^2 takes into account the addition of extra variables in the linear model. Therefore, adding the sentiment-related variables into these predictor models while they have shown to have no relationship with Spotify streams would unnecessarily influence this measure in a negative direction.

Table 11 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 Spotify streams in week 1 (adj. $R^2 = .607$, $F(3, 24) = 14.92$, $p < .001$). Significant predictors were album popularity (ß = .511, $p < .01$) and artist popularity (ß = .378, $p < .05$). Using Twitter predictor variables from week 0 the linear regression model is able to predict 60.7% of Spotify streams in the following week. The Twitter predictor variables from week 1 were able to predict 78.6% of Spotify streams in week 2 (adj. $R^2 = .786$, $F(3, 220) = 29.13$, $p < .001$). Like the previous model album popularity (ß = .684, $p < .001$) and artist popularity (ß = .291, $p < .05$) were significant predictors. Finally, the Twitter predictor variables from week 2 were also able to significantly predict Spotify streams in week 3 (adj. $R^2 = .541$, $F(3, 15) = 8.08$, $p < .01$).

In all weeks album popularity has a significant effect on Spotify streams. However, its predictive power is not always as strong as the relationship between album popularity and Spotify streams in the same week were (see table 9 and 10). On the other hand, artist popularity from week 0 is a significant predictor of Spotify streams in week 1. While artist popularity in week 1 was not a significant predictor of streams in week 1 (see table 9). For all models, the type 2 follower ratio is not a significant predictor of Spotify streams. Comparing the adjusted R^2 measures of all three models shows that the model that uses the Twitter predictor variables from week 1 to predict Spotify streams in week 2 performs best with an adjusted R^2 of .786. While the model predicting week 1 and week 3 had an R^2 of .607 and .541 respectively. The amount of received streams in the release week (week 1) could be more difficult to predict because of other influences related to promotional activities. Although it is unclear why Twitter variables perform much better in predicting week 2 than in predicting week 3. Similar to the previous chapter all models show the importance of the volume-related variables in predicting Spotify streams.

	Week 0 -> Week 1	Week 1 -> Week 2	Week 2 -> Week 3
	ß	ß	ß
Album Popularity	0.511**	0.684***	0.527*
Artist Popularity	0.378*	0.291*	0.318
Type 2 Follower Ratio	-0.233	-0.204	-0.158
Adjusted R2	0.607	0.786	0.541
Degrees of Freedom	24	20	15
F	14.92	29.13	8.08
n	28	24	19

*Table 11: Results from multiple linear regression between the Twitter variables and Spotify streams using one-lagged-period values. Significance levels = *** p < 0.001, ** p < 0.01, * p < 0.05*

5.3 Additional Analyses

This section will explore additional analyses next to the main relationships between our variables. First, additional sentiment analysis is performed and tweets regarding albums and artists are compared. Second, a daily time series is performed on the week leading up to the album release. Third, exploratory analysis explores demographic influences like age and gender. Finally, the additional variables like Spotify Monthly Listeners are included and the best performing models are presented.

5.3.1 Further Sentiment Analysis

After the negative results of adding the sentiment-related variables into the previous models, more options were explored regarding sentiment. Different methods that were earlier discarded during the methodology phase (qdap, sentiment140) were used again to calculate the sentiment of tweets. However, multiple regression analysis similarly showed no significant relationship between these sentiment scores and Spotify streams. Therefore, we stick with the syuzhet method for sentiment analysis. The sentiment-related variables have been calculated over tweets that represented the album (album popularity) because we expected those tweets to have the biggest impact on Spotify streams. However, people might not always mention the album name when tweeting their sentiment about it. For example, a Twitter user might tweet 'I love Mac Miller's new album', instead of 'I love the Divine Feminine'. Therefore, this section explores the sentiments contained in tweets about the artist.

Furthermore, the type 2 follower ratio also performed differently than expected in the previous regression models. In the calculation of the type 2 follower ratio a cut-off point of 75% was used, which was 1500 followers. This means that 25% of the tweets in the dataset came from users with 1500 or more followers. In the study by Rui et al. (2013) the type 2 follower ratio was a significant predictor with a different cut-off point of 85%. The cut-off point for our dataset using the 85% ratio is 2500 followers. Therefore, the type 2 follower ratio were also calculated again for this section (5.3.1) using the new cut-off point of 2500 followers.

Next, the sentiment-related variables and type 2 follower ratio for tweets containing the album name are compared to these ratios for tweets containing the artist. The comparison is made for the multiple linear regression that achieved the highest adjusted R^2 in the previous analyses, which is the model that uses Twitter variables from week 2 and Spotify streams from week 2. Table 12 presents the results. The model that contains the album-related variables performs much better (adj. R^2 = .868, $F(4, 14) = 30.67$, $p < .001$) than the model containing the artist-related variables (adj. R^2 = .350, $F(4, 14) = 3.42$, $p < .05$). These results suggest that our choice to calculate the sentiment-related variables and the type 2 follower ratio over tweets containing the album name was the proper choice. In both cases the sentiment-related variables are not significant and neither is the type 2 follower ratio.

Therefore, it seems like there is absolutely no relationship between the sentiment of tweets and the amount of Spotify streams in our dataset.

	Week 2	
	Album	Artist
	ß	ß
Album Popularity	0.874***	-
Artist Popularity	-	0.672**
Positive Ratio	0.073	-0.167
Negative Ratio	-0.053	-0.047
Type 2 Follower Ratio	-0.182	-0.145
Adjusted R2	0.868	0.350
Degrees of Freedom	14	14
F	30.67	3.419
n	19	19

Table 12: Comparison of album-related Twitter variables and artist-related Twitter variables; results from linear regression between the Twitter variables and Spotify streams performed using variables from week 2.
*Significance levels = *** p < 0.001, ** p < 0.01, * p < 0.05*

5.3.2 Time Series Analysis

Asur and Huberman (2010) were able to predict box-office revenues of movies with high accuracy using volume-related variables. Adding sentiment to their linear regression model only slightly improved its performance. Thus far our results have shown the importance of volume-related variables as well. In this way the relationship between Twitter predictor variables and performance outcomes seems similar in the movie industry and the music industry. After using a weekly measure of tweet volume, Asur and Huberman (2010) built a time series of days in the week before the movie release and used this time series to predict box-office revenues in the opening weekend of the movie. In this section, similar to Asur and Huberman (2010), we also built a time series of daily tweet volume in the week before the album release (week 0). This allows us to compare our results to those of Asur and Huberman (2010) and to compare the movie and music industries. The same methodology is used as Asur and Huberman (2010), which means that each day in the week till release is included as a separate variable in the regression model. Performance of the individual variables (the days) is much less important in this model. Instead, the focus switches to the overall performance of the model, which is measured by the adjusted R^2. The time series of album popularity is shown in figure 8. Day 7 is the day before the album release. Generally, the tweet volume seems to start increasing around one or two days before the album release. Figure 8 shows that two albums do not follow this trend, which is probably caused by promotional activities on those days or by other influences.

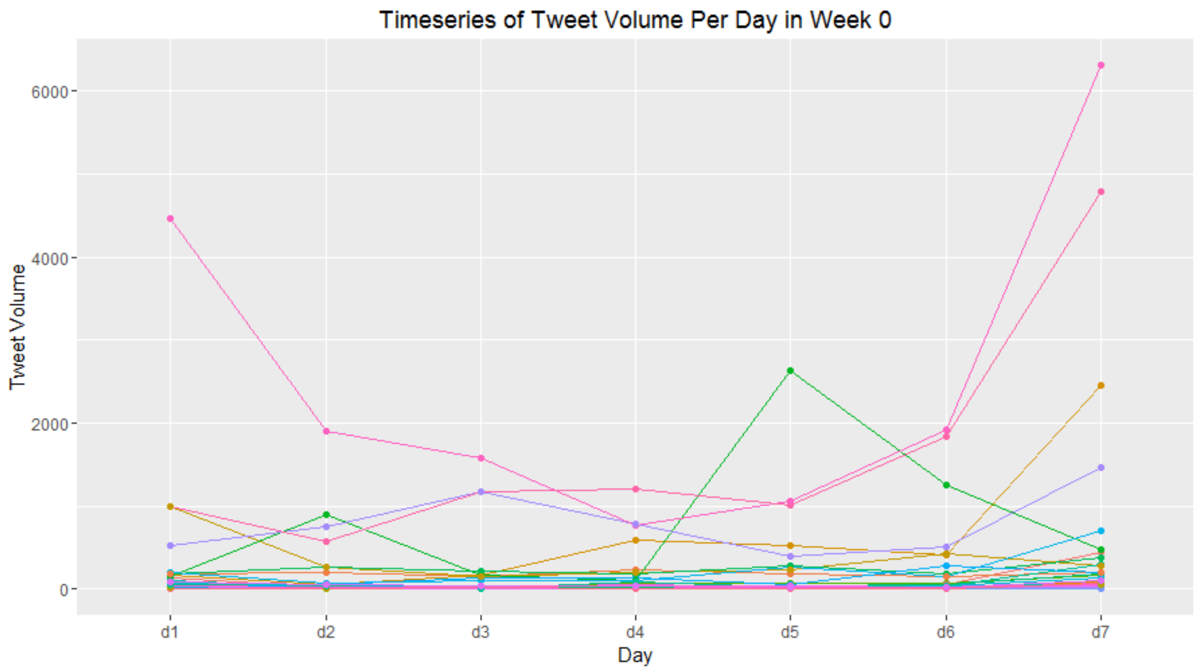

Figure 8: Time series of the tweet volume per day for album popularity. Each line refers to an album. D7 is the day before the album release.

Table 13 presents a comparison of our results and those of Asur and Huberman (2010). Using a weekly measure of tweet volume, Asur and Huberman's linear model achieved an adjusted R^2 of .80. In predicting the Spotify streams in week 1 using the average tweet volume of week 0 our model achieved an adjusted R^2 of .47. The time series achieves a considerably better performance (adj. R^2 = .84). In order to capture the availability of each movie Asur and Huberman (2010) included the variable 'theaters' in their model which represents the amount of theaters the movie played in. In our model we add the variable 'Spotify monthly listeners' which represents the amount of unique listeners that listened to the artist who released the album in the last 28 days. While this variable does not perfectly capture 'availability' it does capture how many people would probably receive notifications about the release of the album. Similar to Asur and Huberman (2010) our model improves by adding this variable. Concluding, building a daily time series of days before the release sees a higher predictive power than using the weekly average. While our model does not reach the heights of the adj. R^2 of Asur and Huberman (2010) it still achieves a very high adjusted R^2.

	Asur & Huberman (2010)	**This study**
	Adj. R^2	Adj. R^2
Average weekly volume	0.80***	0.47***
Time series	0.93***	0.84***
Time series + Theaters/Spotify Monthly Listeners	0.97***	0.88***

*Table 13: Comparison of time series results. *** p < .001 for the significance of the overall model*

5.3.3 Demographic Influences

After analyzing the general effects of social media predictor variables on Spotify streams, this section explores whether these effects could be different for subsets of the data based on demographic influences. While our dataset does not contain enough albums to perform rigorous subsetting it is interesting to explore if the effects of social media could be different for various groups of artists. The model we use is that of week 1. This model uses a higher amount of albums than the model of week 2 and is therefore more resistant to sub-setting. First, the dataset is subsetted based on age. The dataset is split along the median of the age of the musicians resulting in two subsets of the same size, young and old. Table 14 shows that the Twitter predictor variables significantly predict Spotify streams for albums released by younger musicians (adj. R^2 = .821, F(3, 8) = 17.82, p < .001), while this is not the case for older musicians (adj. R^2 = .279, F(3, 8) = 29.88, ns). Compared to previous models, artist popularity seems to play a more important role for the younger group of musicians (ß = 1.080, ns). Although none of the individual variables reach significance. These results suggest that the effect of Twitter predictor variables on Spotify streams could be different depending on the age of the musician.

Second, the dataset was split-up based on the type of performer. Here the categories are: male, female and band. Table 14 shows that the regression models perform well for males (adj. R^2 = .939, F(3, 5) = 42.24, p < .001) and bands (adj. R^2 = .967, F(3, 5) = 78.61, p < .001). For females the model performs less good in predicting Spotify streams and is not significant (adj. R^2 = .531, F(3, 2) =2.887, p < .001). However, we have to be very careful in analyzing this regression model. Because there are only six albums released by females included in this dataset, the degrees of freedom becomes smaller than the amount of variables in the model. This is probably the main reason that this overall model does not reach significance. Because the amount of female performers in our dataset is small it is difficult to say something conclusive about the effects of gender in our dataset. For this same reason it is not possible to subset the dataset based on record label (Dewan & Ramaprasad, 2012), because only six albums in our dataset were released by independent record labels.

Week 1

	All	Young	Old	Male	Female	Band
	ß	ß	ß	ß	ß	ß
Album Popularity	0.722***	-0.226	0.303	0.745***	-0.047	0.120
Artist Popularity	0.211	1.080	0.398	0.334*	1.212	0.838*
Type 2 Follower Ratio	-0.173	-0.192	0.227	-0.241*	-0.390	-0.119
Adjusted R^2	0.734	0.821	0.279	0.939	0.531	0.967
Degrees of Freedom	20	8	8	5	2	5
F	22.19	17.82	29.88 (ns)	42.24	2.887 (ns)	78.61
n	24	12	12	9	6	9

Table 14: Results from multiple linear regression of Twitter predictor variables on Spotify streams in week 1. 'All' represents the data without subsetting. Afterwards, data has been subsetted based on age and type of performer.
*Significance levels = *** p < 0.001, ** p < 0.01, * p < 0.05, ns = not significant*

5.3.4 Additional Variables

As explored in the chapter on time series (5.3.2) Asur and Huberman (2010) added the amount of theaters a movie played in to their regression model. Kim et al. (2014) added the amount of weeks an album had been on the billboard album chart to their model. In this final section of additional analyses the effect of adding additional variables to our model is explored. Table 15 presents the mean values of the additional variables.

	Mean value
Twitter Followers	973,300
Spotify Monthly Listeners	2,053,000
Tracks	12.29
Record Label	0.79

Table 15: Mean Values of Additional variables

After separately entering these variables in our multiple regression models, Twitter followers, tracks and record label were found to not be significant predictors of Spotify streams. However, as was shown by the time series analysis, Spotify monthly listeners (SML) is a significant predictor of Spotify streams. The Spotify metric 'Spotify Monthly Listeners' measures the amount of unique listeners who listened to an artist in the last 28 days (Spotify, 2016). We collected the Spotify monthly listener numbers for each artist in our dataset the day before their newest album was released. It should be noted that this variable was thus measured before the release of the album and is therefore not influenced by people who streamed the album. In order to explore the effect of adding SML the variable is added to our best performing multiple regression models. These models are the models that tested the relationships between variables in week 2 and the model that used Twitter predictor variables from week 1 in order to predict Spotify streams in week 2. Variables that were found to not be significant predictors in these models are left out in order to explore the best performance of these models. The results of adding SML to these models can be found in table 16. This table first presents the models before and then after adding SML. Adding the Spotify Monthly Listeners metric greatly improved these models ((adj. R^2 = .944, $F(3, 16)$ = 152, $p < .001$); (adj. R^2 = .931, $F(3, 20)$ = 104.2, $p < .001$)). Spotify Monthly Listeners is a significant predictor in both models (ß = .400, $p < .001$; ß = .647, $p < .001$).

	Week 2		**Week 1 -> Week 2**	
	Baseline	Baseline + SML	Baseline	Baseline + SML
	ß	ß	ß	ß
Album Popularity	0.924***	0.672***	0.722***	0.630***
Artist Popularity	-	-	0.265*	-0.187*
Spotify Monthly Listeners	-	0.400***	-	0.647***
Adjusted R^2	0.845	0.944	0.752	0.931
Degrees of Freedom	17	16	21	20
F	98.91	152	35.81	104.2
n	19	19	24	24

Table 16: Linear regression models with the highest performance without (baseline) and with Spotify Monthly Listeners (Baseline + SML)

6. Analysis

After performing the regression and additional analyses in the previous chapter, this chapter summarizes the results and compares them with the findings from previous literature. Based on the analyses it is now possible to either accept or reject our hypotheses. The multiple regression analyses in chapter 5.2 were solely meant to test our hypotheses. Additional analyses were performed in chapter 5.3 in order to gain extra understanding into these relationships and explore other explanations behind them. The sub-questions formulated in chapter 1.3 will also be answered in this chapter. First, the results on our hypotheses are reported. Table 17 provides a summary of the results on each hypothesis.

H1a: The volume of tweets for each album is positively associated with Spotify streams.

The volume of tweets for each album has been measured through the variable 'album popularity'. In every multiple linear regression analysis performed in chapter 5.2, album popularity was found to be a significant predictor of Spotify streams. In the models that explored the relationship between album popularity and Spotify streams in the same week album popularity was significant at the < .001 level in both week 1 and week 2. In all cases the relationship between album popularity and Spotify streams was positive. A daily time series of days leading up to the release week of the album also showed a positive and significant relationship between album popularity and Spotify streams. This model of daily time series was able to predict Spotify streams in the first week with high accuracy ($R^2 = .84$). Based on these results H1a is accepted.

H1b: The volume of tweets for each artist is positively associated with Spotify streams.

The tweet volume for each artist has been measured through the variable 'artist popularity'. Similar to Kim et al. (2014), it as expected that there would be a positive association between artist popularity and Spotify streams. In all regression models in chapter 5.2 artist popularity had a positive relationship with Spotify streams. In three out of five of these models, artist popularity was a significant predictor of Spotify streams. These three models are: the model that used both Twitter predictor variables and Spotify streams from week 2, the model that used Twitter predictor variables from week 0 to predict Spotify streams in week 1 and the model that used Twitter predictor variables from week 1 to predict Spotify streams in week 2. Therefore, H1b is accepted.

H2a: Positive sentiment in tweets is positively associated with Spotify streams.

Multiple linear regression analyses for both week 1 and week 2 showed no association between positive sentiment in tweets and Spotify streams. The regression models for prediction also showed no significant association between positive sentiment and Spotify streams. Further sentiment analyses were performed in our additional analyses. The effects of positive sentiment contained in tweets about the album and tweets about the artist were compared. However, none of these models showed a significant relationship with Spotify streams. These results are not in line with our expectations. Ceron et al. (2014, 2015) and Rui et al. (2013) suggested a positive association between positive sentiment contained in tweets and 'positive outcomes'. However, Asur and Huberman (2010) described that the addition of sentiment in their linear model only slightly improved its performance. Liu et al. (2016) was also sceptical about the relationship between sentiment and box-office scores. Based on our findings H2a is rejected.

H2b: Negative sentiment in tweets is negatively associated with Spotify streams.
Similar to positive sentiment, the analyses showed no association between negative sentiment contained in tweets and Spotify streams. The ratio of negative sentiment did always have a negative association with Spotify streams. Although this relationship was very small and not significant in any of our regression models. Additional analyses confirmed this by further showing that there was no significant relationship between negative sentiment contained in tweets and Spotify streams. Therefore, H2b is rejected.

H3: Amount of followers is positively associated with Spotify streams.
Based on previous research (Rui et al., 2013) a positive relationship between the amount of followers of the user that posted a tweet containing the album title and Spotify streams was expected. The type 2 follower ratio measured the ratio of tweets posted by users with more than 1500 followers. Despite the expected positive relationship, the type 2 follower ratio showed a negative relationship with Spotify streams in every regression. In the model that used Twitter predictor variables from week 2 on Spotify streams from week 2, the type 2 follower ratio was found to have a significant negative relationship with Spotify streams. Therefore, our results provide no evidence for a positive relationship between the type 2 follower ratio and Spotify streams. Based on our findings H3 is rejected.

Hypotheses	**Results**
H1a: The volume of tweets for each album is positively associated with Spotify streams.	Accepted
H1b: The volume of tweets for each artist is positively associated with Spotify streams.	Accepted
H2a: Positive sentiment in tweets is positively associated with Spotify streams.	Rejected
H2b: Negative sentiment in tweets is negatively associated with Spotify streams.	Rejected
H3: Amount of followers is positively associated with Spotify streams.	Rejected

Table 17: Results on hypotheses

After performing the literature review, processing the data and reporting on the analyses and hypotheses it is now possible to answer our sub-questions formulated in chapter 1.3. The answers to these sub-questions will provide a blueprint for answering the central research question in the conclusion chapter.

What Twitter predictor variables has previous literature identified as having significant predictive power?
Previous literature has shown that the significance of social media predictor variables depends on the context they are studied in. For example, volume-related variables are often used in a health context (Young et al., 2014), and have been shown to significantly predict box-office revenues of movies (Asur & Huberman, 2010; Rui et al., 2013). On the other hand, volume-related variables do not perform well in predicting political elections (Gayo-Avello, 2013). In the case of political elections, sentiment-related variables have been shown to be significant predictors (Ceron et al., 2015). User profile characteristics of online users received much less attention in academic literature. Although

de Choudhury et al. (2013) and Rui et al. (2013) do find that the amount of Twitter followers is a significant predictor variable in predicting postpartum changes in mothers' behavior and box-office revenues of movies. The literature review showed the importance of taking into account contexts when using social media data for predictions. Not each one of the three types of predictor variables is a significant predictor in each context. Additionally, most research has focused on volume-related variable and sentiment-related variables. User profile characteristics of online users are often overlooked or not included in the analyses.

To what extent has the predictive power of social media in the music industry been demonstrated before?

Much research on the predictive power on social media in the music industry has focused on volume-related variables. Volume-related variables obtained from blogs (Dewan & Ramaprasad, 2012) and Myspace (Chen et al., 2015; Sharma et al., 2012) have been shown to be significant predictors of album sales. Kim et al. (2014) researched whether Twitter could be used to predict album sales using volume-related variables. Their results showed that both the volume of tweets containing the album's name and the volume of tweets containing the artist's name could be used to predict album sales. The opportunity to sample music online has been shown to negatively affect music sales (Dewan & Ramaprasad, 2012). While comments on social media have been shown to have an increasingly positive effect on music sales (Saboo et al., 2014). The literature search showed a lack of sentiment analysis used in these studies. Although Dewan and Ramaprasad (2012) and Dhar and Chang (2009) used customer reviews as a type of sentiment-related variable. Additionally, Dhar and Chang (2009) found that the weekly change in Myspace friends was not a significant predictor of album sales. However, Saboo et al. (2014) did find that social following significantly predicted a song's billboard hot 100 ranking.

How should data from Twitter be collected and preprocessed for further analysis?

The Twitter streaming API turned out to be an accurate and secure way to collect data from Twitter. However, researchers should be very careful in the selection of keywords that are used to collect this data. Using multiple keywords allows for choosing the most relevant keywords to include in the analysis. The biggest part in processing the data was found in classifying the sentiment of the tweets. As previous literature has already shown, sentiment analysis is a complex and immature topic. This is especially the case when using social media texts which are often using informal language, slang or contain sarcasm. In this study sentiment was classified using lexicon-based approaches. The 'qdap' and 'sentiment140' R package were first explored. Finally, we settled on the 'syuzhet' package which uses lexicons validated through academic research (Liu et al., 2005; Mohammad & Turney, 2010; Nielsen, 2011). A combination of four lexicons was used and tweets were only classified as containing a certain sentiment when multiple lexicons agreed on the classification. It is possible that these lexicons are not well fitted for the use on social media messages. It could also be possible that machine learning approaches would have shown a different relation between the sentiment of tweets and Spotify streams. Kalampokis et al. (2013) found that studies that used lexicon-based approaches for their sentiment analysis were less supportive of the predictive power of social media than studies that used machine-learning. Machine learning approaches might simply be more appropriate for classifying the sentiment of social media texts.

Which Twitter predictor variables are significant predictors of album streams on Spotify?

There is a strong relationship between the volume-related Twitter variables and Spotify streams. This relationship is observed for both tweets about the album and tweets about the artist. There is both a strong linear relationship between these variables on Spotify streams in the same week and on Spotify streams in the next week. Sentiment-related variables and user profile characteristics of online users were not found to be significant predictors of album streams on Spotify.

7. Conclusion

This section will present the main conclusions of the research and will answer the central research question. Following, limitations of the research will be discussed. Finally, suggestions for future research are presented.

Central research question: *What is the relationship between volume-related, sentiment-related and profile-related Twitter variables and the volume of Spotify streams of newly released music albums?*

The results of this study heavily suggest that there is a positive relationship between volume-related variables derived from Twitter and Spotify streams. Using these volume-related variables it was possible to predict the streams of newly released albums during the early weeks of release. Building a daily time series of the tweet-volume increased the prediction accuracy compared to using the weekly average volume. However, for sentiment-related variables and profile-related variables no significant relationship with Spotify streams was found. Volume-related variables performed well in both predicting Spotify streams in the same week and in the next week. Additional analyses showed that the results could be different depending on demographic differences. There seems to be a difference in how the Twitter predictor variables are related to Spotify streams for younger and older artists. The type of performer could also play a role in the predictive power of Twitter. Furthermore, the inclusion of the number of Spotify Monthly Listeners per artist improved the performance on all models.

What exactly do these results mean for the music industry? Currently the amount of digital music sales is constantly decreasing while streaming keeps growing. As streaming keeps growing it becomes increasingly important to understand where these streams come from. In this research we found a relationship between volume-related variables and Spotify streams. However, this relationship does not necessarily imply causation. Tweets could lead to streams, but streams could also lead to more tweets. However, the volume of tweets can be used to predict and monitor the amount of Spotify streams. Artists and record labels can use this information to actively track Twitter in order to test if the album will meet the streaming goals set by the record label. If Twitter suggests that the album might underperform, additional promotional appearances and performances can be organized. This also provides a window for analytical companies and data scientists to design smart applications that can be applied in the music industry. For younger musicians the predictive relationship between Twitter and Spotify seems to be very strong. Although this relationship might not be as strong for older musicians. Industry projections suggest that by 2020 physical and digital sales will be close to extinction and streaming will be by far the biggest component of music revenues. This does bring forward a question for future research as well as the music industry: How are older musicians going to be able to promote and monitor the success of their music?

Furthermore, the generalizability of this research towards other types of social media plays an important role here. Previous research showed a significant relationship between volume-related variables from blogs and Myspace on album sales. This study found a similar relationship between Twitter and Spotify streams. The most popular social networking sites nowadays are Facebook, YouTube, Instagram and Twitter. Results from this research are probably applicable to social networking sites that focus on posts that mostly consist of texts. For example, Instagram has a big focus on pictures which makes it difficult to count the volume of Instagram posts containing a certain artist. An option could be to check the descriptions and hashtags included with each picture. YouTube focuses on videos, although each video also comes with its own description and keywords. On the other hand Facebook is more difficult to monitor because most user profiles are private. This would especially make it difficult to collect variables related to the user profile characteristics of online users. However, judging from previous research on volume-related variables similar effects can be observed across multiple forms of social media; blogs, Myspace and Twitter. It seems reasonable that these effects would be observed across other types of social media as well. Nonetheless, for each specific type of social media the specific metrics would still have to be chosen. For example, volume-related variables on Facebook could be the amount of posts, likes or comments.

Additionally, recent research has started to explore the concept of social media strategy (Effing & Spil, 2016). A well-rounded social media strategy consists of various key elements. Monitoring is one of these key elements. For most companies monitoring does not happen until the later stages of implementing the social media strategy. Therefore, many companies do not yet have a big focus on monitoring their social media channels. Insights from this study hopefully show social media managers that monitoring social media volume can provide important real-time updates and that it can also be extended to a powerful tool for prediction. Using prediction models derived from Twitter can assist social media managers in monitoring progress on goals as well. Once future research has explored the predictive power of different social media channels in more detail, social media managers have a clearer view of what social media channels to focus on.

Perhaps the most interesting application of predictive social media research would be the development of predictive models that can process real-time data from multiple types of social media. With the growing popularity of machine learning these models could achieve high degrees of accuracy. Big data offers ways to understand music listening behavior that we were not able to decipher before. On one hand, academic researchers will probably focus on the collection of panel data in order to draw conclusions on the causality between Twitter variables and the number of streams. While on the other hand, data scientists in practice will probably focus on the tuning the specific features and parameters in machine learning algorithms in order to build the best prediction models. Machine learning can have many applications. For example, clustering algorithms are often used in identifying market segments and target groups.

With streaming becoming the dominant factor in the music industry it is certainly interesting to be able to understand and predict streaming behavior. Variables derived from Twitter and other social media can play an important role in the prediction of said streaming behavior. This study build forward on the literature review of Kalampokis et al. (2013). Their social media framework for predictions provided a robust way to design research regarding the predictive power of social media. However, while their model does distinguish between lexicon-based approaches and machine learning in sentiment analysis it does not provide much information on the different types of approaches to be used. For example, there are many different lexicons as well as machine learning algorithms for sentiment classification. We hope that this study has also provided an updated understanding regarding the use of social media in predictive analysis, especially regarding sentiment analysis.

7.1 Limitations

This section will discuss the limitations of the study. First, the results of this study seem to be exclusive to the music industry. For example, the volume-related variables used in this research consist of the volume of tweets containing the name of the artist and the volume of tweets containing the album title. When researching the predictive power of Twitter in different industries, different types of keywords would have to be contained in the tweets. Within the movie industry only the movie title is generally used to determine the tweet volume. In a political context the name of the party or the name of the political leader could both be used. Our analysis showed high similarity between the behavior of social media predictor variables in the music industry and the movie industry. Building a time series of the volume of tweets containing the album title or movie title already is enough be used to build a strong predictive model. Both the music industry and movie industry are part of the entertainment industry, while the health industry and politics are not. A possible explanation could be that social media prediction works similarly within entertainment industries. In this case, the results of this study are not totally exclusive to the music industry but could be extended to other parts of the entertainment industry.

Second, the data collection was collected over five weeks and focused on the release period of the album. Because of this timeframe it is not possible to draw conclusions on causation between Twitter and Spotify streams. The main goal of this study was to capture whether there was a relationship between the social media predictor variables and Spotify streams. Our results do suggest a relationship between the two. The results also show that it is possible to predict Spotify streams using Twitter data with good accuracy. However, it is very difficult to prove causation in this case. An increase of streams on Spotify could also lead to more tweets, which then leads to more Spotify streams. Therefore, we are unable to draw conclusions regarding the causality between tweets and Spotify streams. However, it should be noted that this also was not the intention of this research.

Third, the lack of sophisticated sentiment algorithms proved to be a difficulty. Finally, we settled on a lexicon-based approach using the ‘syuzhet’ package. The ‘qdap’ and ‘sentiment140’ R packages did not perform well on classifying the sentiment in tweets dataset. Machine learning seemed like an alternative approach in classifying the sentiment in tweets. However, time and budget constraints did not allow the use of machine learning approaches in this research. Additionally, academic research has been unclear in what machine learning algorithms work best for sentiment classification.

7.2 Future Research

Based on our findings the following suggestions for future search have been formulated. First, we suggest future research to study the possibility of social media predictor variables to predict Spotify streams over a longer period of time. We suggest a panel data approach which would allow researchers to test for Granger causality between variables. Granger causality assumes that when one time series can be used to predict another time series this can be seen as a form of causality. This would also allow researchers to study whether the relationships observed in this study are different for weeks further away from the release week. It could be possible that sentiment does not play a role close to the release date because most streamers are merely influenced by the volume of tweets (hype). The sentiment of these tweets could play a bigger role when the initial hype slows down.

Second, more research surrounding the user profile characteristics of online users is necessary. Our literature review showed that this variable is often overlooked. Previous literature that did include this variable suggested that this variable was able to be a significant social media predictor variable. In this study no significant relationship was found between the follower count of users who tweeted about an album and Spotify streams received by that album. Although in most cases there did seem to be a negative relationship between the amount of followers and Spotify streams. Previous research suggested that this relationship would be positive. Therefore, we encourage more research into this variable to further understand this type of social media predictor variable.

Third, our exploratory analysis showed that the relationship between social media predictor variables and Spotify streams could be influenced by age. While the linear regression model that used albums released by younger musicians was significant, the same model that used albums released by older musicians was not. These additional analyses also suggested that the type of performer could also play a role. However, because our dataset had to be split-up the sample sizes for the individual regression analyses were not big enough to draw strong conclusions. Therefore, we suggest more research into demographic influences of performers. Similar to our first suggestion, panel data and the inclusion of more albums in the dataset could provide more robust results regarding these demographic influences.

Fourth, future research should explore the use of sentiment analysis on social media texts. As discussed before, these texts can be difficult to classify due to the use of informal language, slang and sarcasm. Future research should explore the appropriateness of lexicon-based methods, but should ultimately switch its focus on the use of machine learning in classifying the sentiment of social media texts. Lexicon-based approaches are easier and quicker to use. However, their application on social media texts might not be optimal based on the results of previous literature and of this research. For the use of machine learning it is important that research into natural language processing identifies the best machine algorithms regarding their efficiency and effectiveness in classifying sentiment. For example, Naïve Bayes is often mentioned for its relative simplicity while support vector machines seem to achieve a slightly higher accuracy.

Finally, the use of machine learning can also be extended to building the actual prediction models. In sentiment analysis machine learning algorithms are often used for classification. However, machine learning algorithms can also be used for regression. Statistical models, like the ones used in this research, focus on finding significant variables and significant models. On the other hand, machine learning focuses more on building the best possible prediction model. Therefore, for future research we suggest to also focus on building the best possible prediction models. After identifying significant type of predictor variables using statistical methods it seems relevant to see the highest possible accuracy that predictive models can achieve using social media data

8. References

8.1 Literature Review

Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*.

Bandari, R., Asur, S., & Huberman, B. A. (2012). The pulse of news in social media: Forecasting popularity. *In ICSWM*.

Bennett, L. (2012). Patterns of listening through social media: online fan engagement with the live music experience. *Social Semiotics*, *22*(5), 545-557.

Bischoff, K., Firan, C. S., Georgescu, M., Nejdl, W., & Paiu, R. (2009, August). Social knowledge-driven music hit prediction. In *International Conference on Advanced Data Mining and Applications* (pp. 43-54). Springer Berlin Heidelberg.

Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., & Voss, A. (2014). Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining*, *4*(1), 1-14.

Burnap, P., Gibson, R., Sloan, L., Southern, R., & Williams, M. (2016). 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies*, *41*, 230-233.

Caldarelli, G., Chessa, A., Pammolli, F., Pompa, G., Puliga, M., Riccaboni, M., & Riotta, G. (2014). A multi-level geographical study of Italian political elections from Twitter data. *PloS one*, *9*(5), e95809.

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 & Society*, *16*(2), 340-358.

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.

Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367-1403.

Chen, H., De, P., & Hu, Y. J. (2015). IT-enabled broadcasting in social media: An empirical study of artists' activities and music sales. *Information Systems Research*, *26*(3), 513-531.

De Choudhury, M., Counts, S., & Horvitz, E. (2013, April). Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3267-3276). ACM.

Denecke, K., Krieck, M., Otrusina, L., Smrz, P., Dolog, P., Nejdl, W., & Velasco, E. (2013). How to exploit twitter for public health monitoring. *Methods Inf Med*, *52*(4), 326-39.

Dewan, S., & Ramprasad, J. (2009). Chicken and egg? Interplay between music blog buzz and album sales. In *Proceedings of PACIS 2009*, 87.

Dewan, S., & Ramaprasad, J. (2012). Research note-music blogging, online sampling, and the long tail. *Information Systems Research*, *23*, 1056-1067.

Dewan, S., & Ramaprasad, J. (2014). Social Media, Traditional Media, and Music Sales. *Mis Quarterly*, *38*(1), 101-121.

Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, *23*(4), 300-307.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, *13*(4), e123.

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from Twitter data. *Social Science Computer Review*, 31, 649-679.

Hanson, C. L., Burton, S. H., Giraud-Carrier, C., West, J. H., Barnes, M. D., & Hansen, B. (2013). Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *Journal of medical Internet research*, *15*(4), e62.

Highfield, T., Harrington, S., & Bruns, A. (2013). Twitter as a technology for audiencing and fandom: The #Eurovision phenomenon. *Information, Communication & Society*, *16*(3), 315-339.

Hong, L., Dan, O., & Davison, B. D. (2011, March). Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World Wide Web* (pp. 57-58). ACM.

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,* 1-9.

Jin, X., Gallagher, A., Cao, L., Luo, J., & Han, J. (2010, October). The wisdom of social multimedia: using Flickr for prediction and forecast. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1235-1244). ACM.

Jungherr, A. (2013, October). Tweets and votes, a special relationship: the 2009 federal election in Germany. In *Proceedings of the 2nd workshop on Politics, elections and data* (pp. 5-14). ACM.

Kaplan, A. M., & Haenlein, M. (2012). The Britney Spears universe: Social media and viral marketing at its best. *Business Horizons*, *55*(1), 27-31.

Kim, E. K., Seok, J. H., Oh, J. S., Lee, H. W., & Kim, K. H. (2013). Use of hangeul twitter to track and predict human influenza infection. *PloS one*, *8*(7), e69305.

Kim, Y., Suh, B., & Lee, K. (2014, July). #nowplaying the future billboard: mining music listening behaviors of twitter users for hit song prediction. In *Proceedings of the first international workshop on Social media retrieval and analysis* (pp. 51-56). ACM.

Lerman, K., & Hogg, T. (2010, April). Using a model of social dynamics to predict popularity of news. In *Proceedings of the 19th international conference on World Wide Web* (pp. 621-630). ACM.

Li, Q., Zhou, B., & Liu, Q. (2016, July). Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion. In *Cloud Computing and Big Data Analysis (ICCCBDA), 2016 IEEE International Conference on* (pp. 359-364). IEEE.

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,

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.

Luo, X., & Zhang, J. (2013). How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems*, *30*(2), 213-238.

Maecker, O., Grabenströer, N. S., Clement, M., & Heitmann, M. (2013). Charts and demand: Empirical generalizations on social influence. *International Journal of Research in Marketing*, *30*(4), 429-431.

Oghina, A., Breuss, M., Tsagkias, M., & de Rijke, M. (2012, April). Predicting IMDB movie ratings using social media. In *European Conference on Information Retrieval* (pp. 503-507). Springer Berlin Heidelberg.

Oh, I., & Park, G. S. (2012). From B2C to B2B: Selling Korean pop music in the age of new social media. *Korea Observer*, *43*(3), 365.

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.

Saboo, A. R., Kumar, V., & Ramani, G. (2015). Evaluating the impact of social media activities on human brand sales. *International Journal of Research in Marketing*, 33(3), 524-541.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science, 311*(845), 854-856.

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.

Sharma, R., Morales-Arroyo, M. & Pandey, T. (2012). The impact of electronic word-of-mouth in the distribution of digital goods. *Journal of Information, Information Technology, and Organizations, 6,* 41-61.

Stoové, M. A., & Pedrana, A. E. (2014). Making the most of a brave new world: Opportunities and considerations for using Twitter as a public health monitoring tool. *Preventive medicine*, *63*, 109-111.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010, August). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 IEEE Second International Conference on Social Computing* (pp. 177-184). IEEE.

White, K. (2016, May). Forecasting Canadian Elections Using Twitter. In *Canadian Conference on Artificial Intelligence* (pp. 186-191). Springer International Publishing.

Won, H. H., Myung, W., Song, G. Y., Lee, W. H., Kim, J. W., Carroll, B. J., & Kim, D. K. (2013). Predicting national suicide numbers with social media data. *PloS one*, *8*(4), e61809.

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.

Zhang, Y., & Pennacchiotti, M. (2013, May). Predicting purchase behaviors from social media. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1521-1532). ACM.

Zhu, J., Xiong, F., Piao, D., Liu, Y., & Zhang, Y. (2011, October). Statistically modeling the effectiveness of disaster information in social media. In *Global Humanitarian Technology Conference (GHTC), 2011 IEEE* (pp. 431-436).

8.2 References oustide of Literature Review

Athanasopoulou, C., Suni, S., Hätönen, H., Apostolakis, I., Lionis, C., & Välimäki, M. (2016). Attitudes towards schizophrenia on YouTube: a content analysis of Finnish and Greek videos. *Informatics for Health and Social Care*, *41*(3), 307-324.

BuzzAngle (2016). *United States 2016 Mid-Year Music Report*. Retrieved from http://www.lidestri.com/files/buzzanglemusic/BuzzAngleMusic-2016-MidYear-Report.pdf

Effing, R., & Spil, T. A. (2016). The social strategy cone: Towards a framework for evaluating social media strategies. *International journal of information management*, *36*(1), 1-8.

Franklin, K. (2013). *Social media is revolutionising the music industry*. Retrieved from https://www.brandwatch.com/2013/08/social-media-the-music-industry

Friedlander, J.P. (2016). *News and Notes on 2015 RIAA Shipment and Revenue Statistics*. Retrieved from http://www.riaa.com/wp-content/uploads/2016/03/RIAA-2015-Year-End-shipments-memo.pdf

Hopkins, D. J., & King, G. (2010). A method of automated nonparametric content analysis for social science. American Journal of Political Science, 54(1), 229-247.

Hopkins, D., King, G., Knowles, M., & Melendez, S. (2010). ReadMe: Software for automated content analysis. *Institute for Quantitative Social Science*.

IFPI. (2015). *Digital Music Report 2015*. Retrieved from http://www.ifpi.org/downloads/Digital-Music-Report-2015.pdf

IFPI. (2016). *Global Music Report 2016*. Retrieved from http://www.ifpi.org/downloads/GMR2016.pdf

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the predictive power of social media. *Internet Research*, *23*(5), 544-559.

Kallas, E. (2016, November 28). *Top 15 Most Popular Social Networking Sites.* Retrieved from http://www.dreamgrow.com/top-15-most-popular-social-networking-sites

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.

Leenders, M. A., Farrell, M. A., Zwaan, K., & ter Bogt, T. F. (2015). How are young music artists configuring their media and sales platforms in the digital age? *Journal of Marketing Management*, *31*(17-18), 1799-1817.

Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web* (pp. 342-351). ACM.

Mohammad, S. M., & Turney, P. D. (2010, June). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26-34). Association for Computational Linguistics.

Mulligan, M. (2016). *After the download: When Apple turns off the iTunes store*. Retrieved from https://musicindustryblog.wordpress.com/2016/05/06/after-the-download-when-apple-turns-off-the-itunes-store

Nielsen (2016). *2016 Nielsen music U.S. mid-year report*. Retrieved from http://cdn.mbw.44bytes.net/files/2016/07/NielsenMidYearMusicReport_2016.pdf

Nielsen, F. Å. (2011, March). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In ESWC2011 Workshop on Making Sense of Microposts.

PASM (2015). *Value Creation At Vivendi*. Retrieved from http://cdn.mbw.44bytes.net/files/2015/03/Vivendi_White_Paper.pdf

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011, October). Our Twitter profiles, our selves: Predicting personality with Twitter. In *Proceedings of Privacy, Security, Risk and Trust (PASSAT)*.

Robertson, A. (2016). Effects of a social media website on primary care givers' awareness of music therapy services in a neonatal intensive care unit. *The Arts in Psychotherapy*, *50*, 17-21.

Shubber, K. (2014). *Music analytics is helping the music industry see the future*. Retrieved from https://www.theguardian.com/technology/2014/apr/09/music-analytics-is-helping-the-music-industry-see-into-the-future

Schneider, M. (2016, September 14). *Spotify Hits 40 Million Paid Users, But Loses Its Chief Revenue Officer*. Retrieved from http://www.billboard.com/articles/business/7510283/spotify-jeff-levick-departs-subscriber-totals-40-million

Spotify (2016). *Artist Guide.* Retrieved from: https://www.spotifyartists.com/guides

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a Literature Review. *MIS quarterly*, *26*(2), 13-23.

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45-55.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544-2558.

Twitter (2016). Retrieved from https://about.twitter.com/company

9. Appendices

Appendix A

This appendix includes more detailed information regarding the inclusion and exclusion criteria for each set of keywords. The ‘&’ sign is used between different set of keywords, while the keywords themselves use the Boolean operator ‘AND’.

“Social media” AND (forecast OR predict) & “Social media” AND “monitoring”
The main goal of these searches was to expand on the work of Kalampokis et al. (2013) by including papers that were published after their review was conducted. Therefore, the year limit is set to 2013-2016. A comparison of search results between 2006-2012 and 2013-2016 for these search terms showed that much more articles regarding this topic were published between 2013 and 2016. For example, using the exact search combination that Kalampokis et al. (2013) used gives 769 results from 2006-2016 on Scopus, 629 of those results are from 2013 onwards. In order to be selected studies have to monitor social media activity as an independent variable. Studies that focus on ‘personal social media use’, geolocation or psychological traits are excluded.

“Social media” AND (forecast OR predict)
The goal of this keyword combination is to include the papers that focus on ‘online outcomes’ which were excluded from Kalampokis et al. (2013). The main inclusion criteria is that these papers use a dependent variable that is related to an online outcome.

“Social media” AND “music”
The goal here is to get an understanding of how social media has been used in the music industry and how it is changing the industry. In order to be included, research has to focus on the effects of social media for the music industry. For example:

- “*Effects of a social media website on primary care givers' awareness of music therapy services in a neonatal intensive care unit*” (Robertson, 2016) has been excluded.
- *“Social media, traditional media, and music sales*” (Dewan & Ramaprasad, 2014) has been included.
-

“Social media” AND “music” AND (forecast OR predict) & “Social media” AND (music sales OR album sales)
These searches are much more specific (to the music industry) than the previous ones and therefore resulted in much less results. The decision was made to drop the limitation of ten references, because this would result in very few papers to be included. Again, the focus of papers has to be on using social media data as the independent variable. For example, Quercia, Kosinski, Stillwell and Crowcroft (2011), who use social media data to predict personality traits has been excluded.

“Twitter” AND (forecast OR predict) & “Twitter” AND “monitoring”
These search terms were performed for the years 2013-2016 with the minimum of ten citations and for 2016 without this restriction. Papers were included if they specifically focused on the use of Twitter variables to perform a prediction.

Appendix B

The concept matrix constructed during the literature search is included below in figure B-1.

Articles		Context =								Type of Social Media =						Type of SM predictor Variables =		
	SM	Music	Politics	Movies	Health	Business	Disasters	News	Other	Blogs	Twitter	Facebook	YouTube	Myspace	Other	Volume	Sentiment	User Profile
Twitter & Forecast/Monitoring																		
Jungherr (2013)	X		X								X					X	X	
Kim et al. (2013)	X				X						X					X		
Burnap et al. (2016)	X		X								X						X	
Li et al. (2016)	X					X					X						X	
Lipizzi et al. (2016)	X			X							X					X	X	
Liu et al. (2016)	X			X							X						X	
Schumaker et al. (2016)	X								X		X						X	
White (2016)	X		X								X						X	
SM & Forecast/Monitoring																		
Ceron et al. (2014)	X		X								X					X	X	
Ceron et al. (2015)	X		X								X						X	
Caldarelli et al. (2014)	X		X								X					X		
Gayo-Avello (2013)	X		X								X							
Burnap et al. (2014)	X						X				X					X	X	X
Zhang & Pennacchiotti (2013)	X					X						X						X
De Choudhury et al. (2013)	X				X						X					X		X
Luo & Zhang (2013)	X					X									X	X	X	
Chen et al. (2014)	X					X									X		X	
Jin et al. (2010)	X		X			X									X	X		
Won et al. (2013)	X				X					X						X		
Asur & Huberman (2010)	X			X							X					X	X	
Rui et al. (2013)	X			X							X					X	X	X
Hanson et al. (2013)	X				X						X					X		
Jashinsky et al. (2014)	X				X						X					X		
Young et al. (2014)	X				X						X					X		
Denecke et al. (2013)	X				X					X	X					X		
Stoové & Pedrana (2014)	X				X						X							
SM & Monitoring/Predict Online																		
Bandari et al. (2012)	X							X			X							
Oghina et al. (2012)	X			X							X		X			X	X	
Eysenbach (2011)	X								X		X							
Zhu et al. (2011)	X								X		X					X		X
Suh et al. (2010)	X								X		X					X		X
Lerman & Hogg (2010)	X							X							X	X		
Hong et al. (2011)	X								X		X					X		
Social Media & Music																		
Bennett (2012)	X	X									X				X			
Highfield et al. (2013)	X	X									X							
Kaplan & Haenlein (2012)	X	X									X	X	X					
Oh & Park (2012)	X	X										X	X					
Social Media Predicting & Music																		
Dewan & Ramaprasad (2012)	X	X								X						X	X	X
Dewan & Ramaprasad (2014)	X	X								X						X		
Dewan & Ramaprasad (2009)	X	X								X						X		
Kim et al. (2014)	X	X									X					X		
Maecker et al. (2013)		X		X												X		
Salganik et al. (2006)		X														X		
Sharma et al. (2011)	X	X								X			X	X		X		
Chen et al. (2015)	X	X												X		X		
Saboo et al. (2015)	X	X									X	X	X	X	X	X		X
Bischoff et al. (2009)	X	X													X	X		X
Dar & Chang (2009)	X	X								X				X		X	X	X

Figure B-1: Concept matrix. Abbreviations: SM = social media

Appendix C

In this appendix the three contexts regarding the use of volume-related variables derived from Twitter for prediction are discussed in detail. First, the studies that focused on other forms of social media than Twitter are discussed. Then, the health, politics and movie contexts are discussed sequentially.

Volume-related Variables in Social Media

Luo & Zhang (2013) found that the volume of consumer reviews from an electronic product review website had a significant positive effect on firm value. Jin et al. (2010) showed that Flickr provided hints that Obama would win the presidential election in 2008. Next to the US election, Flickr also showed the geographical distribution of iPod and iPhone sales and was able to predict quarterly sales before quarterly sales official reports were released with relatively low error (Jin et al., 2010). Finally, Won et al. (2013) built a prediction model that tried to predict the national number of suicides in South Korea. Two social media measures were defined, the 'suicide blog count' and the 'dysphoria blog count'. The first measure refers to the amount of blogs posted with the Korean word for suicide, while the second measure refers to the number of blogs that mention the Korean word used to express being tired, painful or exhausted. Additionally, economic and meteorological measures were collected. Finally, authors included data regarding celebrity suicides. There were six big celebrity suicides during the data collection period, 2008-2010. Suicide blog count seems more related to short term effects and heavily relates to celebrity suicides. On the other hand, dysphoria blog count showed more long term effects. In the final model, dysphoria blog count outperformed the celebrity suicide variable as well as recent suicide blog count (Won et al., 2013). The other studies included in this literature review that used volume-related variables all focus on Twitter. Following, these studies are discussed per context they explore. The final context discussed is the movie industry, which shows the most similarities to the music industry.

Health

In the context of health the main focus of many papers is not strictly on predicting, but more on monitoring. A common practice is the gathering of Twitter data by using health related keywords and then matching the volume of tweets to official disease reports per area. In these cases, it is important that tweets are geolocated or can be traced to a certain area. Hanson et al. (2013) explored the use of Adderall among college students. After collecting tweets containing the word Adderall, the tweets were matched with clusters of colleges. Textual analysis showed that Adderall was often mentioned next to alternative motives like study aid or in combination with other substances. Furthermore, mentions of the drug increased significantly around traditional exam periods. After determining whether a tweet came from a student and checking its geolocation, only around one percent of tweets could be used for further analysis. Most Adderall tweeters were located in the northeastern part of the United States as well as some of the southern states. In another study, strong correlations were found between high risk tweets and actual suicide data per US state (Jashinsky et al., 2014). Results suggest that Twitter can be used to monitor people who are at a high risk of committing suicide. However, there are some limitations to this study. For example, tweets were collected in 2012 while the most recent available suicide numbers were from 2009. Young et al. (2014) studied the relation between geolocated tweets and HIV cases. Similar to Hanson et al. (2013), only 0.4% of collected HIV risk behaviour tweets could be used since these were geolocated in the United States. A significant relationship was found between the volume of tweets

containing HIV risk behaviour and actual HIV cases in the United States (Young et al., 2014). One of the limitations of this study is that the most recent data on HIV cases the authors could use was from 2009, while tweets were collected in 2012. In a commentary on this paper by Stoové and Pedrana (2014) this issue is discussed as well as other limitations. People are less likely to tweet about shamed behaviours and symptoms than they are to tweet about more casual things like headaches. Also, social media users might not be representative of the total population, because the mean age of users is lower than that of the overall population. Overall, Stoové and Pedrana (2014) encourage more research into social media and its potential for surveillance and monitoring. Kim et al. (2013) do not focus on the matching of geolocated data but rather build a country-wide model for monitoring the spread of influenza in South Korea. The authors are able to present an algorithm based in linear regression that has a reasonably good prediction power and includes about 30 influenza related keywords (Kim et al., 2013).
Instead of focusing on monitoring, de Choudhury et al. (2013) tried to predict postpartum changes in behavior and mood of new mothers on Twitter. The goal of the study is to identify mothers who are at risk for postpartum depression. The authors build a model including volume, sentiment, user-profile and linguistic variables. Among the volume-related variables, the amount of twitter posts and amount of replies are the best predictors of extreme changes in behavior and mood of new mothers (de Choudhury et al., 2013). After incorporating Twitter with blog and forum posts and TV and radio channels, Denecke et al. (2013) build a health monitoring method called M-Eco system. Most health signals within this method are received from Twitter. Users of the method reported that despite the high amount of irrelevant messages, this system still provided a better solution than manually tracking the huge social media stream (Denecke et al., 2013).

Politics

Gayo-Avello (2013) provides a critical review of literature that only uses volume-related social media data to predict political elections. Although some papers provide evidence that this simple measure of the volume of tweets might be enough to predict political elections, there are just as many papers that do not achieve these results. Therefore, Gayo-Avello (2013) concludes that in a political context, only counting the mentions of a particular candidate or party is not reliable enough to use for predicting electoral results. However, this measure can be used as a good baseline. When it comes to sentiment analysis it becomes clear that a large number of previous researchers have relied on simplistic measures when performing a sentiment analysis. Therefore, Gayo-Avello (2013) assumes it is unclear whether sentiment is feasible to predict elections. However, the method and results achieved by Ceron et al. (2014) are described by Gayo-Avello (2013) as promising. However, sentiment analysis is not the only option one has when using social media data to predict elections. Another option is the relative support parameter introduced by Caldarelli et al. (2014). Next to the volume of tweets, these authors include the ratio of time variation between the volume of tweets in the comparison of two parties. However, results show that this approach is still in its infancy and might not be the best predictor in this case. Nonetheless, the relative strength parameter could be used to assess the relative strength between two parties. Jungherr (2013) reported similar issues as Gayo-Avello (2013). While trying to predict the outcome of the German 2009 federal election Jungherr (2013) found that the volume of tweets was a bad measurement to solely use when predicting election outcomes. Also, the amount of hashtags used to refer to a party alone has a similar fate. In the case of Jungherr (2013) there was one political party that was very popular online but did not receive many votes in the actual election.

Movies

One of the studies that initiated the research trend into the predictive power of social media was a study related to the movie industry. Asur and Huberman (2010) used the volume and sentiment of tweets to predict box-office scores with a higher accuracy than the Hollywood Stock Exchange (HSE). Their goal is to not monitor all movies mentioned on Twitter, but to predict box-office scores in the opening weekends of newly released movies. Over 2.8 million tweets for 24 movies were collected over a period of three months. The authors then created a simple volume-related metric, the 'tweet-rate'. Which is the amount of tweets that mentioned a movie per hour. By only using this simple measure in a linear regression model the authors were able to significantly predict box-office scores (R^2 = 0.80). Next, the authors construct a time series from seven days before the release of a movie from the tweet-rates and also include an additional variable; the amount of theatres a movie was released in. This new model is able to predict 97.3% of the variance in box-office scores. The methods used in this paper can be extended to other products and industries (Asur & Huberman, 2010). Rui et al. (2013) perform a similar but more comprehensive research by extending the period of data collection, the amount of movies and the amount of variables used. Because of the longer period of data collection, the authors are able to perform a panel data analysis using one-period lagged values of the tweet variables. Rui et al. (2013) determine whether a tweet contains the clear intention to go see a movie and create the variable 'intention tweets'. The dynamic panel shows that the total number of tweets, tweets from users with a high following, intention tweet ratio and the ratio of tweets with a positive sentiment all have a significant and positive influence on box-office scores. Results suggest that next to the volume of tweets, valence is also important. Also, tweets from users with more followers and tweets posted before the release of a movie seem to have a significantly higher effect than their counterparts (Rui et al., 2013). Using textual features of tweets and the likes-dislikes ratio from YouTube Oghina et al. (2012) were able to predict IMDB scores with high accuracy. Liu et al. (2016) focused on three Twitter metrics when predicting box-office scores: purchase intention, tweet volume and sentiment. For purchase intention, tweets are classified using a Support Vector Machine. Liu et al. (2016) received a better result than Asur and Huberman (2010) whose model was used as a baseline in this analysis. When comparing linear methods to support vector regression the authors find that linear models perform better in general circumstances. However, when uncertainties are introduced the SVM model performs better. The model that performs best in predicting box-office revenues uses a combination of purchase intention, sentiment, the amount of theatres the movie was released in, and the popularity of the movie's director (Liu et al., 2016). Twitter variables were not shown to have any relation to the opinions of movie critics (Lipizzi et al., 2016). However, various combinations of models based on different combinations of sentiment, conversational, traffic and analytical Twitter variables were able to successfully predict the box-office revenues in the release weekend (Lipizzi et al., 2016).

Concluding, volume-related variables are often used as the baseline or main variable in many studies. While the use of volume-related metrics has received good results in a health context, there are some limitations related to data availability that pop-up in almost every health study. The low amount of geotagged tweets, outdated regional disease information and shame possibly restricting Twitter users are often mentioned. In a political context there is a more critical opinion on the sole use of volume-related variables. People have a tendency to not always openly express what party they are actually planning to vote for and some elections deal with controversial parties which are popular online but do not receive many actual votes. Finally, the movie context achieved much better results than the previous contexts when using volume-related variables. Like the music industry, the movie industry is part of the entertainment industry. When tweeting about what movies or music people are interested in they generally do not face the shame that comes with disease symptoms or the controversial nature of politics. Also, data on box-office revenues is always updated and limitations related to geolocation also play no role in these studies.

Appendix D

Social Media and Music

The impact of social media within the music industry goes beyond monitoring and forecasting to predict sales. In this part we explore some of the other areas where social media has played a big role in the music industry. For example, the development of social media is greatly impacting the live experience of fans (Bennett, 2012). Whereas fans previously had to be present at the venue to experience a live performance, they can now experience the show through social media. Fans are often engaged in sharing the set list and expressing their sentiment towards it. Other uses of social media include tweeting about the concert, posting pictures and videos and even streaming the whole show. These advancements allow fans at home to join the live experience and connect with other fans in the fandom. Bennett (2012) argues that this is a positive development and that it should be seen as fans trying to reshape and redefine the boundaries of the traditional live experience. However, with technology developing at a fast rate and even the introduction of "tweet seats", we should be careful not to take things too far.

Twitter is not only used to discuss live performances, the medium also extends itself to the discussion of big music events. For example, widely televised music events like the BET Awards and the MTV Video Music Awards created some of the highest rates of tweets per second in 2011 (Highfield et al., 2013). Another big music event is the Eurovision Song Contest, which Highfield et al. (2013) analysed. Tweets were collected if they obtain one of the hashtags #eurovision, #sbseurovision or #esc. Similar to Bennett (2012) the authors point to the power of Twitter to facilitate communication and making connections between fans. During these events Twitter is mostly used by fans to promote their favourite artists and connect with other fans. Twitter can therefore be seen as a backchannel to television because it allows users to run their own commentary to the broadcasting. For example, the Australian broadcaster SBS encouraged viewers to use a specific hashtag so their tweets could appear in the broadcast. Although SBS ran a re-run of the event this made the experience seem much more live to viewers. Highfield et al. (2013) also discuss the opportunities that social media like Twitter offer to broadcasters and other stakeholders. Social media allows for the immediate tracking of audience reactions, as well as the possibility of fan communities to increase visibility by taking over a hashtag. The importance of Twitter also shines through in Kaplan and Haenlein's (2012) case study of Britney Spears her social media efforts. For example, Britney Spears was the first Twitter account to exceed five million followers. From the case study it becomes clear that Britney Spears and here team are excellent integrators of Twitter, Facebook, YouTube and her official website. During the release of the single 'Hold It Against Me' a viral marketing strategy included the radio release, leaked demo and daily teasers of the music video on social media. Besides, fans were stimulated to create user-generated content in order to appear on Britney's website. Kaplan and Haenlein (2012) conclude by stating the cheap use and increased reach of social media compared to traditional media.

Social media can also be used to obtain new fans in different geographical locations. Oh and Park (2012) examine the spread of Korean pop music (K-pop) through Europe and conclude that there is some support that social media is responsible for this phenomenon. YouTube has played a critical role in spreading K-pop throughout Europe. Compared to other mediums YouTube is free and has no physical restrictions. Facebook allow users to post music videos from YouTube on their channels and has thereby also helped the expansion of K-pop. However, most new K-pop fans have discovered the genre through Japanese and Chinese culture. Here, social media works more as a facilitator of K-pop. Another interesting development is the switch in business models in the K-pop industry partly caused by social media. There is a tendency for firms to switch from a B2C approach to a B2B approach. In that case, music distributors like YouTube play an important role. On the one hand Korean entertainment firms align themselves with global MNEs through sponsorships and advertisements, while on the other hand they use YouTube as their man distributor of music.

Appendix E

Prediction of Online Outcomes

A popular online outcome in academic research seems to be the popularity of a tweet, measured by its amount of retweets. Important features to determine to amount of retweets a tweet will receive seem to be 'degree distribution' and whether a tweet has been retweeted before (Hong, Dan & Davison, 2011). Suh et al. (2010) find that the amount of URLs and hashtags have a positive relationship with retweetability. When looking at the contextual features, the age of the account, as well as the number of followers and followees has a strong effect on retweetability. Although the age of the account had a significant effect, the amount of past tweets did not (Suh et al., 2010). Hong et al. (2010) were able to predict whether tweets would fall into the categories of 'not retweeted' or 'retweeted more than 10000'. However, their model acquired low accuracy when trying to predict the two categories in between those. Zhu, Xiong, Piao, Liu and Zhang (2011) were able to design a model that achieved 93.27% accuracy in predicting retweets using influences from the content, network and time of tweets. While all these studies assume that retweets are a good measure of tweet popularity, none of them actually show that tweets with high retweets have more effect or influence. On the other hand, past research actually shows that retweets have relatively low predictive value (Asur & Huberman, 2010). Burnap et al. (2014) monitored Twitter surrounding the Woolwich terrorist attack. The frequency of retweets and the duration between the first and last retweet were taken as the social media measures 'size' and 'survival'. Negative tweets were found to result in a smaller size and survival, while positive tweets were linked to a higher survival. The combination of a URL and a hashtag positively increased the retweet rate, but a tweet containing only one of the two decreased the retweet rate. Also, the amount of followers was positively and significantly associated with size, while the number of previous tweets was negatively associated with both size and survival (Burnap et al., 2014). Judging from these results it seems like retweet metrics are useful for understanding the spread of information surrounding a disaster event (Burnap et al., 2014; Zhu et al., 2011). However, retweet metrics are less useful when researching the predictive power of social media (Asur & Huberman, 2010).

Bandari, Asur and Huberman (2012) study the spread of online news articles on Twitter. After scoring each news article on four characteristics the authors use both regression and classification methods. While the predictive power of the regression method is rather low, the classification model is able to predict whether a new news article will receive a low, medium or high amount of tweets with 84% accuracy. On the social news website Digg, Lerman and Hogg (2011) accurately predict whether a news story would be promoted and whether it would receive a high number of votes based on the analysis of early votes.

Finally, there were two studies that used social media data for predicting an online outcome in a similar way to other studies described in this literature review did for 'real-world outcomes'. Eysenbach (2011) researched the relation between the volume of tweets linking to an academic article and the citations the article received. If tweets contained a link to an article in the Journal of Medical Internet they were collected. After calculating different volume-related metrics, these metrics were then compared to the amount of citations featured on Scopus and Google Scholar. Within the first three days of publication tweets were able to predict highly cited articles. Most of the tweets were sent on the day of publication or on the following day. Eysenbach (2011) also mentions that just using popularity, or tweet-rate, has pitfalls when it is used without other metrics in fields like health and science. However, Eysenbach also confirms its usefulness in other industries like we have seen earlier in this literature review: “popularity … is an extreme useful (and revenue-predicting) measure for commercial enterprises such as the entertainment industry” (Eysenbach, 2011, p. 15). By combining the sentiment on Twitter with the likes-dislikes ratio on YouTube, Oghina et al. (2012) were able to predict IMDB scores of movies with high accuracy in their best model. Other models that incorporated more YouTube features like the number of views and comments did not improve the model. After manually curating the textual tweets and including the likes-dislikes ratio, this model achieved an accuracy of 89.15%.

Appendix F

This appendix presents the 28 albums included in the dataset as well as their release dates.

Artist	Album	Release Date
A Day to Remember	Bad Vibrations	02-09-2016
Eluvium	False Readings on	02-09-2016
James Vincent McMorrow	We Move	02-09-2016
Sophie Ellis-Bextor	Familia	02-09-2016
Till Brönner	The Good Life	02-09-2016
Bastille	Wild World	09-09-2016
Jacob Whitesides	Why?	09-09-2016
Kt Tunstall	Kin	09-09-2016
Local Natives	Sunlit Youth	09-09-2016
M.I.A.	A.I.M.	09-09-2016
Nick Cave & the Bad Seeds	Skeleton Tree	09-09-2016
Of Mice & Men	Cold World	09-09-2016
Okkervil River	Away	09-09-2016
The Head and the Heart	Sings of Light	09-09-2016
Aaron Lewis	Sinner	16-09-2016
Against me!	Shape Shift With Me	16-09-2016
Mac Miller	The Divine Feminine	16-09-2016
Taking Back Sunday	Tidal Wave	16-09-2016
Usher	Hard II Love	16-09-2016
Dwight Yoakam	Swimming pools, Movie stars…	23-09-2016
Idina Menzel	Idina.	23-09-2016
Kristin Chenoweth	The Art of Elegance	23-09-2016
Passenger	Young as the Morning, Old as the Sea	23-09-2016
Skylar Grey	Natural Causes	23-09-2016
Alex & Sierra	As Seen on TV	30-09-2016
Regina Spektor	Remember us to Life	30-09-2016
Bob Weir	Blue Mountain	30-09-2016
Epica	The Holographic Principle	30-09-2016

Table F-1: The 28 albums and their respective artists and release dates contained in the dataset

Appendix G

The 'social media analysis framework for predictions' by Kalampokis et al. (2013) has been completed for this study and can be found below in table G-1.

Collection and Filtering of Raw Data	
Determination of time window	Data has been collected for a period of 5 weeks. Using panel data is unfortunately not possible because of time, budget and availability restrictions. However, shorter timeframes have been used many times before in academic research. For example, 8 weeks (Dhar & Chang, 2009) and 10 weeks (Kim et al., 2014).
Identification of location	This study focuses on tweets and streams on a global level. Therefore, identifying the location of users is not important in this case. Studies that focus on linking tweets to geolocation like Williams et al. (2013) are concerned with this measure.
Identification of user profile characteristics	Next to the content of the tweet, the account name, amount of followers and time stamp of the tweet will also be collected.
Selection of search terms	Search terms are set by the researcher using a 'manual approach'. Another option is the 'dynamic approach' where search terms are derived from a computational process. The search terms will consists of the corresponding album titles, similar to how Rui et al. (2013) and Asur and Huberman (2010) collected tweets that contained movie titles. As well as the corresponding artist names (Kim et al., 2014).
Computation of Predictor Variables	
Selection of predictor variables	Kalampokis et al. (2013) identify three types of social media predictor variables: volume-related variables, sentiment-related variables and profile characteristics of online users. *Volume-related variables:* Volume of tweets that contain the album title. Volume of tweets that contain the name of the artist. *Sentiment-related variables:* Positive sentiment in tweets, negative sentiment in tweets. *Profile characteristics of online users*: The number of followers from the user that posted the tweet.
Measurement of predictor variables	The volume-related variables have been measured by the number of tweets containing the album title (or artist name) per week. The sentiment-related variables are measured by ratios. For example, positive sentiment has been measured by the ratio of tweets containing positive sentiment. The number of followers from the user has been used to calculate the ratio of tweets from users with more than 1500 followers.
Computation of predictor variables	Custom R functions have been written that calculate the volume-related variables per week using regular expressions. Another R function calculates the ratio of tweets from users with more than 1500 followers. The R package 'syuzhet' has been used for the sentiment analysis.

Creation of Predictive Model	
Selection of predictive method	This study will use linear regression analyses using one-period-lagged values. Asur and Huberman (2010) achieved good results with this method. However, the only study in this sample that focused on Twitter in the music industry (Kim et al., 2014) mentioned that this method might not fit the data and recommend using a support vector machine (SVM). In this study multiple linear regression analyses were performed and good results were achieved using it. Linear regression is the most common method used in predictive social media research (Kalampokis et al., 2013). The relationship between the Twitter predictor variables and Spotify streams is measured in the same week as well as the subsequent week.
Selection and use of non-SM predictor variables	Twitter fans and Spotify monthly listeners, both collected on the night before the album release. Type of record label: major versus independent (Dewan & Ramaprasad, 2009, 2012 & 2014; Dhar & Chang, 2009). Amount of tracks on the album.
Identification of data for evaluation of prediction	The data referred to here are the weekly streams received on Spotify. These streams are manually collected for each week and previously received streams are subtracted to calculate the amount of new streams received each week.
Evaluation of the Predictive Performance	
Selection of the evaluation method	In this study a more exploratory analytics focus is used. The regression analyses are performed using the data in our sample. Another option would be to collect out-of-sample data and test the performance of the linear models on that data. This is certainly an interesting option for future research.
Specification of the prediction baseline	Asur and Huberman (2010) compared their results to those of the Hollywood Stock Exchange. However, in this case there is no official company that predicts global Spotify streams. However, we are able to compare our results with those of Asur and Huberman (2010) in order to compare the movie and music industries. We also compare our results to Kim et al. (2014).

Table G-1: Social media analysis framework for predictions' by Kalampokis et al. (2013) completed for this study.

Appendix H

This appendix contains both the zoomed and original version of figure 7 (tweet volume of tweets containing the album name per day). The original figure is also include here to provide a better comparison. In both charts day 8 is the release date. In the original chart the y-axis ranges to 25,000 while the zoomed version limits the y-axis to 1,500. The same peak on release day can be observed in both figures.

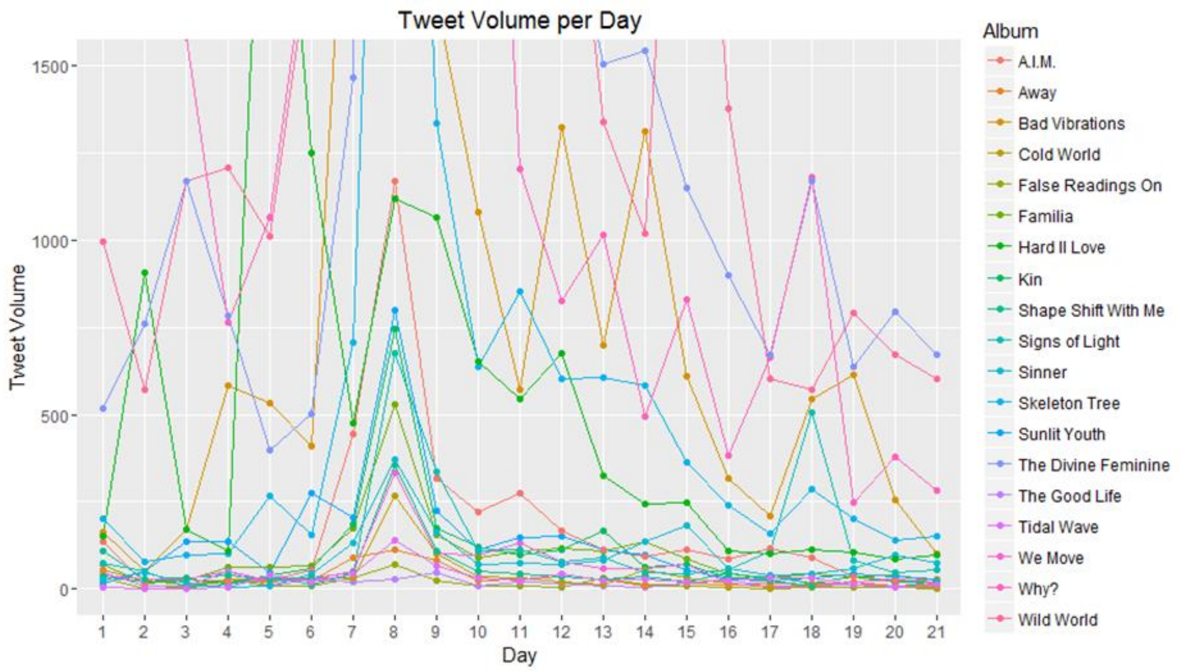

Figure H-1: Zoomed version of the tweet volume of tweets containing the album name (album popularity) per day. The y-axis is limited to 1500.

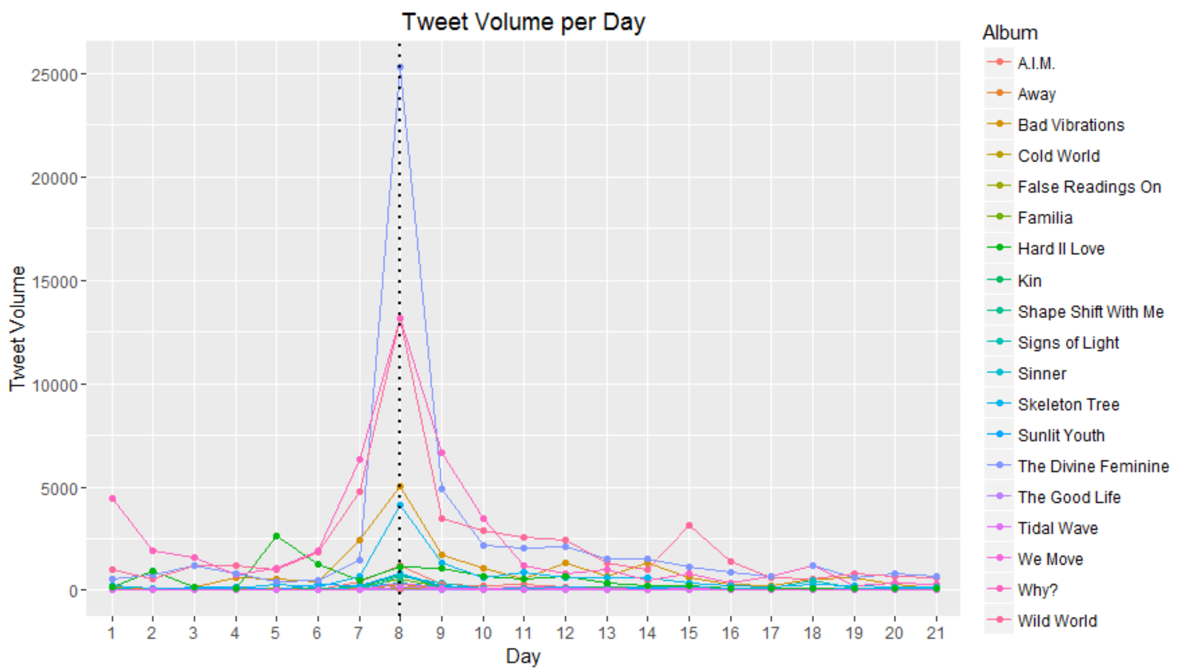

Figure H-2: Tweet volume of tweets containing the album name (album popularity) per day. The dotted line on day 8 represents the album release.