Does Chatter Matter?

Predicting Music Sales with Social Media

Robert E. Vossen

Annenberg School of Communication

University of Southern California

Abstract

This paper evaluates the possibility to employ Twitter as a means for music sales forecasting. Other prediction models for instance, which are built upon music sales data from previous records or evaluate how often an album is shared illegally in P2P networks, differ in accuracy and reliability. In contrast, Twitter is deemed an ideal source of information as it is used constantly, which makes it a limitless focus group consisting of 500 million users publishing one billion tweets every three days. From a theoretical perspective, Twitter fulfills the conditions for crowds to be wise, such as diversity and independence, and it also allows aggregating the various percolating opinions, while the very structure of the social network allows observations about the diffusion of information about forthcoming records. For the purpose of this study, two million tweets that were published two weeks before and one week after the release dates of 25 albums by the Universal Music Group were scrutinized, creating 255 distinct explanatory variables. In a comparison of 288 different linear regression models, I find that daily data outperforms weekly data, and that the cumulative reach of the tweets sent by unique users has the greatest predictive potential. By collecting this kind of data for a one-week period seven days prior to the release, album sales figures for three consecutive sales weeks following the release date can be forecasted with an accuracy of 68.9 %. When evaluating Twitter data two weeks prior to the album release, including the actual release date, the prediction accuracy rises to 95.0 %. We can therefore conclude that Twitter data exhibit meaningful and reliable correlations with music sales, and that Twitter is indeed a sufficient means to forecast album sales figures.

TABLE OF CONTENTS

Introduction

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

Pierre-Simon Laplace (1749 – 1827)

Mathematician and Astronomer

For centuries, men have sought to predict the future. The knowledge about the future was – and still is – deemed to be a tremendous advantage to those who hold it. Mankind tried almost everything one could imagine to be able to look into the future from consulting fortunetellers to dreaming of time travelling machines. Fast-forwarding into the digital era of the 21st century, we may have come closer to fulfilling this wish.

Digital technologies produce a vast amount of data and provide tools for analysis that Laplace could only dream of. For instance, users turn to Twitter to get information and share it with others, which can provide insights about how information percolate in this particular social system. I propose that this diffusion can be understood as a form of collective wisdom and that it correlates, to a certain degree, with real-world outcomes. Since entertainment topics, much like politics, are highly discussed on Twitter, we may hence assume that the chatter about musical products on Twitter allows for the forecasting of their economic success.

Predicting whether a future product is successful or not is of incredible value to the producer and seller. Given that four out of five records do not recoup their costs, finding a reliable correlation and discovering predictive patterns makes this even more meaningful to the music industry (Bradley et al., 2010). In addition, marketing and promotion costs account for about one third of the investment in newly signed artists and record companies invest 16 percent of their revenue in A&R[1] activity (IFPI, 2012). An accurate and reliable predictive model would thus help these companies spend and invest their money more efficiently. Further, knowing in advance the likelihood of whether or not a new talent has the potential to become popular could be of value for A&R managers as well as for radio stations and music services of any kind. Thus, I propose and will investigate in this paper that Twitter chatter is potentially an ideal source of information that can predict sales for forthcoming music albums.

[1] Artist & Repertoire.

Literature Review

The wish to predict musical success is not new. There are two general strands of thought in this regard: One strand focuses on forecasting actual sales figures, whereas it is also tried to predict whether or not a particular song will be a hit based on its musical characteristics. For instance, by analyzing UK Top 40 Single Chart songs from the past 50 years, researchers of the University of Bristol have claimed to have cracked the *hit song formula* and subsequently offer their services online (Ni et al., 2011). Similarly, The Echo Nest, a project that evolved out of the MIT Media Lab, provides such song analysis to industry executives (Echo Nest, n. d.). Because the respective formulas are kept secret their predictive performance remains unknown to the wider public.

But if all commercial compositions were to follow such a formula, all these songs would be popular hits – which does not appear to be the case since hits are by definition more popular than other songs. Nevertheless, there are indeed patterns and heuristics in popular music – having a catchy melody and avoiding overly dissonant chords and odd rhythm – that should be followed. One attempt to quantify these patterns, and to thereby determine the musical beauty of a given song, is to measure its ability to compress well. As a generalization, music consisting of too simple or too complex patterns lacks ease of compression and hence sounds boring to the human ear (Hudson, 2011).

Lee & Hunningham (2012) took another approach and looked at how listening patterns in one city influence listening patterns in other cities. With a social network analysis approach they could show that, indeed, music flows geographically and that some cities play a crucial role in diffusing new songs to the wider public. For instance, Hip Hop songs listened to in Atlanta at first may generally become popular in the US and Canada later, while listening patterns in Montreal serve as an indicator for Indie music. In Western Europe, Paris is deemed to be guide for Indie Music, while Oslo is a good indicator for all kinds of music genres.

While this approach tries to examine the geographical flow of music across a social network, Berns et al. (2009) sought to understand how group dynamics and peer pressure drive song popularity among adolescents. By conducting functional magnetic resonance imaging scans of the participant's neural activity, the researchers were able to show that the song popularity, determined by online user ratings, affects how a participant rates the song him- or herself, and that there is a positive feedback loop at work. Although the two studies reveal interesting insights, they come with a major caveat, that is, a song has to be somewhat popular in the first place to become even more popular later among peer groups or other regional networks.

Apart from determining a song's general likeability more or less vaguely, several researchers have tried to forecast actual music sales. The Bayesian Model proved to be a meaningful approach in this regard. The model, developed by Lee et al. (2003), is based on prelaunch data such as the success of previous records, and gets updated sequentially when the first sales data of the particular record is available. One shortcoming of this approach, however, is that prelaunch data is only available for well-known artists. Furthermore, as Silver reminds us, "past performance is not indicative of future results" (2012, 339) or, as Nikoleava & Hinz put it, "3 platinum albums do not prevent an artist from future failures" (2012, 3), which is why prior performance is a questionable source of information for an initial forecast.

As the Internet became more ubiquitous, researchers have pursued various avenues for estimating future music sales. Dewan & Ramaprasad (2009) and Dhar & Chang (2007) have analyzed the impact of word-of-mouth effects on music blogs on sales figures, whereas Hann et al. (2011) have considered how widely a forthcoming album circulates in P2P networks. Instead, Nikolaeva & Hinz (2012) have looked at how often a song is tagged with the popular smartphone app Shazam[2].

---

[2] Shazam is a smartphone app that records an audio sample and compares its digital fingerprint against a database on a central server.

The above-mentioned studies differ both in accuracy and reliability. Lee et al. (2003) report a MAPE[3] of 52 percent twelve weeks prior to the album release, which is reduced to 29 percent after one week of observed sales data. While it is remarkable to be able to cut the statistical error almost in half, the adjusted forecast is of little value to music managers since the album is already released at that point. In contrast, Nikolaeva & Hinz (2012) could show that Shazam charts indeed precede the actual charts by two weeks, which thus serve as a good indicator for chart predictions.[4] This model lacks reliability since it is only applicable to songs, which remain in the charts for at least four consecutive weeks. In addition, it is unknown whether or not the song's position in the Shazam charts can be translated into actual sales figures. For example, while the most tagged song may only have a marginal lead over the second-most tagged song, their chart position still differs by the magnitude of one. Also, it may be easier to get in the charts when fewer songs are releases versus in a period with more competition for the scarce chart positions.

Another shortcoming can be found in the attempt to estimate music sales with blog buzz, although Dewan & Ramparasad (2007) find correlations between blog mentions and music sales, which are stronger for mainstream music. Dhar & Chang (2007), despite confirming these findings, raise the question whether buzz on music blogs *causes* popularity or *is* caused by the popularity of the artist. This doubt is also inherited in sales forecasts based on P2P download data, which, given the progressive attempts to curtail piracy, might not be a sustainable approach anyway (Hann et al., 2011).[5]

In contrast to the above-mentioned studies, I propose Twitter as a source for music sales prediction for four reasons: First, users tweet constantly, which might, at least in theory, allow for predictions at any time. Second, given the construction and usage of the micro-blogging platform we

---

[3] Mean Average Percentage Error.
[4] The Shazam charts are determined by how often a song was tagged by users in a given week.
[5] The IFPI claims that anti-piracy laws such as the HADOPI law, which was introduced 2010 in France, have reduced illegal P2P file sharing significantly while other countries are about to introduce similar legislation (IFPI, 2012a).

might understand the network as a source for collective wisdom. Third, by tweeting and following others users get aware of new albums, which is not only a prerequisite for a later purchase but also observable through analyzing tools. In contrast, most proposed models lack this observation, and fourth, and finally, entertainment in general and music in particular are common topics for user discussion on Twitter.

In 2012, Twitter surpassed 500 million users, more than 100 million of which living in the US (Lunden, 2012; Semiocast, 2012). Despite the fact that only about one third of these are considered to be active that still amounts to more than 150 million users who send more than one billion tweets every three days (Weber, 2012). This makes Twitter, along with other social media websites, what Northwestern University Professor Alok Choudhary calls a limitless focus group (Smith, 2012). In addition, Twitter seems to be an ideal resource to elicit the wisdom of crowds. Although Twitter users do not come together explicitly to solve problems or to find the right answer to a particular question, but rather enjoy discussing topics of their interest and publishing their opinions, the aggregate of the user generated content satisfies the four conditions of wise crowds, that is diversity, independence, decentralization, and aggregation (Surowiecki, 2004).

It can be assumed that 150 million active Twitter users living in different countries have diverse backgrounds, but more importantly, we may believe that they also have diverse opinions. Further, Twitter users supposedly form their opinions independently, meaning that they are not directly influenced by the opinions of their peers.[6] Also, Twitter users are presumably decentralized, that is, they draw on tacit knowledge and publish their tweets themselves instead of having a central proxy acting on their behalf. Finally, the swarm of tweets gets aggregated through the use of so-called

[6] This, of course, could trigger a sociological controversy whether and to what extent people adjust their opinion when they utter them publicly. While the theory of the spiral of silence suggests that those who perceive their opinion to be in the minority do not speak up publicly (Noelle-Neumann, 1993), Pariser (2011) argues that we tend to be surrounded by like-minded opinions online, creating a so-called ‘filter bubble’. Here, however, we may assume that especially the use of nicknames on Twitter allows individuals to express their opinion without fearing social isolation.

hashtags and may be aggregated even further by ranking lists such as the Twitter Trending Topics or other analytical tools including sentiment analysis algorithms.

Another important characteristic of Twitter is how users receive and disseminate information. For Twitter users the purpose of the network is twofold: They *seek* information and *share* their views with others (Java et al., 2007). On average, Twitter users follow more than 100 other users and get followed by more than 200 others. Considering that these averages include inactive accounts as well and that ten percent of Twitter users do not follow anyone at all, we might expect these numbers to be even higher for active users (Beevolve, 2012).

By definition, users need to gain knowledge about an innovation or a new product before they can form an attitude towards it and eventually adopt or buy it (Rogers, 2003). Therefore, following others on Twitter makes users aware of new products – forthcoming music albums in this regard – while tweeting about such a product might be an expression of the user's attitude towards a particular song, album or artist. In a sense, every Twitter user can be regarded as a virtual sensor with each tweet consisting of sensory information, which allows us to draw conclusions about the diffusion of new information on Twitter (Sakaki et al., 2010). The next step of the innovation-decision process is the adoption decision, which we might evaluate through music sales data (Rogers, 2003).[7]

The two main advantages of Twitter data as a source for forecasts are that it satisfies the conditions for wise crowds and that it continuously and instantaneously provides information about the diffusion and percolation of new information in the social system. As such, recent research has focused on employing Twitter as a means to predict future events in various areas. For instance, Sakaki et al. (2010) have shown that it is possible to detect earthquakes in Japan by monitoring and analyzing tweets in real-time. More importantly, the proposed system is able to notify citizens much faster than the

[7] After having made a positive adoption decision, innovations get implemented and confirmed (Rogers, 2003). However, the two last stages of the innovation-decision process are irrelevant for the purpose of this study.

Japan Meteorological Agency. Moreover, several studies have revealed the potential of Twitter to detect pandemics and disease outbreaks (Chew & Eisenbach, 2010; Krieck & Dreesmann, 2011). In the field of politics and economics, Tumasjan et al. (2010) could demonstrate that a sentiment analysis of tweets can be indicative of the outcome of political elections, while other studies substantiate the hypothesis that stock market outcomes can be predicted by analyzing the mood and sentiment of tweets (Bollen et al., 2011; Zhang et al., 2011).

Although the breadth of these studies is noteworthy, two similar studies are of remarkable relevance to this paper, one of which was released recently by market research company Nielsen, who provides the official music sales as well as TV ratings in the US and has acquired SocialGuide, a Twitter analytics specialist. Having analyzed tweets about live TV consumption, Nielsen confirms a strong relationship between Twitter activity and TV ratings. The market research company found that, in the age group of 13-34 year olds, an 8.5 % increase in Twitter volume translates to a 1 % increase in TV ratings for premiere episodes (Nielsen, 2013). Apart from predicting TV ratings, researchers at HP Labs and the University of Southern California independently predicted box office revenues of Hollywood movies by looking at the number of tweets regarding a certain movie and their sentiment in the weeks prior to the theater release (Asur & Huberman, 2010; Keegan, 2011). Even more notable, Asur & Huberman (2010), in an analysis of 24 feature film releases, outperformed the Hollywood Stock Exchange, a sophisticated prediction market for box office performance (HSX, n. d.). In fact, Asur & Huberman (2010) proposed a model, which incorporates the tweet rate[8] of seven days prior to the release date and the number of theaters that show a particular movie, which scored a .97 correlation with first weekend box office revenues.

[8] The tweet-rate is defined as the number of tweets referring to a particular movie per hour (Asur & Huberman, 2010).

This astonishing accuracy suggests that Twitter data is indeed a sufficient tool to predict music sales correctly, considering that music and movies share some similar characteristics. For instance, both are experiential goods and are promoted in a way that shall ensure the most buzz on their particular release date. For example, releasing movie trailers or lifting singles serve as promotional means that can attract potential buyers and moviegoers in advance. One might presume, however, that predictions for musical success are even superior given the differences between films and albums, although it is obviously challenging to outperform an accuracy of 97 %. Notwithstanding, apart from sequels, movies are a singular product, whereas albums released by the same artist can be seen as a product series of the same brand. As such, incorporating data of previous records, as done in some of the previously mentioned studies, might improve the prediction model proposed in this study. In addition, 26 of the 40 most followed Twitter users are musical artists, who account for about two thirds of all Top 40 user followers (Twitter Counter, 2013). It is worth noting, however, that while a similar amount of tweets are related to music and movies, it seems that fewer users engage in discussions about music (Romero et al., 2011).

In addition to this difference there is one major shortcoming of the approach to estimate economic success by analyzing Twitter data, that is, one should not mistake correlation for causation (Silver, 2012: 187). Such a predictive algorithm may be weakened once music managers try to tweak promotion campaigns in order to increase Twitter mentions. Also, other reasons such as scandals about a particular band or singer may lead to increased Twitter activity that might distort the prediction model and its accuracy. In order to account for such effects at least to some extent, I suggest not only looking at the volume of tweets, but also how many different users sent them and how many followers were reached with these tweets.

This investigation into previous research leaves us with three research questions: (1) To what extent is Twitter a reliable source of information for predicting music sales, (2) to what extent are more and different data points improving the sales forecast, (3) and to what extent does sentiment analysis improve sales forecasts? I hypothesize that (1) Twitter data is indeed a reliable source of information and that (2) more variables such as the number of unique users and the reach improve the accuracy of the prediction model. However, I assume that (3) sentiment analysis does not improve the forecast model significantly in accordance with the study by Asur & Hubermann (2010), where sentiment analysis improved the prediction only marginally.

Methodology

In order to test the hypotheses quantitatively I am comparing two data points: Twitter data and music sales data. The beauty of music sales is that unit sales figures are collected by Nielsen and made available to third parties through Nielsen SoundScan. This data sample is the official source for sales records in the US music industry and also utilized for the official Billboard music charts. Since the data are not publicly available my thanks go to Universal Music Group Santa Monica, who provided the sales data for their album releases from end of January to February 2013. The release period was chosen in mutual agreement with James Hill, Director Business Analysis at Universal Music, in order to avoid odd sales patterns due to the holiday period, which for digital music sales lasts until the end of January. Of the 118 albums Universal Music released in this period, 30 albums were chosen for the purpose of this analysis. These 30 albums were released either on January 29, February 5 or February 12 and had at least three reported sales weeks. Note that albums are usually released Tuesdays, the first sales week ends the following Sunday, or 12 days after the release. The last sales period in the data set ended March 10. The remaining 88 albums were excluded for various reasons. Some of the albums were released on February 19 or later and thus had only one or two weeks of sales data. Further, compilations, 'best of' albums, movie soundtracks, albums with generic titles such as 'romances' or 'ballads' and those with prereleases or without reported sales data at all were excluded from the analysis. For the remaining 30 album titles, Twitter data was collected through DiscoverText, an analytics solution that can capture data from various social media outlets, including Twitter. DiscoverText has GNIP-enabled access to Twitter, which allows them to pull historical data from the complete Twitter archive. Tweets were collected when they featured either the artist name or the album title in the tweet, were released two weeks prior or one week after the particular album release, and were written in English. The search queries for the 30 albums produced 2.98 million tweets in total. Four albums had to be excluded from the analysis as it turned out that the reported release date was wrong. Technical issues with DiscoverText led to the exclusion of another album so that 2.05 million tweets about 25 albums were analyzed.[9]

Two distinct models were developed: In the first model, tweets had to feature *either* the artist name *or* the album name to be considered. Tweets with obvious typos such as '*Marianne Faithful*' instead of '*Marianne Faithfull*' and excerpts of the album title such as '*Flat Top Guitar*' instead of '*Queen of the Flat Top Guitar*' were included as well. In the second model, tweets had to feature the artist name *and* the album name. In addition, tweets mentioning the artist name *and* popular track names of the album or words such as '*album*', '*record*', '*release*' or '*CD*' were considered as well. In a sense, one could describe Model 1 to be artist-related, whereas Model 2 focuses on album-related tweets. However, many tweets in Model 1 were not related to the artist at all. For instance, search queries for albums such as '*Passion*' from Andrea Bocelli, '*Icon*' from the Allman Brothers Band or the re-release of '*Broken English*' by Marianne Faithfull resulted in many irrelevant tweets. Therefore, tweets collected in Model 1 and Model 2 were manually coded for relevance with a binary coding scheme (*0=Relevant*, *1=Irrelevant*), creating Model 1b and Model 2b. The average sample size for relevance coding was 18.5 % of the total tweets per album in Model 1, and 43.3 % of the total tweets in Model 2, or roughly 18,000 tweets in total. On average, 61.8 % of all tweets of the samples of Model 1 were relevant, i.e. related to the artist, whereas the relevance score jumped to 88.3 % in Model 2. In Model 2 two albums had only a relevance of 2.0 % ('*Chasing the Saturdays*' by '*The Saturdays*') and 6.0 % ('*Icon'* by '*Allman Brothers Band*'). Without these two albums the average relevance of Model 2 had been 95.7 %. Since Model 1a (all tweets featuring *either* artist name *or* album title) produced many irrelevant tweets it was not considered for further analysis. We were thus left with three models: (1)

---

[9] See Appendix, Figure A, for a list of all 25 albums.

Model 1b, relevant artist-related tweets, (2) Model 2a, total album-related tweets and (3) Model 2b, relevant album-related tweets, with Model 2a and Model 2b overlapping for the most part.

For each of the models I calculated the daily volume of tweets, the amount of daily unique users[10] and the total reach of the unique users per day. I will call these 'variable categories'. Since the tweets ranged from two weeks prior the release date to one week after the release date this, or 22 days in total, I arrived at 198 distinct variables for the three variable categories[11]. Also, the daily values were summed up to weekly aggregates (Week 1 and Week 2 prior to the release, Week 3 after the release). Besides calculating daily Twitter data and aggregating the weekly volumes, I conducted a manual sentiment analysis on each album in Model 1b and Model 2b, both before and after its release. The average sample size per album was 18 % in Model 1b and 40 % in Model 2b, or 5,500 tweets in total. The tweets were coded *Neutral*, *Positive*, or *Negative*. Sentiment analysis for Model 2a was excluded since Model 2a and Model 2b exhibit a strong overlap and are thus assumed to have a similar sentiment.

In total, I arrived at 255 independent variables plus three dependent variables with the weekly sales data, and 6,450 data points for the 25 albums that were analyzed with a linear regression model.

---

[10] A Twitter user who has sent multiple tweets about the same album per day is only considered as one unique user.
[11] 3 models (1b, 2a, 2b) x 3 categories (tweets, unique users, reach) x 22 days.

Discussion

Sales, Twitter and Sentiment Pattern Analyses

The analysis of the data revealed some interesting patterns. In terms of the sales data it is observed that the second and third week produce significantly less sales than the first week. One extreme outlier, whose sales data might have been misreported, had ten times as many sales in the second week than in the first week. When excluding this particular album, the sales at the end of the second sales week decline to 57 % of first week sales, and to 48 % at the end of the third week. The median is 50 % for second week sales and 35 % for third week sales[12].

In terms of the daily Twitter data there is an obvious pattern that emerges. The volume of tweets is growing in the prerelease period, peaking at the release date, and then declining significantly. However, there are some exceptions to the rule. For example, the album '*Chasing the Saturdays*' by '*The Saturdays*' has a peak every seven days, but not necessarily at the release date, which can be explained by the fact that a weekly reality TV show about the English-Irish girl group premiered two weeks before the album release date. These artifacts may complicate the prediction process, especially in Model 1b, where relevant artist-related tweets are considered. Since in this specific case the album title equals the title of the TV series, even Model 2b is likely to be inaccurate.

Nonetheless, it is interesting to see that the total tweets of the whole observation period of Model 1b and Model 2a correlate by .90. The R^2 for total tweets in Model 1b and Model 2b is .70, Model 2a and Model 2b correlate by .78[13]. When excluding the two albums with low relevance scores[14] the correlation jumps to .99. Analyzing the ratio between tweets and unique users, I found that on average, a particular user tweets roughly 1.1 times about the same album per day. This is true in all

---

[12] See Appendix, Figure B, for a graphic representation.
[13] See Appendix, Figure C, for a graphic representation.
[14] 'Chasing the Saturdays' by 'The Saturdays' and 'Icon' by the 'Allman Brothers Band'.

models. On average, a tweet sent in Model 1b reaches 1,920 followers compared to 2,770 users in Model 2a and 4,180 users in Model 2b. There is no clear explanation for this difference. One factor might be that in Model 2a and 2b many tweets are published by music magazines or media outlets who link to reviews about a new album, and these professional Twitter users tend to have many more followers than the average Twitter user.

This brings us to sentiment analysis, where I found that the vast majority of tweets are not subjective. Instead, in Model 1b, 83 % of the tweets were neutral compared to 82 % in Model 2b. Positive tweets account for 17 % and 18 %, respectively. Negative tweets were basically nonexistent. It is observed, though, that subjectivity increases after the release date due to more positive tweets. While in both models neutral tweets declined from 86 % in the prerelease period to 80 % (Model 1b) or 77 % (Model 2a), positive tweets rose from 14 % to 20 % (Model 1b) and 22 % (Model 2b). However, negative tweets were rare both before and after the album release, regardless of the model.

	Model 1b		Model 2b	
	Pre-Release	Post-Release	Pre-Release	Post-Release
Neutral	86 %	80 %	86 %	77 %
Positive	14 %	20 %	14 %	22 %
Negative	0 %	0 %	0 %	1 %

Figure 1: Average sentiment for 25 albums

This is in contrast to findings related to sentiment analysis of tweets in political contexts (Parker, 2013; Tumasjan, 2010) where positive tweets dominate, but are accompanied by negative and sarcastic tweets. One factor for this is that many tweets, often sent by media outlets, consist only of a link to an album review. However, the absence of negative tweets suggests that, at least for the albums scrutinized in this study, tweeting is something that is done predominantly by fans. In fact, one is oftentimes left with the feeling that every album that was tweeted about must be nominated for the best album of the year. Also, more and more services let users share seamlessly what they are listening to at the moment, using the hashtag #nowplaying or #np. The use of this hashtag, and thus neutral tweets, might increase with Twitter #Music, a mobile music discovery app Twitter released recently (Crook, 2013).

Correlation Analysis

As mentioned previously, the creation of the three models, three variable categories (tweets, unique users, reach) and sentiment analysis created 255 explanatory variables for the dependent sales data variables. It is observed that for cumulative weekly reach variables generally achieve higher correlations with sales data than tweets or unique users. For instance, in Model 1b, prerelease reach correlation with first week sales is .70*[15] whereas prerelease tweets and prerelease unique users correlation is .50* and .49**, respectively. While the correlation scores are decreasing when looking at longer sales periods, (for instance cumulative sales through the second or third week) it still holds true that reach outperforms tweets and unique users. Surprisingly, Model 2a scores lower on correlations throughout all cumulative weekly variables when compared with Model 1b. However, for weekly data Model 2b achieves the strongest correlations, for instance between prerelease reach in Model 2b and Week 1 sales. The correlation is .83** and declines to .82** when considering sales from Week 1 and Week 2, and to .79** when considering sales from Week 1 through Week 3. Sentiment variables seem not to be related with sales at all, regardless of model or sentiment value (*Neutral*, *Positive* or *Negative*).

[15] *Correlation is significant at the 0.05 level (2-tailed)
**Correlation is significant at the 0.01 level (2-tailed)

Pearson Correlation	Week 1 Sales	Week 1+2 Sales	Total Sales (W1-3)
M1b_Tweets	0,505*	0,489*	0,451*
M1b_Unique Users	0,494*	0,479*	0,441*
M1b_Reach	0,705**	0,698**	0,669**
M2a_Tweets	0,068	0,065	0,056
M2a_Unique Users	0,081	0,078	0,068
M2a_Reach	0,317	0,314	0,301
M2b_Tweets	0,717**	0,697**	0,647**
M2b_Unique Users	0,710**	0,690**	0,639**
M2b_Reach	0,827**	0,819**	0,786**

Figure 2: Correlation Coefficients for weekly pre-release variables

It is observed that some daily data variables occasionally correlate higher with sales variables than the cumulative weekly variables. This finding was unexpected. However, there is not a single day that is strongly correlated with sales in every model, but rather a fair amount of variance between models, variable categories and days. For instance, the strongest correlation overall was observed between Model 1b reach five days before the release and sales data. For the first week sales, the correlation is .95**, for the cumulative sales data (W1+2 as well as W1-3) correlation is .96**. There is no explanation for this strong correlation, but there are six more daily reach variables out of 22 days in this particular model (Daily reach variables in Model 1b) that achieve correlations higher than .70**. The fact that correlations are sometimes stronger for longer sales periods can be explained by less variance in the sales pattern over time. Interestingly, Model 1b exhibits the strongest correlations for daily data, whereas Model 2b is superior in daily data.

Linear Regression Models with Daily and Weekly Data

Linear Regression was performed with IBM SPSS Statistics 21 Automatic Linear Modeling. Besides the three different models and variable categories, linear models were limited to either weekly or daily data and performed either with or without sentiment. In addition to the analysis of the three distinct variable categories, a ‘complete model’ was performed, including tweets, unique users and reach at the same time. However, the complete model may be subject to multicollinearity. Also, the linear models were limited to different time periods such as Week 1 Only (data ranging from 14 to 7 days before the release date), the Prerelease Period (data ranging from 14 days before the release date until the album release, including the actual release day) and the Total Period (data ranging from 14 days before the release to 7 days after the release). Dependent variables were either Week 1 Sales or Total Sales (W1-3 after the release). In total, 288 models were processed[16].

With weekly data, the complete Model 2a including sentiment and using total sales as output variable achieved the highest accuracy of 94.6 %. This result is likely to be flawed due to multicollinearity. Other non-complete linear models with weekly data achieved a prediction accuracy of about 65 %, including Model 1b (Prerelease Period Reach, without Sentiment, Total Sales) and Model 2b (Total Period Reach, with Sentiment, Week 1 Sales). Even when considering Week 1 Only data one week before the release date, prediction accuracy is 55.2 % for Week 1 Sales and 57.3 % for Total Sales, respectively. As a generalization, analyzing the reach usually produces the most predictive model and, as expected, sentiment usually does not improve the models. Model 2b produces slightly better results than Model 1b and Model 2a, although this is not always the case. Also, a generalization about the different output variables (Week 1 Sales versus Total Sales) is difficult as sometimes accuracy is higher when considering Total Sales, but sometimes it is not. Nonetheless, the linear regression model confirms the first hypothesis that Twitter data is to some degree a reliable source of information for predicting music sales.

It is remarkable that daily data improves the accuracy of the linear regression model significantly. The highest correlation of 99.9 % is achieved in a complete model (Total Period Complete Model 2b, with Sentiment, Total Sales), which, again, seems to be corrupted by

---

[16] See Appendix, Figure D, for a table of all linear regression models.

multicollinearity. However, other non-complete models score high on accuracy as well, such as the Total Period reach of Model 2b, which produces an R^2 of .97 on Total Sales. For Week 1 Sales the R^2 of this model is 90.7 %. Similarly high scores are also achieved when only considering Twitter data of the Prerelease Period. Analyzing the prerelease reach produces accuracies of about 95 % on Total Sales, regardless of the model or sentiment. For Twitter data conducted of Week 1 only one week before the release, the linear model is 68.9 % accurate (Model 2b, Week 1 Only reach, Total Sales, regardless of Sentiment). When the albums by '*The Saturdays*' and the '*Allman Brothers Band*' are excluded, accuracy jumps to 77.8 % in Model 2b and to 72.9 % in Model 2a. Considering all albums for the Week 1 Only reach, Model 2a accuracy is as low as 15.0 % on Total Sales.

Much like in the linear models with weekly data, reach variables also produce significantly better results in linear models fueled with daily data. More often than not, sentiment analysis does not affect the prediction accuracy at all. Also, Model 2b usually produces the best results with daily data, especially the shorter the scrutinized period is. Therefore, the analysis confirms the second hypothesis that more data points improve the sales forecast, as well as the third hypothesis that sentiment analysis generally does not improve predictions.

Model Comparison and Evaluation

One can conclude that daily data has a significantly better predictive potential than weekly data. Also, the album-related models tend to produce more accurate predictions than the artist-related model. This has positive economic implications since the efforts of analysis increase the more tweets are to be considered. Note that in this study the artist-related search queries delivered more than two million tweets, which were often irrelevant, whereas the album-related search queries only generated 100,000 results. Especially when conducting analyses of historical Twitter data this difference is crucial since data providers have to charge extra fees for historical tweets. Since the artist-related tweets were only relevant in about 60 % of the cases, the artist-related model would only exhibit predictive power if it were coded for relevance, which was done manually in this project. It is unclear to what extent algorithms can produce comparable and reliable results.

It is telling that the performance of Model 2a with Week 1 Only reach data gets improved significantly when excluding two albums that score low on relevance. This suggests that relevance coding is helpful even in an album-related model or that the model should only be applied to artist names and album titles that are to a large degree unique. While it is understandable that search queries for '*The Saturdays*' produce irrelevant tweets, this was not expected for the '*Allman Brothers Band*'.

Also, reach usually outperforms tweets and unique users as predictor variables significantly. While this finding was unexpected one explanation might be that the reach model takes into account the popularity of a user and the impact of his or her tweets whereas looking at the tweets or unique users levels their importance. It is assumed that a tweet sent by a user with a large following has a greater impact on sales, which is why it seems to be helpful for studies of this kind to consider reach instead of tweet or unique user figures.

Limitations and Future Research

Despite high prediction accuracies the study has several limitations. First, many albums with generic names as well as compilations and movie sound tracks were excluded from the analysis and it is important to evaluate the forecast accuracy when these albums are included as well, especially when considering the issues with the albums by '*The Saturdays*' and the '*Allman Brothers Band*'. Second, in this study albums were excluded when their release date was misreported. While this source of error can be fixed easily, it needs to be assessed how the prediction model is affected by not entirely accurate data. In this study, two search queries deviated slightly from the original album title – *Andrea Bocelli's* album was titled '*Passione*', not '*Passion*', and the correct album title of '*Rock Candy Funk Party*' is

‘*We want to Groove*’, not ‘*We Want Groove*’. Third, the impact of prereleases or pre-streaming is unclear. However, it is believed that an exclusive online pre-streaming of an album only affects its sentiment analyses scores, which, in this study, had no impact on the prediction accuracy. Fourth, manual coding for relevance and sentiment is not always easy. It is not always clear whether a tweet such as ‘*Love Saturdays*’ or ‘*Passion is great*’ is related to the artist or the album at all. This is why Model 2a is the most economical and most reliable one, while Model 2b is the most accurate one. Fifth, tweets were collected when they were written in English, which is not the same as tweets sent by US users. Remember that the sales data only accounts for US sales. Further, the followers that had seen the tweets presumably live in all parts of the world, which makes it a little bit surprising that reach outperformed other models significantly. Sixth, the meta data that was collected through DiscoverText occasionally lacked the follower count, which was then assumed to be zero. Also, I observed several anomalies in the data. On the one hand, one Twitter user tweeted 40 times about an album on a single day; on the other hand, I encountered a few tweets that had exactly the same wording, but were sent – not retweeted – by dozens of different Twitter users. This lets one assume that some Twitter users are fraudulent. In fact, it was reported recently that, according to estimates, 20 million of the Twitter accounts are fake, which are sold for commercial reasons to brands and musicians who want to improve the number of their followers (Perloth, 2013; Perloth, 2013a). However, it is assumed that the reach variable is less susceptible to fraud as it is difficult to fake users with a large followership.

Nonetheless, one has to be generally cautious about the power of big data. Boyd & Crawford (2012) express their doubts about the objectivity and accuracy and argue that bigger data is not always better data, especially when taken out of context. Hence, it is important not to mistake correlation for causation. Especially when marketers try to tweak social media campaigns and create more chatter about an album the prediction model will eventually deteriorate significantly.

This study is, to the knowledge of this author, the first one to juxtapose Twitter data and music sales data. As such, more research is needed. For instance, it would be interesting to see how the chatter about pre-released singles is predictive of later album sales, and whether long-term observations allow for accurate forecasts of album sales.

Conclusion

I have shown that Twitter is a useful source of information for the forecast of music sales, with daily Twitter data outperforming weekly data. Also, the reach of album-related tweets sent by unique users per day is consistently the most predictive explanatory variable, which could be a helpful insight for other studies of this kind. In contrast, sentiment analysis does not improve predictions, but it is useful to look at longer sales periods. For instance, by only looking at the daily reach from January 22 to January 29 for an album that was released on February 5 the predicted values correlate with actual sales data until March 3 by .69. By expanding the observed Twitter period to February 5, prediction accuracy rises to 95.0 %.[17]

This finding suggests that the economic success of music album can indeed be forecasted with a remarkable accuracy by only looking at how often it was tweeted about prior to its release. However, one should not mistake correlation for causation since the prediction model will deteriorate when marketers try to tweak social media campaigns in order to create more Twitter chatter. In addition, this is only a small study consisting of 25 albums and many albums with generic titles were excluded purposefully, which is why this kind of research needs to be extended.

This is even more important since the media landscape is constantly changing. While users might change their communication habits or might communicate on other social media networks, music listeners seem to increasingly access music streaming services, which would ask for other dependent variables.

[17] See Appendix, Figure E, for a scatter plot of predicted versus actual total sales.

Bibliography

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *arXiv:1003.5699*.

Beevolve. (2012, October 10). An Exhaustive Study of Twitter Users Across the World. *Beevolve*. Retrieved May 20, 2013 from http://www.beevolve.com/twitter-statistics/#b1

Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage*, *49*(3), 2687-2696.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.

Bradley, S., Cespedes, F. & Herman, K. (2010). Live Nation Faces the Music. *Harvard Business School Case 709-441*. Boston, MA. (Revised from original January 2009 version.)

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*, *5*(11), e14118.

Crook, J. (2013, April 18). Twitter Goes To Good Morning America To Confirm Twitter #Music, Aiming Straight For The Mainstream. *TechCrunch*. Retrieved April 19, 2013, from http://techcrunch.com/2013/04/18/twitter-music-app-goes-live/

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

Dhar, V., & Chang, E. (2007). Does chatter matter? The impact of user-generated content on music sales. NYU Working Paper No. CEDER-07-06.

Echo Nest. (n. d.). *Our Company*. Retrieved April 20, 2013, from: http://echonest.com/company/

Hann, I. H., Oh, J., & James, G. (2011). *Forecasting the Sales of Music Albums: A Functional Data Analysis of Demand and Supply Side P2P Data*. Working paper.

HSX. (n. d.). *About HSX*. Retrieved April 23, 2013, from: http://www.hsx.com/about/

Hudson, N. J. (2011). Musical beauty and information compression: Complex to the ear but simple to the mind?. *BMC research notes*, *4*(1), 9.

IFPI. (2012, November 12). Record labels invest US$4.5 billion in new music. *IFPI*. Retrieved April 22, 2013, from: http://www.ifpi.org/content/section_news/investing_in_music.html

IFPI. (2012a, January 23). IFPI publishes Digital Music Report 2012. *IFPI*. Retrieved April 22, 2013, from http://www.ifpi.org/content/section_resources/dmr2012.html

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). ACM.

Keegan, R. (2011, July 14). USC launches Twitter movie box office predictor. *Los Angeles Times*. Retrieved April 25, 2013, from http://latimesblogs.latimes.com/movies/2011/07/usc-launches-a-new-twitter-box-office-predictor.html

Krieck, M., Dreesman, J., Otrusina, L., & Denecke, K. (2011). A new age of public health: Identifying disease outbreaks by analyzing tweets. In *Proceedings of Health Web-Science Workshop, ACM Web Science Conference*.

Lee, J., Boatwright, P., & Kamakura, W. A. (2003). A Bayesian model for prelaunch sales forecasting of recorded music. *Management Science*, *49*(2), 179-196.

Lee, C., & Cunningham, P. (2012). The Geographic Flow of Music. *arXiv:1204.2677*.

Lunden, I. (2012, July 30). Analyst: Twitter Passed 500M Users In June 2012, 140M Of Them in US; Jakarta 'Biggest Tweeting' City. *TechCrunch*. Retrieved April 29, 2013, from: http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-

140m-of-them-in-us-jakarta-biggest-tweeting-city/

Ni, Y., Santos-Rodríguez, R., McVicar, M., & De Bie, T. (2011). Hit song science once again a science?. In *Proceedings of 4th International Workshop on Machine Learning and Music: Learning from Musical Structure* (MML 2011).

Nielsen. (2013). *New Study Confirms Correlation Between Twitter and TV Ratings*. Retrieved April 20, 2013, from http://www.nielsen.com/us/en/newswire/2013/new-study-confirms-correlation-between-twitter-and-tv-ratings.html

Nikolaeva, I., & Hinz, O. (2012). How Smartphone Apps Can Help Predicting Music Sales. *ECIS 2012 Proceedings.* Paper 102.

Noelle-Neumann, E. (1993). *The Spiral of Silence: Public Opinion – Our Social Skin*. Chicago, IL: University of Chicago Press.

Parker, G. (2013, January 30). Tool Developed By USC Annenberg Innovation Lab Explores True Sentiment Of Social Media. *USC Annenberg News*. Retrieved April 28, 2013, from http://annenberg.usc.edu/News%20and%20Events/News/130130TaplinSentiment.aspx

Pariser, E. (2011). *The Filter bubble. How the new, personalized web is changing what we read and how we think*. London, England: Penguin Press.

Perloth, N. (2013, April 5). Fake Twitter Followers Become Multimillion-Dollar Business. *New York Times*. Retrieved May 2, 2013, from http://bits.blogs.nytimes.com/2013/04/05/fake-twitter-followers-becomes-multimillion-dollar-business/

Perloth, N. (2013a, April 25). Researchers Call Out Twitter Celebrities With Suspicious Followings. *New York Times*. Retrieved May 2, 2013, from http://bits.blogs.nytimes.com/2013/04/25/researchers-call-out-twitter-celebrities-with-suspicious-followings/

Rogers, E. M. (2003). *Diffusion of innovations*. New York, NY: Simon and Schuster.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 695-704). ACM.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851-860). ACM.

Semiocast. (2012, January 31). *Brazil becomes 2nd country on Twitter, Japan 3rd, Netherlands most active country. Geolocation analysis of Twitter accounts by Semiocast*. Retrieved May 2, 2013, from: http://semiocast.com/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan

Silver, N. (2012). *The Signal and the Noise. Why so many predictions fail – but some don't*. New York, NY: Penguin Press.

Smith, D. P. (2012, April 10). A limitless focus group: Big data in social media has transformed marketing. *Kellogg School of Management, Northwestern University*. Retrieved April 18, 2013, from: http://www.kellogg.northwestern.edu/News_Articles/2012/choudhary-focus-group.aspx

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Doubleday.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010, May). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the fourth international aaai conference on weblogs and social media* (pp. 178-185).

Twitter Counter. (2013, February 23). *Twitter top 100: most followers*. Retrieved April 20, 2013, from: http://twittercounter.com/pages/100

Weber, H. (2012, March 21). With 140 million active users & 340 million tweets per day, Twitter is officially mainstream. *TheNextWeb.* Retrieved April 23, 2013, from:

http://thenextweb.com/socialmedia/2012/03/21/twitter-has-over-140-million-active-users-sending-over-340-million-tweets-a-day/

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*, *26*, 55-62.

Appendices

Figure A

List of all albums.

ARTIST	ALBUM TITLE
Allman Brothers Band	Icon
Andrea Bocelli	Passione
Ballake Sissoko	At Peace
Brotha Lynch Hung	Mannibalector
Canton Jones	CJ Talks
Circle Ii Circle	Season Will Fall
Coheed & Cambria	Afterman Decension
Cult Of Luna	Vertikal
Dave Koz	Live At The Blue Note Tokyo
Eels	Wonderful Glorious
Eric Burdon	'Til Your River Runs
Jonas Kaufmann	Wagner
Lena Hughes	Queen Of The Flat Top Guitar
Marianne Faithfull	Broken English
Modestep	Evolution Theory
Otis Taylor	My World Is Gone
RDGLDGRN	Red Gold Green
Rock Candy Funk Party	We Want To Groove
Saturdays	Chasing The Saturdays
Steeldrivers	Hammer Down
Stone Foxes	Small Fires
Terri Lyne Carrington	Money Jungle
Tim McGraw	Two Lanes Of Freedom
Tomahawk	Oddfellows
Wake Owl	Wild Country

Figure B

Sales Pattern of 25 albums for three weeks (Adjusted Mean, n=23).

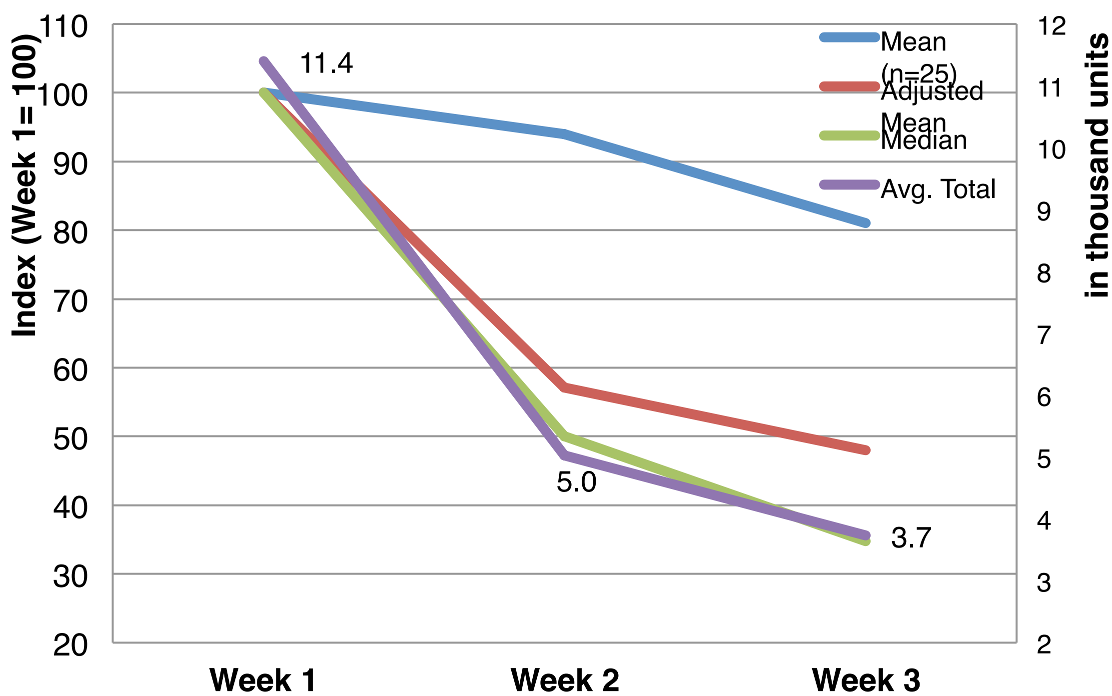
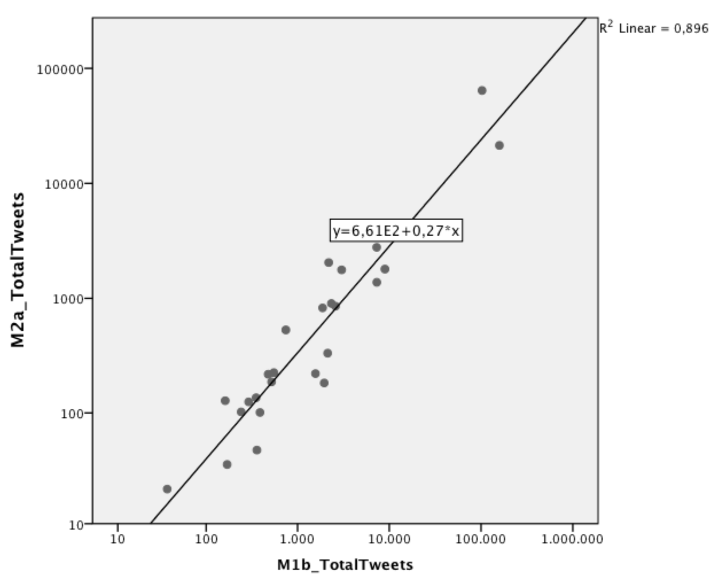
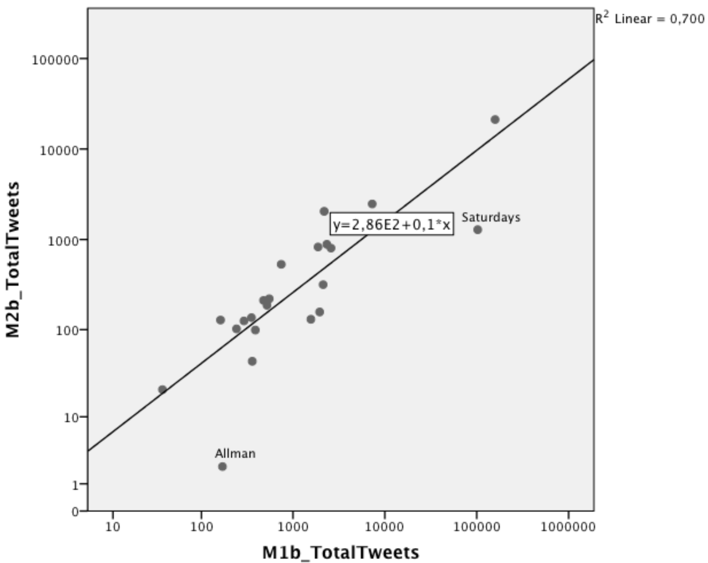


Figure C

Scatter plot comparison of Total Period Tweets in Model 1b, Model 2a and Model 2b.




R² Linear = 0,700
y=2,86E2+0,1*x
Saturdays
Allman
M2b_TotalTweets
M1b_TotalTweets
100000
10000
1000
100
10
1
0
10
100
1000
10000
100000
1000000

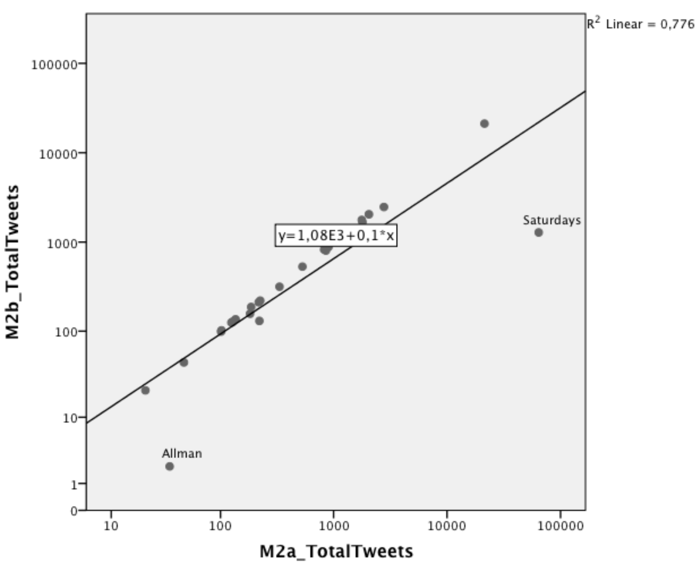

R² Linear = 0,776
y=1,08E3+0,1*x
Saturdays
Allman
M2b_TotalTweets
M2a_TotalTweets
100000
10000
1000
100
10
1
0
10
100
1000
10000
100000

Figure D

Comparison of 288 linear regression models.

WEEK 1 ONLY TWITTER DATA	WEEKLY DATA					DAILY DATA				
	No Sentiment		Sentiment**			No Sentiment		Sentiment**		
	W1 Sales	Total Sales	W1 Sales	Total Sales		W1 Sales	Total Sales	W1 Sales	Total Sales	
M1b_CompleteModel	59,1%	68,3%	59,1%	68,3%		31,9%	36,6%	31,9%	36,6%	
M1b_Tweets	21,6%	16,6%	21,6%	16,6%		60,5%	67,2%	60,5%	67,2%	
M1b_Unique Users	21,3%	16,3%	21,9%	16,9%		61,8%	68,6%	61,8%	68,6%	
M1b_Reach	55,2%	57,3%	55,2%	57,3%		31,9%	36,6%	31,9%	36,6%	
M2a_CompleteModel	7,1%	11,1%	7,1%	11,1%	*	76,5%	78,4%	76,5%	78,4%	*
M2a_Tweets	0,0%	0,0%	0,0%	0,0%	*	20,2%	19,9%	20,2%	19,9%	*
M2a_Unique Users	0,0%	0,0%	0,0%	0,0%	*	47,9%	45,7%	47,9%	45,7%	*
M2a_Reach	7,1%	11,1%	7,1%	11,1%	*	9,5%	15,0%	9,5%	15,0%	*
M2b_CompleteModel	10,2%	14,6%	10,2%	14,6%		39,2%	81,3%	39,2%	81,3%	
M2b_Tweets	0,0%	0,0%	0,0%	0,0%		33,8%	32,9%	33,8%	32,9%	
M2b_Unique Users	0,0%	0,0%	0,0%	0,0%		33,9%	31,9%	33,9%	31,9%	
M2b_Reach	10,2%	14,6%	10,2%	14,6%		65,3%	68,9%	65,3%	68,9%	

PRERELEASE PERIOD (W1+W2) TWITTER DATA	WEEKLY DATA					DAILY DATA				
	No Sentiment		Sentiment			No Sentiment		Sentiment		
	W1 Sales	Total Sales	W1 Sales	Total Sales		W1 Sales	Total Sales	W1 Sales	Total Sales	
M1b_CompleteModel	59,4%	69,1%	59,7%	69,2%		91,4%	96,0%	91,4%	96,0%	
M1b_Tweets	23,3%	18,4%	23,9%	19,1%		62,4%	71,3%	62,4%	71,3%	
M1b_Unique Users	23,1%	18,3%	23,8%	18,9%		63,2%	72,3%	63,2%	72,3%	
M1b_Reach	58,1%	65,1%	58,0%	64,6%		91,4%	96,0%	91,4%	96,0%	
M2a_CompleteModel	55,0%	61,2%	55,0%	61,2%	*	91,1%	95,3%	91,1%	95,3%	*
M2a_Tweets	20,8%	16,1%	20,8%	16,1%	*	62,3%	66,7%	62,3%	66,7%	*
M2a_Unique Users	19,9%	15,1%	19,9%	15,1%	*	62,0%	67,6%	62,0%	67,6%	*
M2a_Reach	34,7%	32,1%	34,7%	32,1%	*	88,9%	94,3%	88,9%	94,3%	*
M2b_CompleteModel	54,0%	55,9%	54,0%	55,9%		90,7%	95,3%	90,7%	95,3%	
M2b_Tweets	29,9%	27,9%	29,9%	27,9%		87,7%	87,7%	87,7%	87,7%	
M2b_Unique Users	26,1%	23,6%	26,1%	23,6%		89,8%	89,3%	89,8%	89,3%	
M2b_Reach	54,0%	55,9%	54,0%	55,9%		90,7%	95,0%	90,7%	95,0%	

TOTAL PERIOD (W1-3) TWITTER DATA	WEEKLY DATA					DAILY DATA				
	No Sentiment		Sentiment			No Sentiment		Sentiment		
	W1 Sales	Total Sales	W1 Sales	Total Sales		W1 Sales	Total Sales	W1 Sales	Total Sales	
M1b_CompleteModel	59,7%	69,2%	59,7%	69,2%		91,4%	96,0%	91,4%	96,0%	
M1b_Tweets	23,9%	19,1%	23,9%	19,1%		62,4%	71,3%	62,4%	71,3%	
M1b_Unique Users	30,0%	31,6%	30,0%	31,6%		63,2%	72,3%	63,2%	72,3%	
M1b_Reach	58,0%	64,6%	58,0%	64,6%		91,4%	96,0%	91,4%	96,0%	
M2a_CompleteModel	71,3%	79,8%	91,2%	94,6%	*	93,8%	96,1%	93,8%	96,1%	*
M2a_Tweets	21,4%	16,6%	42,4%	31,4%	*	62,3%	66,7%	88,7%	66,7%	*
M2a_Unique Users	19,9%	15,1%	42,4%	31,4%	*	62,0%	67,6%	80,5%	67,6%	*
M2a_Reach	34,7%	32,1%	60,5%	32,1%	*	92,4%	96,3%	92,4%	96,3%	*
M2b_CompleteModel	54,0%	55,9%	65,0%	55,9		90,7%	99,9%	90,7%	99,9%	
M2b_Tweets	30,9%	26,1%	48,1%	38,3%		95,7%	88,4%	92,0%	88,4%	
M2b_Unique Users	26,1%	23,6%	42,4%	31,4%		91,3%	89,3%	84,2%	89,3%	
M2b_Reach	54,0%	55,9%	65,0%	55,9%		90,7%	97,0%	90,7%	97,0%	
*SENTIMENT ANALYSIS DATA TAKEN FROM MODEL M2B										
**SENTIMENT ANALYSIS DATA TAKEN FROM PRERELEASE PERIOD										

Figure E

Actual Total Sales versus Predicted Values (Model 2b Prerelease Reach, Daily Data, Without Sentiment)